Today: Segmentation + Targeting

Part 1: Segmentation and Cluster Analysis

1. The basic framework: segmentation, targeting, positioning
2. Bases for segmentation: what data should we use?
3. Data-driven segmentation: cluster analysis
    1. Hierarchical clustering + implementation in Python
    2. K-means + implementation in Python

Part 2: Targeting

1. Choosing a target segment
2. Nicorette discussion

Today’s Goals

Understand:

- How segmentation is used in practice (STP)
- Common types of data used for segmentation, and how to choose between them
- The hierarchical + k-means clustering algorithms for data-driven segmentation
- How segmentation is used within a marketing strategy

Be able to:

- Choose a suitable number of clusters for segmentation
- Interpret the results of a cluster analysis
- Implement cluster analysis in Python

Marketing Strategy & STP

Segmentation, Targeting, Positioning (STP)

Basic goal of marketing:

*Deliver the **right products**, to the **right people**, in the **right way***

Targeting Segmentation Positioning

Why does this matter?

For the Customer

- Customized products and services
- Personalized experiences
- Higher customer satisfaction

Loyalty and retention

For the Firm

- Identify high-value customers
- Targeted marketing actions
- Greater price premium
- Higher CLV

Sustainable profit growth

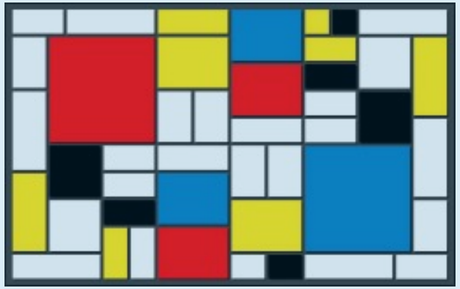
Where will we play?

How will we win?

Segmentation

S

Discovering and profiling groups of customers with similar needs and preferences



Targeting

T

Evaluating segment attractiveness and targeting most attractive ones

Positioning

P

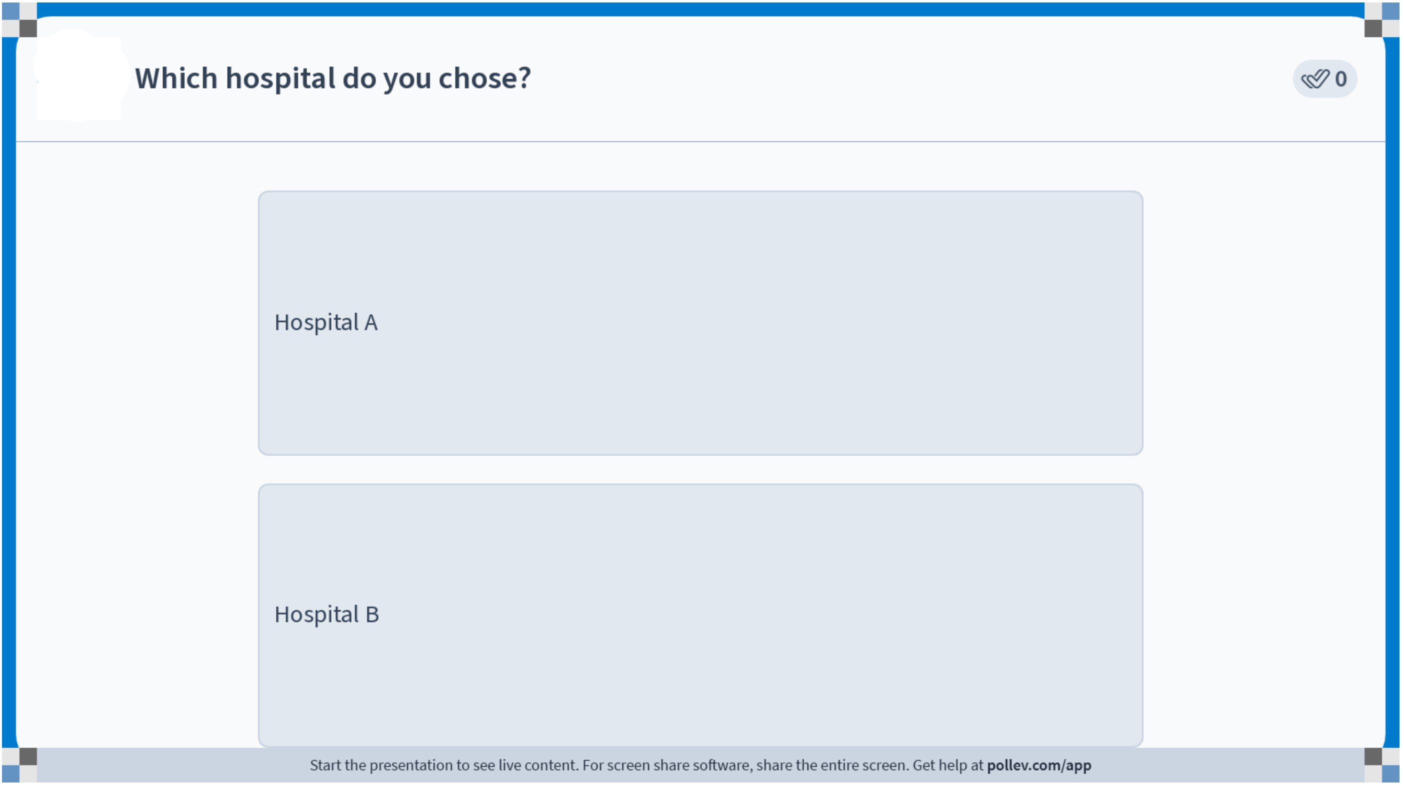
Defining value proposition for target segments and developing a marketing plan

Segmentation

Motivation

Poll Title: Which hospital do you choose?

	Hospital A	Hospital B
Died	123	72
Survived	3777	3744
Total	3900	3816
Death rate	0.032	0.019



Which hospital do you chose?

Hospital A

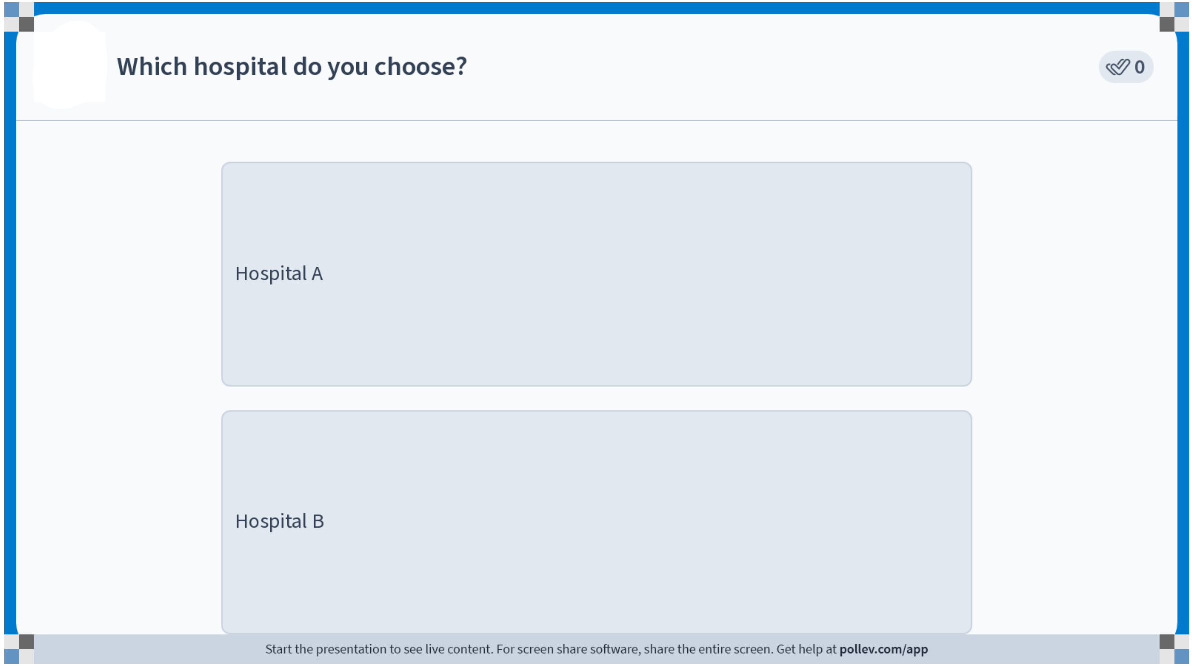
Hospital B

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Poll Title: Which hospital do you choose?

	Hospital A	Hospital B
Died	123	72
Survived	3777	3744
Total	3900	3816
Death rate	0.032	0.019

	Good Condition		Bad Condition	
	Hospital A	Hospital B	Hospital A	Hospital B
Died	9	56	114	16
Survived	891	3552	2886	192
Total	900	3608	3000	208
Death rate	0.010	0.016	0.038	0.077



Which hospital do you choose?

Hospital A

Hospital B

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Motivation 1 – A Tale of Averages

135°F

57 °F

How do you like your tea?

Motivation 1 – A Tale of Averages

96 °F

“An average describes everyone…and no one…”

Motivation 2 - Heterogeneity

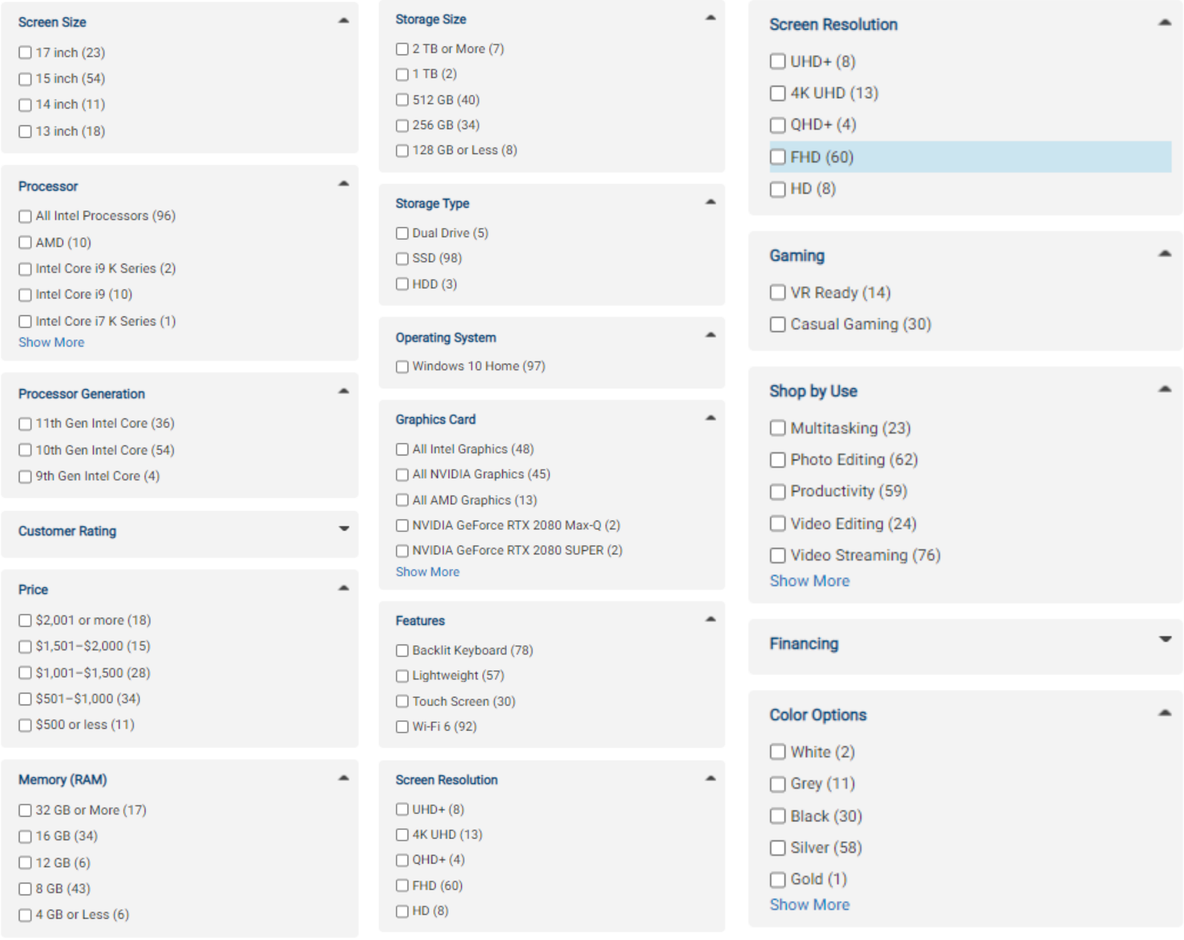
Without segmentation, we treat customers as averages.

Motivation 3 – Perils of Hyper-Customization

How many different Dell laptops can be manufactured?

https://www.dell.com/en-us/shop/dell-laptops/sc/laptops

Motivation 3 – Perils of Hyper-Customization



Screen Size
- ☐ 17 inch (23)
- ☐ 15 inch (54)
- ☐ 14 inch (11)
- ☐ 13 inch (18)

Processor
- ☐ All Intel Processors (96)
- ☐ AMD (10)
- ☐ Intel Core i9 K Series (2)
- ☐ Intel Core i9 (10)
- ☐ Intel Core i7 K Series (1)

Show More

Processor Generation
- ☐ 11th Gen Intel Core (36)
- ☐ 10th Gen Intel Core (54)
- ☐ 9th Gen Intel Core (4)

Customer Rating

Price
- ☐ $2,001 or more (18)
- ☐ $1,501–$2,000 (15)
- ☐ $1,001–$1,500 (28)
- ☐ $501–$1,000 (34)
- ☐ $500 or less (11)

Memory (RAM)
- ☐ 32 GB or More (17)
- ☐ 16 GB (34)
- ☐ 12 GB (6)
- ☐ 8 GB (43)
- ☐ 4 GB or Less (6)

Storage Size
- ☐ 2 TB or More (7)
- ☐ 1 TB (2)
- ☐ 512 GB (40)
- ☐ 256 GB (34)
- ☐ 128 GB or Less (8)

Storage Type
- ☐ Dual Drive (5)
- ☐ SSD (98)
- ☐ HDD (3)

Operating System
- ☐ Windows 10 Home (97)

Graphics Card
- ☐ All Intel Graphics (48)
- ☐ All NVIDIA Graphics (45)
- ☐ All AMD Graphics (13)
- ☐ NVIDIA GeForce RTX 2080 Max-Q (2)
- ☐ NVIDIA GeForce RTX 2080 SUPER (2)

Show More

Features
- ☐ Backlit Keyboard (78)
- ☐ Lightweight (57)
- ☐ Touch Screen (30)
- ☐ Wi-Fi 6 (92)

Screen Resolution
- ☐ UHD+ (8)
- ☐ 4K UHD (13)
- ☐ QHD+ (4)
- ☐ FHD (60)
- ☐ HD (8)

Screen Resolution
- ☐ UHD+ (8)
- ☐ 4K UHD (13)
- ☐ QHD+ (4)
- ☐ FHD (60)
- ☐ HD (8)

Gaming
- ☐ VR Ready (14)
- ☐ Casual Gaming (30)

Shop by Use
- ☐ Multitasking (23)
- ☐ Photo Editing (62)
- ☐ Productivity (59)
- ☐ Video Editing (24)
- ☐ Video Streaming (76)

Show More

Financing

Color Options
- ☐ White (2)
- ☐ Grey (11)
- ☐ Black (30)
- ☐ Silver (58)
- ☐ Gold (1)

Show More

Based on website, at least
4*5*3*5*5*5*3*5*4*5*2*5*5=112.5M!

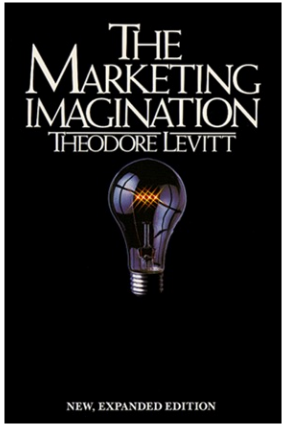
Solution: Segmentation

Market segmentation is the **subdividing** of a market into **distinct subsets of customer**

> To think segments means you have to think about what drives customers, … , and the choices that are or might be available to them.
>
> —Levitt, The *Marketing Imagination*

- **What does this mean?**

Earth GEOX Tech



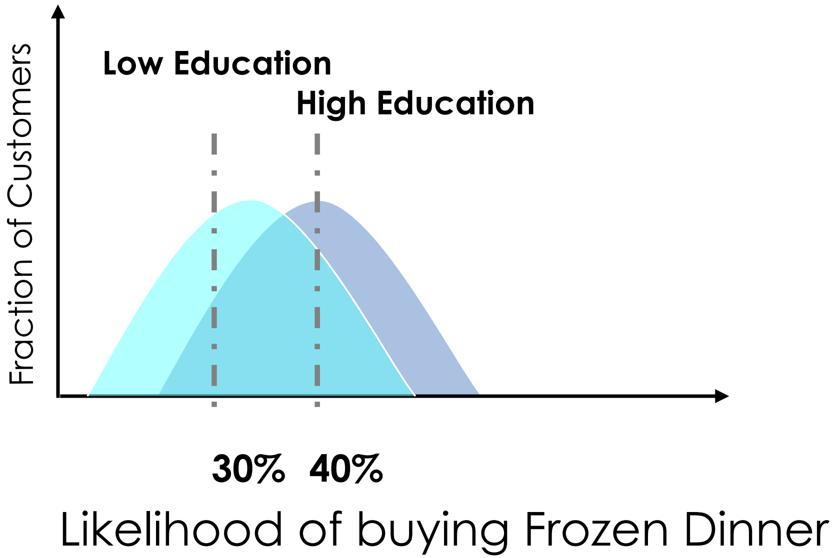
Segmentation

Data

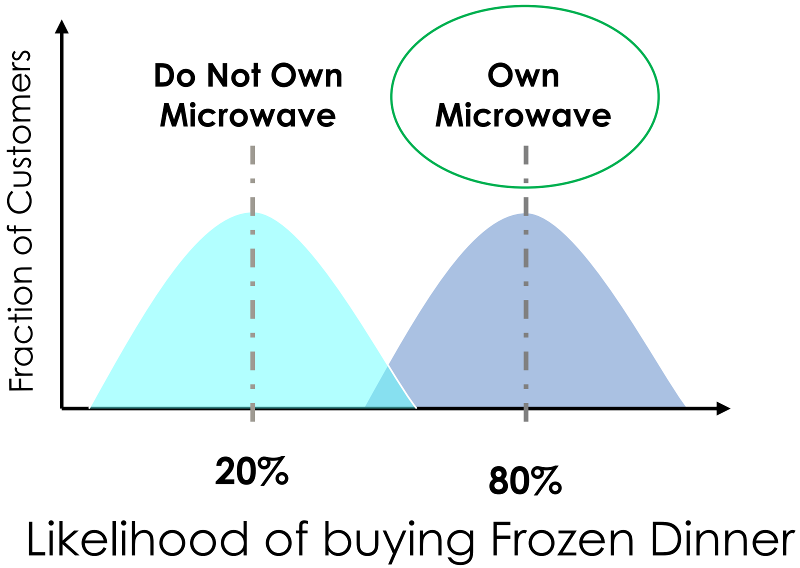
Choosing Data for Segmentation

Microwaveable Meals

Ineffective Segmentation



Effective Segmentation



Data Types for Segmentation

- **Geodemographics**
- Psychographics
- Behavioral
- Benefits & Needs

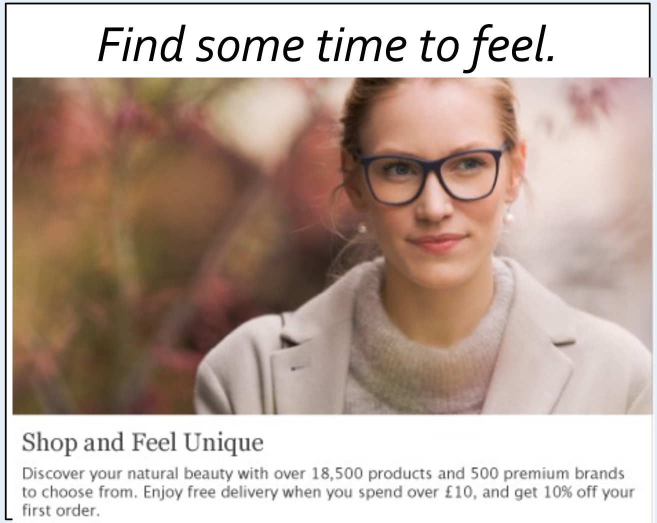
Who a person is (statistically),
and where they live.

- **Age, Gender, Family composition**
- **Education and Income**
- **Geography (urban vs rural)**
- **Affiliations, Group Identity**
- …

Data Types for Segmentation

- Geodemographics
- **Psychographics**
- Behavioral
- Benefits & Needs

How a person thinks: attitudes, interests, likes, opinions



Find some time to feel.

Shop and Feel Unique

Discover your natural beauty with over 18,500 products and 500 premium brands to choose from. Enjoy free delivery when you spend over £10, and get 10% off your first order.

Data Types for Segmentation

- Geodemographics
- **Psychographics**
- Behavioral
- Benefits & Needs

How a person thinks: attitudes, interests, likes, opinions

Love the spotlight, feel the moment.

Shop and Feel Unique

Discover your natural beauty with over 18,500 products and 500 premium brands to choose from. Enjoy free delivery when you spend over £10, and get 10% off your first order.

Data Types for Segmentation

- Geodemographics
- **Psychographics**
- Behavioral
- Benefits & Needs

How a person thinks: attitudes, interests, likes, opinions

- **Attitude**
- **Interests**
- **Personality**
- **Values**
- …

Data Types for Segmentation

- Geodemographics
- Psychographics
- **Behavioral**
- Benefits & Needs

What a customer <u>does</u>;
When? Where? How Much?

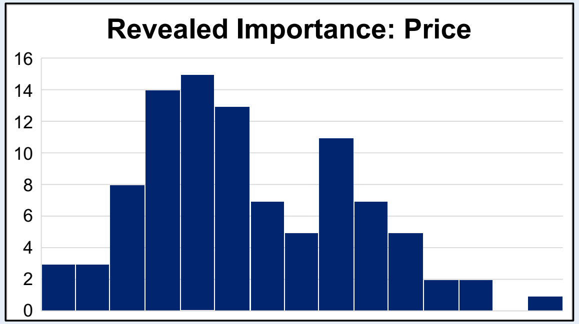
- Past purchases
- Channels (online/offline)
- Product usage
- Customer service interactions
- Browsing behavior

Data Types for Segmentation

- Geodemographics
- Psychographics
- Behavioral
- **Benefits & Needs**

What customers <u>want</u> (need) from your product

- How much customers value each attribute?



- *Needs-based* segmentation

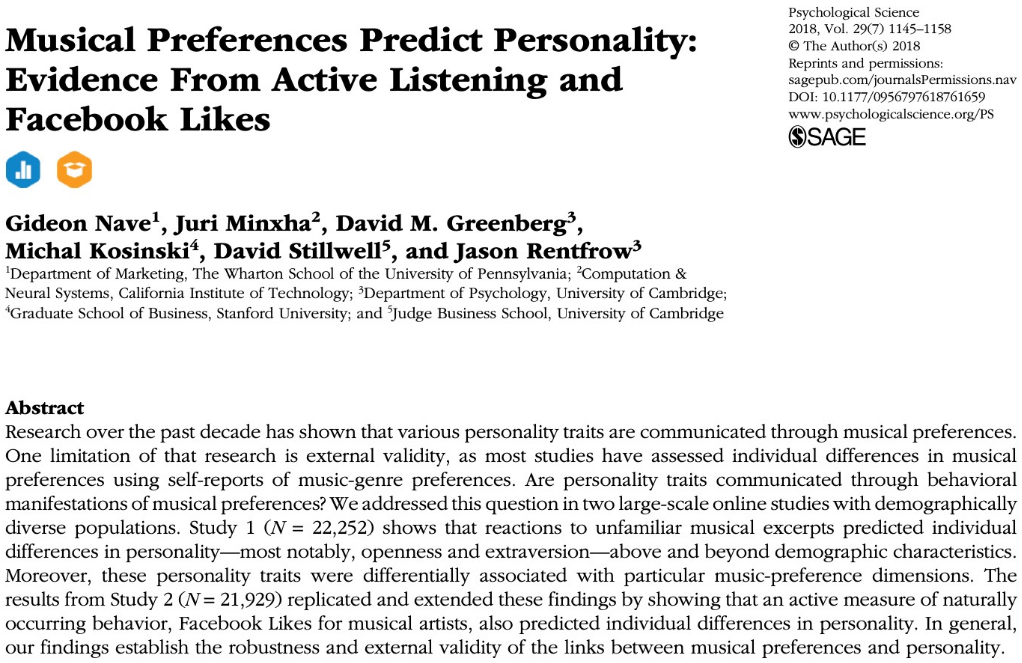
Which Data Type to Use?

- **It depends!**

- Is the data type predictive of future behavior?
  - Very often: behavior > psycho / demographics

- Is it actionable for developing / implementing strategy?
  - Segment-level: Can you reach the specific segment?
  - Individual-level: Can you identify segment membership beyond the study sample? (And does it matter?)
  - Psychographics are great for targeting / positioning, but difficult to identify "in the wild"

Which Data Type to Use?



Psychological Science
2018, Vol. 29(7) 1145–1158
© The Author(s) 2018
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797618761659
www.psychologicalscience.org/PS
SAGE

Musical Preferences Predict Personality: Evidence From Active Listening and Facebook Likes

Gideon Nave[1], Juri Minxha[2], David M. Greenberg[3], Michal Kosinski[4], David Stillwell[5], and Jason Rentfrow[3]
[1]Department of Marketing, The Wharton School of the University of Pennsylvania; [2]Computation & Neural Systems, California Institute of Technology; [3]Department of Psychology, University of Cambridge; [4]Graduate School of Business, Stanford University; and [5]Judge Business School, University of Cambridge

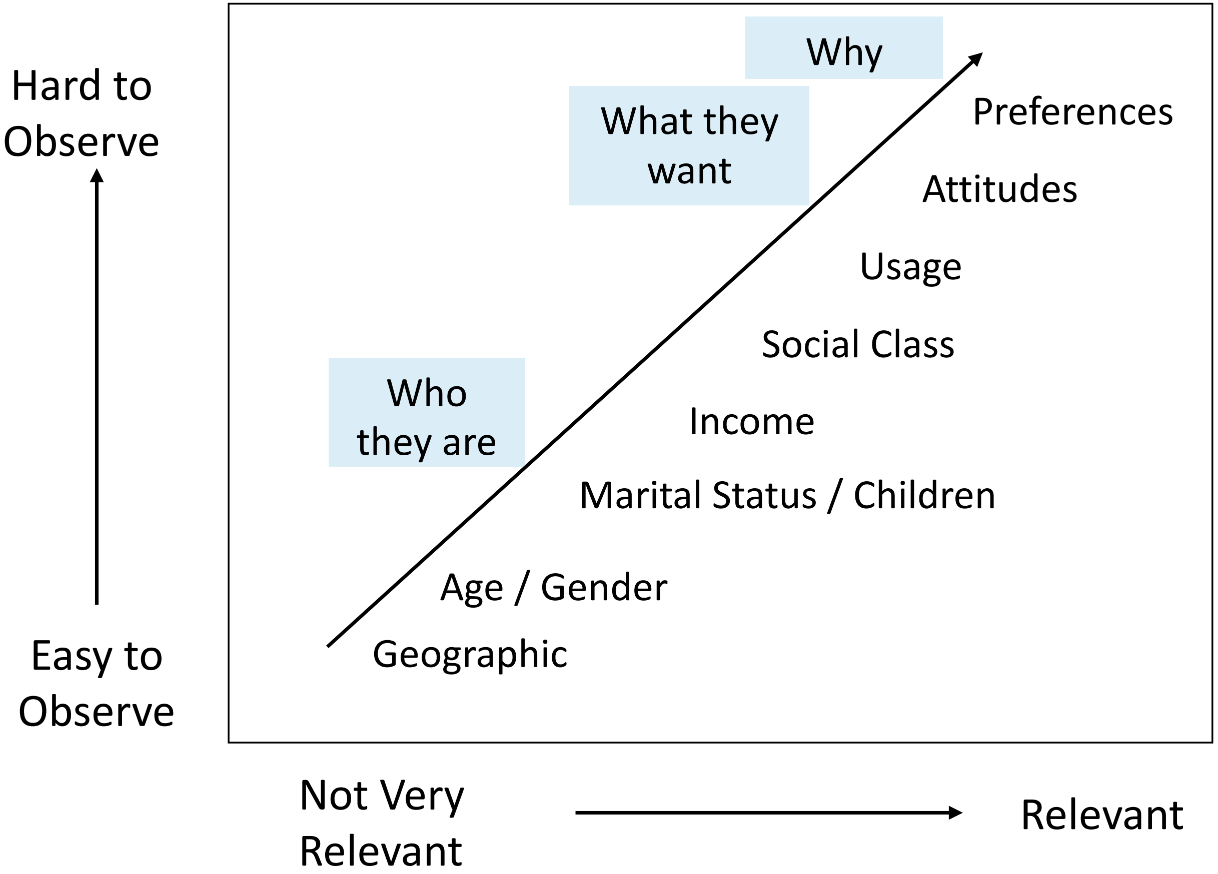
Abstract

Research over the past decade has shown that various personality traits are communicated through musical preferences. One limitation of that research is external validity, as most studies have assessed individual differences in musical preferences using self-reports of music-genre preferences. Are personality traits communicated through behavioral manifestations of musical preferences? We addressed this question in two large-scale online studies with demographically diverse populations. Study 1 (N = 22,252) shows that reactions to unfamiliar musical excerpts predicted individual differences in personality—most notably, openness and extraversion—above and beyond demographic characteristics. Moreover, these personality traits were differentially associated with particular music-preference dimensions. The results from Study 2 (N = 21,929) replicated and extended these findings by showing that an active measure of naturally occurring behavior, Facebook Likes for musical artists, also predicted individual differences in personality. In general, our findings establish the robustness and external validity of the links between musical preferences and personality.

- Psychographics are great for targeting / positioning, but difficult to identify “in the wild”

https://journals.sagepub.com/doi/pdf/10.1177/0956797618761659

Foundations of Segmentation



Hard to Observe

Easy to Observe

Why

What they want

Who they are

Preferences

Attitudes

Usage

Social Class

Income

Marital Status / Children

Age / Gender

Geographic

Not Very Relevant

Relevant

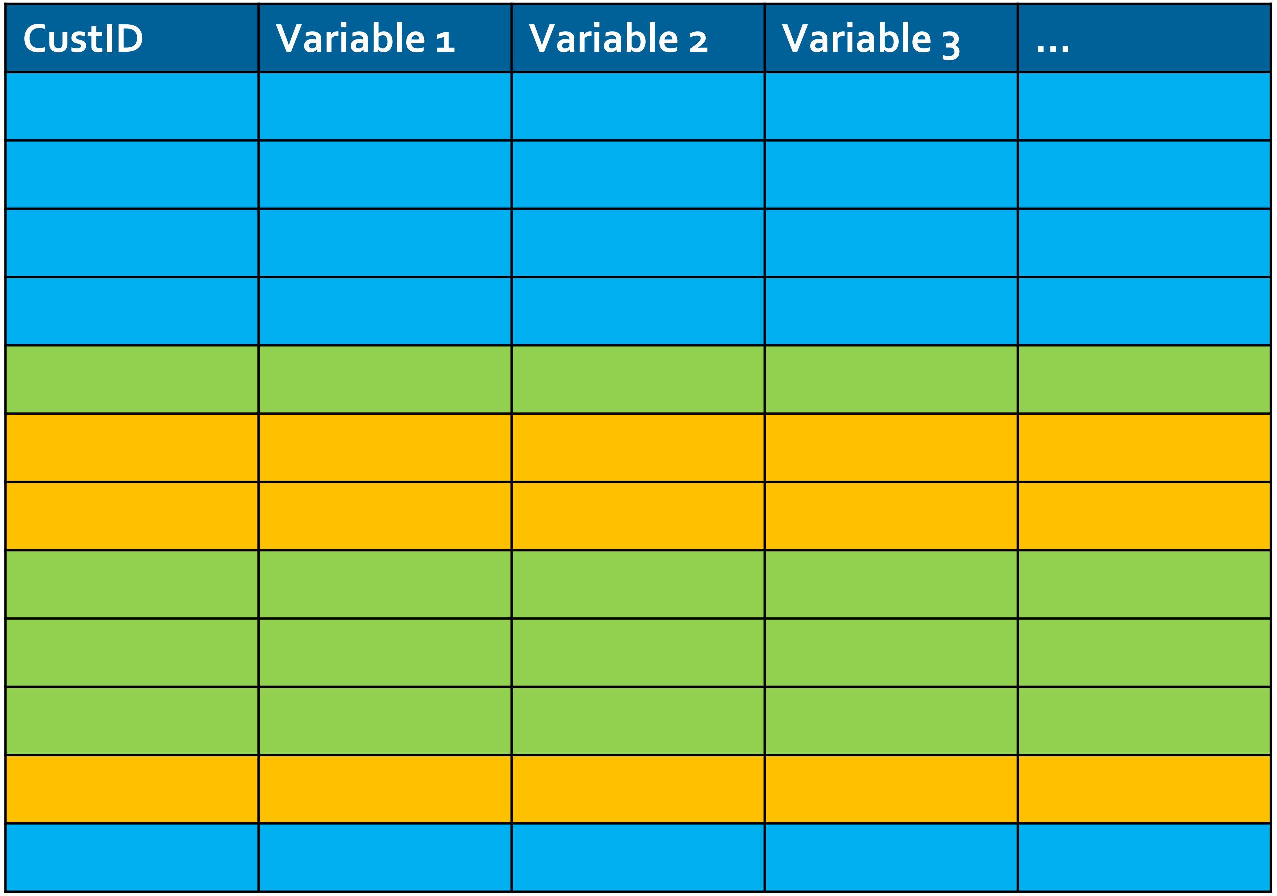
Segmentation

Cluster Analysis

Data-Driven Segmentation

CustID	Variable 1	Variable 2	Variable 3	…

Data-Driven Segmentation



CustID	Variable 1	Variable 2	Variable 3	…
🟦	🟦	🟦	🟦	🟦
🟦	🟦	🟦	🟦	🟦
🟦	🟦	🟦	🟦	🟦
🟦	🟦	🟦	🟦	🟦
🟩	🟩	🟩	🟩	🟩
🟨	🟨	🟨	🟨	🟨
🟨	🟨	🟨	🟨	🟨
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟨	🟨	🟨	🟨	🟨
🟦	🟦	🟦	🟦	🟦

Segment 1

Segment 2

Segment 3

Basic idea:
use the columns to
group the rows

Data-Driven Segmentation: How? Cluster Analysis

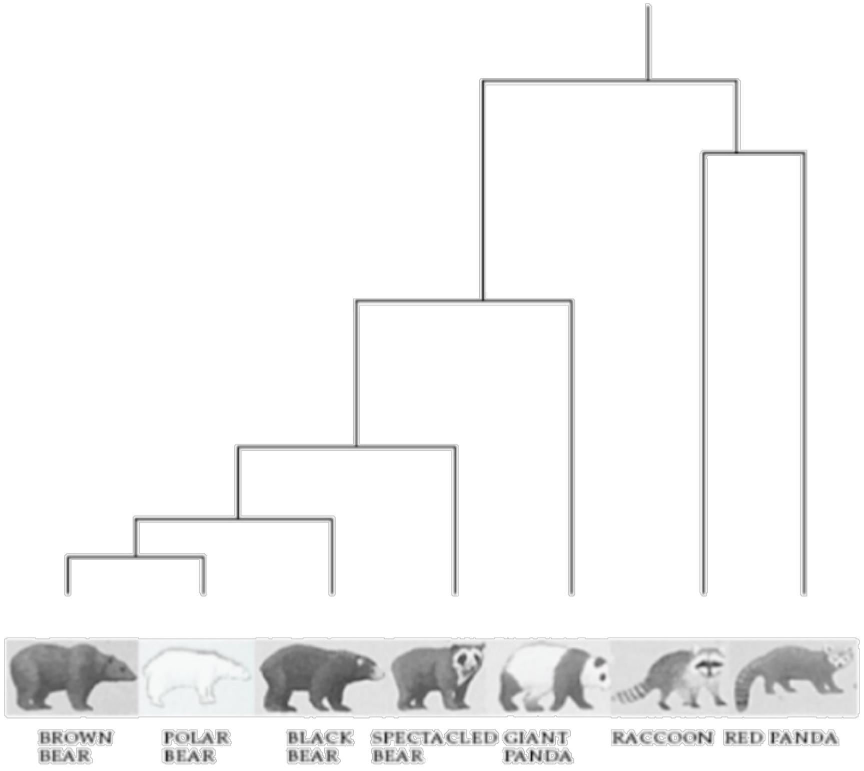
- A class of techniques used to divide objects into groups
  - Objects within a group should be as similar as possible
  - Objects belonging to different groups should be as dissimilar as possible

- Two very commonly used techniques:
  - Hierarchical Clustering
  - K-Means

Hierarchical Clustering

Hierarchical Clustering

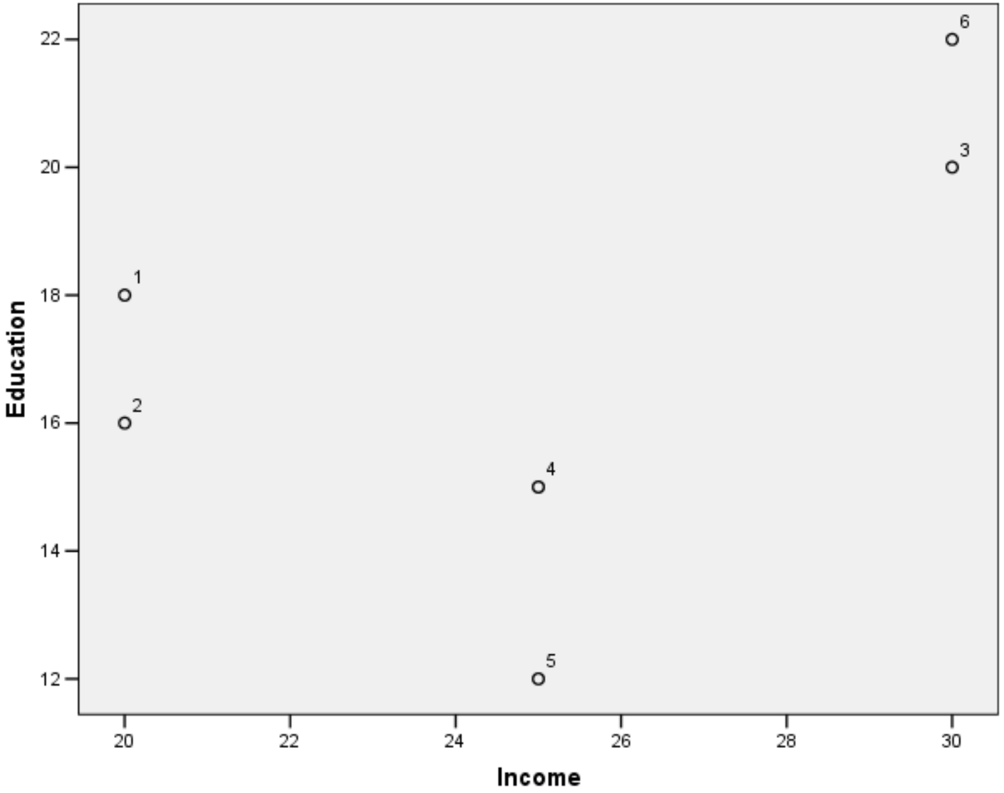
- Goal: Recursively group entities based on how similar they are
- Each entity starts in its own cluster
- All the clusters are then grouped to form bigger and bigger clusters
- Typically, visualized through a dendrogram



Example: Demographic Segmentation

ID	Income in $K	Education in Yrs
1	20	18
2	20	16
3	30	20
4	25	15
5	25	12
6	30	22

Data Plot



We need to measure the distance between each point

Euclidean Distance

- $D_{12} = \sqrt{(20-20)^2+(18-16)^2}=2$

ID	Income in $K	Education in Yrs
1	20	18
2	20	16
3	30	20
4	25	15
5	25	12
6	30	22

The Distance Matrix: D

ID	1	2	3	4	5	6
1	0	2.0	10.2	5.8	7.8	10.8
2		0	10.8	5.1	6.4	11.6
3			0	7.1	9.4	2.0
4				0	3.0	8.6
5					0	11.2
6						0

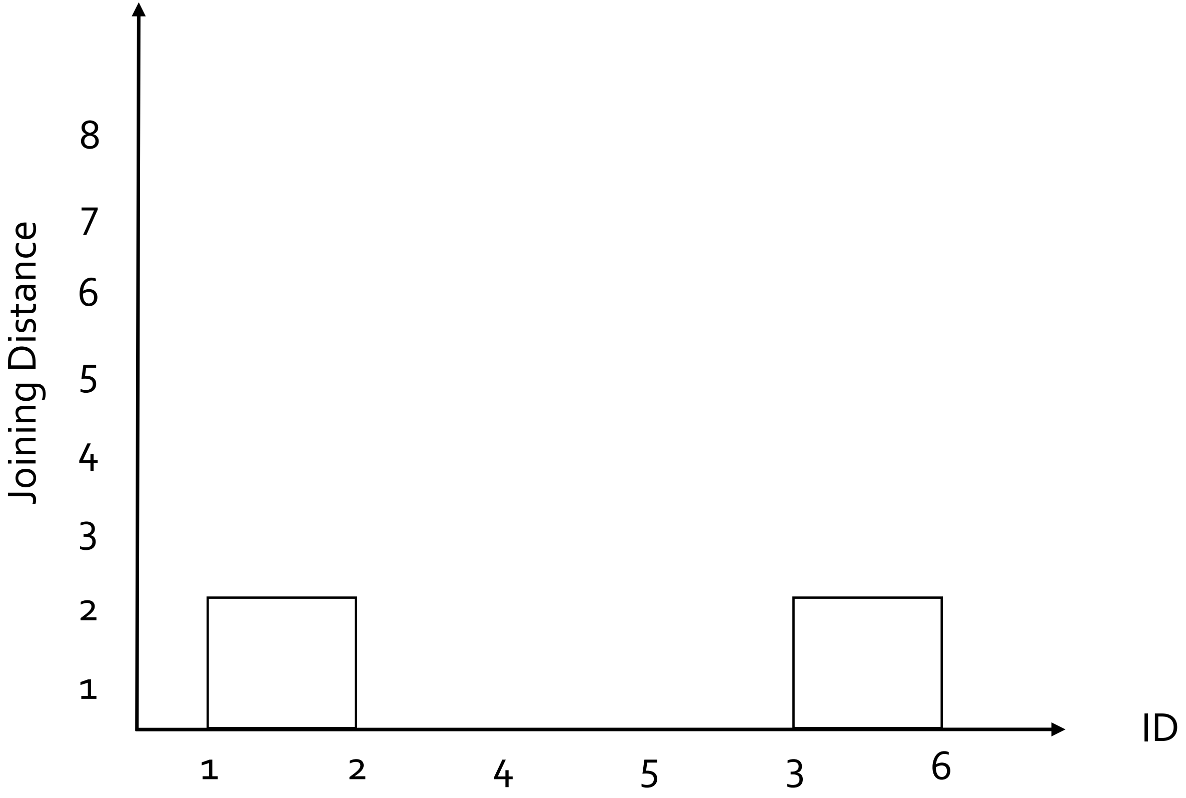
Hierarchical Clustering Algorithm

STEP 1: Select Min $\{D_{ij}\}$ and join i and j at that distance

- $D_{12} = 2.0$ → join subjects 1 and 2 in one group (cluster)
- $D_{36} = 2.0$ → join 3 and 6 in another cluster

ID	1	2	3	4	5	6
1	0	2.0	10.2	5.8	7.8	10.8
2		0	10.8	5.1	6.4	11.6
3			0	7.1	9.4	2.0
4				0	3.0	8.6
5					0	11.2
6						0

Dendrogram Construction



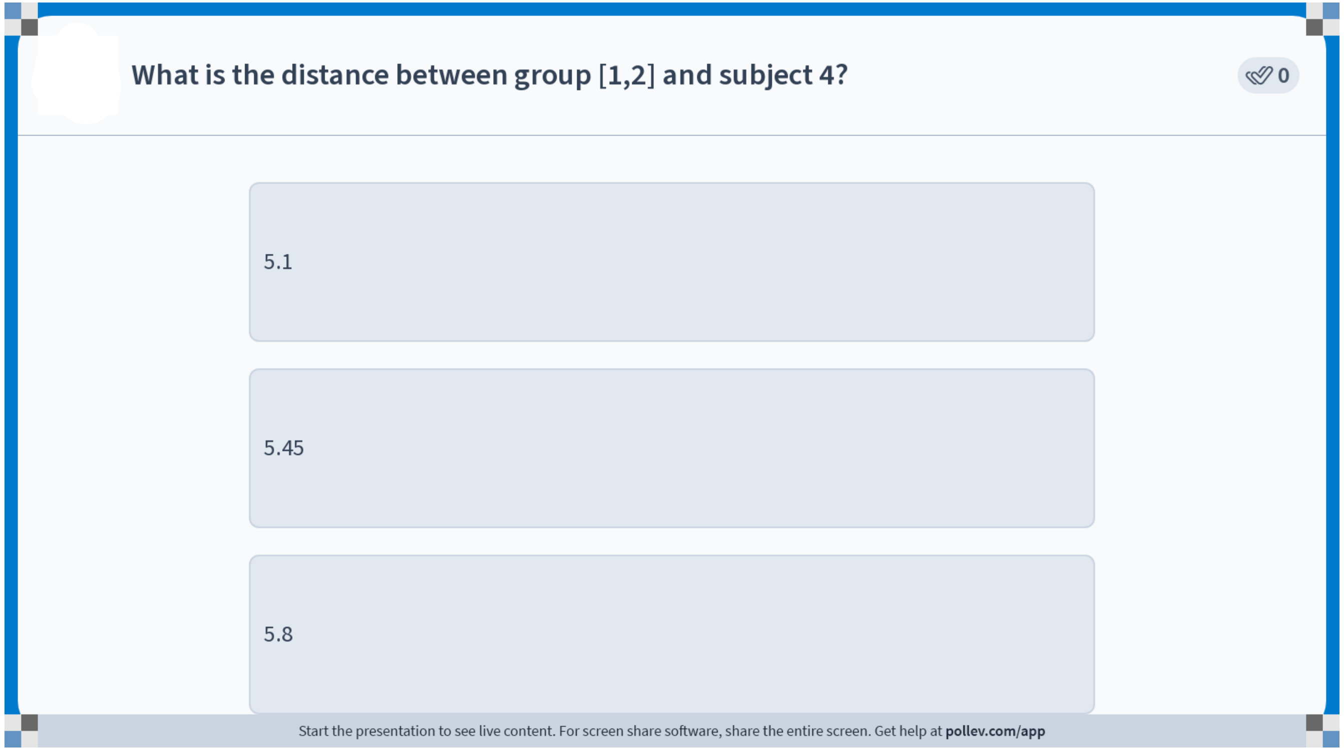
Step 2: Update Distance Matrix

What is the distance between group [1,2] and subject 4?

	[1,2]	[3,6]	4	5
[1,2]	0			
[3,6]		0		
4			0	3.0
5				0

Poll Title: What is the distance between group [1,2] and subject 4?

ID	1	2	3	4	5	6
1	0	2.0	10.2	5.8	7.8	10.8
2		0	10.8	5.1	6.4	11.6
3			0	7.1	9.4	2.0
4				0	3.0	8.6
5					0	11.2
6						0



What is the distance between group [1,2] and subject 4?

5.1

5.45

5.8

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

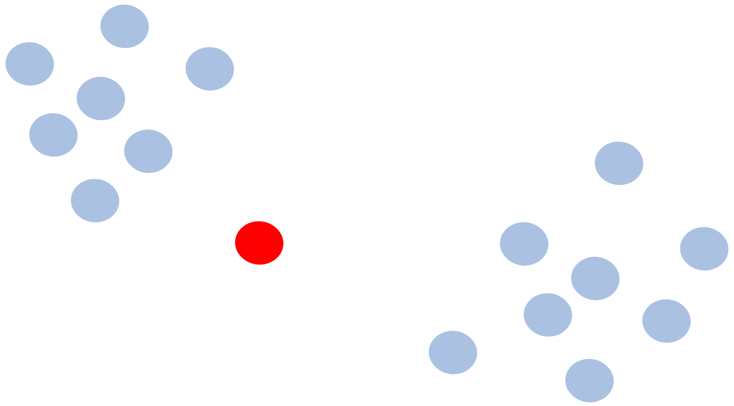
Linkage Rules

- Minimum (single) linkage

- Average linkage

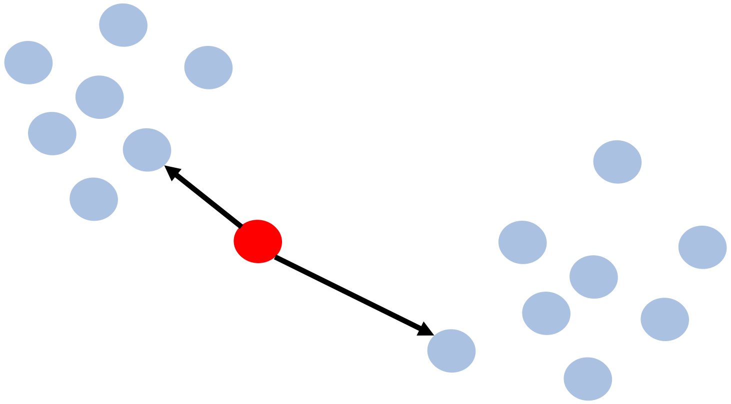
- Maximum (complete) linkage

- Ward linkage



Minimum (Single) Linkage

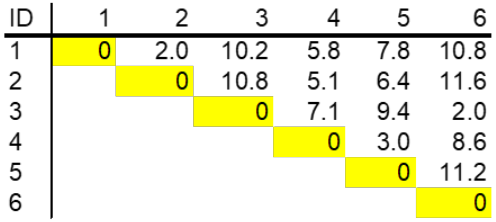
We compare the point to the closest point in each cluster



Updated Distance Matrix Using Min Linkage

What is the distance between group [1,2] and subject 4?

$\min(D_{14} = 5.8, D_{24} = 5.1)=5.1$

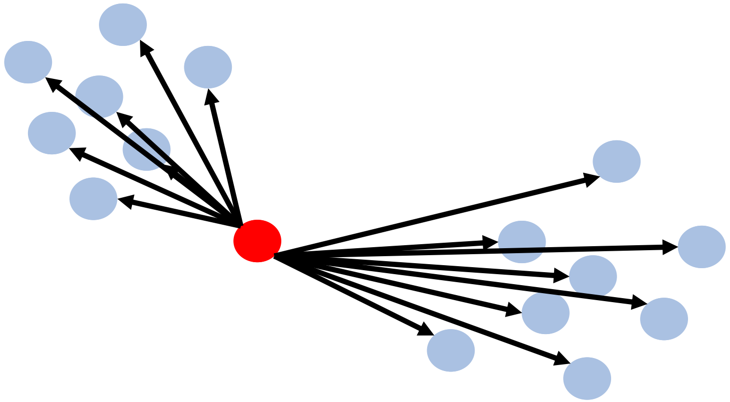


ID	1	2	3	4	5	6
1	0	2.0	10.2	5.8	7.8	10.8
2		0	10.8	5.1	6.4	11.6
3			0	7.1	9.4	2.0
4				0	3.0	8.6
5					0	11.2
6						0

	[1,2]	[3,6]	4	5
[1,2]	0		5.1	
[3,6]		0		
4			0	3.0
5				0

Average Linkage

We average the distance over all pairs of points between two clusters



Updated Distance Matrix Using Average Linkage

What is the distance between group [1,2] and subject 4?

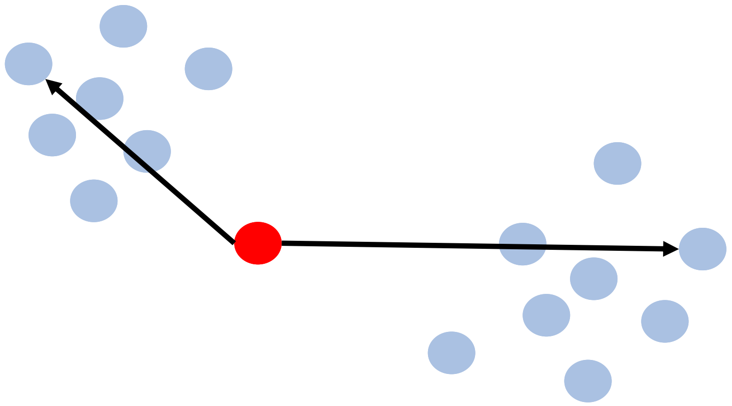
$\text{ave}(D_{14} = 5.8,\ D_{24} = 5.1) = 5.45$

ID	Income in $K	Education in Yrs
1	20	18
2	20	16
3	30	20
4	25	15
5	25	12
6	30	22

	[1,2]	[3,6]	4	5
[1,2]	0		5.45	
[3,6]		0		
4			0	3.0
5				0

Maximum (Complete) Linkage

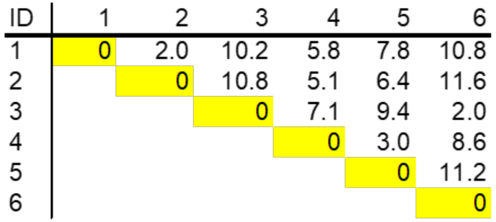
We compare the point to the furthest point in each cluster



Updated Distance Matrix Using Max Linkage

What is the distance between group [1,2] and subject 4?

$\max(D_{14} = 5.8, D_{24} = 5.1)=5.8$



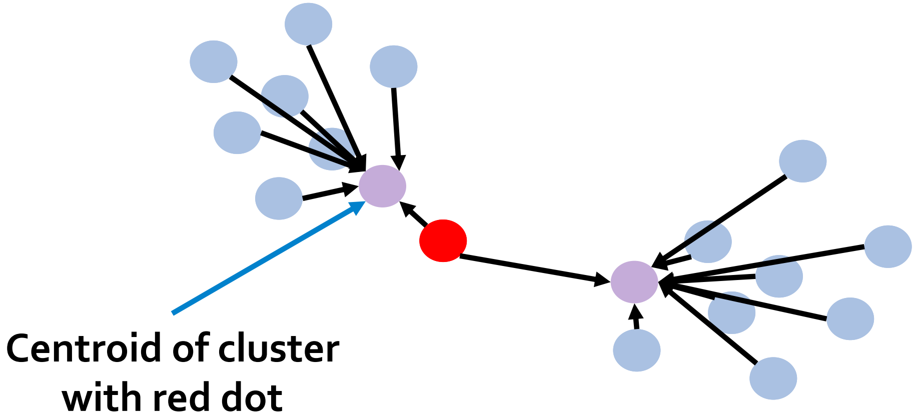
ID	1	2	3	4	5	6
1	0	2.0	10.2	5.8	7.8	10.8
2		0	10.8	5.1	6.4	11.6
3			0	7.1	9.4	2.0
4				0	3.0	8.6
5					0	11.2
6						0

	[1,2]	[3,6]	4	5
[1,2]	0		5.8	
[3,6]		0		
4			0	3.0
5				0

Ward Linkage

We minimize the within-cluster variance

- Add [4] to [1,2] to form cluster [1,2,4]
- Distance = variance of [1,2,4] – (variance of [1,2] + variance of [4])



Centroid of cluster with red dot

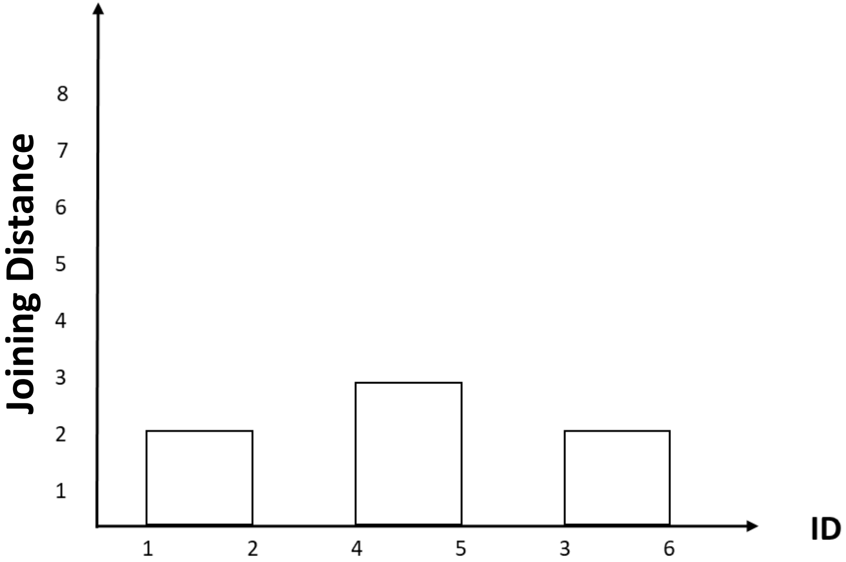
Updated Distance Matrix Using Min Linkage

	[1,2]	[3,6]	4	5
[1,2]	0	10.2	5.1	6.4
[3,6]		0	7.1	9.4
4			0	3.0
5				0

Step 3: Pick Min D_{ij} and Join i and j

D_{45} = 3.0 → 4 and 5 are joined

	[1,2]	[3,6]	4	5
[1,2]	0	10.2	5.1	6.4
[3,6]		0	7.1	9.4
4			0	3.0
5				0



Step 3: Pick Min D_{ij} and Join i and j

What is the distance between [1,2] and [4,5]?

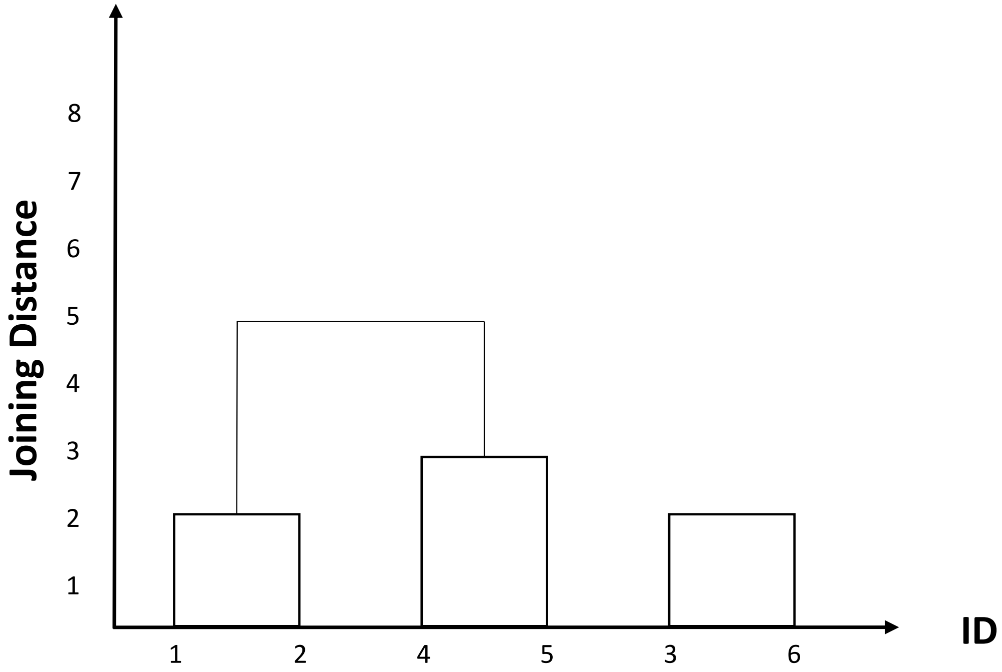
	[1,2]	[3,6]	[4,5]
[1,2]	0	10.2	
[3,6]		0	7.1
[4,5]			0

ID	1	2	3	4	5	6
1	0	2.0	10.2	5.8	7.8	10.8
2		0	10.8	5.1	6.4	11.6
3			0	7.1	9.4	2.0
4				0	3.0	8.6
5					0	11.2
6						0

- Solution: Min($D_{1,4}$, $D_{2,4}$, $D_{1,5}$, $D_{2,5}$) = 5.1

Step 4: Update Distance Matrix as in Step 2 and pick Min D_{ij} and Join i and j

	[1,2]	[3,6]	[4,5]
[1,2]	0	10.2	5.1
[3,6]		0	7.1
[4,5]			0

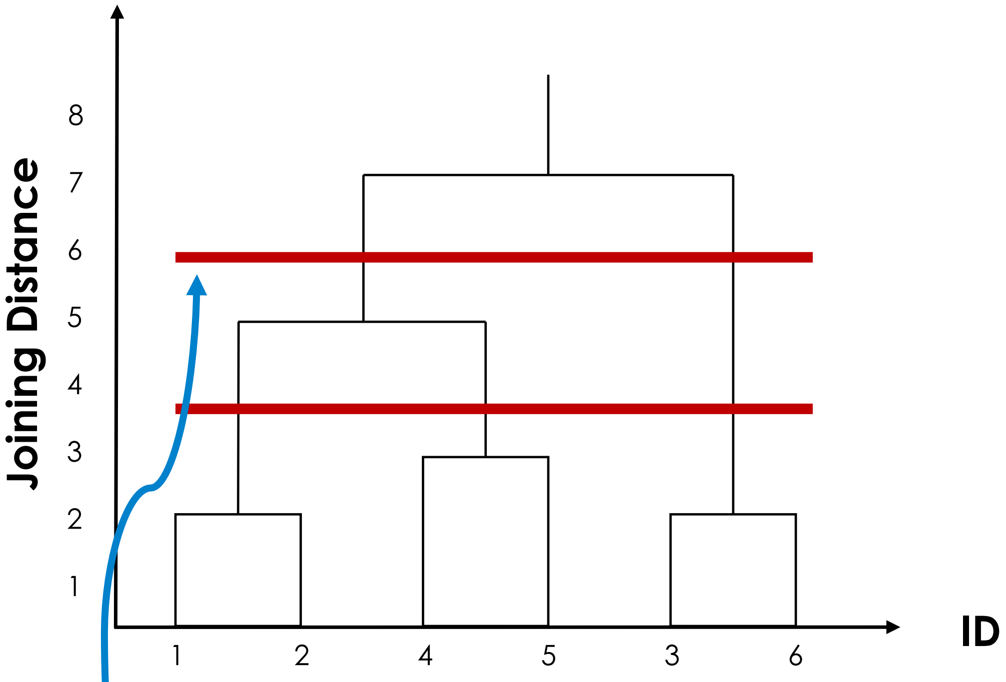


Step 5: Update Distance Matrix

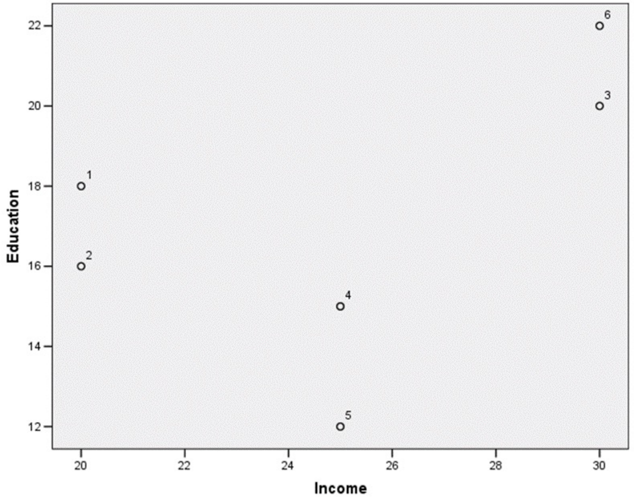
Join [1,2,4,5] and [3,6] and Stop

	[1,2,4,5]	[3,6]
[1,2,4,5]	0	7.1
[3,6]		0

Final Dendrogram



Clusters = how many branches a horizontal line intersects



How many segments?

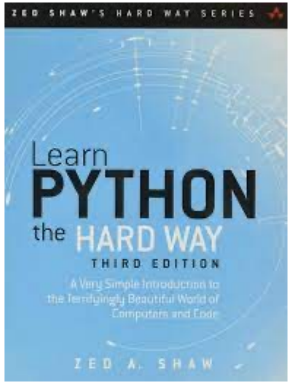
- Decided through threshold decision
- **Judgment**

Let’s go to Python

Hierarchical Clustering

Python Practice

- Why Python?
  - What's most frequently used in industry
  - Easy to read
  - Fast
  - Integrates well with other programs
  - Large community – support, many packages



- Go to https://colab.research.google.com/

- Download B9651_Segmentation_Fall2024_Share.ipynb and coffee_data.csv from Canvas

Coffee Data

There are 282 individuals in the data. For the clustering, we will focus on variables A1-A5, which are answers to the following five questions:

Rate the following on a scale from Strongly Disagree (0) to Strongly Agree (8):

* Q1: I pay close attention to the origin and sourcing of my coffee.

* Q2: I do my best work at coffee shops.

* Q3: A good coffee shop has free Wi-fi.

* Q4: Good food is important in a coffee shop.

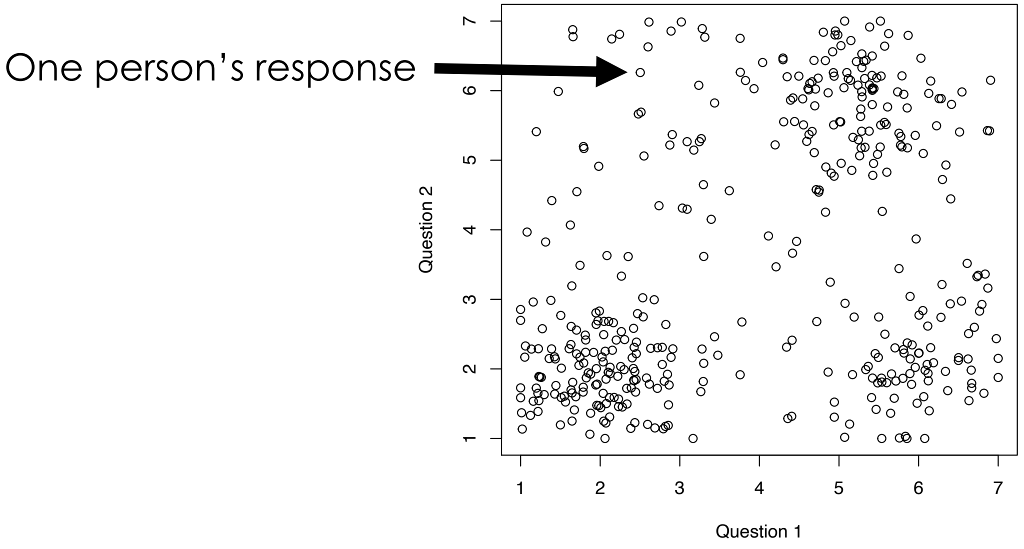
* Q5: I enjoy drinking espresso.

Coffee Data

Rate the following on a scale from Strongly Disagree to Strongly Agree:

Question 1: I pay close attention to the origin and sourcing of my coffee.

Question 2: I do my best work at coffee shops.



One person's response

Summary: Hierarchical Cluster Analysis

- A numerical procedure which attempts to separate a set of observations into groups/clusters
- Members of the same group/cluster are more similar than members of different clusters
- Agglomerative – seeks to join objects sequentially until gets one large cluster
- Obtains a tree or "dendrogram" representation
- Very popular technique!

K-Means Clustering

Approach

Cluster Analysis: Coffee Example

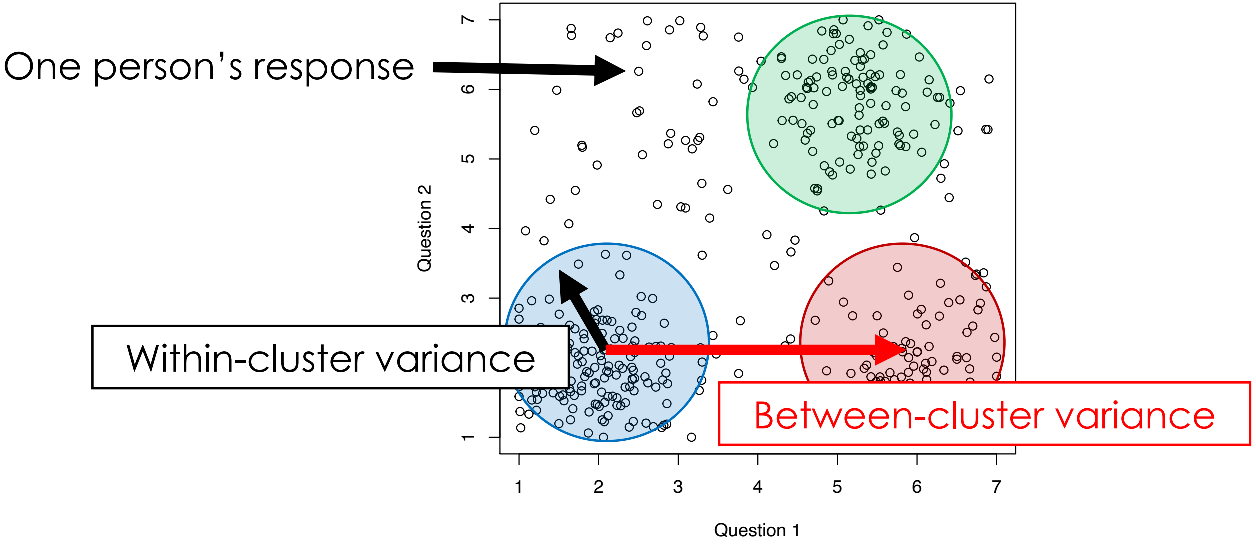
Find groups of data points that look the **same within** groups, and **distinct across** groups

Statistically: minimize **within-cluster variance**, maximize **between-cluster variance**

Rate the following on a scale from Strongly Disagree to Strongly Agree:

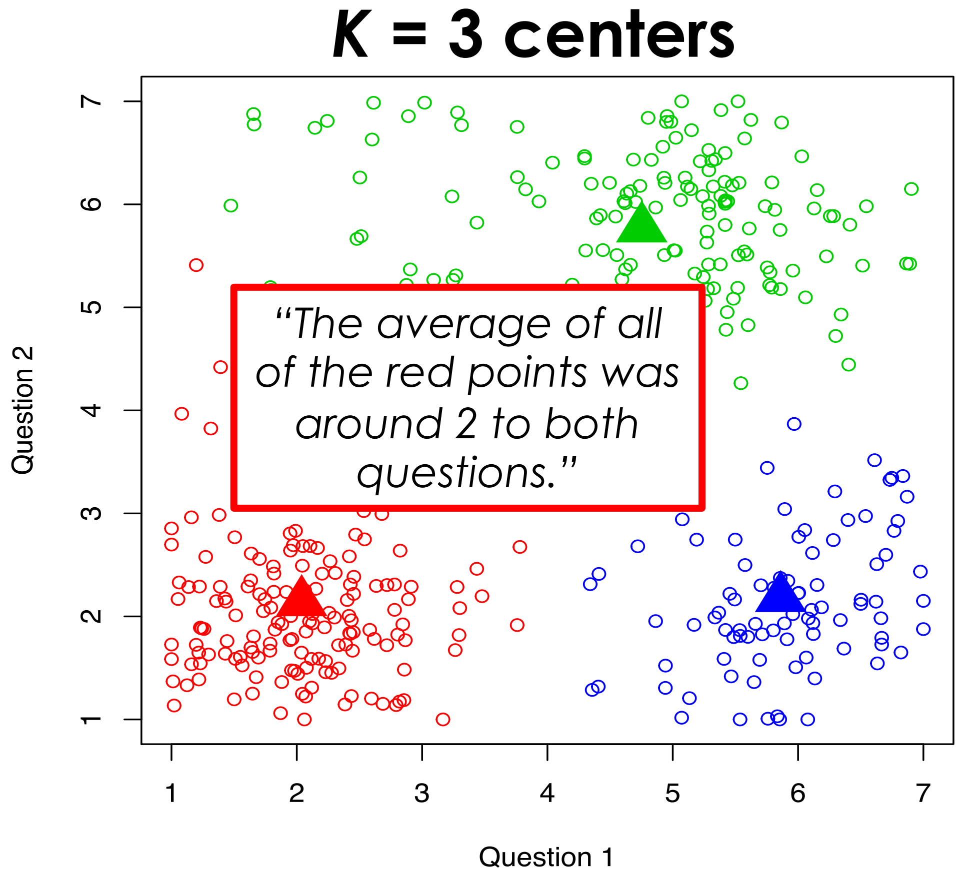
Question 1: I pay close attention to the origin and sourcing of my coffee.

Question 2: I do my best work at coffee shops.



K-means Clustering

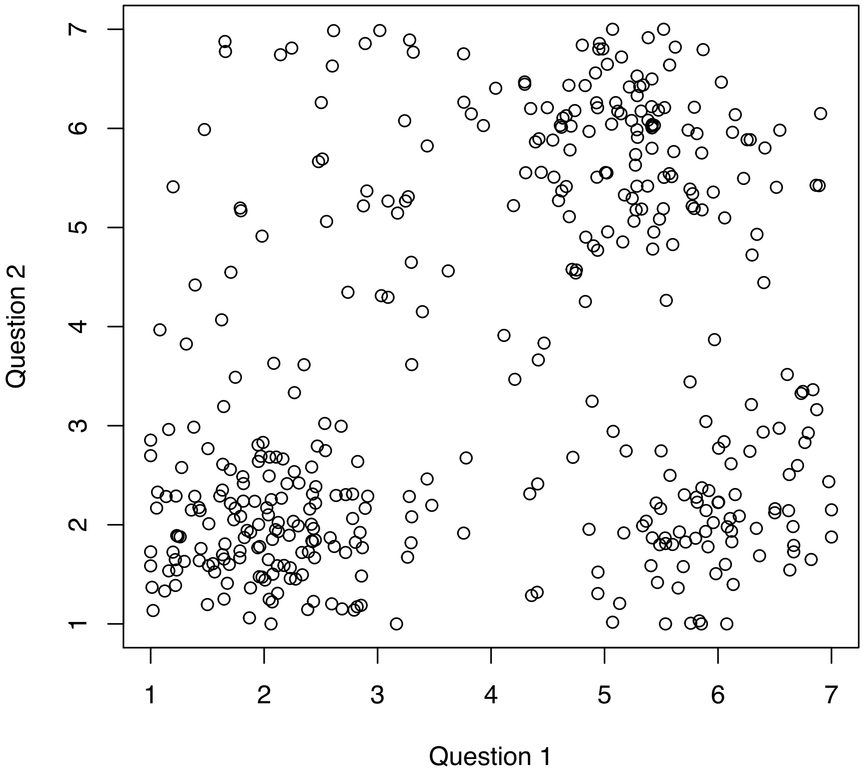
- Most widely used clustering algorithm
- Goal: summarize all the data using a set of *K* **centers** (**centroids**)
- Each observation is assigned to the nearest center
- Requires you to pre-specify the number of clusters, K



***K* = 3 centers**

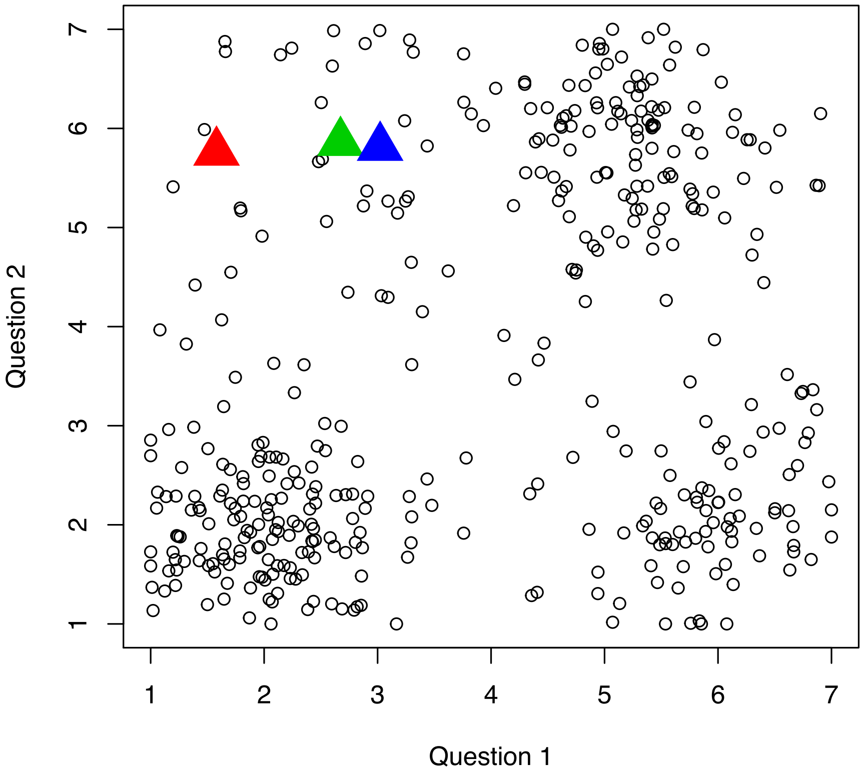
"The average of all of the red points was around 2 to both questions."

Visualizing K-means



Visualizing K-means

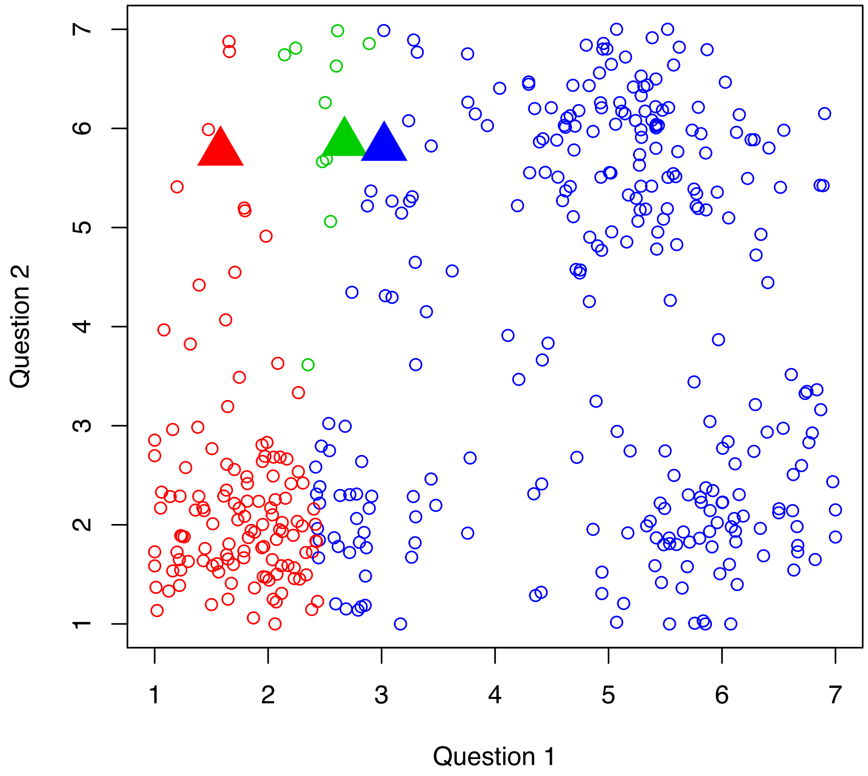
Step 1: Initialize centroids



Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

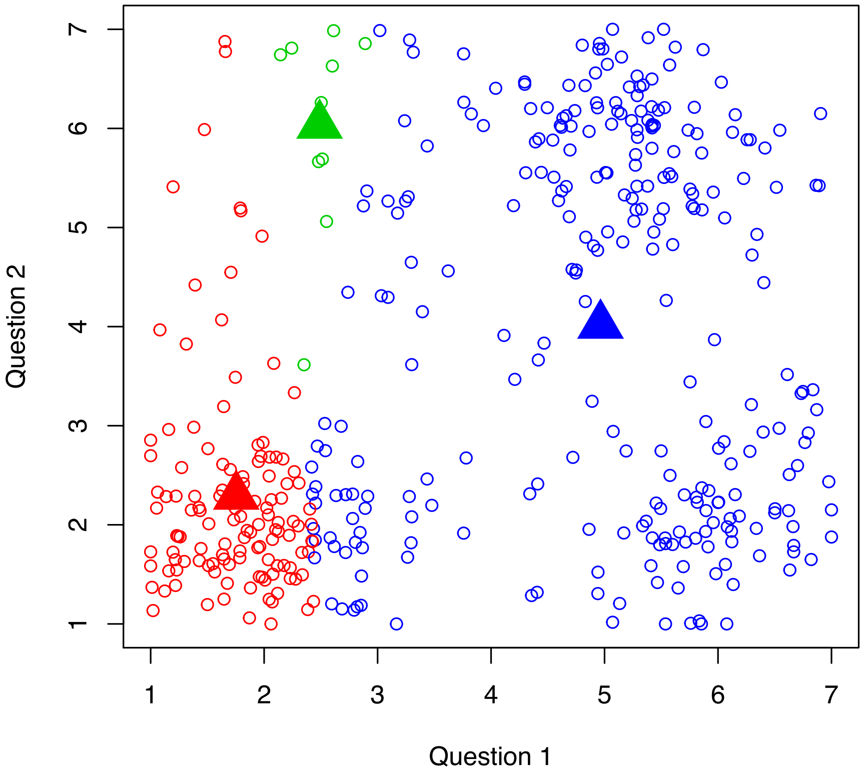


Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers

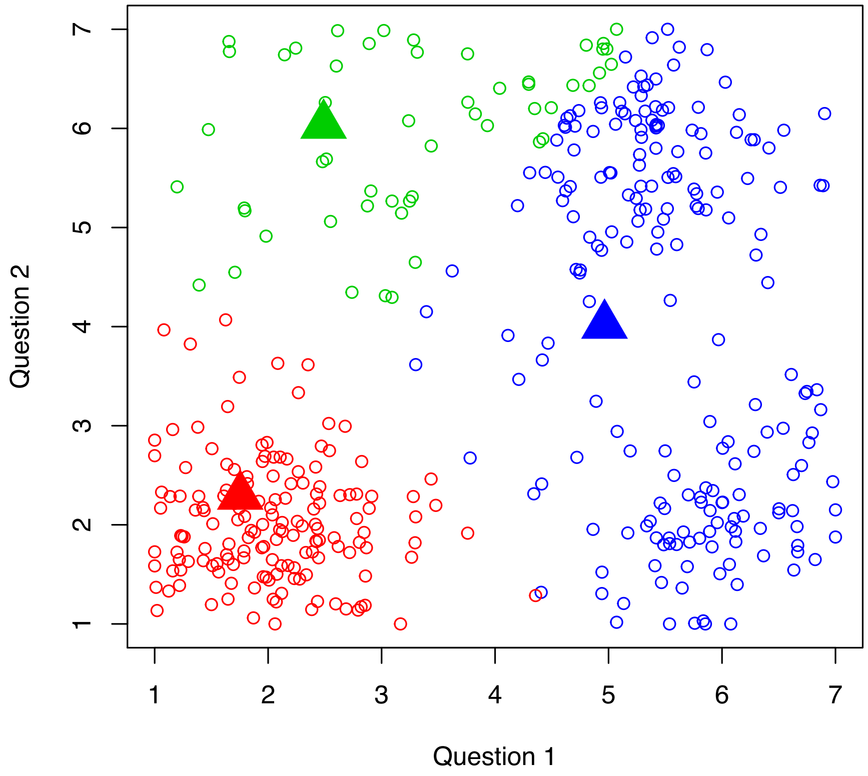


Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers

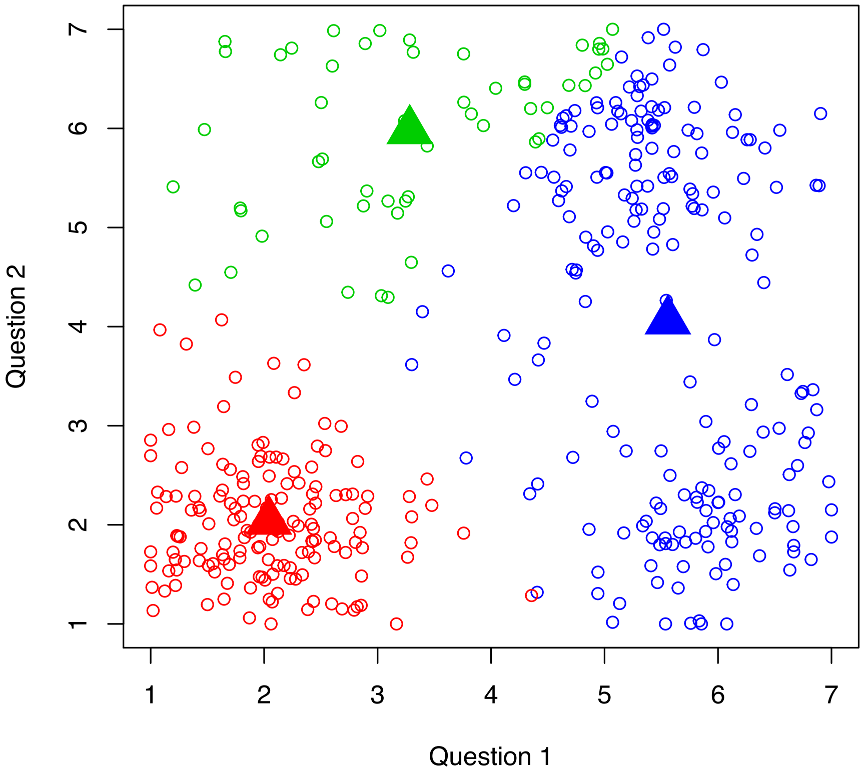


Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers

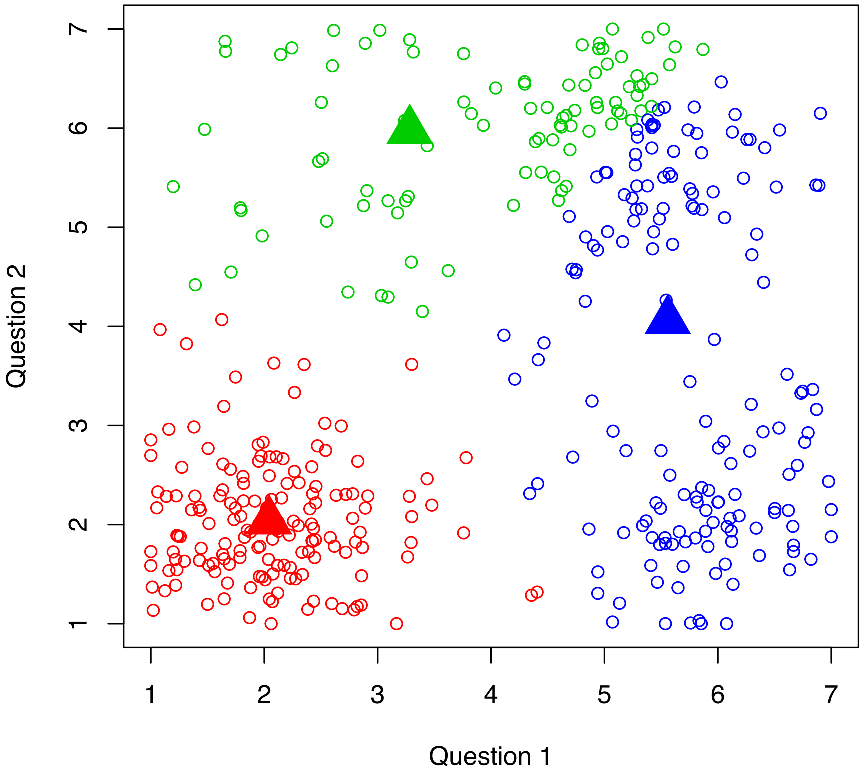


Visualizing K-means

Step 1: Initialize centroids

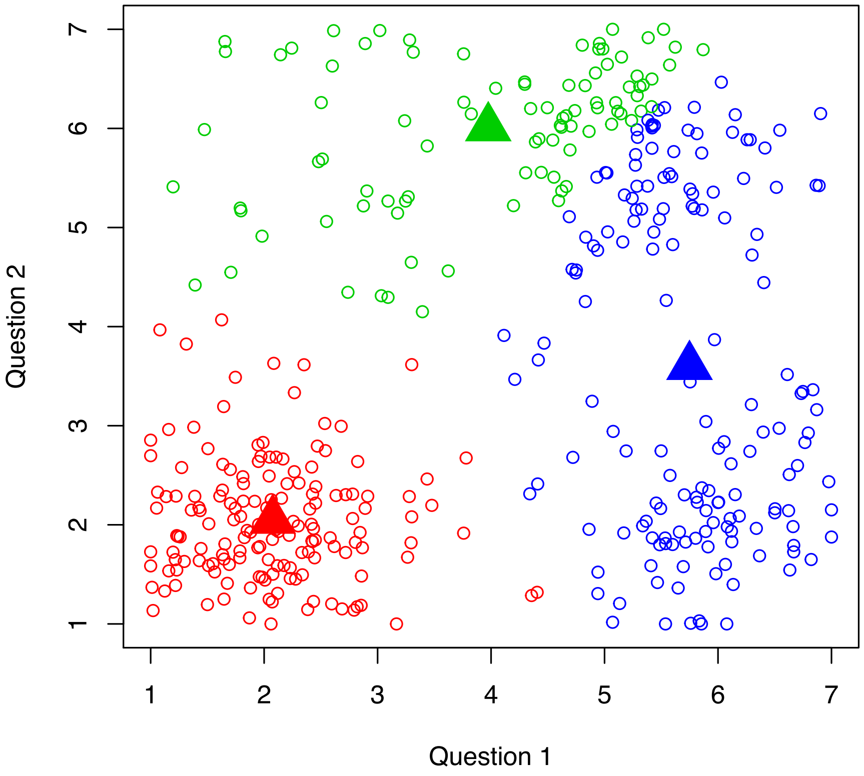
Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers



Visualizing K-means

Step 1: Initialize centroids
Step 2: Assign points (observations) to the nearest centroid
Step 3: Re-compute centers

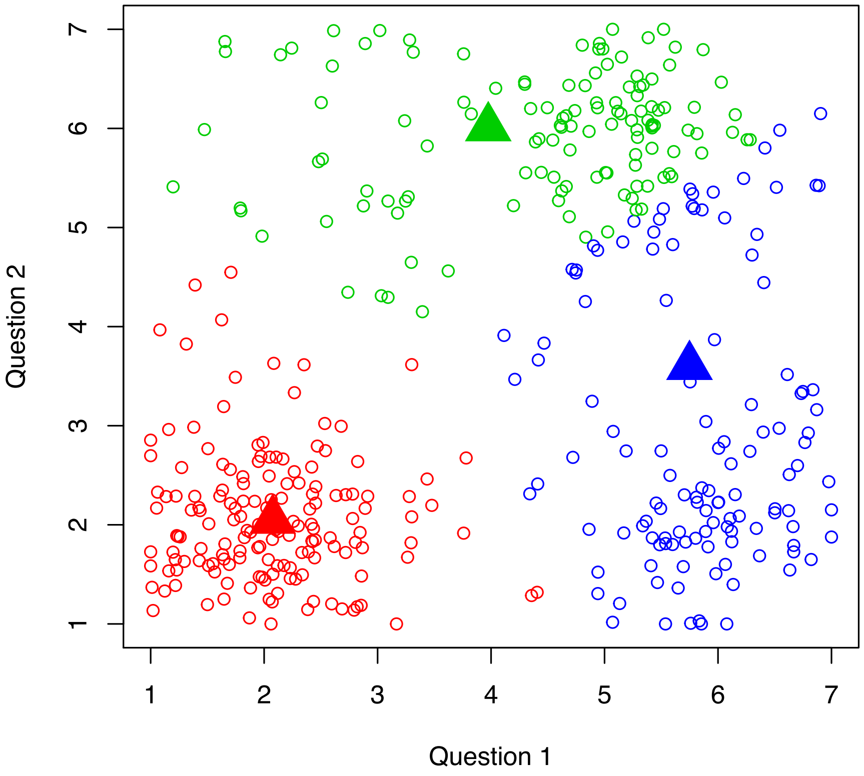


Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers

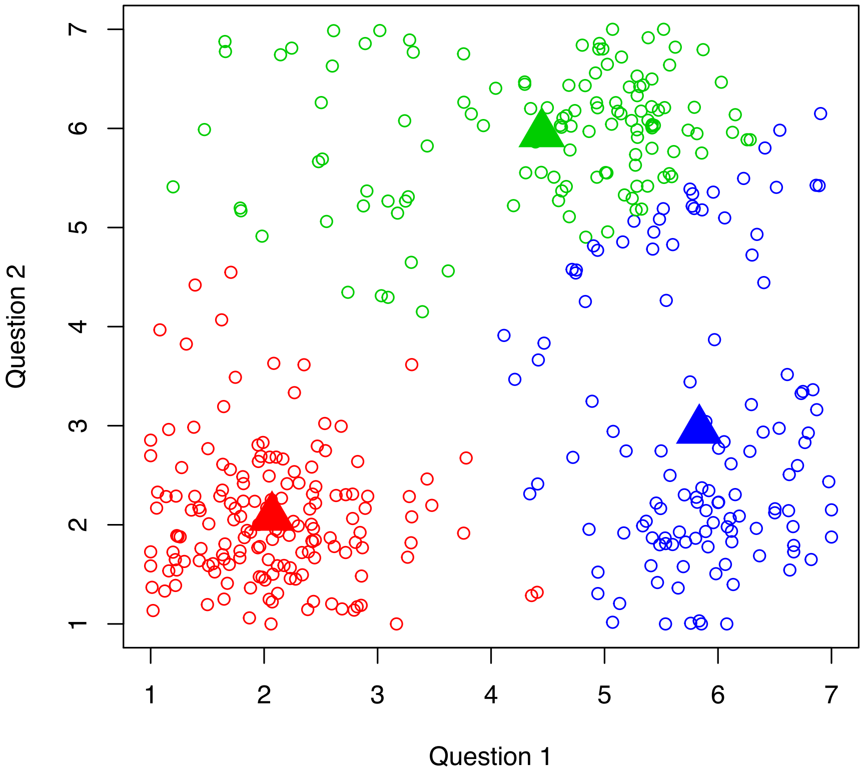


Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers

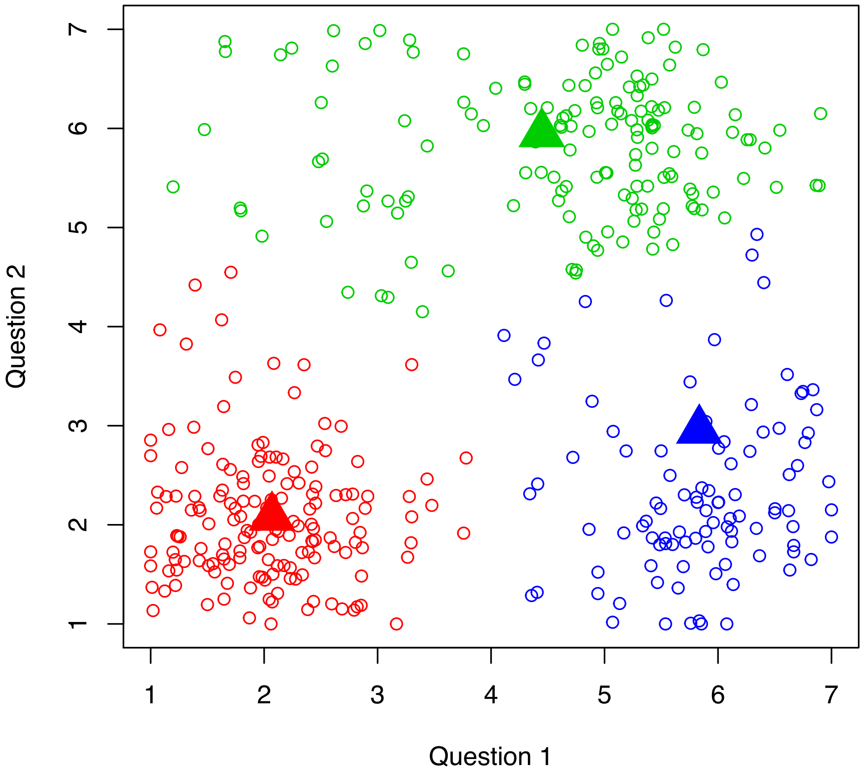


Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers

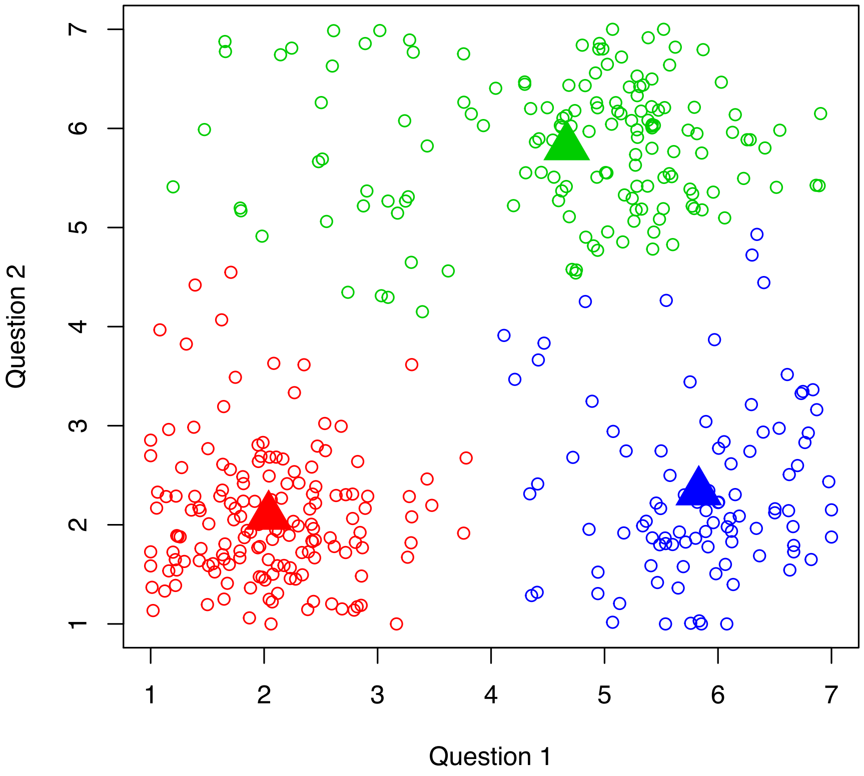


Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers

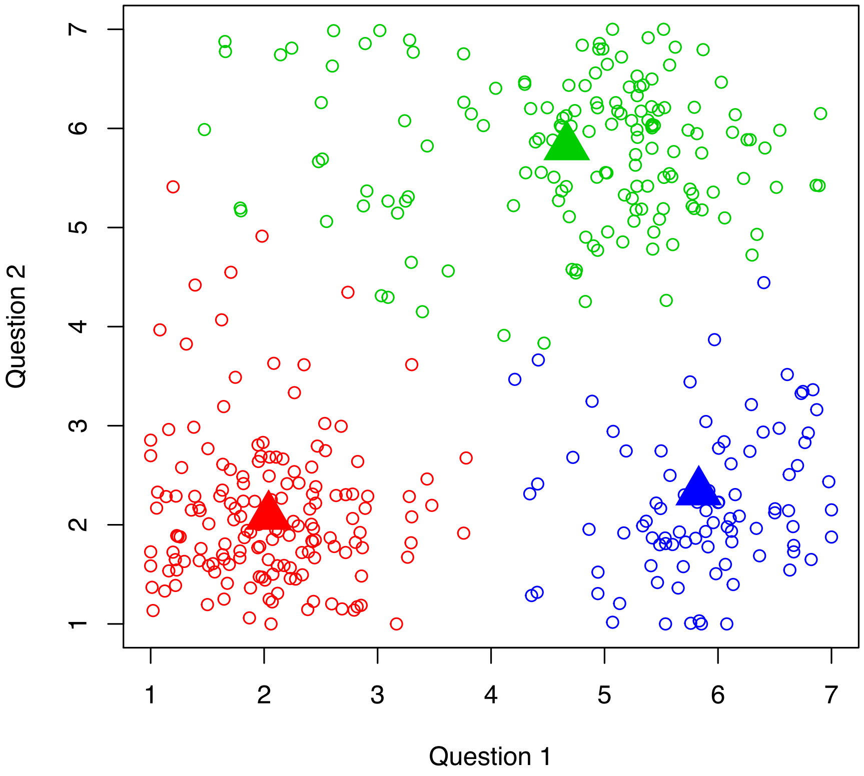


Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers

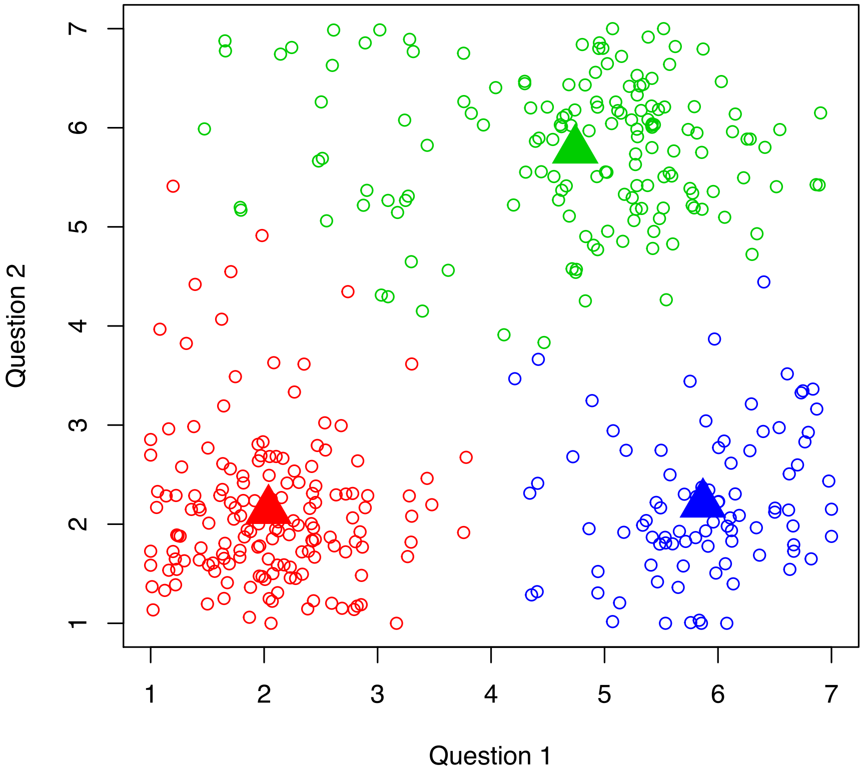


Visualizing K-means

Step 1: Initialize centroids

Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers

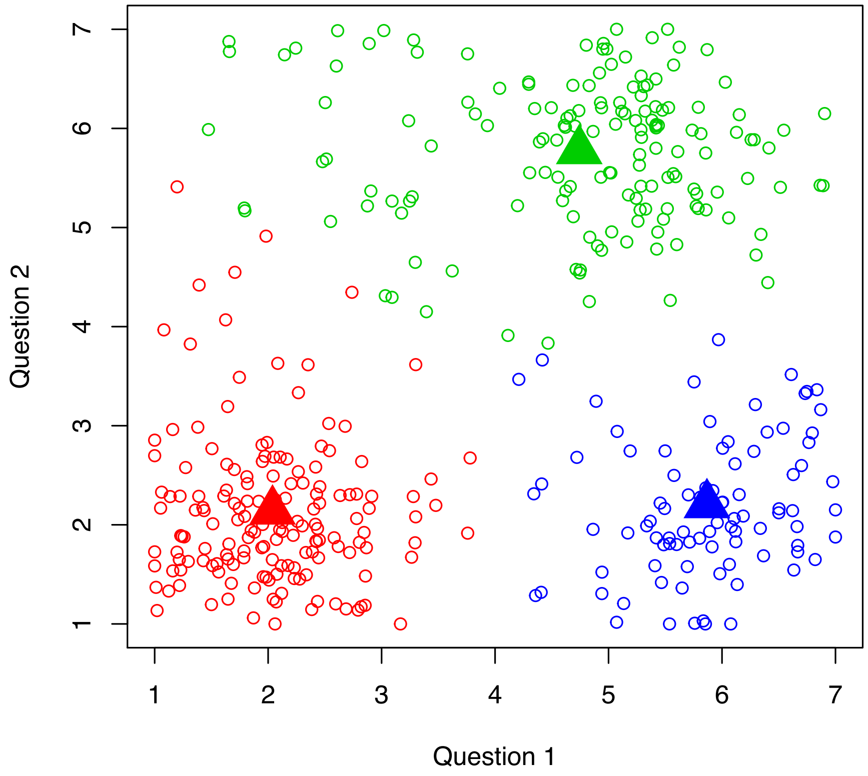


Visualizing K-means

Step 1: Initialize centroids

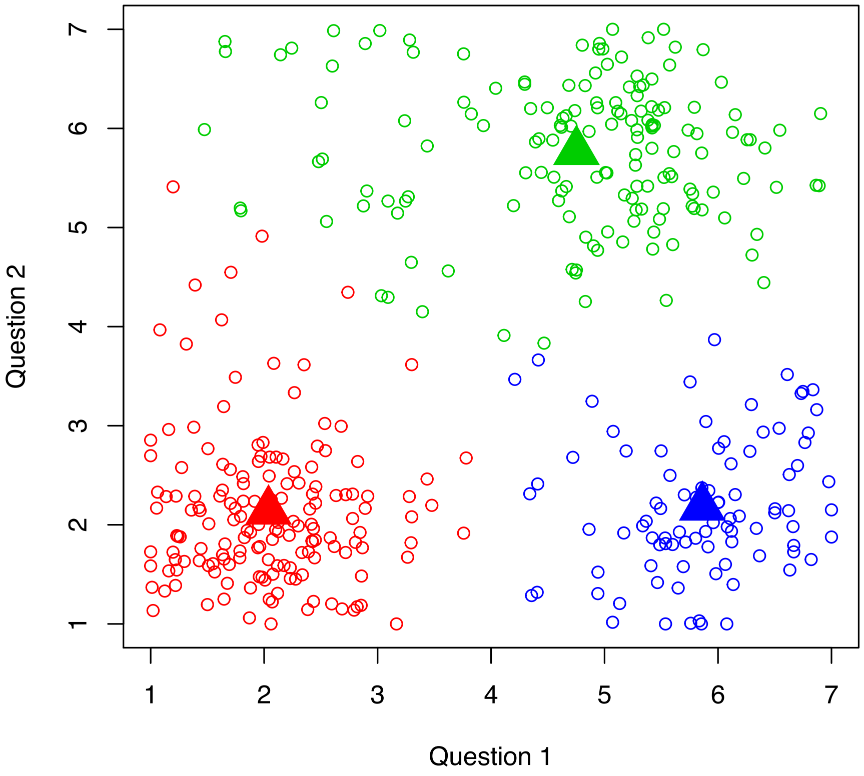
Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers



Visualizing K-means

Step 1: Initialize centroids
Step 2: Assign points (observations) to the nearest centroid
Step 3: Re-compute centers

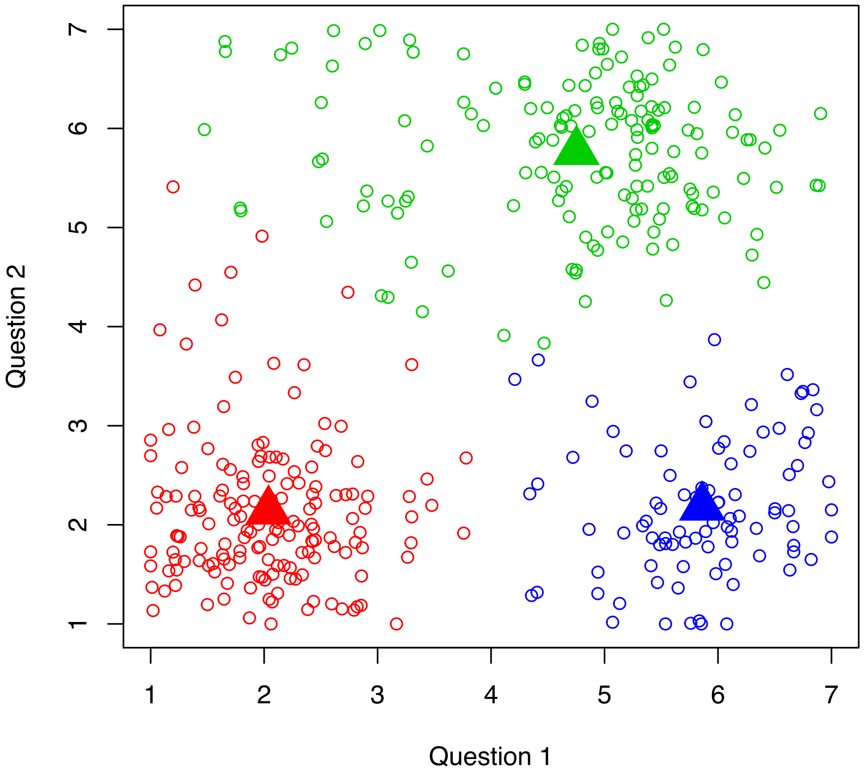


Visualizing K-means

Step 1: Initialize centroids

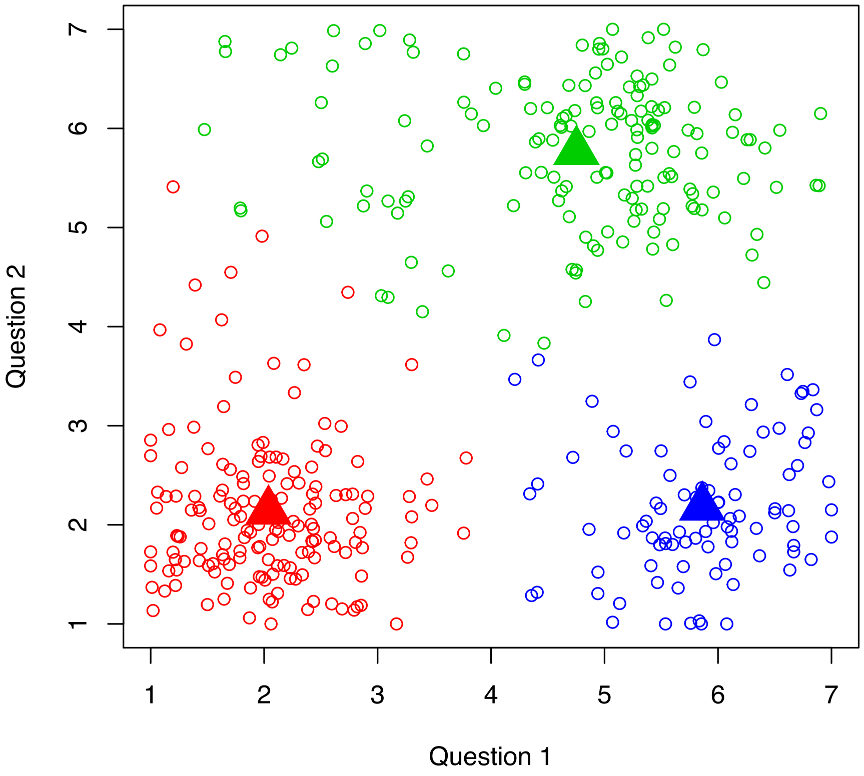
Step 2: Assign points (observations) to the nearest centroid

Step 3: Re-compute centers



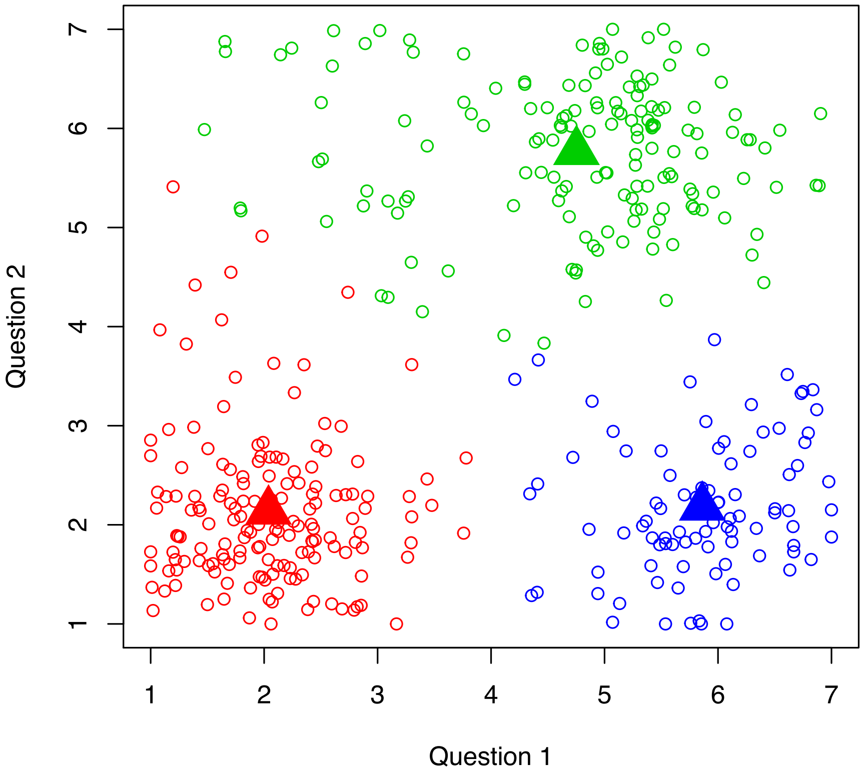
Visualizing K-means

Step 1: Initialize centroids
Step 2: Assign points (observations) to the nearest centroid
Step 3: Re-compute centers



Visualizing K-means

Step 1: Initialize centroids
Step 2: Assign points (observations) to the nearest centroid
Step 3: Re-compute centers
Stop when no change.



K-Means Clustering

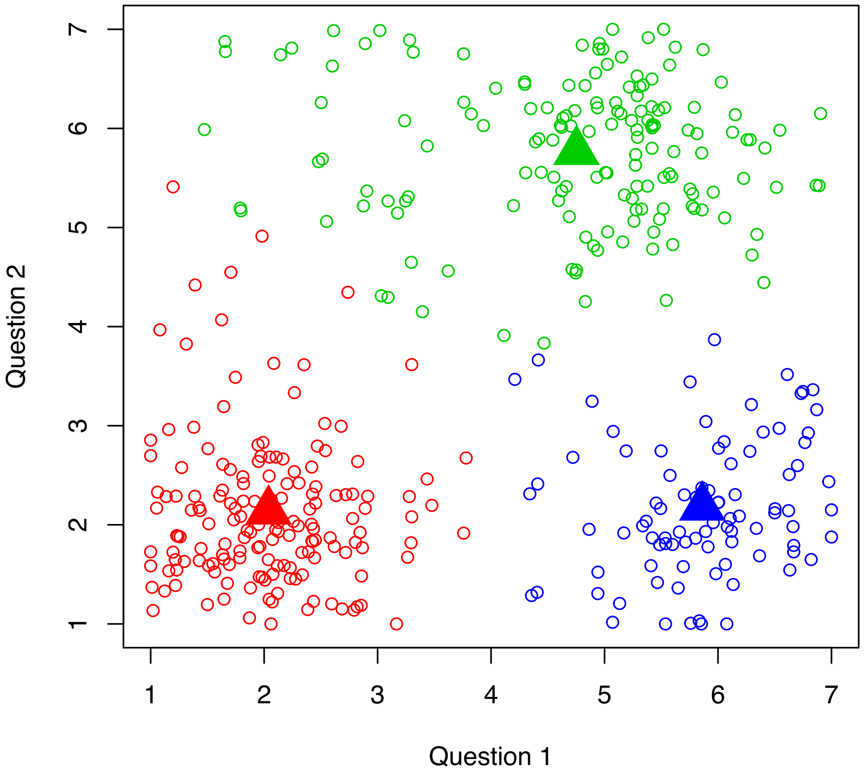
Warnings

Warning 1: Initialization

Step 1: Initialize centroids
Step 2: Assign points (observations) to the nearest centroid
Step 3: Re-compute centers

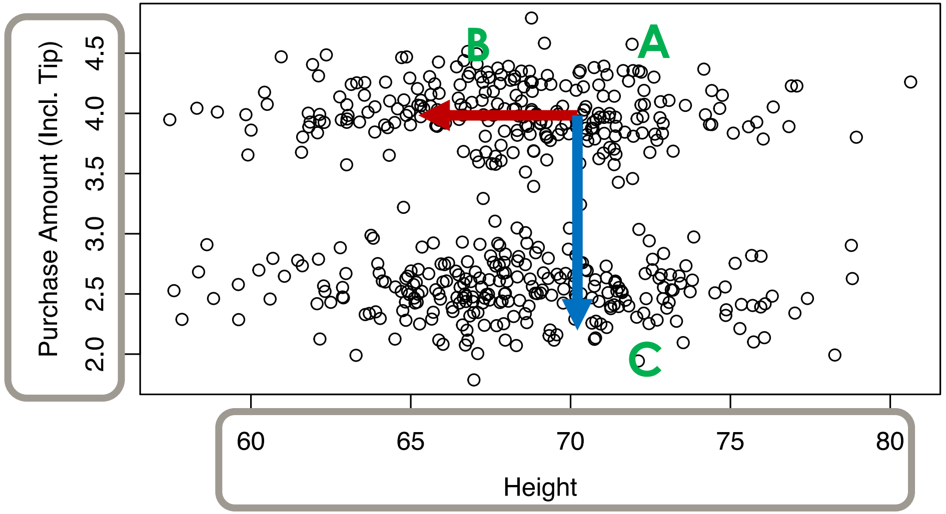
Warning!
The end result depends on the initialization! You will get different results each time you run k-means.
→ *Always run it more than once!*

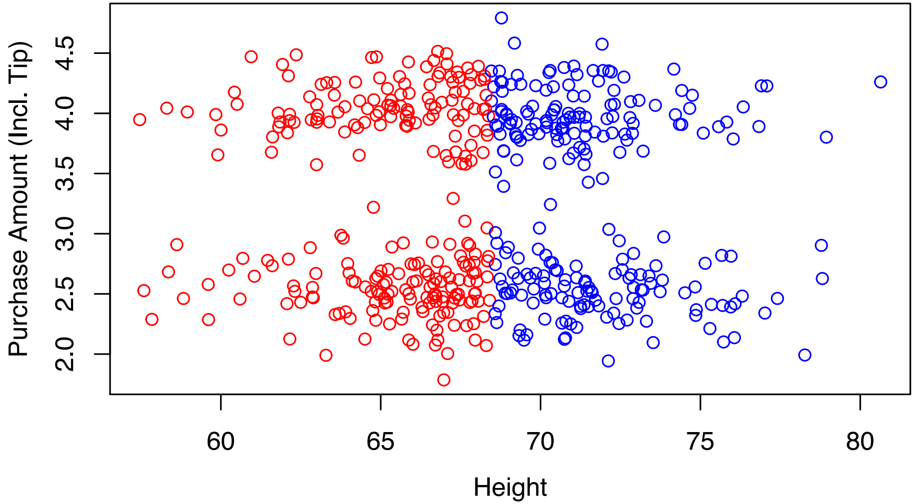
Link: kmeans_initialization



Warning 2: Scaling of Inputs

- K-means relies on computing distances: if columns have different units, results may not make sense





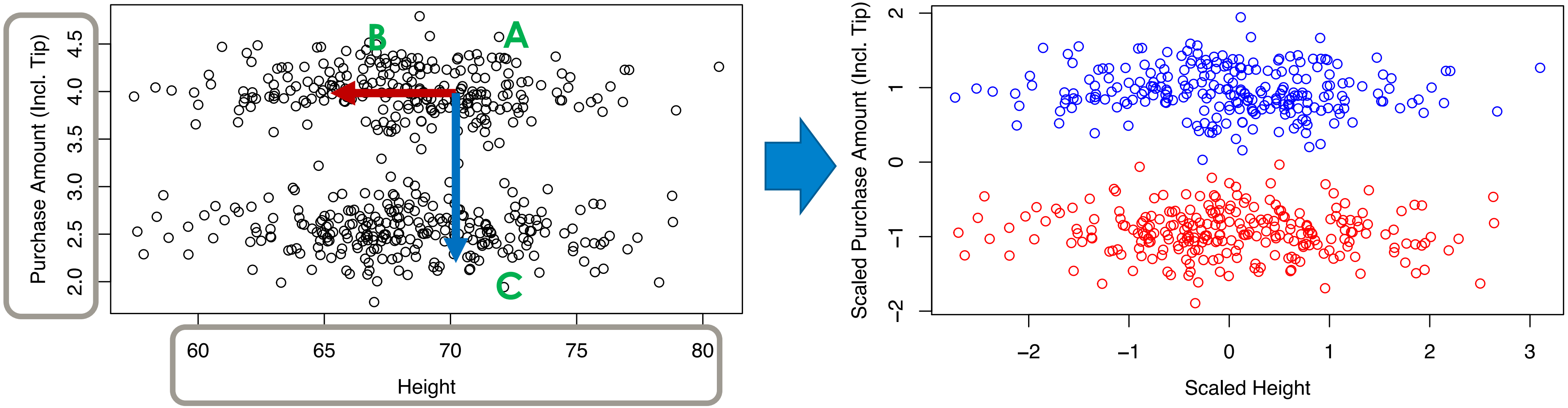
A(70,4); B(65,4); C(70,2)

$$D(A,B) = \sqrt{(70-65)^2 + (4-4){}^\wedge 2} = 5$$

$$D(A,C) = \sqrt{(70-70)^2 + (4-2){}^\wedge 2} = 2$$

Warning 2: Scaling of Inputs

- K-means relies on computing distances: if columns have different units, results may not make sense



- **Solution:** standardize (scale) inputs!

(Subtract mean and divide by standard deviation)

$$x^* = \frac{x - \bar{x}}{s}$$

A(70,4); B(65,4); C(70,2)

$$D(A,B) = \sqrt{(70-65)^2 + (4-4)^\wedge 2} = 5$$

$$D(A,C) = \sqrt{(70-70)^2 + (4-2)^\wedge 2} = 2$$

K-Means Clustering

Interpretation

Interpreting K-means

Rate the following on a scale from Strongly Disagree to Strongly Agree:

Question 1: I pay close attention to the origin and sourcing of my coffee.

Question 2: I do my best work at coffee shops.

Question 3: A good coffee shop has free Wi-Fi.

Question 4: Good food is important in a coffee shop.

Question 5: I enjoy drinking espresso.

```
check_clusters(df_sc,labels)
```

```
[(0, 82), (1, 152), (2, 48)]
```

3 clusters of sizes 82, 152, and 48

	A1	A2	A3	A4	A5
0	0.805415	1.492864	0.449114	0.441528	-1.077835
1	-0.867381	-0.620702	0.409086	0.415840	0.031088
2	1.370790	-0.584752	-2.062677	-2.071104	1.742855

```
labels
```

```
array([0, 1, 1, 1, 0, 2, 1, 0, 1, 1, 0, 1, 2, 1, 2, 1, 1, 0, 0, 1, 1, 1,
       0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 2, 1, 1, 2, 0, 2, 2, 1, 1,
       1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 2, 1, 1, 0, 1, 1, 0, 1, 0,
       0, 0, 2, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 2, 1, 1, 2, 1, 0, 1,
       1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 2,
       1, 1, 2, 1, 0, 1, 2, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 2, 1, 1,
       0, 0, 1, 1, 1, 1, 0, 2, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 2, 1, 1, 2,
       1, 2, 2, 0, 1, 2, 2, 1, 2, 1, 1, 0, 0, 0, 0, 2, 1, 0, 0, 1, 0, 2,
       0, 1, 0, 1, 1, 2, 2, 0, 0, 1, 0, 0, 1, 1, 2, 1, 0, 0, 0, 2, 1, 1,
       1, 1, 0, 2, 0, 1, 2, 0, 0, 0, 1, 0, 1, 0, 0, 2, 0, 0, 1, 1, 1, 1,
       0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 2, 0, 1, 1, 1, 1, 2, 0, 0, 1,
       2, 1, 1, 1, 2, 2, 1, 2, 1, 2, 1, 2, 0, 1, 2, 0, 1, 0, 2, 0, 2, 2,
       1, 2, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 2], dtype=int32)
```

Interpreting K-means

Rate the following on a scale from Strongly Disagree to Strongly Agree:

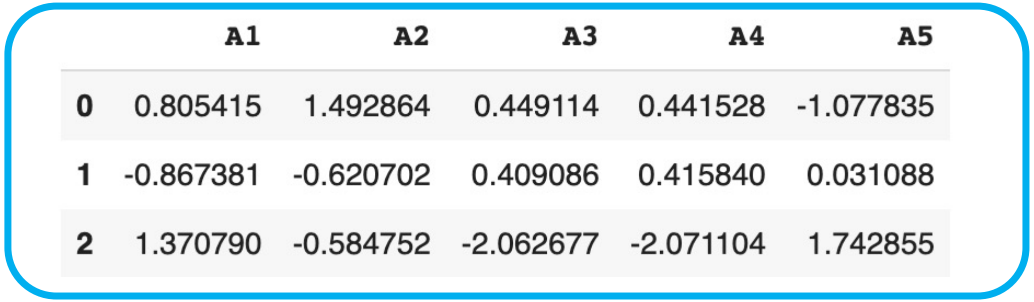
Question 1: I pay close attention to the origin and sourcing of my coffee.

Question 2: I do my best work at coffee shops.

Question 3: A good coffee shop has free Wi-Fi.

Question 4: Good food is important in a coffee shop.

Question 5: I enjoy drinking espresso.



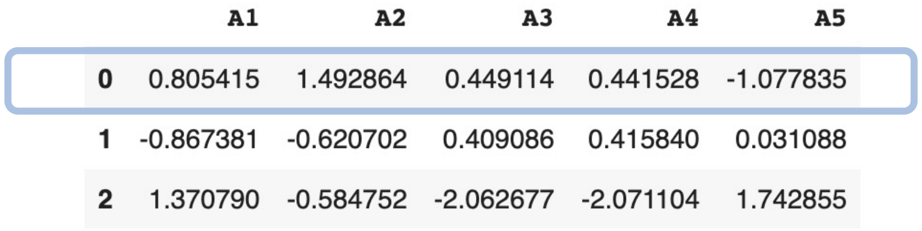
	A1	A2	A3	A4	A5
0	0.805415	1.492864	0.449114	0.441528	-1.077835
1	-0.867381	-0.620702	0.409086	0.415840	0.031088
2	1.370790	-0.584752	-2.062677	-2.071104	1.742855

The cluster centers (centroids):

- Rows = which cluster
- Columns = mean value of that variable in the cluster

If your data is standardized, these represent ***standardized differences from the overall mean***

Poll Title: Name Cluster 1



	A1	A2	A3	A4	A5
0	0.805415	1.492864	0.449114	0.441528	-1.077835
1	-0.867381	-0.620702	0.409086	0.415840	0.031088
2	1.370790	-0.584752	-2.062677	-2.071104	1.742855

Rate the following on a scale from Strongly Disagree to Strongly Agree:

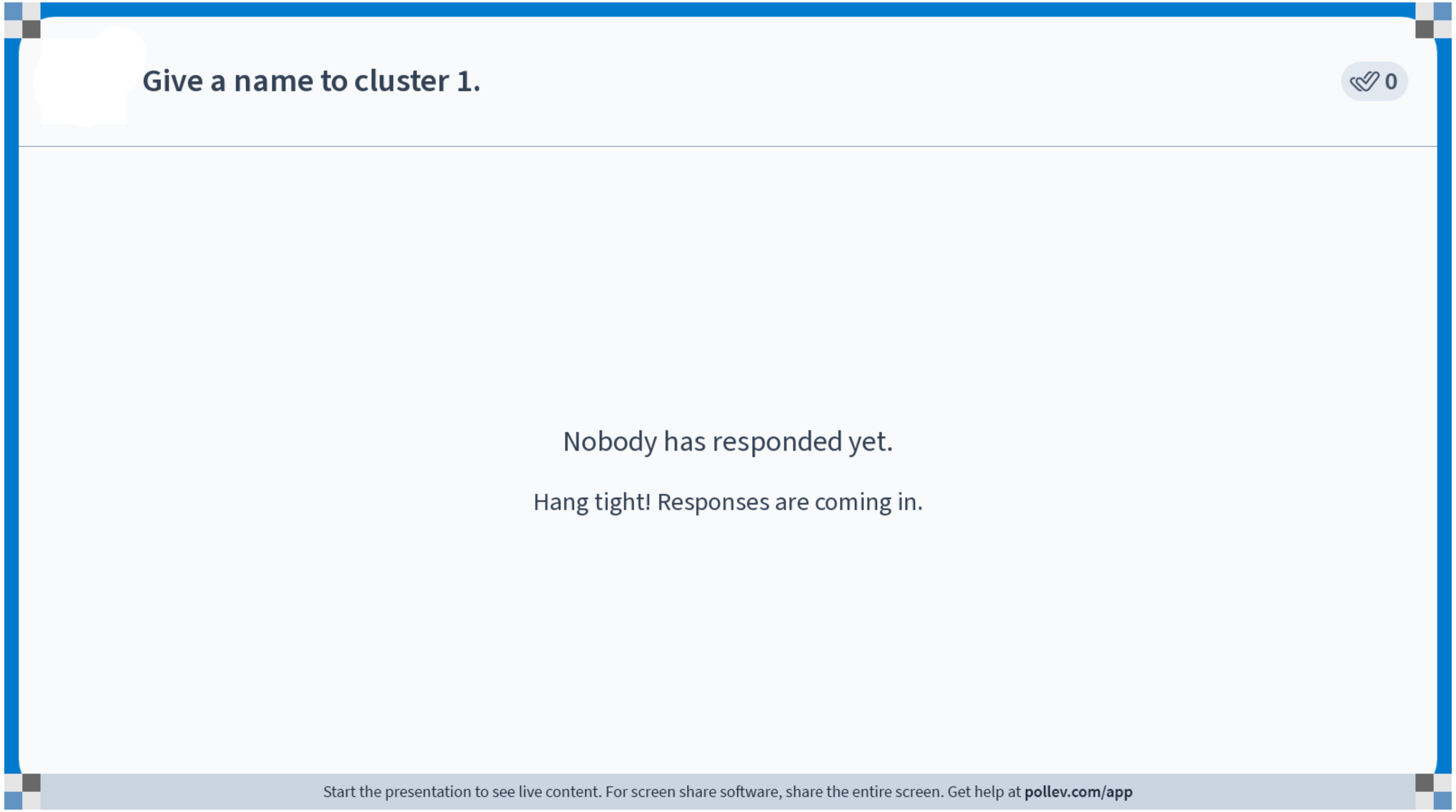
Question 1: I pay close attention to the origin and sourcing of my coffee.

Question 2: I do my best work at coffee shops.

Question 3: A good coffee shop has free Wi-Fi.

Question 4: Good food is important in a coffee shop.

Question 5: I enjoy drinking espresso.



Give a name to cluster 1.

Nobody has responded yet.

Hang tight! Responses are coming in.

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Interpreting K-means

Rate the following on a scale from Strongly Disagree to Strongly Agree:

↑ **Question 1:** I pay close attention to the origin and sourcing of my coffee.

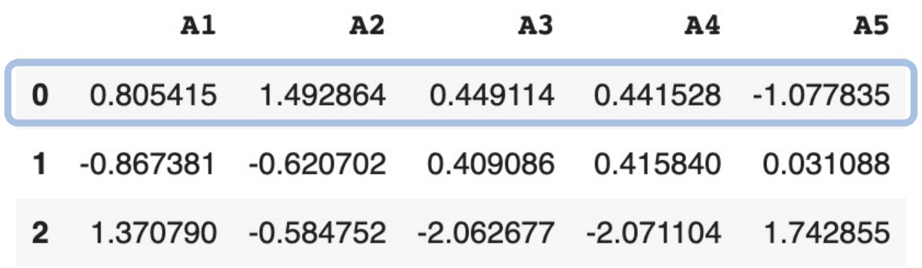
↑ **Question 2:** I do my best work at coffee shops.

↑ **Question 3:** A good coffee shop has free Wi-Fi.

↑ **Question 4:** Good food is important in a coffee shop.

↓ **Question 5:** I enjoy drinking espresso.

"The Coffee Shop Workers"



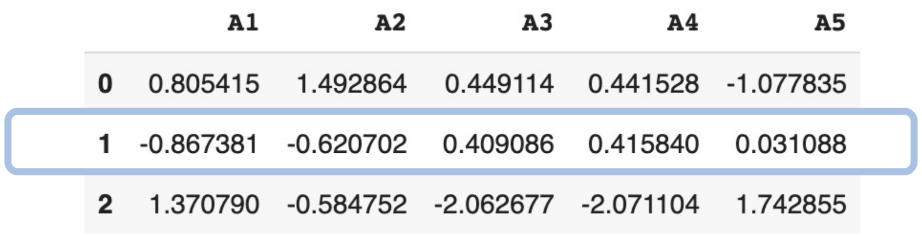
	A1	A2	A3	A4	A5
0	0.805415	1.492864	0.449114	0.441528	-1.077835
1	-0.867381	-0.620702	0.409086	0.415840	0.031088
2	1.370790	-0.584752	-2.062677	-2.071104	1.742855

The cluster centers (centroids):

- Rows = which cluster
- Columns = mean value of that variable in the cluster

If your data is standardized, these represent ***standardized differences from the overall mean***

Poll Title: Name Cluster 2



	A1	A2	A3	A4	A5
0	0.805415	1.492864	0.449114	0.441528	-1.077835
1	-0.867381	-0.620702	0.409086	0.415840	0.031088
2	1.370790	-0.584752	-2.062677	-2.071104	1.742855

Rate the following on a scale from Strongly Disagree to Strongly Agree:

Question 1: I pay close attention to the origin and sourcing of my coffee.

Question 2: I do my best work at coffee shops.

Question 3: A good coffee shop has free Wi-Fi.

Question 4: Good food is important in a coffee shop.

Question 5: I enjoy drinking espresso.

Give a name to cluster 2.

Nobody has responded yet.

Hang tight! Responses are coming in.

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Interpreting K-means

Rate the following on a scale from Strongly Disagree to Strongly Agree:

↓ **Question 1:** I pay close attention to the origin and sourcing of my coffee.

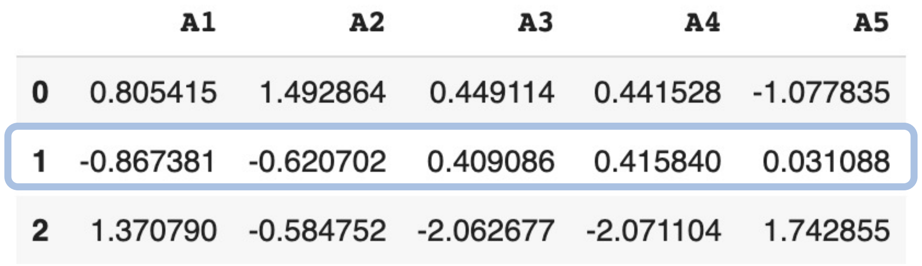
↓ **Question 2:** I do my best work at coffee shops.

↑ **Question 3:** A good coffee shop has free Wi-Fi.

↑ **Question 4:** Good food is important in a coffee shop.

○ **Question 5:** I enjoy drinking espresso.

"The Basic Coffee Drinkers"



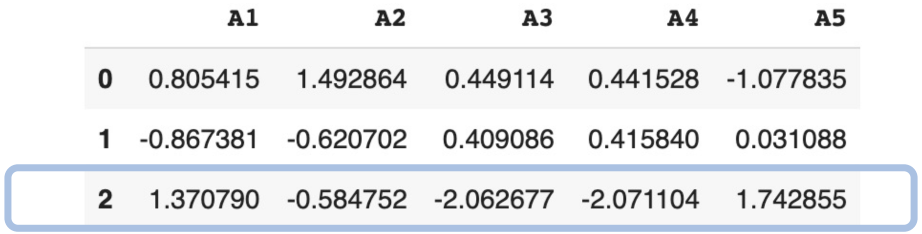
	A1	A2	A3	A4	A5
0	0.805415	1.492864	0.449114	0.441528	-1.077835
1	-0.867381	-0.620702	0.409086	0.415840	0.031088
2	1.370790	-0.584752	-2.062677	-2.071104	1.742855

The cluster centers (centroids):

- Rows = which cluster
- Columns = mean value of that variable in the cluster

If your data is standardized, these represent ***standardized differences from the overall mean***

Poll Title: Name Cluster 3



	A1	A2	A3	A4	A5
0	0.805415	1.492864	0.449114	0.441528	-1.077835
1	-0.867381	-0.620702	0.409086	0.415840	0.031088
2	1.370790	-0.584752	-2.062677	-2.071104	1.742855

Rate the following on a scale from Strongly Disagree to Strongly Agree:

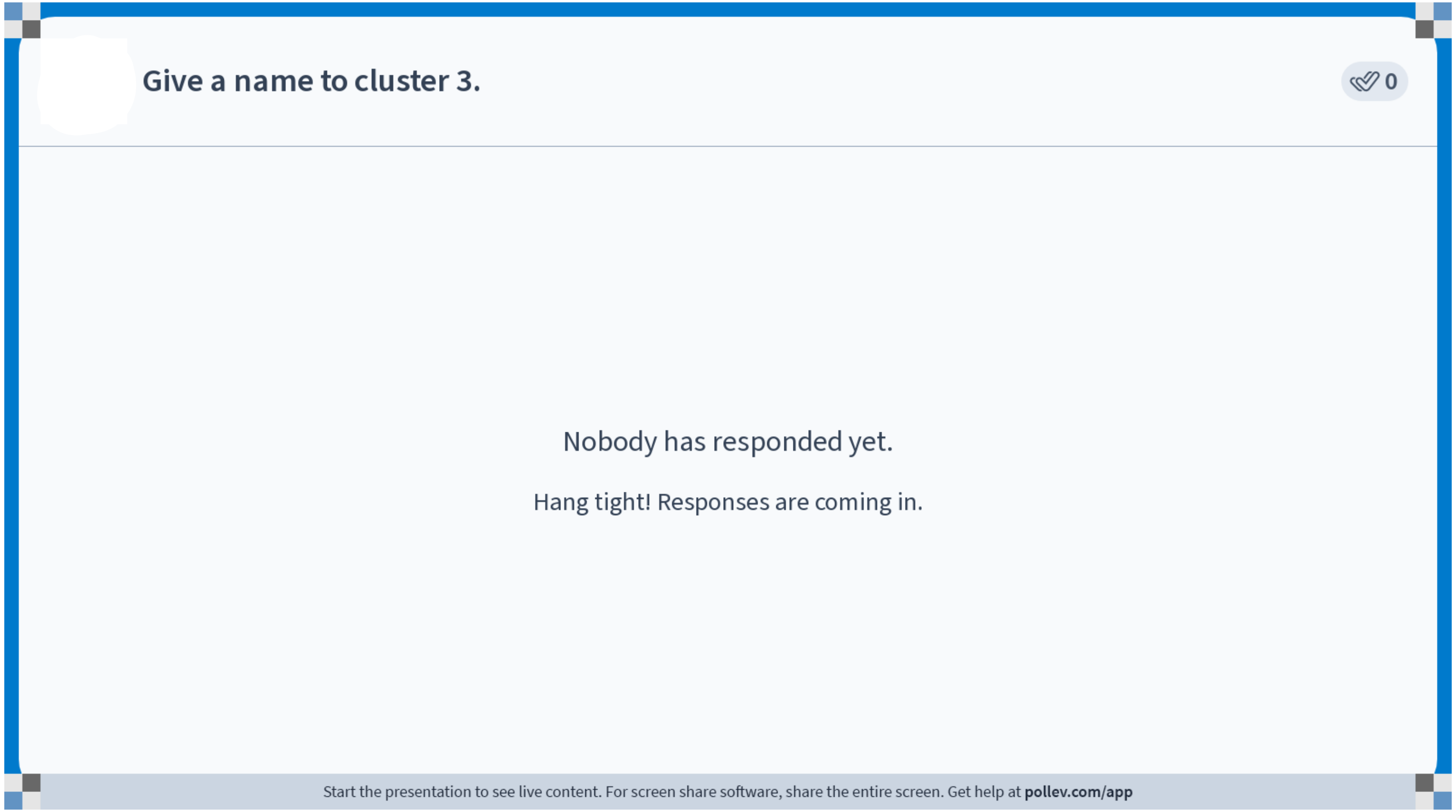
Question 1: I pay close attention to the origin and sourcing of my coffee.

Question 2: I do my best work at coffee shops.

Question 3: A good coffee shop has free Wi-Fi.

Question 4: Good food is important in a coffee shop.

Question 5: I enjoy drinking espresso.



Give a name to cluster 3.

Nobody has responded yet.

Hang tight! Responses are coming in.

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Interpreting K-means

Rate the following on a scale from Strongly Disagree to Strongly Agree:

↑ **Question 1:** I pay close attention to the origin and sourcing of my coffee.

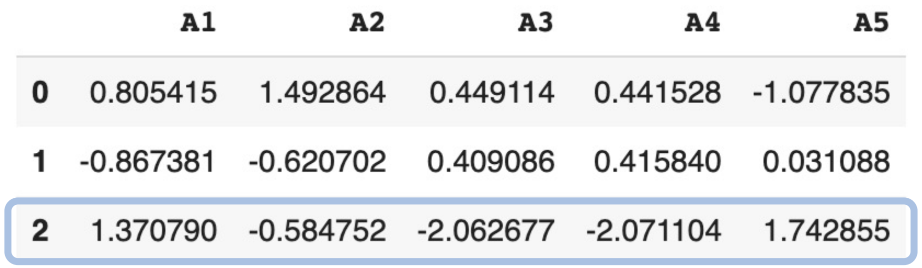
↓ **Question 2:** I do my best work at coffee shops.

↓ **Question 3:** A good coffee shop has free Wi-Fi.

↓ **Question 4:** Good food is important in a coffee shop.

↑ **Question 5:** I enjoy drinking espresso.

"The Espresso Snobs"



	A1	A2	A3	A4	A5
0	0.805415	1.492864	0.449114	0.441528	-1.077835
1	-0.867381	-0.620702	0.409086	0.415840	0.031088
2	1.370790	-0.584752	-2.062677	-2.071104	1.742855

The cluster centers (centroids):

- Rows = which cluster
- Columns = mean value of that variable in the cluster

If your data is standardized, these represent ***standardized differences from the overall mean***

Interpreting K-means

Rate the following on a scale from Strongly Disagree to Strongly Agree:

Question 1: I pay close attention to the origin and sourcing of my coffee.

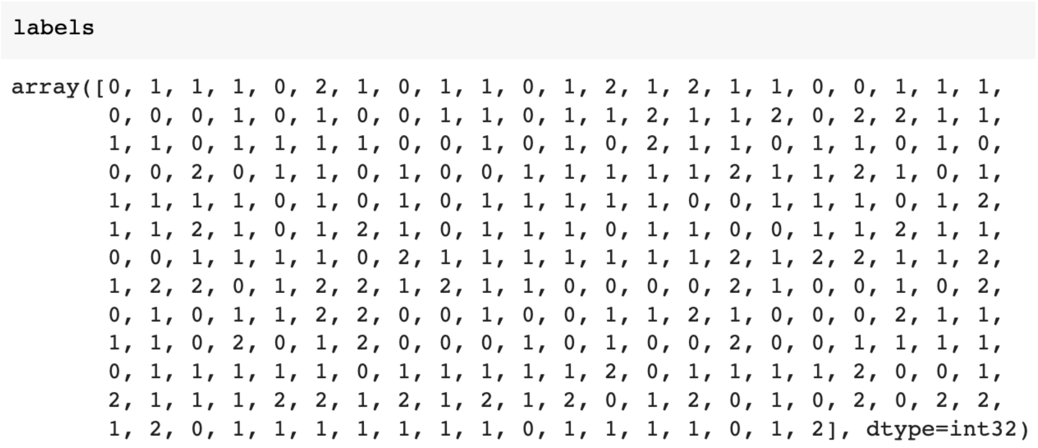
Question 2: I do my best work at coffee shops.

Question 3: A good coffee shop has free Wi-Fi.

Question 4: Good food is important in a coffee shop.

Question 5: I enjoy drinking espresso.

Predicted class label (cluster) for each data point



```
labels
```

```
array([0, 1, 1, 1, 0, 2, 1, 0, 1, 1, 0, 1, 2, 1, 2, 1, 1, 0, 0, 1, 1, 1,
       0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 2, 1, 1, 2, 0, 2, 2, 1, 1,
       1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 2, 1, 1, 0, 1, 1, 0, 1, 0,
       0, 0, 2, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 2, 1, 1, 2, 1, 0, 1,
       1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 2,
       1, 1, 2, 1, 0, 1, 2, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 2, 1, 1,
       0, 0, 1, 1, 1, 1, 0, 2, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 2, 1, 1, 2,
       1, 2, 2, 0, 1, 2, 2, 1, 2, 1, 1, 0, 0, 0, 0, 2, 1, 0, 0, 1, 0, 2,
       0, 1, 0, 1, 1, 2, 2, 0, 0, 1, 0, 0, 1, 1, 2, 1, 0, 0, 0, 2, 1, 1,
       1, 1, 0, 2, 0, 1, 2, 0, 0, 0, 1, 0, 1, 0, 0, 2, 0, 0, 1, 1, 1, 1,
       0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 2, 0, 1, 1, 1, 1, 2, 0, 0, 1,
       2, 1, 1, 1, 2, 2, 1, 2, 1, 2, 1, 2, 0, 1, 2, 0, 1, 0, 2, 0, 2, 2,
       1, 2, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 2], dtype=int32)
```

Interpreting K-means – What is missing?

- K-means provides **meaningful** segments in terms of preferences
- But:
  - Which segment is more important to me?
    - Business decision driven by your values, core customers…
  - How do I reach out to each segment?
    - Customers do not tell you in advance which segment they belong to
    - Usually, you would need to compare segments with other variables (e.g., demographics) → **segments must be interpretable**
  - **Why three segments?**

Interpreting K-means

Rate the following on a scale from Strongly Disagree to Strongly Agree:

Question 1: I pay close attention to the origin and sourcing of my coffee.

Question 2: I do my best work at coffee shops.

Question 3: A good coffee shop has free Wi-Fi.

Question 4: Good food is important in a coffee shop.

Question 5: I enjoy drinking espresso.

Inertia = Total within-cluster sum of squares

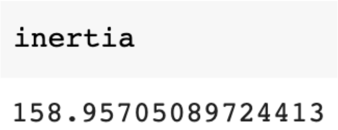
$$\sum_{i=1}^{n} \left\| x_i - \mu_j \right\|^2$$

Each data point Centroid for each cluster

Basically: measure of how internally coherent clusters are, lower = better

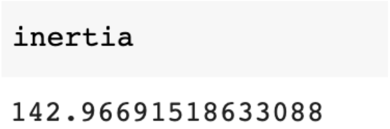
Caveat: Always decreases with the number of clusters!

3 clusters:



inertia
158.95705089724413

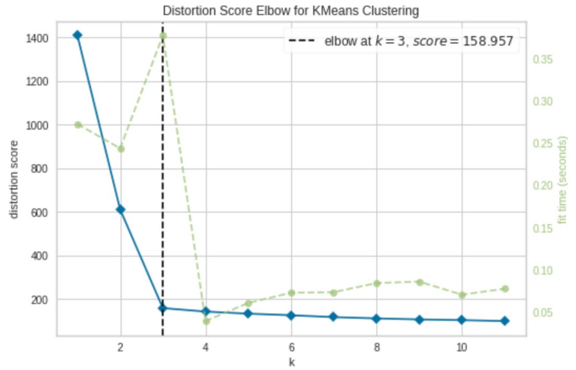
4 clusters:



inertia
142.96691518633088

Determining the Number of Clusters

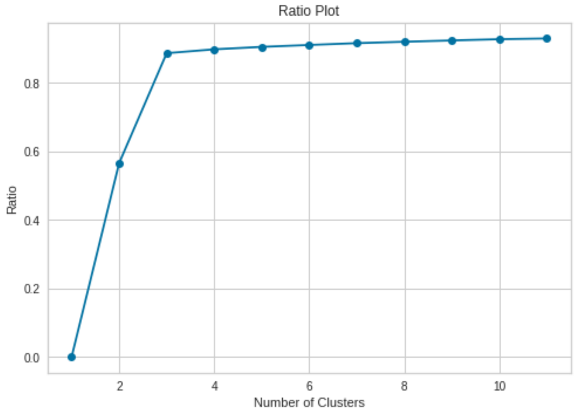
- **Elbow Plot:** Increase the number of clusters and monitor the `inertia (distortion)`
  - When it starts to level off, stop!
  - In Python, you may use KElbowVisualizer



- **Ratio Plot**: Increase the number of clusters and monitor (total between sum of squares/total sum of squares) = $\sum_{j=1}^{J} \left\|\mu_j - \bar{X}\right\|^2 / \sum_{i=1}^{n} \|x_i - \bar{X}\|^2$

Centroid for each cluster Overall mean of all data

  - Total sum of squares = total within sum of squares (i.e., inertia) + total between sum of squares



- Many other criteria… **none** perfect

Determining the Number of Clusters

- **Fit:** Determined by elbow plot or ratio plot (`between_SS / total_SS` statistic)
  - Note the jargon

- **Interpretability:** Are the segments well differentiated, and capturing meaningful differences?
  - Fewer clusters is usually better
  - Caveat: small number of clusters may ignore niches

- No criterion is perfect! You must use it to **inform** your decision rather than **substitute** it

Chi-square Test of Independence

- Is there a relationship between cluster membership and variables we have?
- A statistical test to determine whether a difference between two categorical variables is due to chance or a relationship between them
- Is there a statistically significant difference between the expected frequencies and the observed frequencies in a contingency table?
- H_0: The two variables are independent

	Success	Failure	Total
Group 1	A	B	A+B
Group 2	C	D	C+D
Total	A+C	B+D	A+B+C+D

$$\text{Expected count} = \frac{\text{(row total)(column total)}}{\text{total sample size}}$$

$$\chi^2 = \sum_{i=1}^{rc} \frac{(observed - expected)^2}{expected}$$

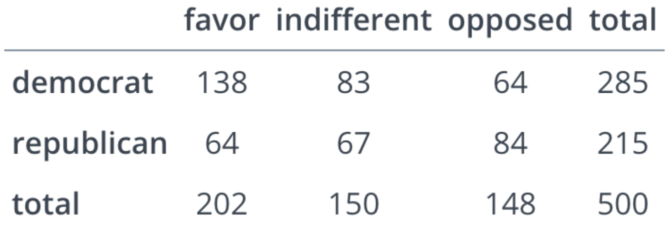
with degrees of freedom = (# of rows - 1)(# of columns - 1)

Reject the null when p-value of χ^2 with df less than 0.05

Chi-square Example

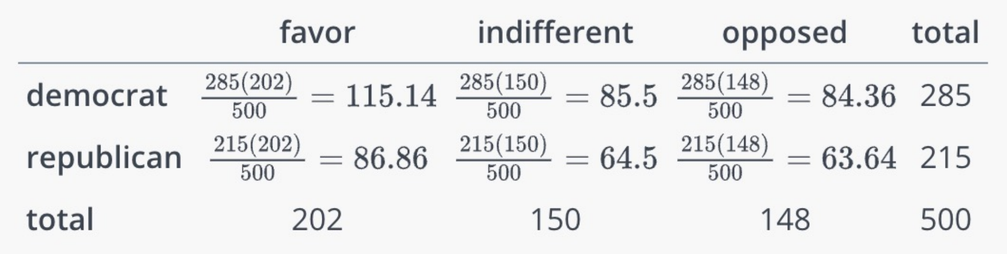
H_0: Political leaning and support for policy are independent

Observed



	favor	indifferent	opposed	total
democrat	138	83	64	285
republican	64	67	84	215
total	202	150	148	500

Expected



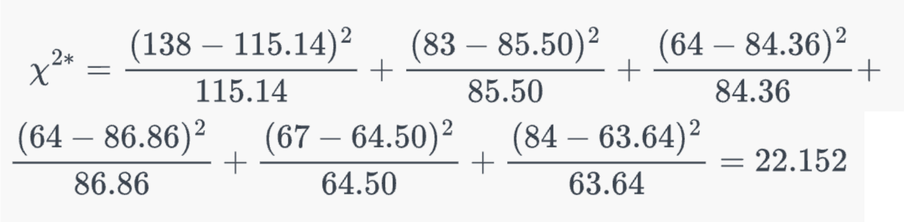
	favor	indifferent	opposed	total
democrat	$\frac{285(202)}{500} = 115.14$	$\frac{285(150)}{500} = 85.5$	$\frac{285(148)}{500} = 84.36$	285
republican	$\frac{215(202)}{500} = 86.86$	$\frac{215(150)}{500} = 64.5$	$\frac{215(148)}{500} = 63.64$	215
total	202	150	148	500

Expected count = $\frac{\text{(row total)(column total)}}{\text{total sample size}}$

$$\chi^2 = \sum_{i=1}^{rc} \frac{(observed - expected)^2}{expected}$$

with degrees of freedom = (# of rows - 1)(# of columns - 1)

Reject the null when p-value of χ^2 with df less than 0.05



$$\chi^{2*} = \frac{(138-115.14)^2}{115.14} + \frac{(83-85.50)^2}{85.50} + \frac{(64-84.36)^2}{84.36} + \frac{(64-86.86)^2}{86.86} + \frac{(67-64.50)^2}{64.50} + \frac{(84-63.64)^2}{63.64} = 22.152$$

df = (2-1)(3-1) = 2 → p-value < 0.05

Let’s go to Python

K-means Clustering

Takeaway: What is a Market Segment?

"Market segmentation is the **subdividing of a market into distinct subsets**, where any subset may conceivably be selected as a marketing target to be reached within a distinct marketing mix."

- Kotler

Takeaway: What is a Market Segment?

- Characteristics of ideal segments: **L**arge, **I**dentifiable, **D**istinctive, **S**table (**LIDS**)
- But even more important… **actionable!**

"If I knew the segments, what would I do with them?"

- The data type used for the segmentation often determines its usefulness
  - Think about the goal of the segmentation
  - Also consider: actionability, accessibility
- Cluster analysis provides a data-driven tool for learning segments from data
  - Hierarchical clustering and K-means are both simple and powerful
  - Determining the number of clusters can be tricky: balance fit, interpretability, and usefulness

Targeting

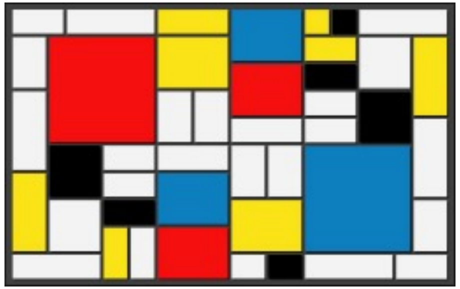
Where will we play?

How will we win?

Segmentation

S

Discovering and profiling groups of customers with similar needs and preferences



Targeting

T

Evaluating segment attractiveness and targeting most attractive ones

Positioning

P

Defining value proposition for target segments and developing a marketing plan

Segmentation & Targeting based on Demographics

Bic/Amazon

'Bic for Her' Pens Deluged With Sarcastic Reviews on Amazon Silly product gets response to match as backlash gathers steam By Tim Nudd

This is an example of bad segmentation and targeting!

"Someone has answered my gentle prayers and FINALLY designed a pen that I can use all month long! I use it when I'm swimming, riding a horse, walking on the beach and doing yoga. It's comfortable, leak-proof, non-slip **and it makes me feel so feminine and pretty!** Since I've begun using these pens, **men have found me more attractive and approachable.** It has given **me soft skin and manageable hair and it has really given me the self-esteem.**"

amazon reviews

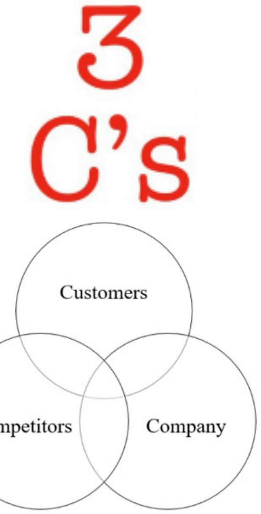
Choosing Which Target Market(s) to Serve

Market Opportunities for Profit:
- Segment size
- Growth rate/potential

Target Market Selection

Competitive Intensity
- Underserved needs?
- Competitors' strengths

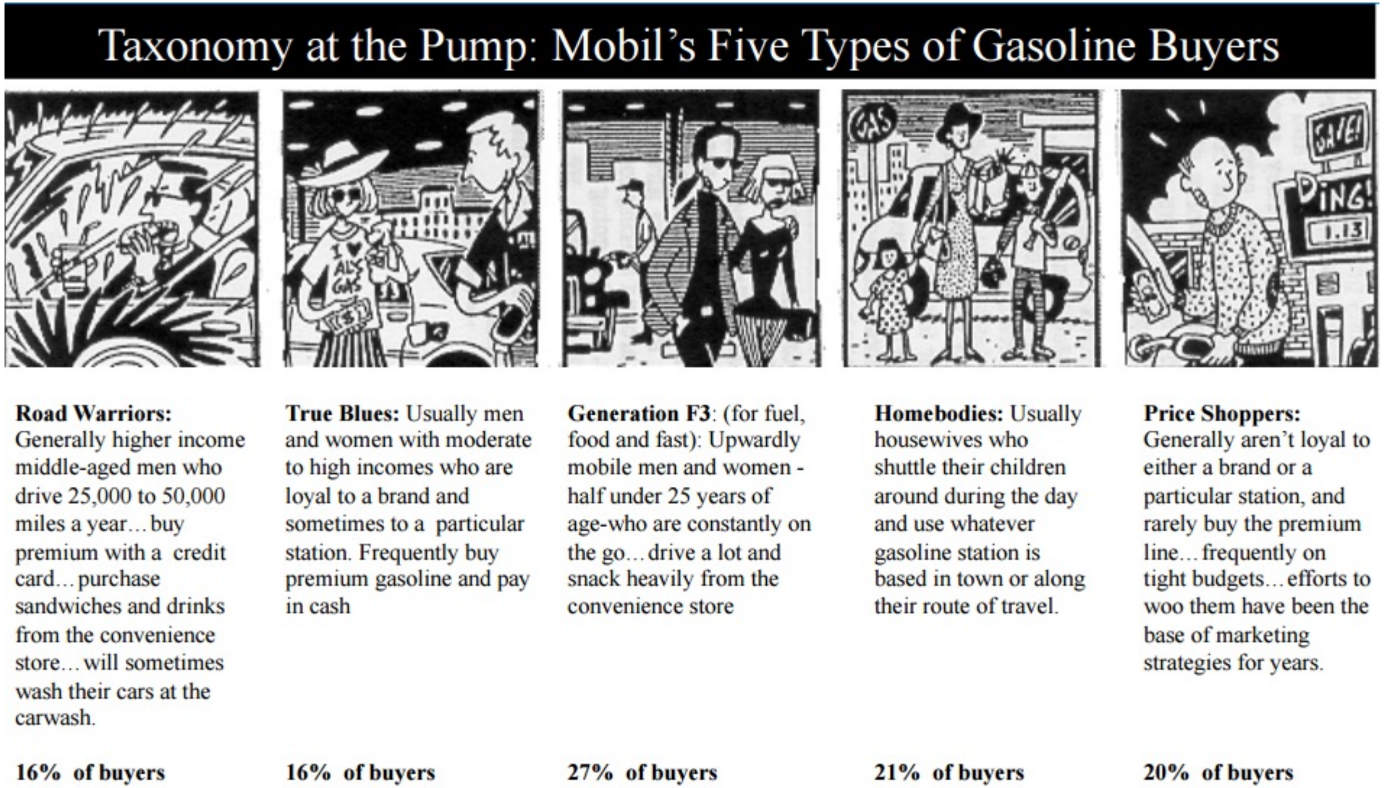
Company "Fit" with
- Objectives
- Competencies / Resources
- Customer Base



The Importance of Saying "No"

Early 1990s: price wars at the gas pump threatened profitability of oil companies. Oil companies assumed consumers were extremely price-sensitive

To better understand customers, Mobil conducted a study of 2,000 customers. Using cluster analysis, they identified 5 segments of gasoline buyers…



Taxonomy at the Pump: Mobil's Five Types of Gasoline Buyers

Road Warriors: Generally higher income middle-aged men who drive 25,000 to 50,000 miles a year… buy premium with a credit card… purchase sandwiches and drinks from the convenience store… will sometimes wash their cars at the carwash.

16% of buyers

True Blues: Usually men and women with moderate to high incomes who are loyal to a brand and sometimes to a particular station. Frequently buy premium gasoline and pay in cash

16% of buyers

Generation F3: (for fuel, food and fast): Upwardly mobile men and women - half under 25 years of age-who are constantly on the go… drive a lot and snack heavily from the convenience store

27% of buyers

Homebodies: Usually housewives who shuttle their children around during the day and use whatever gasoline station is based in town or along their route of travel.

21% of buyers

Price Shoppers: Generally aren't loyal to either a brand or a particular station, and rarely buy the premium line… frequently on tight budgets… efforts to woo them have been the base of marketing strategies for years.

20% of buyers

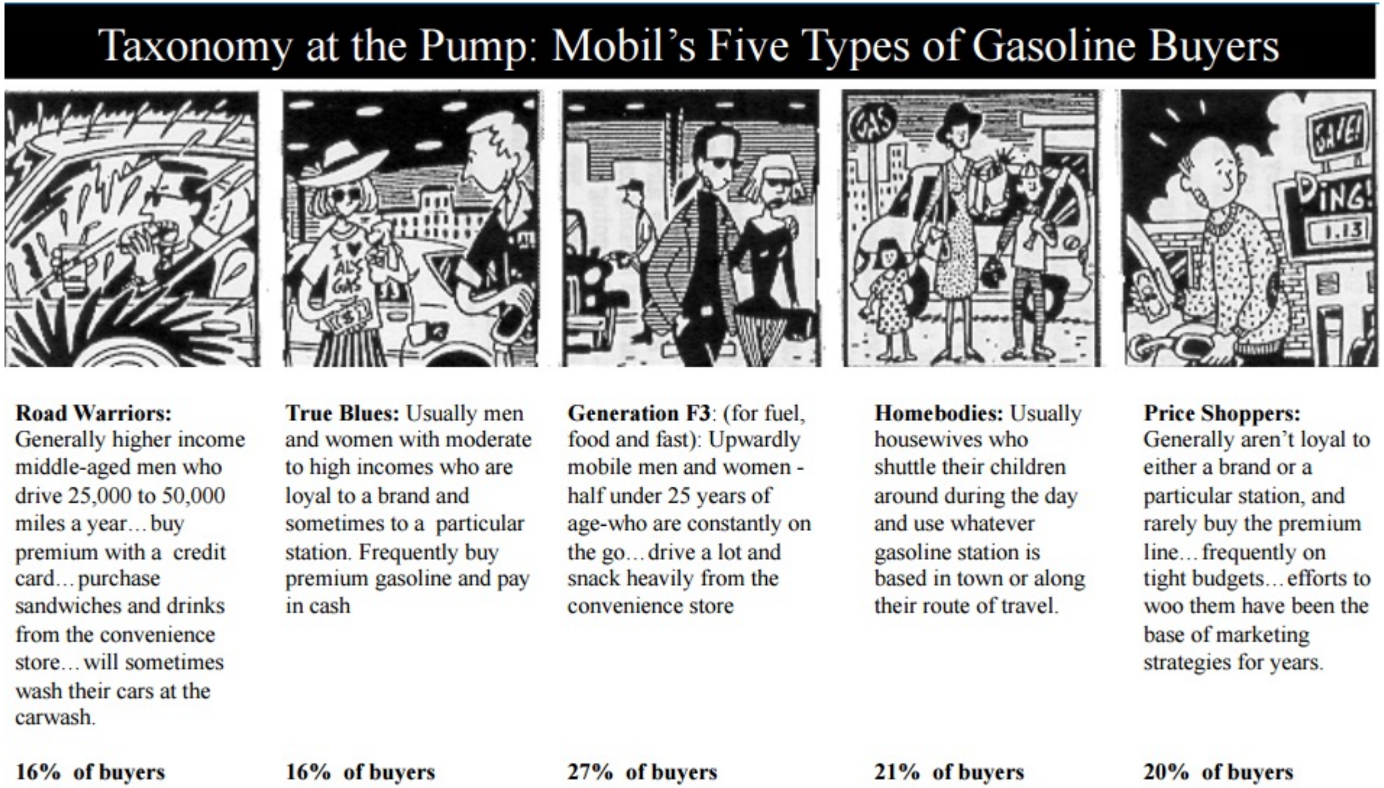
The Importance of Saying “No”

Road warriors: people who used their cars as part of their profession

Middle aged men with higher incomes

Prefer credit to cash

Like buying food and drinks



Taxonomy at the Pump: Mobil’s Five Types of Gasoline Buyers

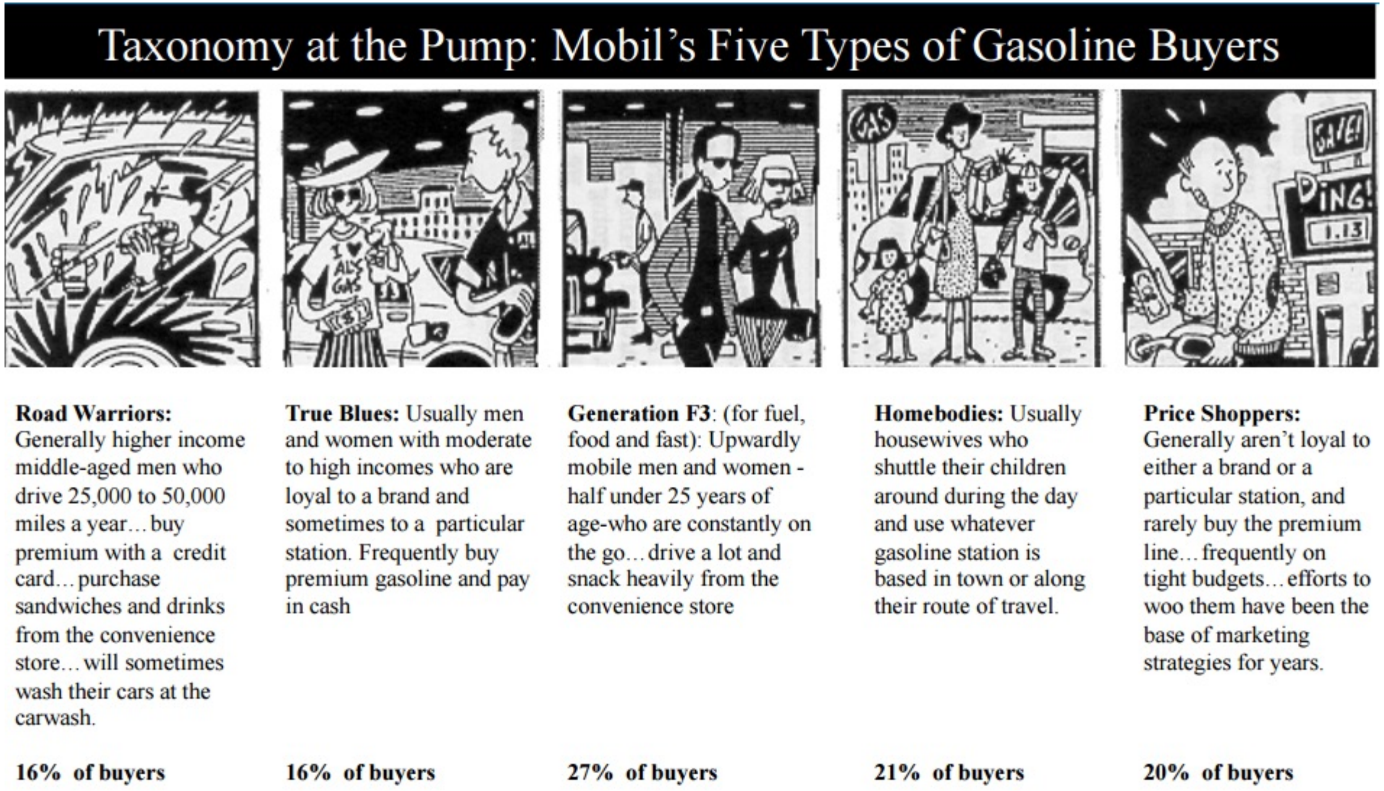
Road Warriors: Generally higher income middle-aged men who drive 25,000 to 50,000 miles a year…buy premium with a credit card…purchase sandwiches and drinks from the convenience store…will sometimes wash their cars at the carwash.	**True Blues:** Usually men and women with moderate to high incomes who are loyal to a brand and sometimes to a particular station. Frequently buy premium gasoline and pay in cash	**Generation F3**: (for fuel, food and fast): Upwardly mobile men and women - half under 25 years of age-who are constantly on the go…drive a lot and snack heavily from the convenience store	**Homebodies:** Usually housewives who shuttle their children around during the day and use whatever gasoline station is based in town or along their route of travel.	**Price Shoppers:** Generally aren’t loyal to either a brand or a particular station, and rarely buy the premium line…frequently on tight budgets…efforts to woo them have been the base of marketing strategies for years.
16% of buyers	**16% of buyers**	**27% of buyers**	**21% of buyers**	**20% of buyers**

The Importance of Saying “No”

True blues: brand loyal and occasionally station loyal

Moderate to high incomes

Prefer to pay in cash and buy premium gas



Taxonomy at the Pump: Mobil’s Five Types of Gasoline Buyers

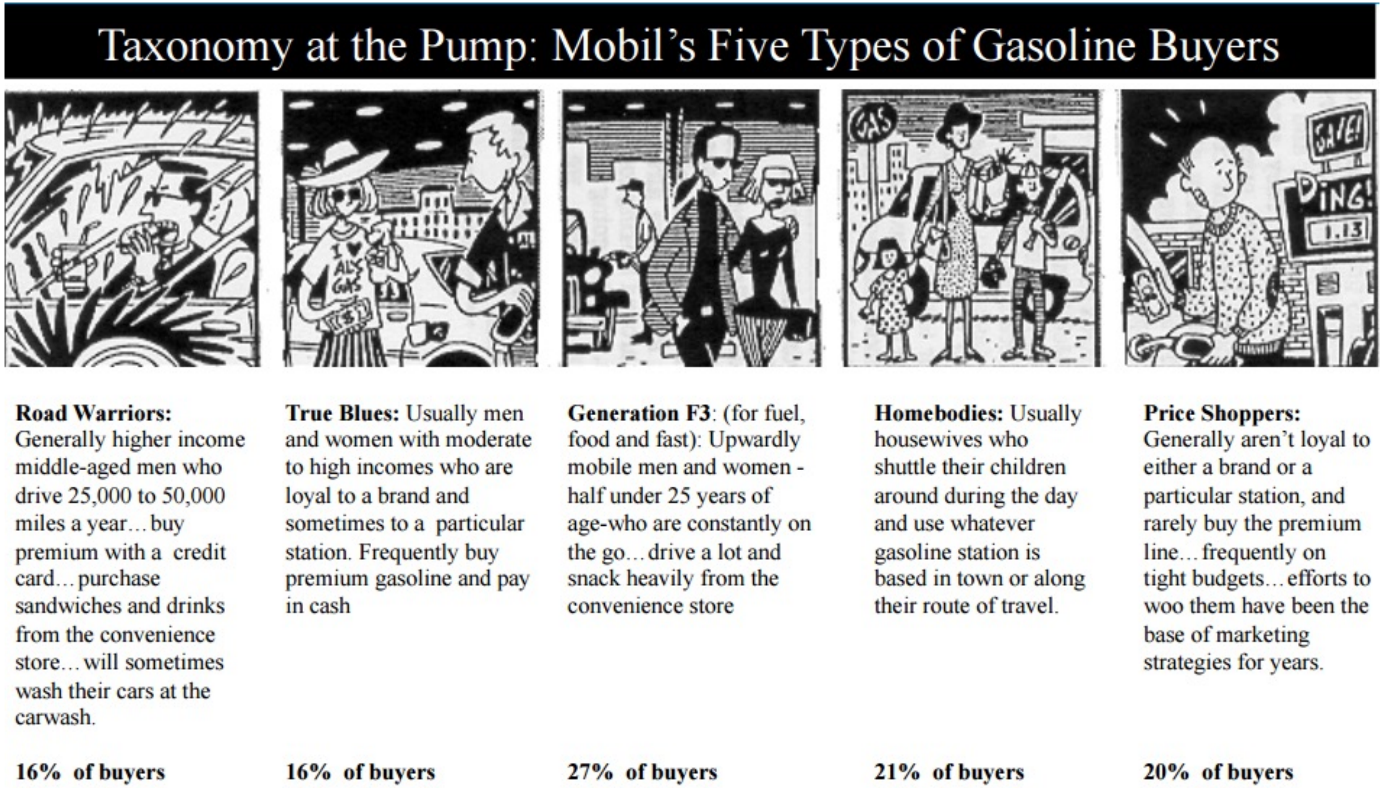
Road Warriors: Generally higher income middle-aged men who drive 25,000 to 50,000 miles a year…buy premium with a credit card…purchase sandwiches and drinks from the convenience store…will sometimes wash their cars at the carwash.	**True Blues:** Usually men and women with moderate to high incomes who are loyal to a brand and sometimes to a particular station. Frequently buy premium gasoline and pay in cash	**Generation F3**: (for fuel, food and fast): Upwardly mobile men and women - half under 25 years of age-who are constantly on the go…drive a lot and snack heavily from the convenience store	**Homebodies:** Usually housewives who shuttle their children around during the day and use whatever gasoline station is based in town or along their route of travel.	**Price Shoppers:** Generally aren’t loyal to either a brand or a particular station, and rarely buy the premium line…frequently on tight budgets…efforts to woo them have been the base of marketing strategies for years.
16% of buyers	**16% of buyers**	**27% of buyers**	**21% of buyers**	**20% of buyers**

The Importance of Saying "No"

Generation F3: upwardly mobile young consumers (50% under 25 years old)

Drove often

Habitually purchased snacks



Taxonomy at the Pump: Mobil's Five Types of Gasoline Buyers

Road Warriors: Generally higher income middle-aged men who drive 25,000 to 50,000 miles a year…buy premium with a credit card…purchase sandwiches and drinks from the convenience store…will sometimes wash their cars at the carwash.

16% of buyers

True Blues: Usually men and women with moderate to high incomes who are loyal to a brand and sometimes to a particular station. Frequently buy premium gasoline and pay in cash

16% of buyers

Generation F3: (for fuel, food and fast): Upwardly mobile men and women - half under 25 years of age-who are constantly on the go…drive a lot and snack heavily from the convenience store

27% of buyers

Homebodies: Usually housewives who shuttle their children around during the day and use whatever gasoline station is based in town or along their route of travel.

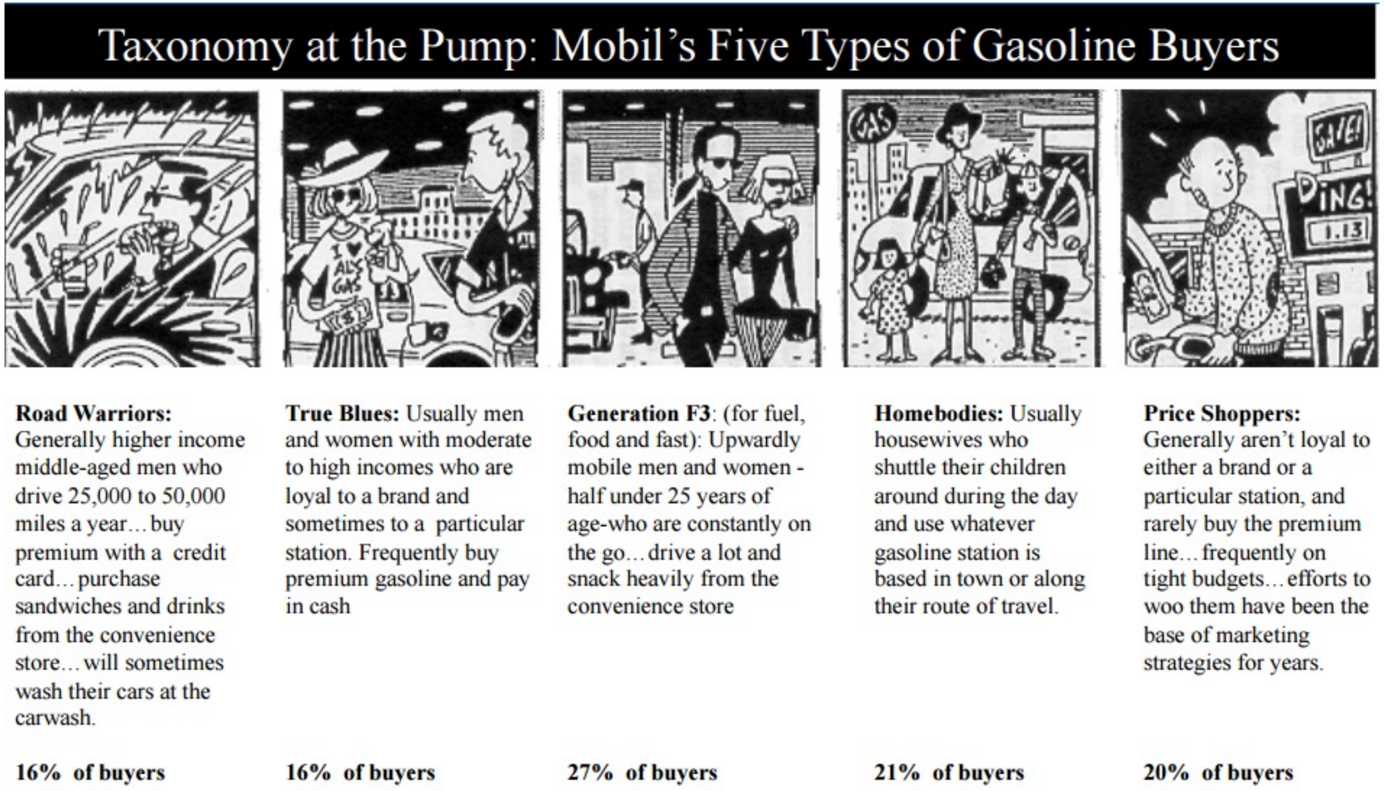
21% of buyers

Price Shoppers: Generally aren't loyal to either a brand or a particular station, and rarely buy the premium line…frequently on tight budgets…efforts to woo them have been the base of marketing strategies for years.

20% of buyers

The Importance of Saying "No"

Homebodies: stay-at-home mothers who valued gas station proximity to home/normal travel routes



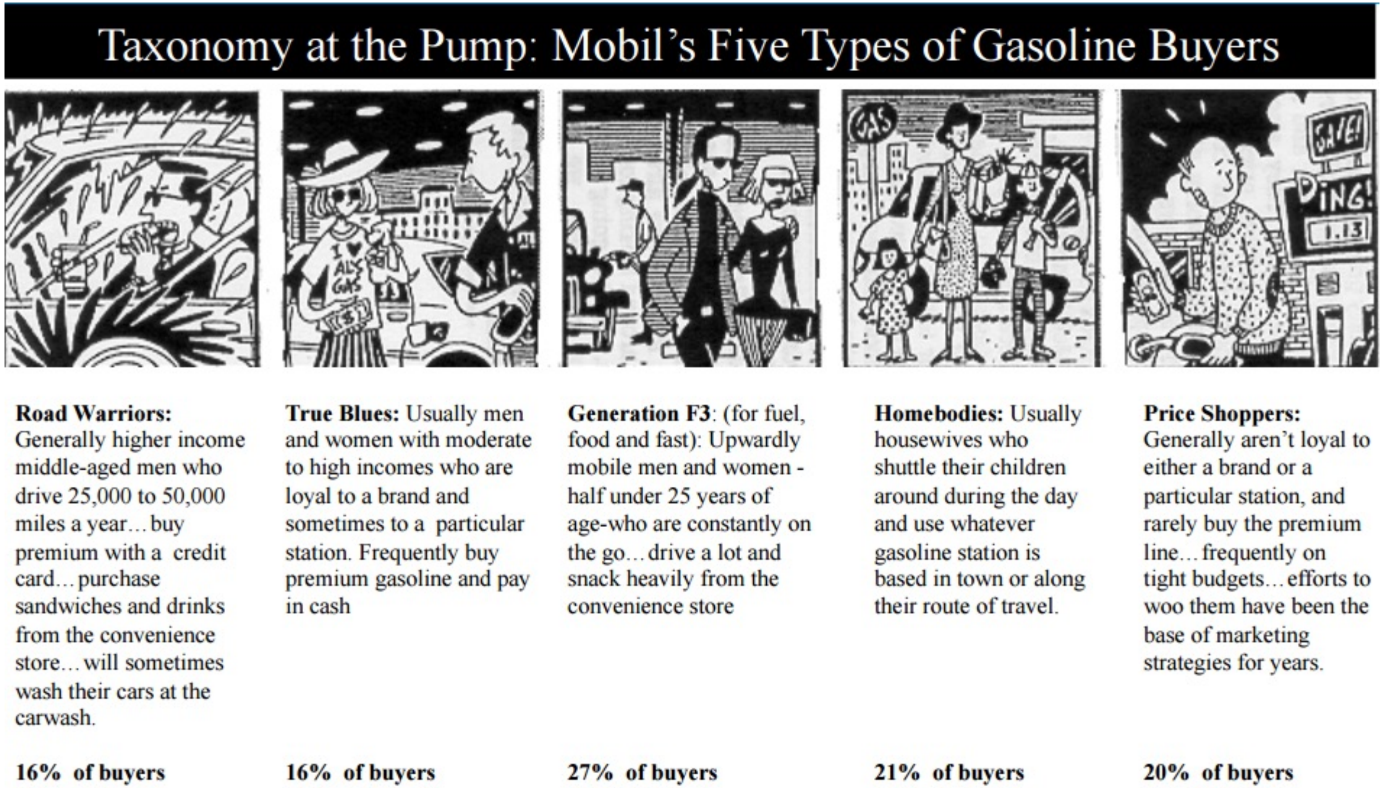
Taxonomy at the Pump: Mobil's Five Types of Gasoline Buyers

Road Warriors:	**True Blues:**	**Generation F3**	**Homebodies:**	**Price Shoppers:**
Generally higher income middle-aged men who drive 25,000 to 50,000 miles a year… buy premium with a credit card… purchase sandwiches and drinks from the convenience store… will sometimes wash their cars at the carwash.	Usually men and women with moderate to high incomes who are loyal to a brand and sometimes to a particular station. Frequently buy premium gasoline and pay in cash	(for fuel, food and fast): Upwardly mobile men and women - half under 25 years of age-who are constantly on the go… drive a lot and snack heavily from the convenience store	Usually housewives who shuttle their children around during the day and use whatever gasoline station is based in town or along their route of travel.	Generally aren't loyal to either a brand or a particular station, and rarely buy the premium line… frequently on tight budgets… efforts to woo them have been the base of marketing strategies for years.
16% of buyers	**16% of buyers**	**27% of buyers**	**21% of buyers**	**20% of buyers**

The Importance of Saying "No"

Price shoppers: customers on a budget who rarely bought premium gas

Not brand or station loyal



Taxonomy at the Pump: Mobil's Five Types of Gasoline Buyers

Road Warriors: Generally higher income middle-aged men who drive 25,000 to 50,000 miles a year… buy premium with a credit card… purchase sandwiches and drinks from the convenience store… will sometimes wash their cars at the carwash.

16% of buyers

True Blues: Usually men and women with moderate to high incomes who are loyal to a brand and sometimes to a particular station. Frequently buy premium gasoline and pay in cash

16% of buyers

Generation F3: (for fuel, food and fast): Upwardly mobile men and women - half under 25 years of age-who are constantly on the go… drive a lot and snack heavily from the convenience store

27% of buyers

Homebodies: Usually housewives who shuttle their children around during the day and use whatever gasoline station is based in town or along their route of travel.

21% of buyers

Price Shoppers: Generally aren't loyal to either a brand or a particular station, and rarely buy the premium line… frequently on tight budgets… efforts to woo them have been the base of marketing strategies for years.

20% of buyers

The Importance of Saying "No"

Taxonomy at the Pump: Mobil's Five Types of Gasoline Buyers

Road Warriors: Generally higher income middle-aged men who drive 25,000 to 50,000 miles a year… buy premium with a credit card… purchase sandwiches and drinks from the convenience store… will sometimes wash their cars at the carwash.

16% of buyers

True Blues: Usually men and women with moderate to high incomes who are loyal to a brand and sometimes to a particular station. Frequently buy premium gasoline and pay in cash

16% of buyers

Generation F3: (for fuel, food and fast): Upwardly mobile men and women - half under 25 years of age-who are constantly on the go… drive a lot and snack heavily from the convenience store

27% of buyers

Homebodies: Usually housewives who shuttle their children around during the day and use whatever gasoline station is based in town or along their route of travel.

21% of buyers

Price Shoppers: Generally aren't loyal to either a brand or a particular station, and rarely buy the premium line… frequently on tight budgets… efforts to woo them have been the base of marketing strategies for years.

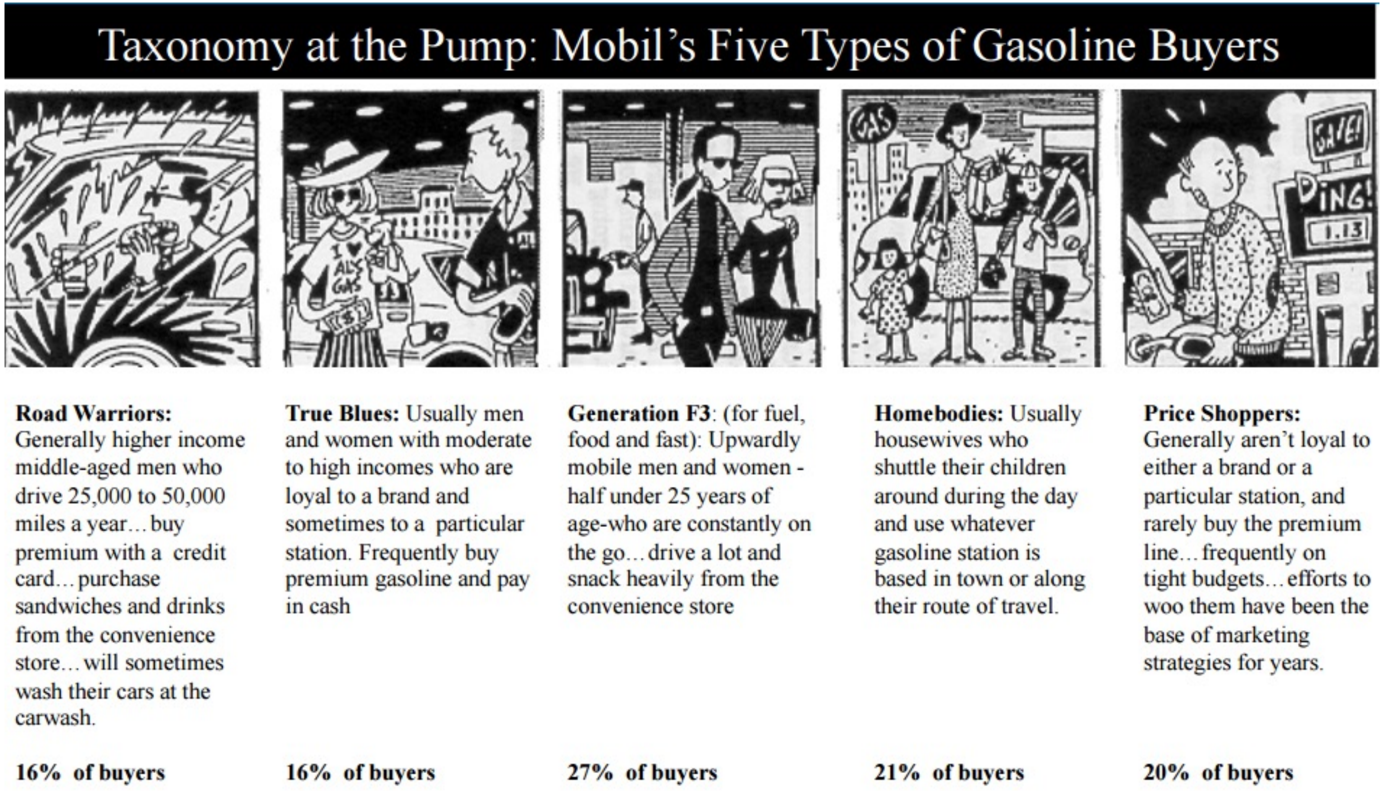
20% of buyers

The Importance of Saying “No”

Focus on the **80%** of the market that was **not** price sensitive…

- Better lighting
- 24-hour stations
- Larger variety snacks and drinks

20% increase in sales



Taxonomy at the Pump: Mobil’s Five Types of Gasoline Buyers

Road Warriors: Generally higher income middle-aged men who drive 25,000 to 50,000 miles a year…buy premium with a credit card…purchase sandwiches and drinks from the convenience store…will sometimes wash their cars at the carwash.

16% of buyers

True Blues: Usually men and women with moderate to high incomes who are loyal to a brand and sometimes to a particular station. Frequently buy premium gasoline and pay in cash

16% of buyers

Generation F3: (for fuel, food and fast): Upwardly mobile men and women - half under 25 years of age-who are constantly on the go…drive a lot and snack heavily from the convenience store

27% of buyers

Homebodies: Usually housewives who shuttle their children around during the day and use whatever gasoline station is based in town or along their route of travel.

21% of buyers

Price Shoppers: Generally aren’t loyal to either a brand or a particular station, and rarely buy the premium line…frequently on tight budgets…efforts to woo them have been the base of marketing strategies for years.

20% of buyers

Targeting Case: Nicorette

Targeting Case: Nicorette

- In clinical trials, the Nicorette patch had proven effective in helping smokers quit

- A study showed that 47.5% of subjects using the nicotine patch abstained from smoking for a period of 3 months or longer … The single most important success factor … was the smoker's **motivation to quit**

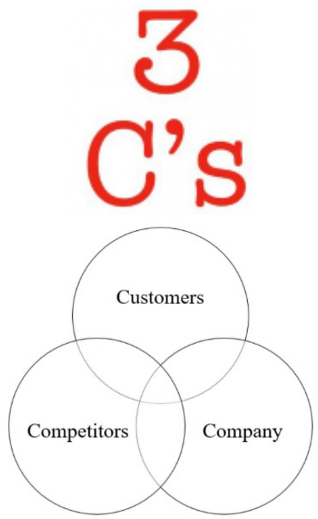
- "Committed quitters" were the most likely to quit smoking successfully, using Nicorette or any other smoking cessation method.

- How would you segment the population?
- Which segment would you target?
- How would you reach the target segment?

Takeaway – Targeting

Market Opportunities for Profit:
- Segment size
- Growth rate/potential

Target Market Selection



Competitive Intensity
- Underserved needs?
- Competitors' strengths

Company "Fit" with
- Objectives
- Competencies / Resources
- Customer Base

Next Class

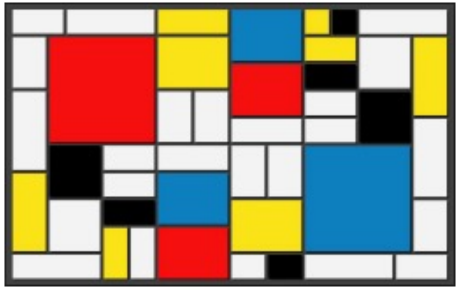
- Market Segmentation concept check due 9AM Sep 17
- Start reading Ford Ka
  - Should be able to answer Questions 1-4
- Read Python for Marketing Research and Analytics (chapter 9)
- We will study segmentation and positioning

Last Time…

Segmentation

S

Discovering and profiling groups of customers with similar needs and preferences



➡

Targeting

T

Evaluating segment attractiveness and targeting most attractive ones

}

Where will we play?

Last Time…

- Segmentation
  - Dividing market into meaningful subsets of customers
  - Cluster Analysis as a technique to group entities such that
    - Objects within a group should be as similar as possible
    - Objects belonging to different groups should be as dissimilar as possible
  - **L**arge, **I**dentifiable, **D**istinctive, **S**table (**LIDS**) and **actionable** (managerially relevant)
- Targeting
  - Choosing attractive segments (customers, company, competitors)
- Key learnings:
  - Segmentation and targeting and type of data used
  - How to segment the market (intuition + implementation)
- Today: segmentation and positioning

Course Roadmap

STP Analytics (Identify Value)	Customer Analytics (Deliver Value)	4P Analytics (Capture Value)
Module 1	**Module 2**	**Module 3**
What datasets can we use? How can we segment and target our customers? How should we position our products/services?	How much are our customers worth? Are our customers leaving? How do our customers make choices?	How do we build a new product? How should we price our products? How do we distribute them? How do we quantify the impact of our promotions?

Today: Segmentation + Positioning

Part 1: Dimension Reduction Techniques

1. Big Data, Companies' Perspectives
2. Factor Analysis + Implementation in Python
3. Novel Techniques
    1. Latent Dirichlet Allocation
    2. (Variational) Autoencoders

Part 2: Application to Segmentation

1. DuPont

Part 3: Application to Positioning

1. Perceptual Maps (Beers)

Today’s Goals

Understand:

- Factor analysis and recent dimension reduction techniques
- How to perform segmentation with a large number of variables
- How to build a perceptual map

Be able to:

- Implement factor analysis in Python
- Choose a suitable number of factors
- Interpret the results of factor analysis

Dimension Reduction

Why?

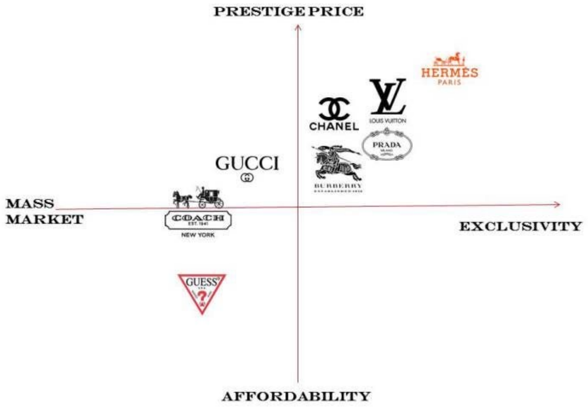
Big Datasets

- Last time: cluster analysis using 5 variables (questions)

- But an important question remains:
  - What if we have 800 variables?
    - Example: How much data can Netflix collect about you?
    - What you watch? When? Do you stop often? “Old-school” watcher vs binge watcher?
  - What is the issue?
    - Hard to interpret
    - Complexity
    - Some variables may be correlated

Company’s Perspective

- Imagine a fashion brand performed a segmentation analysis of luxury brands
  - Segments along low vs high price and mass market vs exclusivity
- What does the brand need?
  - Understand where the brand lives in customers’ minds



Multicollinearity - Bank Service Survey

A national bank wants to create a new "spin-off" brand, to target certain segments of the personal banking market.

To help ***position*** *this new brand, they conducted a survey about* ***consumers' attitudes toward banking****.*

Scale from 1 (Strongly Disagree) to 10 (Strongly Agree):

Q1: Small banks charge less than large banks.

Q2: Large banks are more likely to make mistakes than small banks.

Q3: Tellers do not need to be extremely courteous and friendly; it's enough for them simply to be civil.

Q4: I want to be known personally at my bank and be treated with special courtesy.

Q5: If a financial institution treated me in an impersonal or uncaring way, I would never patronize that organization again.

Regression Analysis: Bank Data

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_j$$

Y =

- Dependent variable
- Outcome variable

Xs =

- Predictor variables
- Features
- Independent variables

βs =

- Coefficients
- Parameters

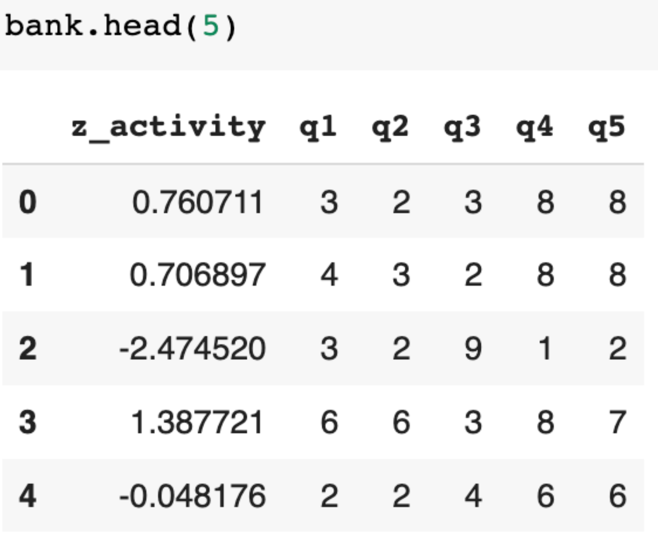
Note the jargon!

Number of times you go to the bank

Survey questions

Linking Attitudes to Behaviors

Question: do consumers' attitudes explain how much customers use the bank?



```
bank.head(5)
```

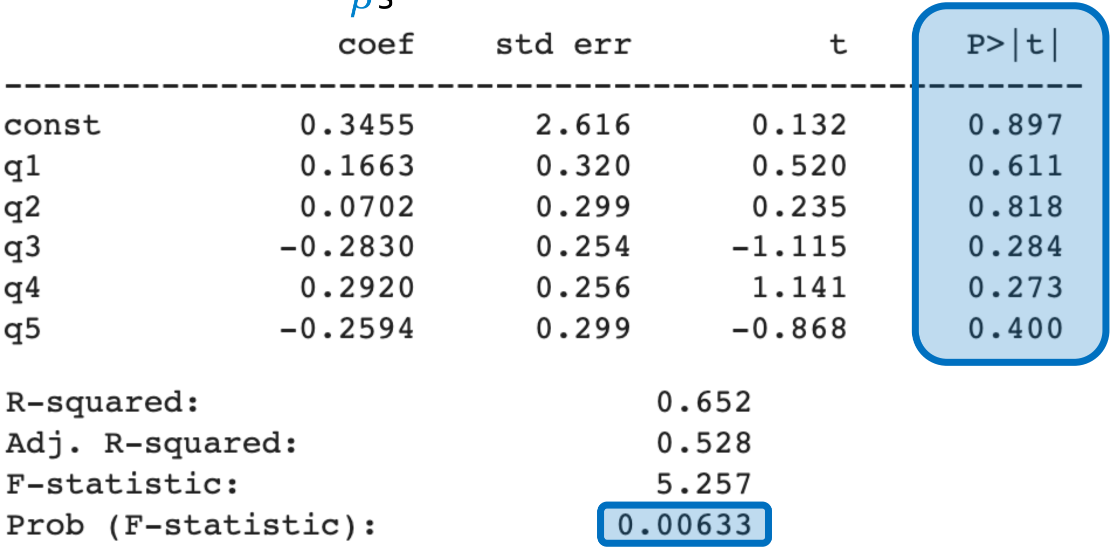
	z_activity	q1	q2	q3	q4	q5
0	0.760711	3	2	3	8	8
1	0.706897	4	3	2	8	8
2	-2.474520	3	2	9	1	2
3	1.387721	6	6	3	8	7
4	-0.048176	2	2	4	6	6

Y

Xs

We can use a linear regression to estimate the linear relationship between these variables and banking activity

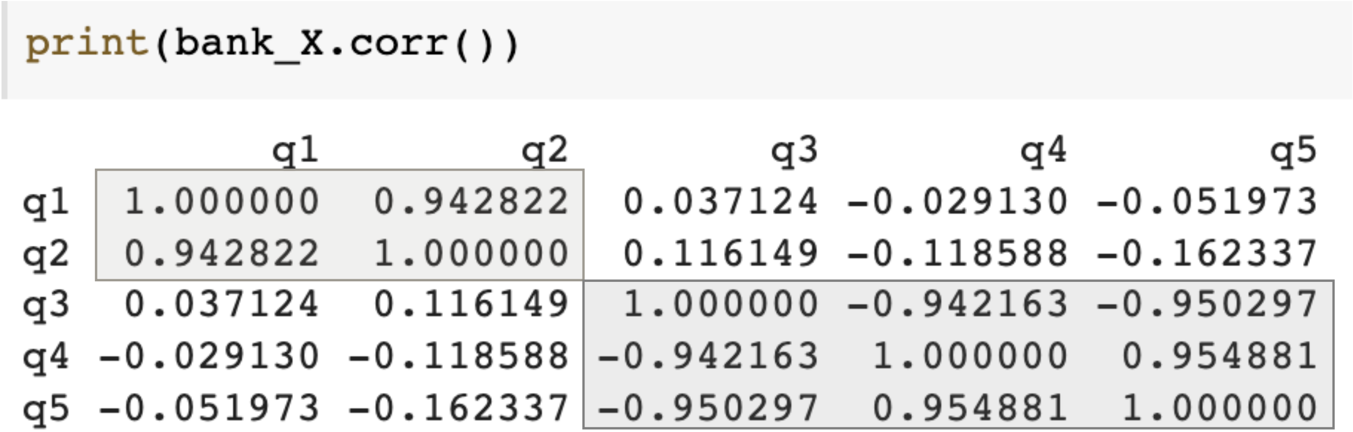
βs



```
                 coef    std err          t      P>|t|
------------------------------------------------------
const          0.3455      2.616      0.132      0.897
q1             0.1663      0.320      0.520      0.611
q2             0.0702      0.299      0.235      0.818
q3            -0.2830      0.254     -1.115      0.284
q4             0.2920      0.256      1.141      0.273
q5            -0.2594      0.299     -0.868      0.400

R-squared:                        0.652
Adj. R-squared:                   0.528
F-statistic:                      5.257
Prob (F-statistic):             0.00633
```

Multicollinearity!

Multicollinearity – Correlated Questions



```
print(bank_X.corr())
```

```
          q1        q2        q3        q4        q5
q1  1.000000  0.942822  0.037124 -0.029130 -0.051973
q2  0.942822  1.000000  0.116149 -0.118588 -0.162337
q3  0.037124  0.116149  1.000000 -0.942163 -0.950297
q4 -0.029130 -0.118588 -0.942163  1.000000  0.954881
q5 -0.051973 -0.162337 -0.950297  0.954881  1.000000
```

- Are these five questions truly independent, or are they measuring the same thing?
- Can we convert many questions into a few independent factors?

Dimension Reduction

How? Factor Analysis

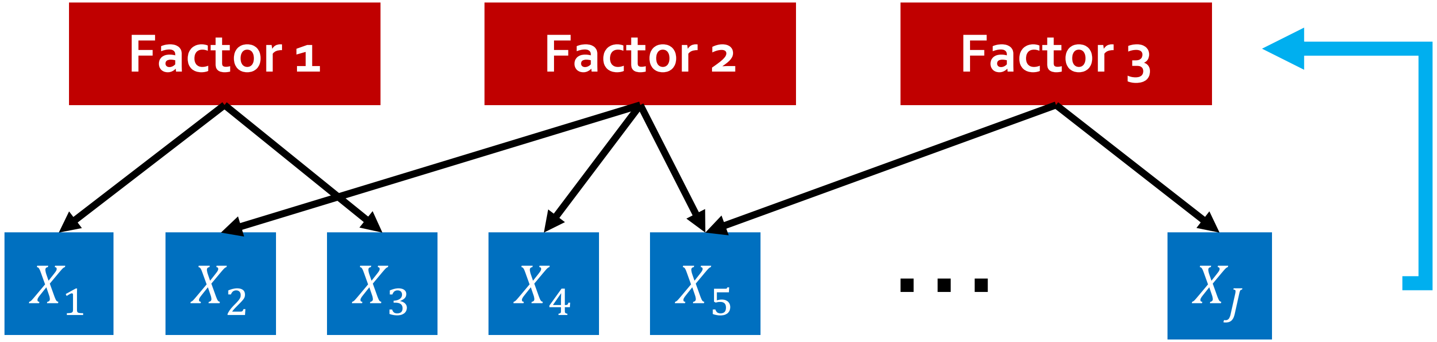
Factor Analysis

- **Factor analysis** is a statistical technique that aims to
  - Replace an initial set of variables with a smaller number of “factors”
  - **Factors** reflect what sets of variables have in common with one another

- One of multiple ***data or dimension reduction*** techniques
  - Multidimensional Scaling
  - T-SNE
  - Autoencoders…

Factor Analysis: Intuition

- Starting point: a set of variables (questions): $X_1, X_2, \ldots, X_J$
- $X_1, X_2, \ldots, X_J$ may be derived from a few underlying "concepts" or ***factors***



We have to work backwards!

Problem: we observe $X_1, X_2, \ldots, X_J$, but not the factors!

→ Goals of factor analysis: <u>what</u> are these factors, <u>how many</u> are there, and <u>how</u> do they relate to the original X's?

What are the Factors?

Our customers say we are doing well on Q1, Q2, Q3, but poorly on Q4, Q5, Q6, and on Q7 and Q8.

vs.

*Our customers say we are doing well on **value** but poorly on **quality** and **experience**.*

These are factors!

Interpretation:

- Q1, Q2, Q3 are measuring value
- Q4, Q5, Q6 are measuring quality
- Q7, Q8 are measuring experience

Back to Banking: Survey

A national bank wants to create a new "spin-off" brand, to target certain segments of the personal banking market.

To help ***position*** *this new brand, they conducted a survey about* ***consumers' attitudes toward banking****.*

Scale from 1 (Strongly Disagree) to 10 (Strongly Agree):

Q1: Small banks charge less than large banks.

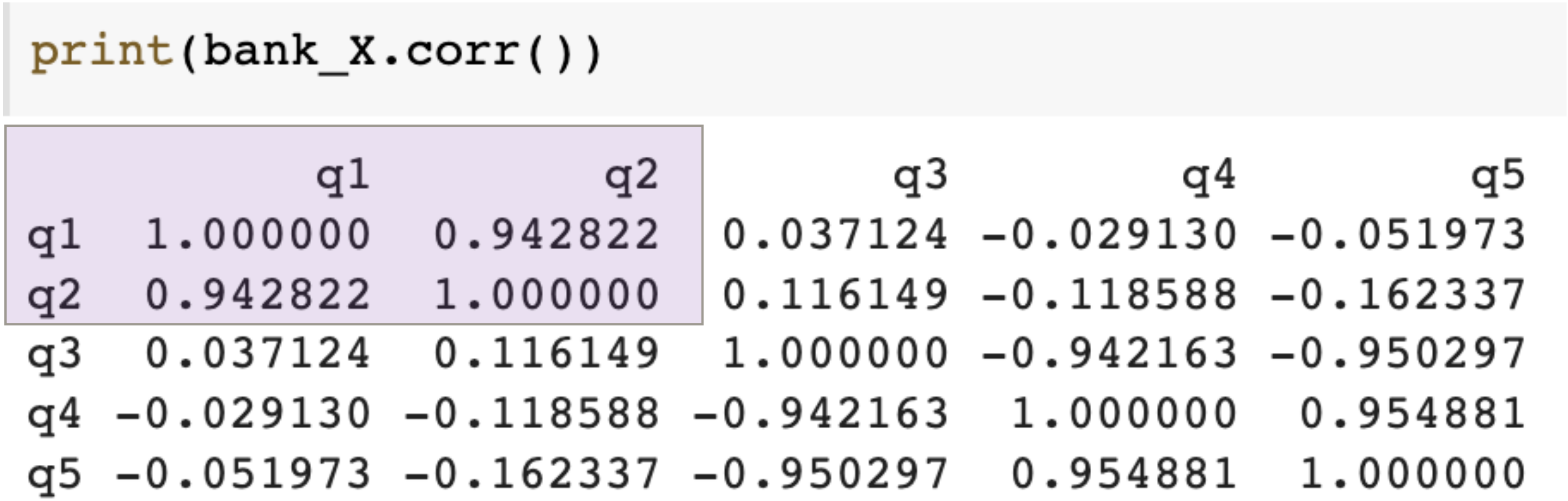
Q2: Large banks are more likely to make mistakes than small banks.

Q3: Tellers do not need to be extremely courteous and friendly; it's enough for them simply to be civil.

Q4: I want to be known personally at my bank and be treated with special courtesy.

Q5: If a financial institution treated me in an impersonal or uncaring way, I would never patronize that organization again.

What is the Factor Underlying Q1 & Q2?



```
print(bank_X.corr())
```

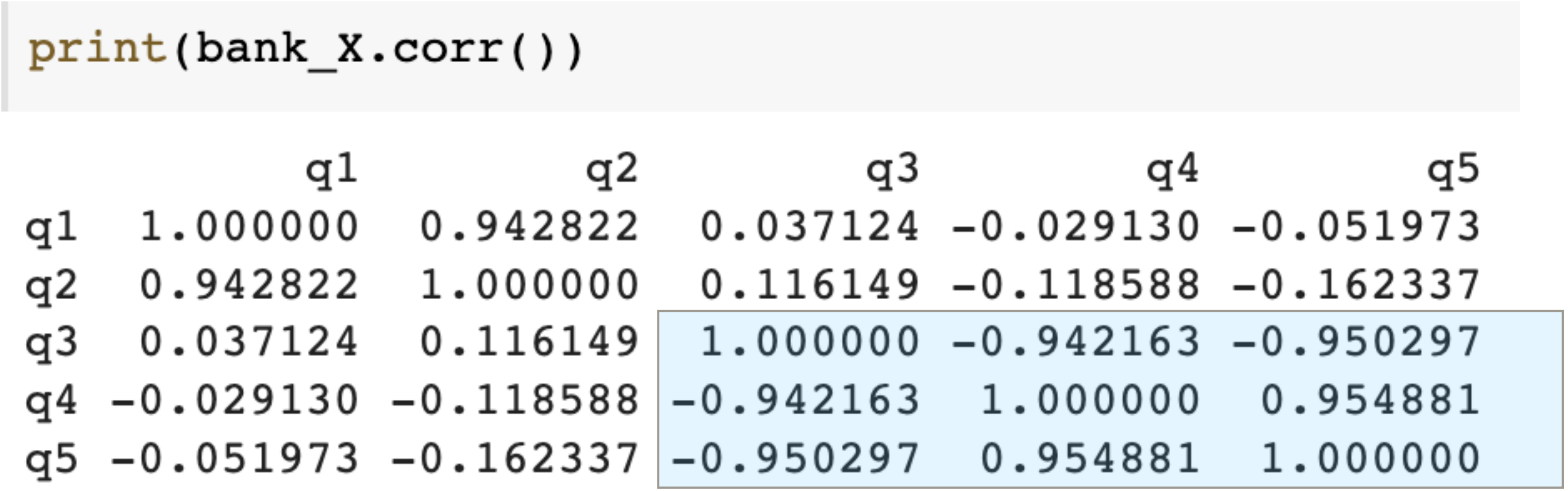
```
          q1        q2        q3        q4        q5
q1  1.000000  0.942822  0.037124 -0.029130 -0.051973
q2  0.942822  1.000000  0.116149 -0.118588 -0.162337
q3  0.037124  0.116149  1.000000 -0.942163 -0.950297
q4 -0.029130 -0.118588 -0.942163  1.000000  0.954881
q5 -0.051973 -0.162337 -0.950297  0.954881  1.000000
```

On a scale from 1 (Strongly Disagree) to 10 (Strongly Agree):

Q1: Small banks charge less than large banks.

Q2: Large banks are more likely to make mistakes than small banks.

What is the Factor Underlying Q3, Q4 & Q5?



```
print(bank_X.corr())
```

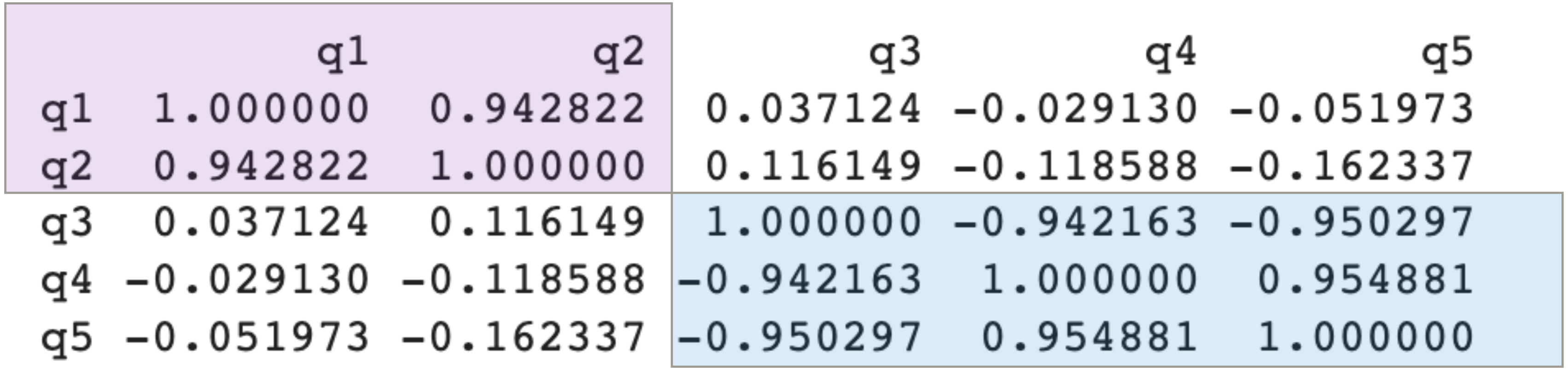
```
          q1        q2        q3        q4        q5
q1  1.000000  0.942822  0.037124 -0.029130 -0.051973
q2  0.942822  1.000000  0.116149 -0.118588 -0.162337
q3  0.037124  0.116149  1.000000 -0.942163 -0.950297
q4 -0.029130 -0.118588 -0.942163  1.000000  0.954881
q5 -0.051973 -0.162337 -0.950297  0.954881  1.000000
```

Q3: Tellers do not need to be extremely courteous and friendly; it's enough for them simply to be civil.

Q4: I want to be known personally at my bank and be treated with special courtesy.

Q5: If a financial institution treated me in an impersonal or uncaring way, I would never patronize that organization again.

Back to Banking…



```
          q1        q2        q3        q4        q5
q1  1.000000  0.942822  0.037124 -0.029130 -0.051973
q2  0.942822  1.000000  0.116149 -0.118588 -0.162337
q3  0.037124  0.116149  1.000000 -0.942163 -0.950297
q4 -0.029130 -0.118588 -0.942163  1.000000  0.954881
q5 -0.051973 -0.162337 -0.950297  0.954881  1.000000
```

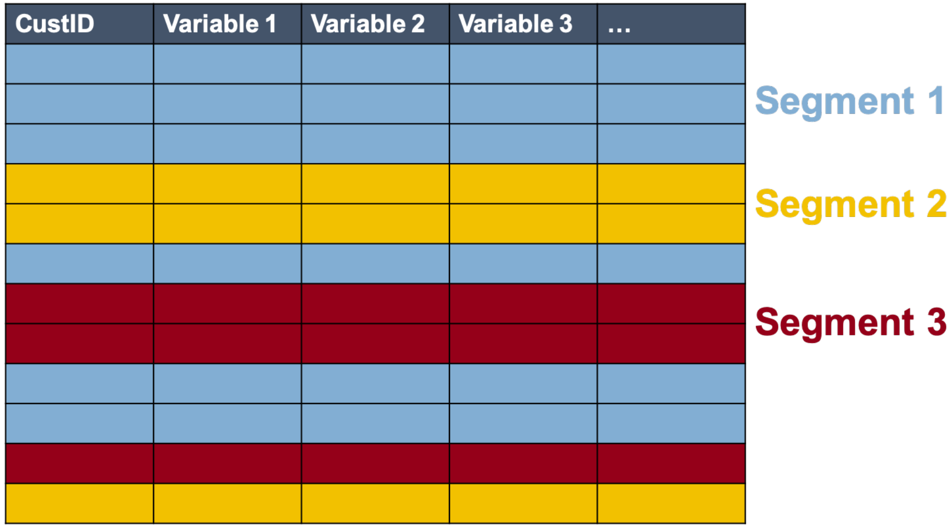
Two factors:

- Q1-2: "Smaller banks are better"
- Q3-5: "Personal touch"

Factor analysis looks for these "blocks" of correlation:

- High correlation ***within*** blocks
- Low correlation ***across*** blocks

Intuition: Clusters vs Factors



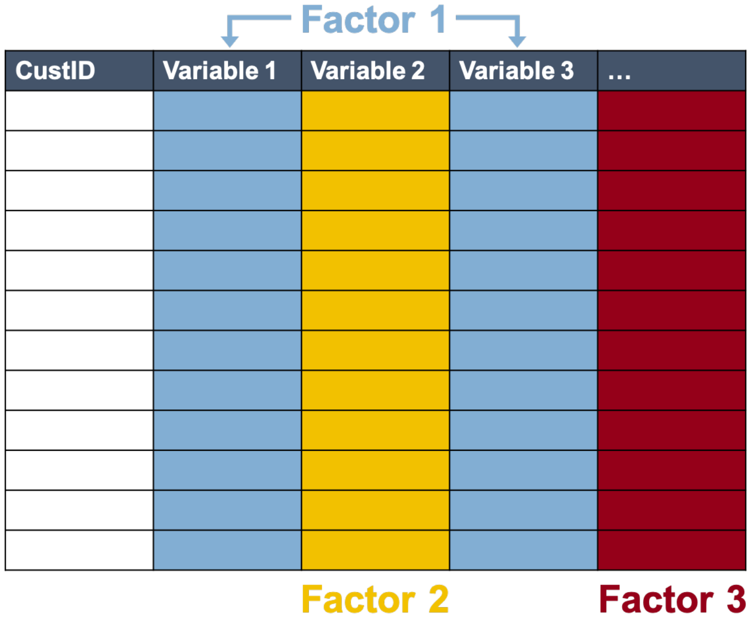
CustID | Variable 1 | Variable 2 | Variable 3 | …

Segment 1

Segment 2

Segment 3

Cluster Analysis:
using the columns to
group the rows



Factor 1

CustID | Variable 1 | Variable 2 | Variable 3 | …

Factor 2

Factor 3

Factor Analysis:
using the rows to
group the columns

Factor Analysis: The Math

Assumption: each variable $(X_1, X_2, \dots, X_J)$ can be represented as a linear combination of K underlying factors, $F_1, F_2, \dots, F_K$

$$
\begin{aligned}
X_1 &= l_{11}F_1 + l_{12}F_2 + \cdots + l_{1K}F_K + \epsilon_1 \\
X_2 &= l_{21}F_1 + l_{22}F_2 + \cdots + l_{2K}F_K + \epsilon_2 \\
&\vdots \\
X_J &= l_{J1}F_1 + l_{J2}F_2 + \cdots + l_{JK}\,F_K + \epsilon_J
\end{aligned}
$$

“coefficients” = **factor loadings** (i.e., how much does that factor explain the X?)

Basically, a regression where X is the dependent variable and the factors are the independent variables! But we do not know the factors.

What Makes a Useful Factor Structure?

- Factors retain as much of the original information as possible, with the fewest number of factors
  - **Dimensionality reduction**: # factors << J, but contains similar information
  - Basically: you could recreate the X's using the factors
- For interpretation: the factors are **uncorrelated**
  - Also called **orthogonal**
  - Basically: independent concepts
- How do we achieve this? Principal Components Analysis

Principal Components Analysis (PCA) & Factor Analysis

	Objectives	Method
Principal Components Analysis	Find uncorrelated linear dimensions that capture maximal variance in the data	Algebra-based
Factor Analysis	Capture variance with a small number of dimensions while aiming to make the dimensions interpretable in terms of the original variables	Optimization-based

Note: In this class, we will use these interchangeably

Principal Component Analysis (PCA)

- Algebra-based technique to determine the factors
  - Looks for uncorrelated "axes" that describe the data - principal components
- **Principal components** are new variables that are constructed as linear combinations of all the initial variables
  - Note that some contribute minimally
- Example: can we reduce Q1 and Q2 into one component?

Q1: Small banks charge less than large banks.

Q2: Large banks are more likely to make mistakes than small banks.

↓

Factor: "Smaller banks are better"

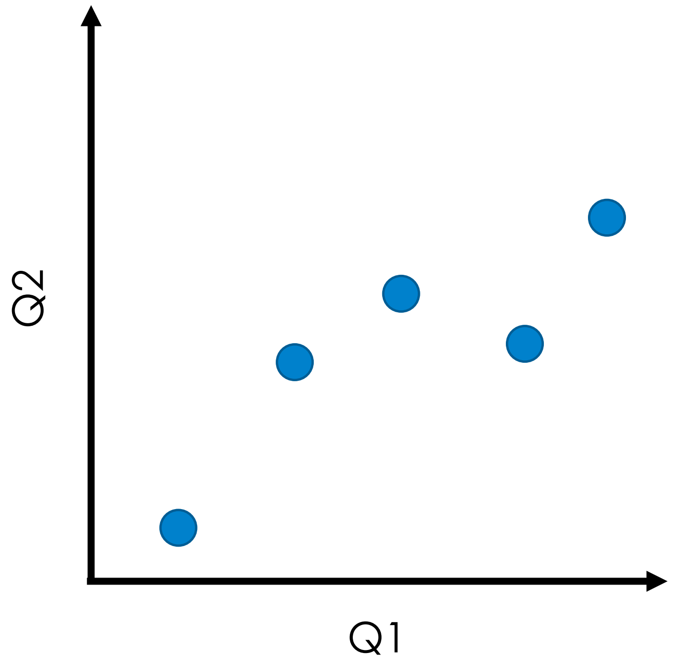
Principal Component Analysis (PCA)

Graphically, we can represent the relationship between these variables:

Q1: Small banks charge less than large banks.

Q2: Large banks are more likely to make mistakes than small banks.

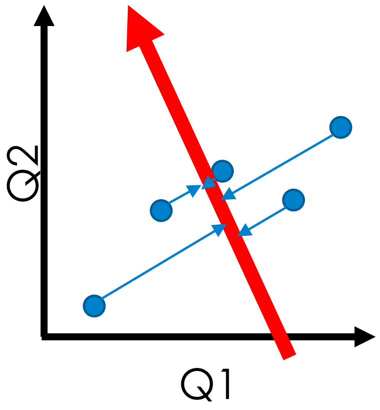
These variables are positively correlated. Can we reduce them into one factor?

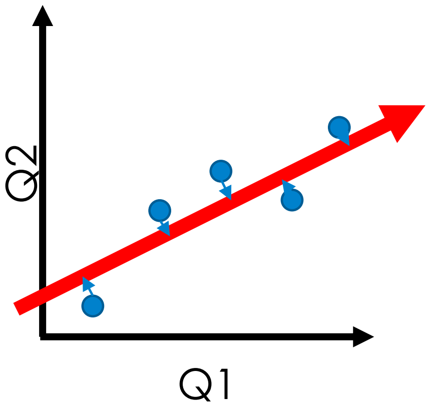


Principal Component Analysis (PCA)

How does PCA work?

- We want to project the data onto a line that captures as much variance (information) as possible
  - We want the projected data points to be as spread out as possible



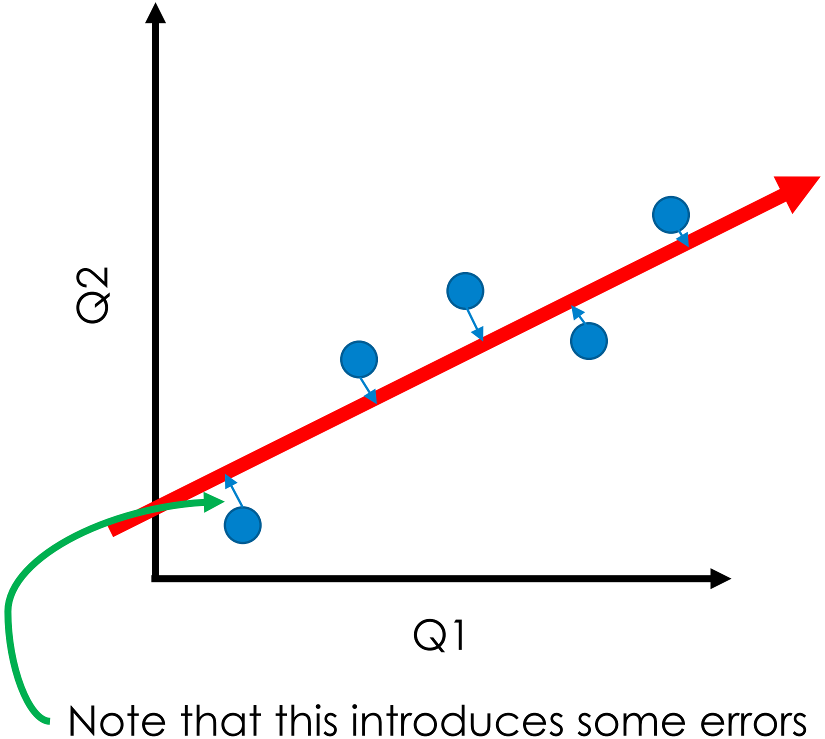


Which captures
more information?

Principal Component Analysis (PCA)

How does PCA work?

- We want to project the data onto a line that captures as much variance (information) as possible
    - We want the projected data points to be as spread out as possible
- We project each observation to the line,

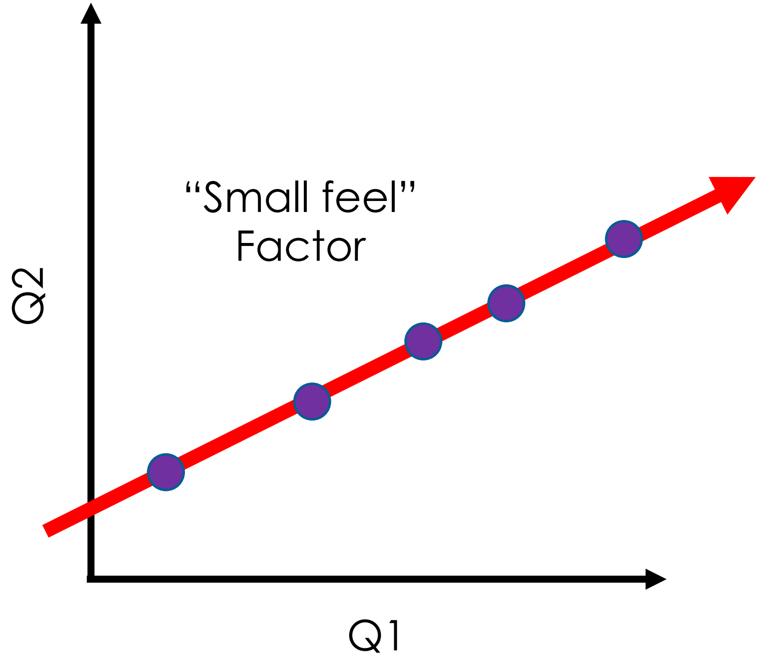


Note that this introduces some errors (gap between line and points)

Principal Component Analysis (PCA)

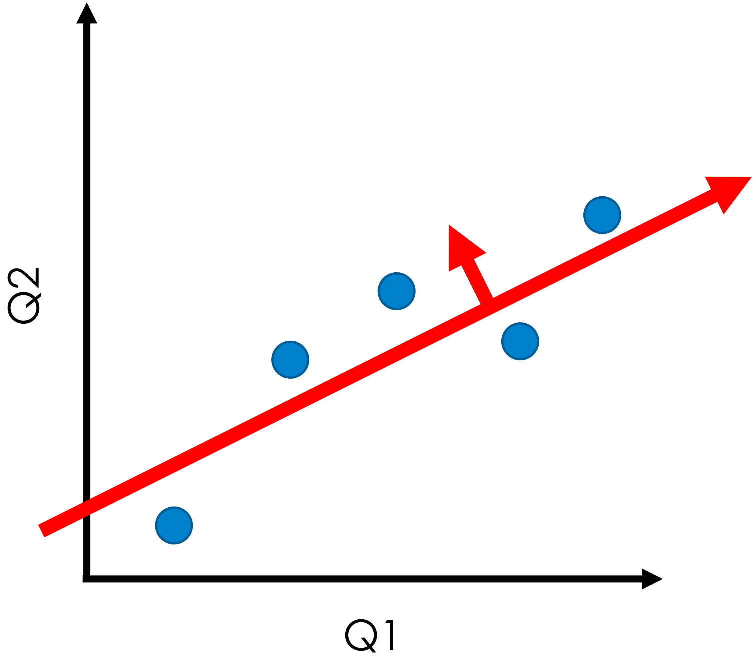
What is PCA doing?

- We want to project the data onto a line that captures as much variance (information) as possible.
  - We want the projected data points to be as spread out as possible
- We project each observation to the line,
- and…
- Our dataset changes a bit, but it is much simpler to interpret
- Now, we only have one axis, which is the "small feel" feature



Principal Components Analysis (PCA)

- PCA is based on the eigen decomposition of the correlation matrix (by definition standardized)
  - If run PCA on covariance matrix, need to first standardize the data
- **Principal component** = scaled **eigenvectors**
- The variance explained by each component is proportional to its **eigenvalue**
- # possible components = # variables
  - First component captures most variance, but the remaining errors (gap) is captured by the second component
  - How to select K (# of factors)? Variance explained!



How do we interpret PCA? Rotation

- Once K (# of factors) is selected, we want to replace the variables by the factors
- Often, the results of PCA are not intuitive:
  - Example: Q1 = -0.3 * F1 + 0.6 * F2
  - What we want: most equations load onto one factor, but not others

- Solution: “rotate” the results
  - Varimax: given K principal components, find the most interpretable rotated components
  - The above example becomes: Q1 = 0 * F1’ + 0.99 * F2’

Varimax

- Varimax is an orthogonal rotation, so it preserves the orthogonal structure of the principal components

- Objective: maximize the variance of the squared loadings of a factor on all the variables to have either large or small loadings on each variable
  - Maximizing the variance makes the distributions of the squared loadings as spread out as possible, so some high and some low

Steps to Factor Analysis with PCA

1. Estimate all the principal components (**without rotation**)

2. Determine the number of components (factors) to keep:
   1. Eigenvalues>1, Cumulative Variance>80%, scree plot, managerial relevance

3. Compute the **rotated** factor loading matrix to understand (and name!) the underlying factors

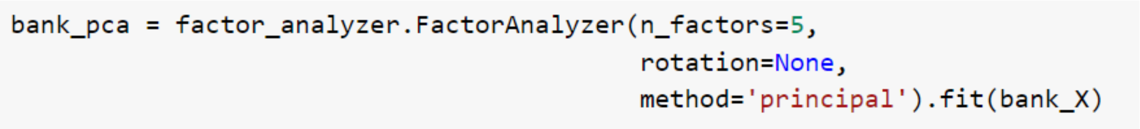
4. Compute the factor scores:
   1. For each observation, what are the values of the factors?

Banking: Full Analysis

We use as many factors as variables

Step 1: PCA, all components, no rotation

We will use the package *factor_analyzer*



```
bank_pca = factor_analyzer.FactorAnalyzer(n_factors=5,
                                          rotation=None,
                                          method='principal').fit(bank_X)
```

Banking: Full Analysis

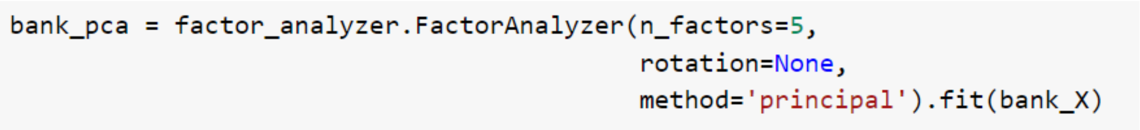
Step 1: PCA, all components, no rotation

Step 2: Determine the number of components to keep

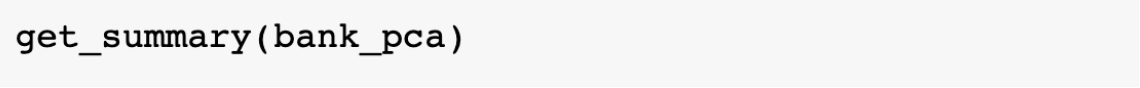
Criteria for keeping a component:

- Variance (Sum of Squares Loadings/Eigenvalues) > 1
- Cumulative var > 80%

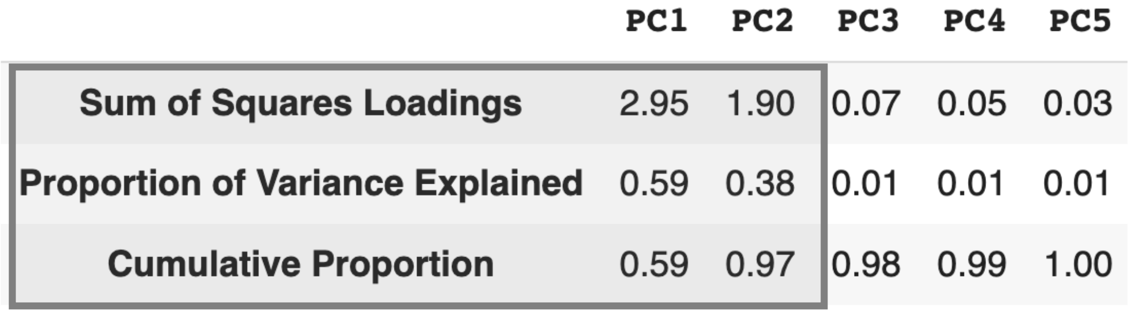
We will use the package *factor_analyzer*



```
bank_pca = factor_analyzer.FactorAnalyzer(n_factors=5,
                                          rotation=None,
                                          method='principal').fit(bank_X)
```



```
get_summary(bank_pca)
```



	PC1	PC2	PC3	PC4	PC5
Sum of Squares Loadings	2.95	1.90	0.07	0.05	0.03
Proportion of Variance Explained	0.59	0.38	0.01	0.01	0.01
Cumulative Proportion	0.59	0.97	0.98	0.99	1.00

Banking: Full Analysis

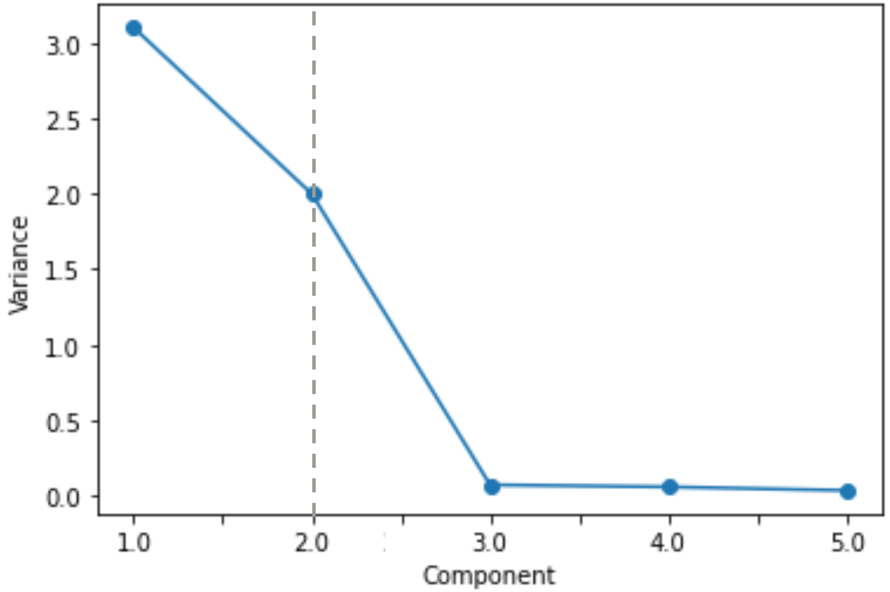
Step 1: PCA, all components, no rotation

Step 2: Determine the number of components to keep

Criteria for keeping a component:

- Variance (Sum of Squares) > 1
- Cumulative var > 80%
- Scree plot elbow (elbow – 1)

Scree Plot



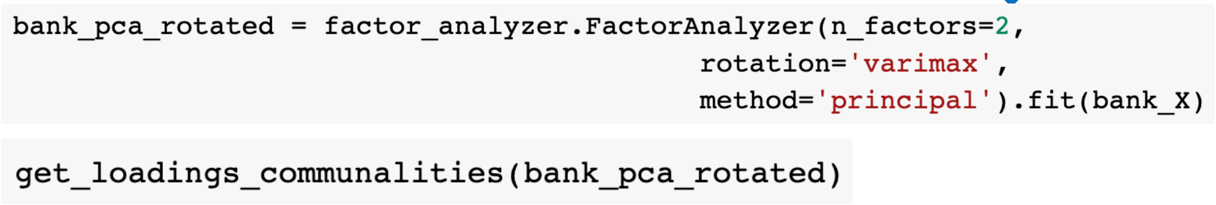
Banking: Full Analysis

Step 1: PCA, all components, no rotation

Step 2: Determine the number of components to keep

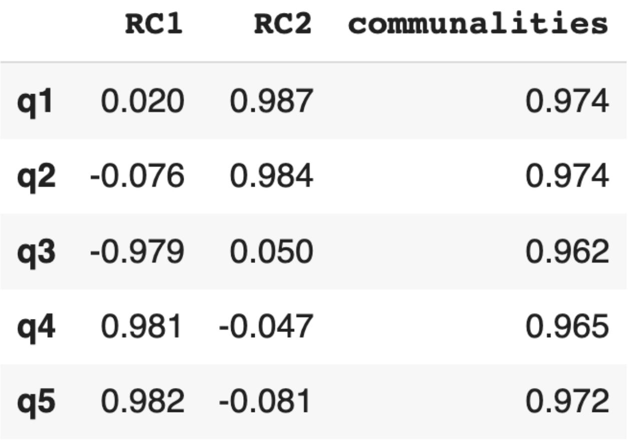
Step 3: Understand the retained, rotated factors

Selected # factors



```
bank_pca_rotated = factor_analyzer.FactorAnalyzer(n_factors=2,
                                                  rotation='varimax',
                                                  method='principal').fit(bank_X)
```

```
get_loadings_communalities(bank_pca_rotated)
```



	RC1	RC2	communalities
q1	0.020	0.987	0.974
q2	-0.076	0.984	0.974
q3	-0.979	0.050	0.962
q4	0.981	-0.047	0.965
q5	0.982	-0.081	0.972

Banking: Full Analysis

Step 1: PCA, all components, no rotation

Step 2: Determine the number of components to keep

Step 3: Understand the retained, rotated factors

Factor loadings

Q1, Q2 load highly on factor 2
The "Small feel" Factor
-Q3, Q4, Q5 load highly on factor 1
The "Personal touch" Factor

Q5 = 0.98 * RC1 – 0.08 * RC2

	RC1	RC2	communalities
q1	0.020	0.987	0.974
q2	-0.076	0.984	0.974
q3	-0.979	0.050	0.962
q4	0.981	-0.047	0.965
q5	0.982	-0.081	0.972

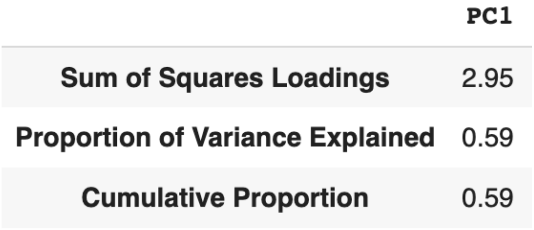
Communalities (h^2)

How much of the variance in the original variable is captured by the common factors?
97.4% of the variance in Q1 is explained by RC1 and RC2

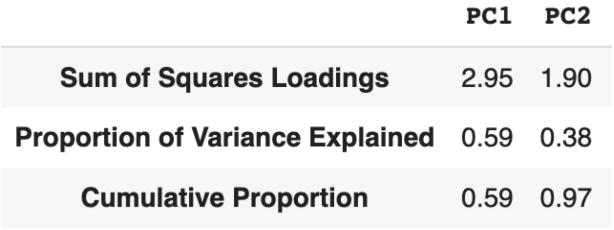
Varimax Clarification

- Varimax does not change the total amount of variance explained

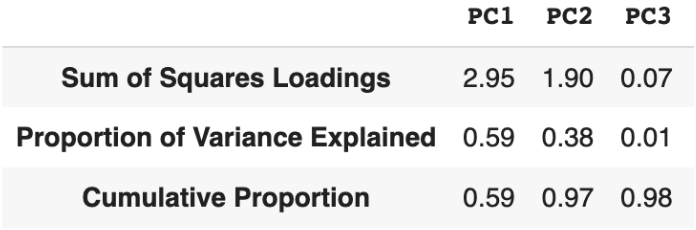
No Rotation



	PC1
Sum of Squares Loadings	2.95
Proportion of Variance Explained	0.59
Cumulative Proportion	0.59

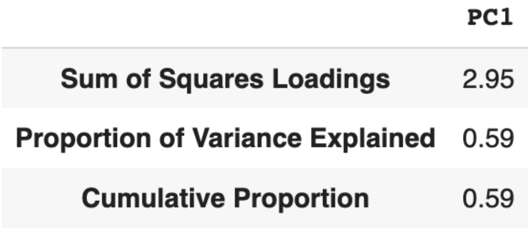


	PC1	PC2
Sum of Squares Loadings	2.95	1.90
Proportion of Variance Explained	0.59	0.38
Cumulative Proportion	0.59	0.97

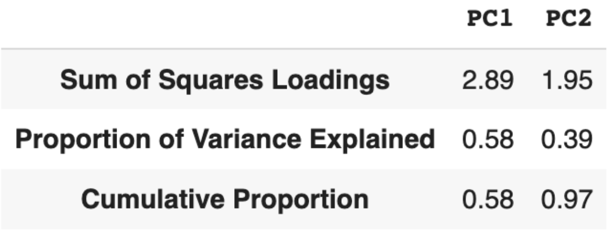


	PC1	PC2	PC3
Sum of Squares Loadings	2.95	1.90	0.07
Proportion of Variance Explained	0.59	0.38	0.01
Cumulative Proportion	0.59	0.97	0.98

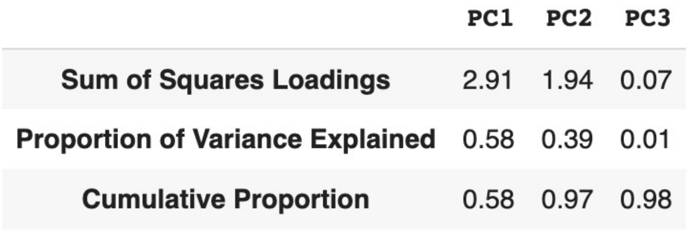
With Varimax Rotation



	PC1
Sum of Squares Loadings	2.95
Proportion of Variance Explained	0.59
Cumulative Proportion	0.59



	PC1	PC2
Sum of Squares Loadings	2.89	1.95
Proportion of Variance Explained	0.58	0.39
Cumulative Proportion	0.58	0.97



	PC1	PC2	PC3
Sum of Squares Loadings	2.91	1.94	0.07
Proportion of Variance Explained	0.58	0.39	0.01
Cumulative Proportion	0.58	0.97	0.98

Important: PCA *without* rotation should be used to determine how many factors there are (once we rotate, we are changing the structure of the data), rotation helps with interpretation

Banking: Full Analysis

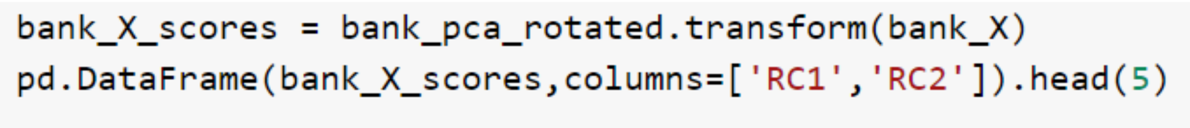
Step 1: PCA, all components, no rotation

Step 2: Determine the number of components to keep

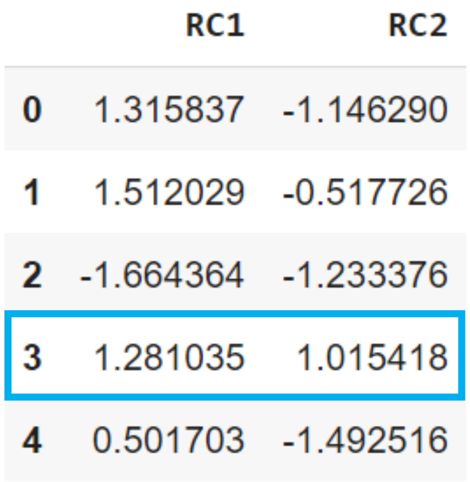
Step 3: Understand the retained, rotated factors

Step 4: Compute the factor scores (translation of original data into factors)

- Multiply standardized data by principal components



```
bank_X_scores = bank_pca_rotated.transform(bank_X)
pd.DataFrame(bank_X_scores,columns=['RC1','RC2']).head(5)
```



```
        RC1        RC2
0  1.315837  -1.146290
1  1.512029  -0.517726
2 -1.664364  -1.233376
3  1.281035   1.015418
4  0.501703  -1.492516
```

Respondent 3’s Factor Scores

These are the standardized scores (z-scores)

For this respondent: RC1 = 1.28, RC2 = 1.02

Interpretation: *Respondent 3 scores 1.28 SD above the mean for RC1, 1.02 SDs above the mean for RC2.*

Jargon Summary

- Loadings = how the original variables relate to the factors
  - E.g., Q5 = 0.98 * RC1 – 0.08 * RC2

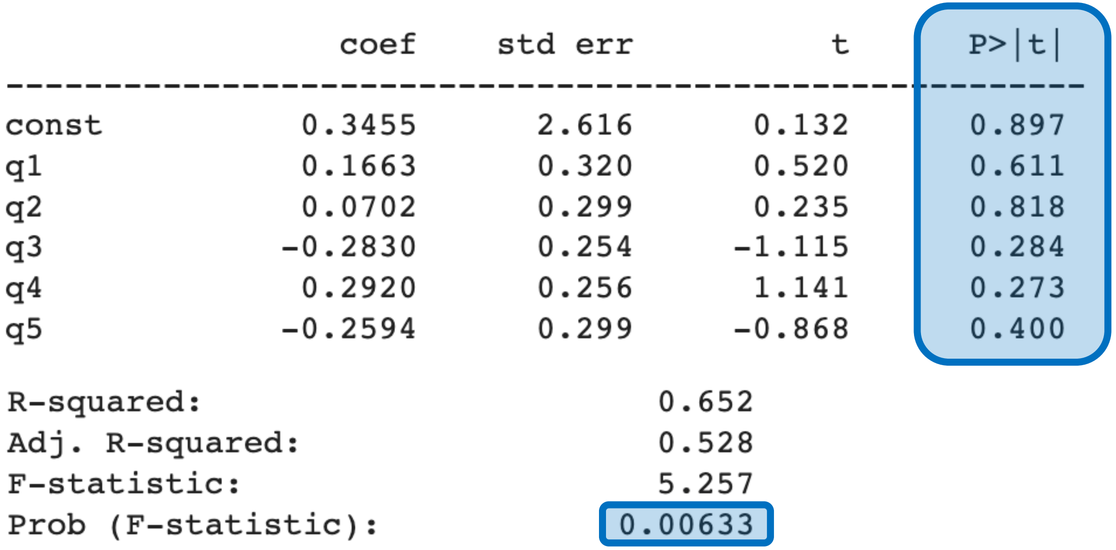
- Communalities = how much variability in the original variables is explained by the factors
  - E.g., communality of Q1 is 0.974 → Using 2 components is sufficient to approximate 97.4% of the variation in Q1 (i.e., the error that remains is small)

- Scores = translation of original data into factors
  - E.g., Respondent 3 scores 1.28 SD above the mean for RC1, 1.02 SDs above the mean for RC2

Let’s go to Python

PCA

Back to our Problematic Regression



```
                 coef    std err          t      P>|t|
------------------------------------------------------
const          0.3455      2.616      0.132      0.897
q1             0.1663      0.320      0.520      0.611
q2             0.0702      0.299      0.235      0.818
q3            -0.2830      0.254     -1.115      0.284
q4             0.2920      0.256      1.141      0.273
q5            -0.2594      0.299     -0.868      0.400

R-squared:                     0.652
Adj. R-squared:                0.528
F-statistic:                   5.257
Prob (F-statistic):          0.00633
```

Multicollinearity!

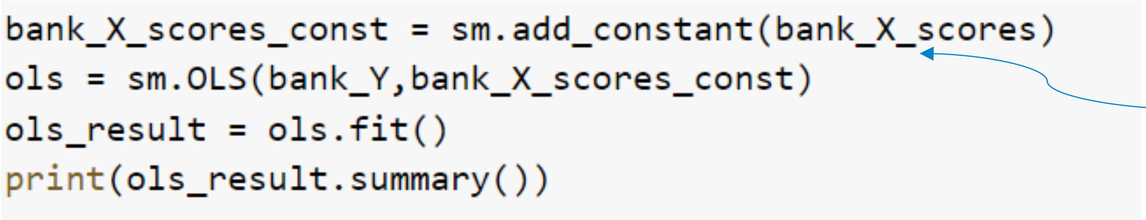
How can we fix this?

One solution: Principal Components Regression

Super easy: replace original X variables with factor scores (reduced dimension X's)!

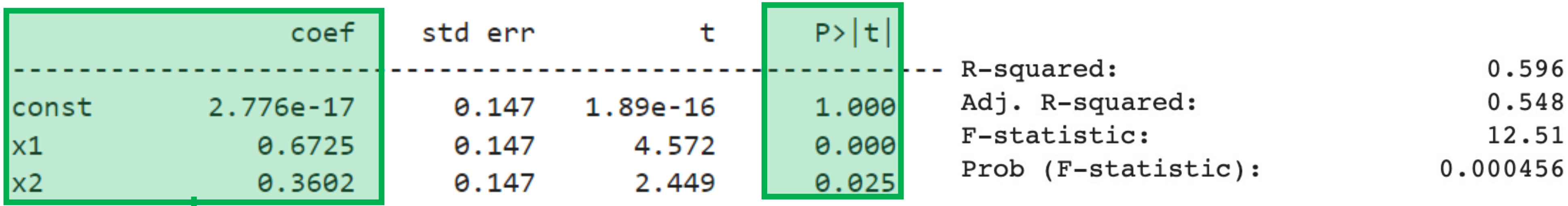
- Remember the goals of factor analysis:
  - Reduce # variables
  - Retain same information
- Factors are uncorrelated → no multicollinearity

Back to our Problematic Regression



```
bank_X_scores_const = sm.add_constant(bank_X_scores)
ols = sm.OLS(bank_Y,bank_X_scores_const)
ols_result = ols.fit()
print(ols_result.summary())
```

Factor scores



```
                 coef    std err          t      P>|t|
------------------------------------------------------------ R-squared:                       0.596
const       2.776e-17      0.147   1.89e-16      1.000       Adj. R-squared:                  0.548
x1             0.6725      0.147      4.572      0.000       F-statistic:                     12.51
x2             0.3602      0.147      2.449      0.025       Prob (F-statistic):           0.000456
```

Interpretation?

Scoring higher on both factors is significantly associated with higher activity → **Favoring small banks** and **need for personal touch** are both significantly associated with higher activity

Takeaway: Factor Analysis Basics

- The goal of factor analysis: uncover underlying structure between many variables

- Good factors: uncorrelated, capture as much of the original variance as possible

- Factors are often intuitive, easier to use, and managerially interesting

Dimension Reduction

Techniques for Unstructured Data

Topic Modeling

- Automatic summarization of documents through **topics**
  - Statistical definition: topic = set of commonly co-occurring words
  - Example: in tablet reviews, “Apple, iPad, iTunes, Mac” = Apple topic
- Intuition: factor analysis for documents!

many words → few interpretable topics

- Uses:
  - Information retrieval and automatic labeling
  - Discovering patterns
  - Predicting outcomes from topics
- Most common model: **Latent Dirichlet Allocation (LDA)**
- In Python: sklearn, nltk, gensim

Latent Dirichlet Allocation (LDA)

One use: as input to regression!
"Which topics are predictive of my outcome?"

Output 1: Which words belong to which topics (i.e., what are the topics)?

Note: You have to set the number of topics in advance!

Topic 2: "kindl" "fire" "amazon" "read" "book"	**Topic 4** "screen" "good" "touch" "nice" "like"	**Topic 7** "great" "product" "love" "purchas" "bought"	**Topic 8** "ipad" "like" "much" "appl" "use"	**Topic 10** "problem" "work" "day" "back" "tri"

Output 2: Which topics best describe each document (i.e., what percentage of the words in a given document are from topic 1, topic 2, …)?

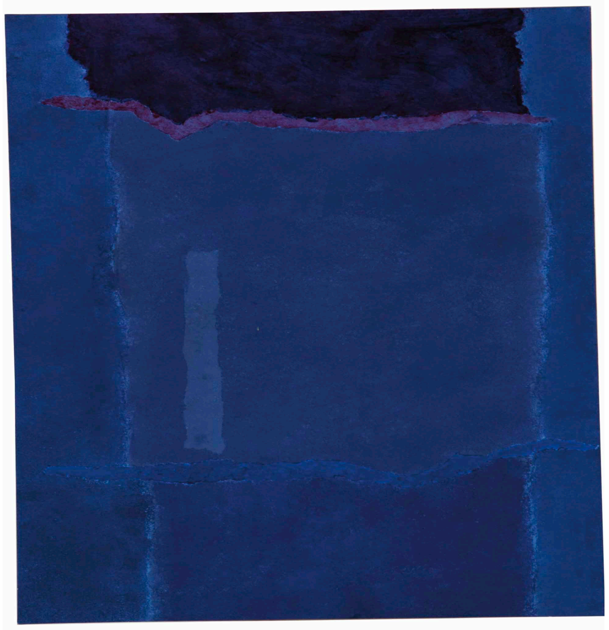
I love my fire and highly recommend it to anyone who wants to watch videos (netflix, hulu, amazon), read ebooks (purchased or from the local library), surf the net and play games. I work in the tech field and I LOVE apple entertainment products (I own many apple products and at work I work with several). I am very thrilled with my fire (I LOVE IT TOO!) because it works great as an entertainment product (and more affordable than my apple products). I also think the fire is a great product because of Amazons cloud and support

Topic	Proportion
1	0.09
2	0.15
3	0.10
4	0.05
5	0.05
6	0.07
7	0.17
8	0.12
9	0.07
10	0.12

Image Analysis

What Makes Art Valuable?

(1957)
Mark Rothko (1903-1970)
No. 17
$32,645,000 (Christies 2016)
Post-War and Contemporary Art Evening Sale

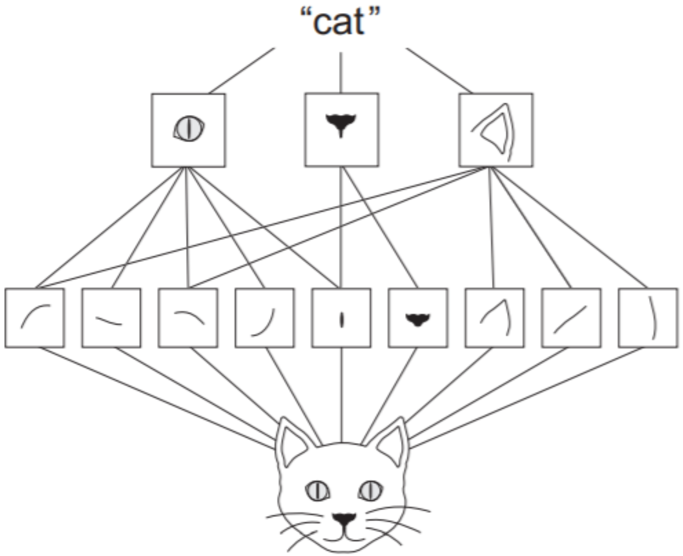


(1978)
Theodoros Stamos (1922-1997)
Infinity Field, Lefkada Series #4
$10,625 (Christies 2013)
Interiors

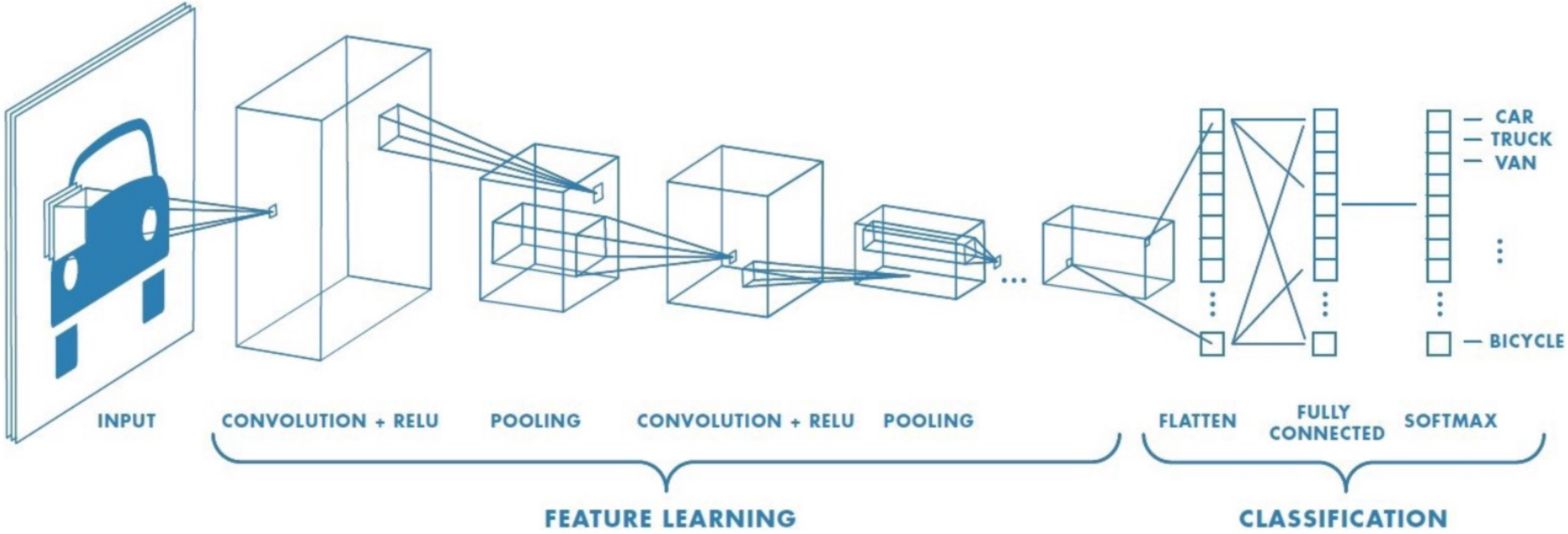
Convolutional Neural Networks

- Go-to algorithms for computer vision tasks
  - Dominates ImageNet competition

- “Convnets” learn:
  - translation invariant patterns
  - spatial hierarchy of patterns

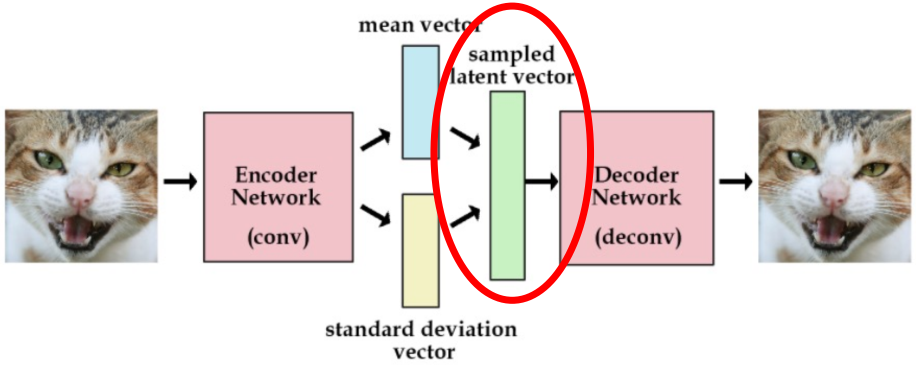


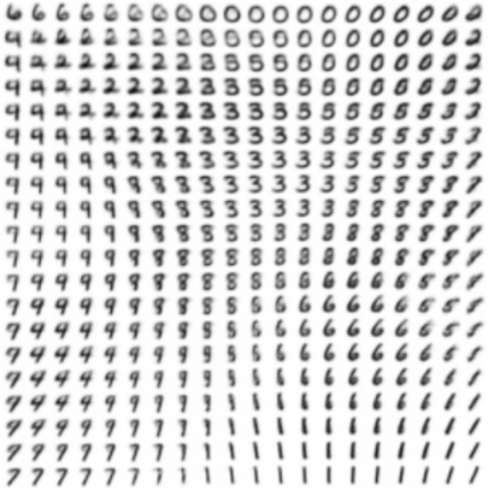
How Do “Convnets” Work?



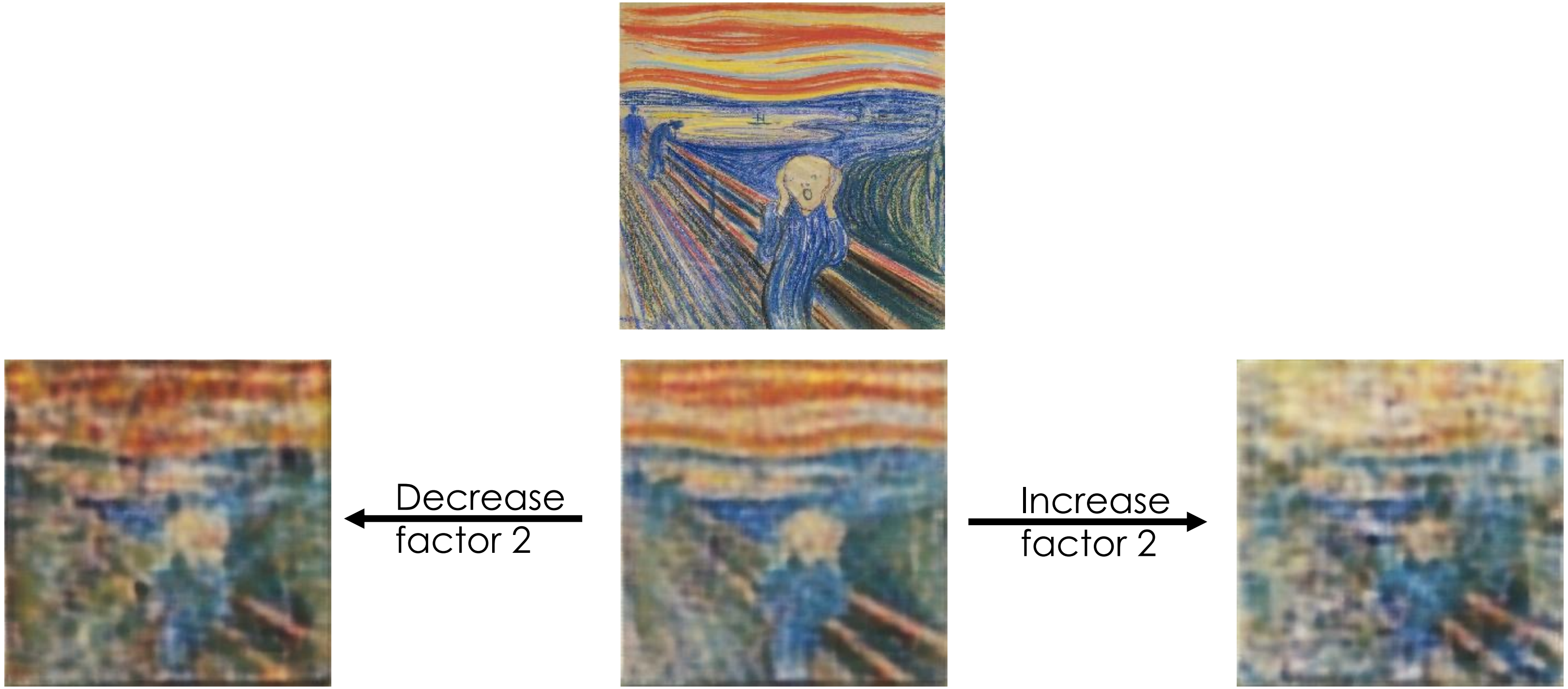
Variational Auto-Encoders

- Deep generative model assumes images generated by statistical process
- VAE contains two parts:
  - Encoder: takes image as input and compress its information in latent parameters
  - Decoder: takes latent space representation as input and outputs a reconstitution of the original image
- Latent parameters are used as predictors





Example – 100 Factors - The Scream



Decrease factor 2

Increase factor 2

Parameter captures reddish-hue in upper part of the painting

What Can We Do With This Information?

- For art:
  - See which features correlate with higher prices
  - See which paintings were most influential and creative over time
- For Marketing?

Application of Factor Analysis to Segmentation

Context: DuPont and B2B Marketing

DuPont is a large chemical company that sells industrial chemicals, synthetic fibers, pharmaceuticals, building materials, agricultural chemicals, etc. as inputs to other businesses' manufacturing.

Context: DuPont and B2B Marketing

- DuPont collected mail survey data for 58 respondents
- One set of questions was about interests of the company, their size, etc.
- One set of questions involved satisfaction with DuPont in 5 primary areas

Demographic/Background

Exp1 = interest in exporting (1=L, 2=M, 3=H)

Size = Number of employees in thousands

Revenue = Amount sold to that company by DuPont in $MM

Years = Number of years as a DuPont customer

Numprod = Number of products that they buy from DuPont

Survey

Q1-Q4 = questions about quality

TS1-TS3 = questions about tech support

SM1-SM2 = questions about sales and marketing support

SD1-SD7 = questions about supply and delivery

INN1-INN3 = questions about innovation

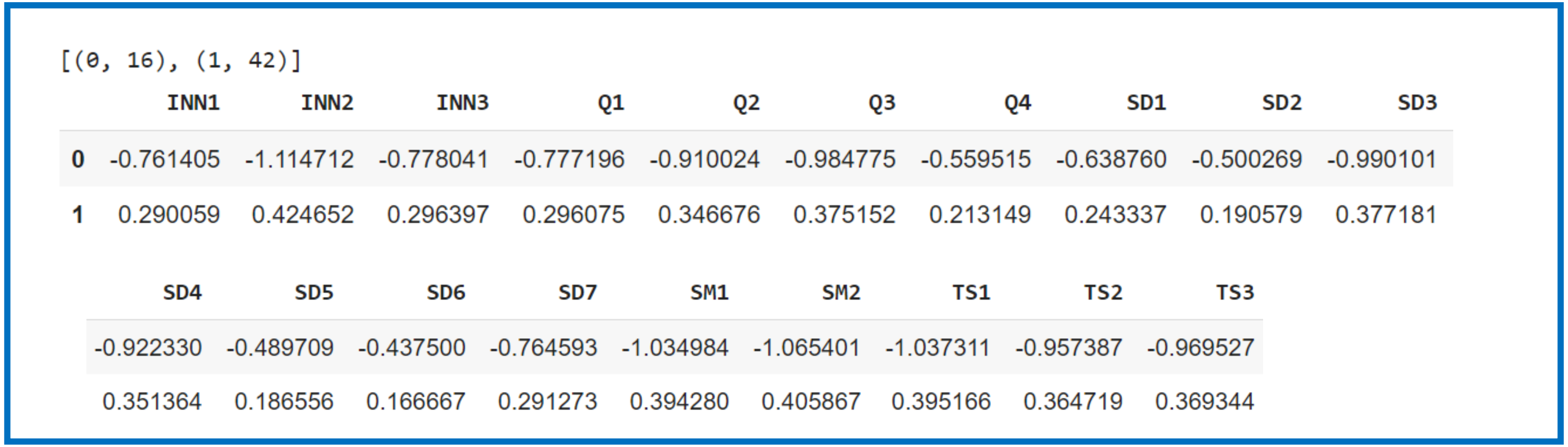
DuPont Analysis Goals

- **Segmentation:** are there different segments in the existing customer base, and if so, how do they differentially drive revenue?
- **Feedback:** Can DuPont make this survey better next time?

Cluster Analysis

Perhaps the drivers of revenue differ by <u>segment</u>

K-means clustering with 2 clusters



```
[(0, 16), (1, 42)]
```

	INN1	INN2	INN3	Q1	Q2	Q3	Q4	SD1	SD2	SD3	SD4	SD5	SD6	SD7	SM1	SM2	TS1	TS2	TS3
0	-0.761405	-1.114712	-0.778041	-0.777196	-0.910024	-0.984775	-0.559515	-0.638760	-0.500269	-0.990101	-0.922330	-0.489709	-0.437500	-0.764593	-1.034984	-1.065401	-1.037311	-0.957387	-0.969527
1	0.290059	0.424652	0.296397	0.296075	0.346676	0.375152	0.213149	0.243337	0.190579	0.377181	0.351364	0.186556	0.166667	0.291273	0.394280	0.405867	0.395166	0.364719	0.369344

Too many variables!

It won't be easy to come up with clear segment descriptions.

Solution: factor analysis

Example Question

		Rating
	PRODUCT QUALITY	
01	THE RANGE OF CHOICES IN THE (product) PRODUCT LINE	________
02	THE CONSISTENCY OF (product) QUALITY FROM LOT TO LOT	________
03	THE WAY (product) PROCESSES IN YOUR MANUFACTURING OPERATIONS	________
04	THE WAY (product) PERFORMS IN YOUR FINISHED PRODUCTS	________

Why do we need factor analysis?

Let’s go to Python

DuPont Factor Analysis

In groups, perform k-means clustering with 2 groups on DuPont survey data.

15 minutes

Set random_state = 1690

Steps to Factor Analysis with PCA

1. Estimate all the principal components (**without rotation**)

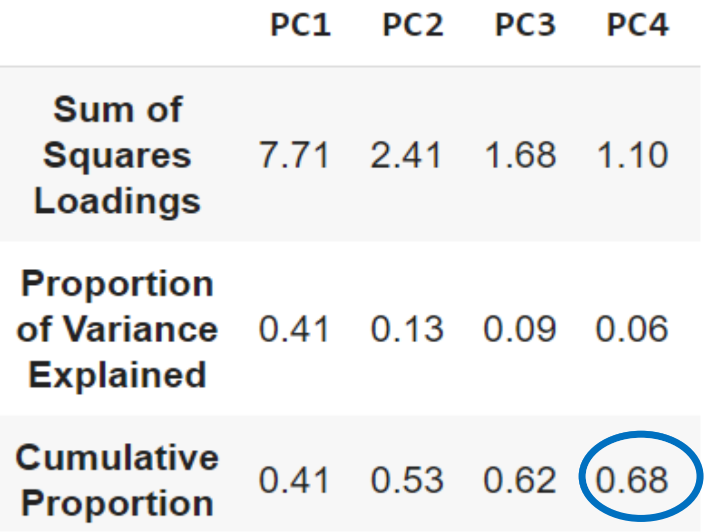
2. Determine the number of components (factors) to keep:
    1. Eigenvalues>1, Cumulative Variance>80%, scree plot, managerial relevance

3. Compute the **rotated** factor loading matrix to understand (and name!) the underlying factors

4. Compute the factor scores:
    1. For each observation, what are the values of the factors?

Four Factor Solution

- Unrotated PCA to determine # of factors

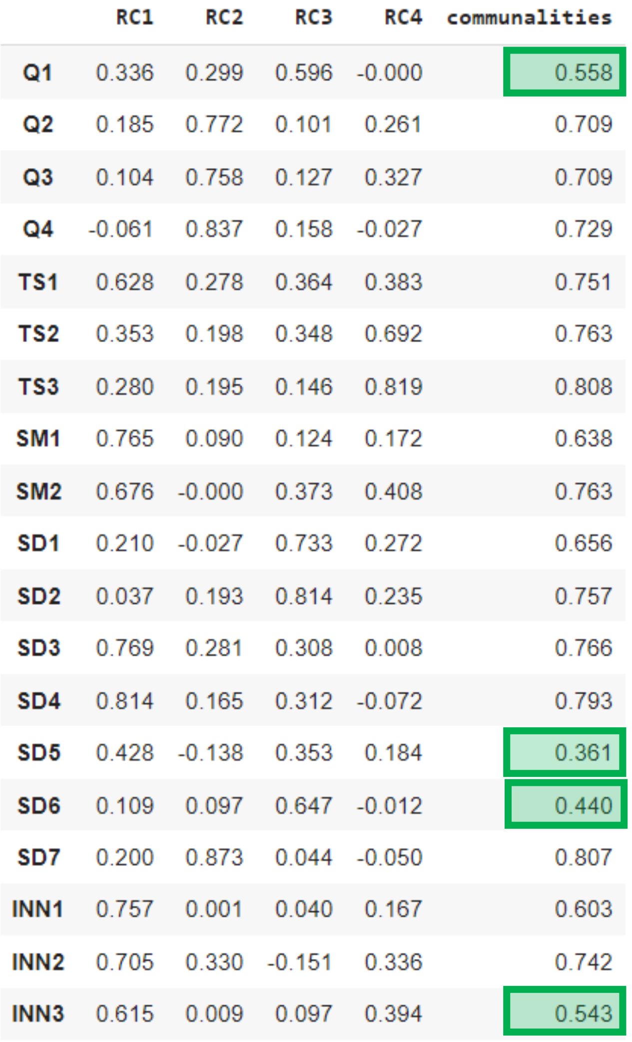


	PC1	PC2	PC3	PC4
Sum of Squares Loadings	7.71	2.41	1.68	1.10
Proportion of Variance Explained	0.41	0.13	0.09	0.06
Cumulative Proportion	0.41	0.53	0.62	0.68

68% variance explained

Four Factor Solution

- Notice: several communalities are pretty low (< 0.6)
  - Don't give high weight to these questions when interpreting factors
  - Consider including these separately in analysis



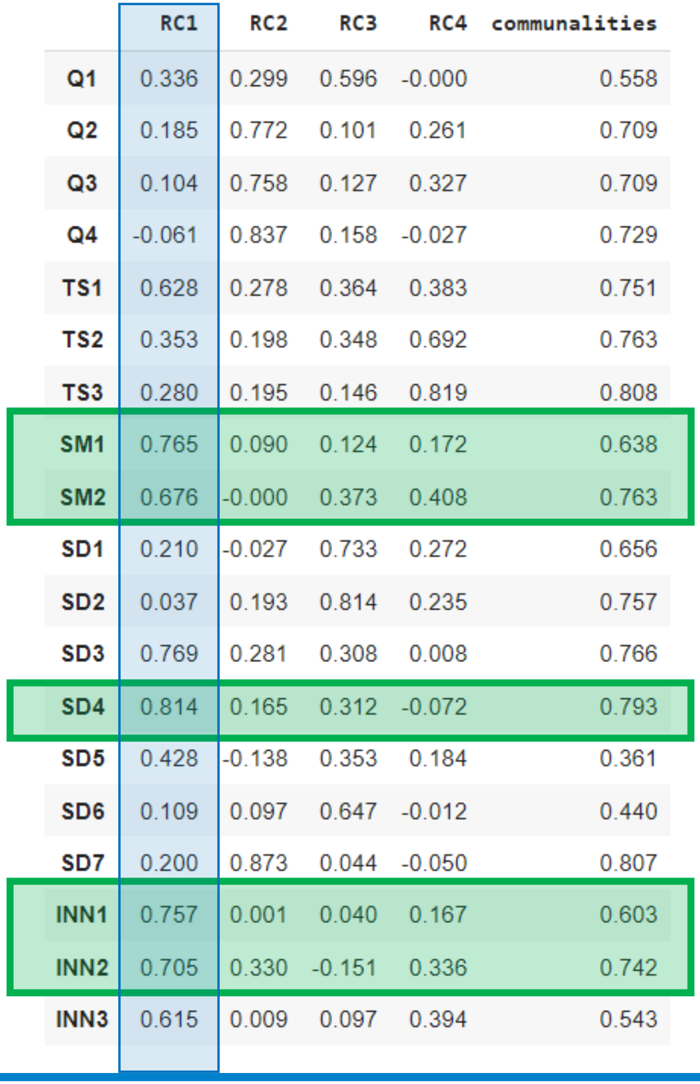
	RC1	RC2	RC3	RC4	communalities
Q1	0.336	0.299	0.596	-0.000	0.558
Q2	0.185	0.772	0.101	0.261	0.709
Q3	0.104	0.758	0.127	0.327	0.709
Q4	-0.061	0.837	0.158	-0.027	0.729
TS1	0.628	0.278	0.364	0.383	0.751
TS2	0.353	0.198	0.348	0.692	0.763
TS3	0.280	0.195	0.146	0.819	0.808
SM1	0.765	0.090	0.124	0.172	0.638
SM2	0.676	-0.000	0.373	0.408	0.763
SD1	0.210	-0.027	0.733	0.272	0.656
SD2	0.037	0.193	0.814	0.235	0.757
SD3	0.769	0.281	0.308	0.008	0.766
SD4	0.814	0.165	0.312	-0.072	0.793
SD5	0.428	-0.138	0.353	0.184	0.361
SD6	0.109	0.097	0.647	-0.012	0.440
SD7	0.200	0.873	0.044	-0.050	0.807
INN1	0.757	0.001	0.040	0.167	0.603
INN2	0.705	0.330	-0.151	0.336	0.742
INN3	0.615	0.009	0.097	0.394	0.543

Four Factor Solution

Interpreting the factors:

1. Combination of:
    - sales and marketing support
    - complaints handling
    - innovativeness

   Basically: a **“general competence factor”** or maybe **“customer centricity”**



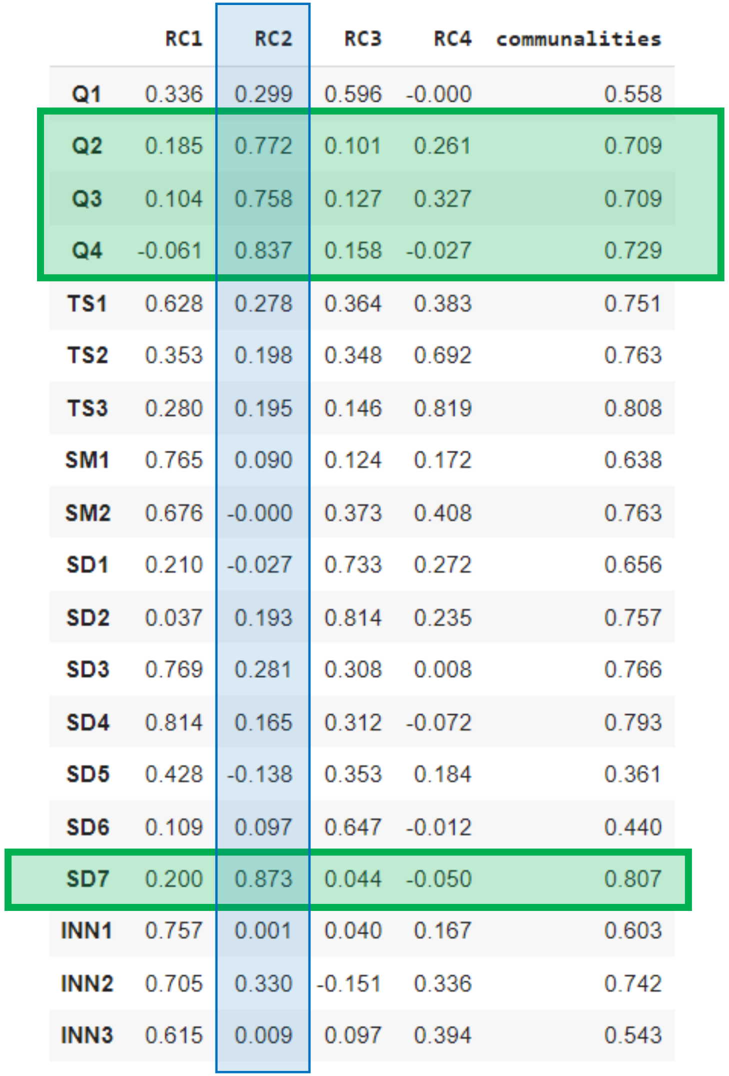
	RC1	RC2	RC3	RC4	communalities
Q1	0.336	0.299	0.596	-0.000	0.558
Q2	0.185	0.772	0.101	0.261	0.709
Q3	0.104	0.758	0.127	0.327	0.709
Q4	-0.061	0.837	0.158	-0.027	0.729
TS1	0.628	0.278	0.364	0.383	0.751
TS2	0.353	0.198	0.348	0.692	0.763
TS3	0.280	0.195	0.146	0.819	0.808
SM1	0.765	0.090	0.124	0.172	0.638
SM2	0.676	-0.000	0.373	0.408	0.763
SD1	0.210	-0.027	0.733	0.272	0.656
SD2	0.037	0.193	0.814	0.235	0.757
SD3	0.769	0.281	0.308	0.008	0.766
SD4	0.814	0.165	0.312	-0.072	0.793
SD5	0.428	-0.138	0.353	0.184	0.361
SD6	0.109	0.097	0.647	-0.012	0.440
SD7	0.200	0.873	0.044	-0.050	0.807
INN1	0.757	0.001	0.040	0.167	0.603
INN2	0.705	0.330	-0.151	0.336	0.742
INN3	0.615	0.009	0.097	0.394	0.543

Four Factor Solution

Interpreting the factors:

1. Combination of:
   - sales and marketing support
   - complaints handling
   - innovativeness

   Basically: a "general competence factor"
2. **Quality**



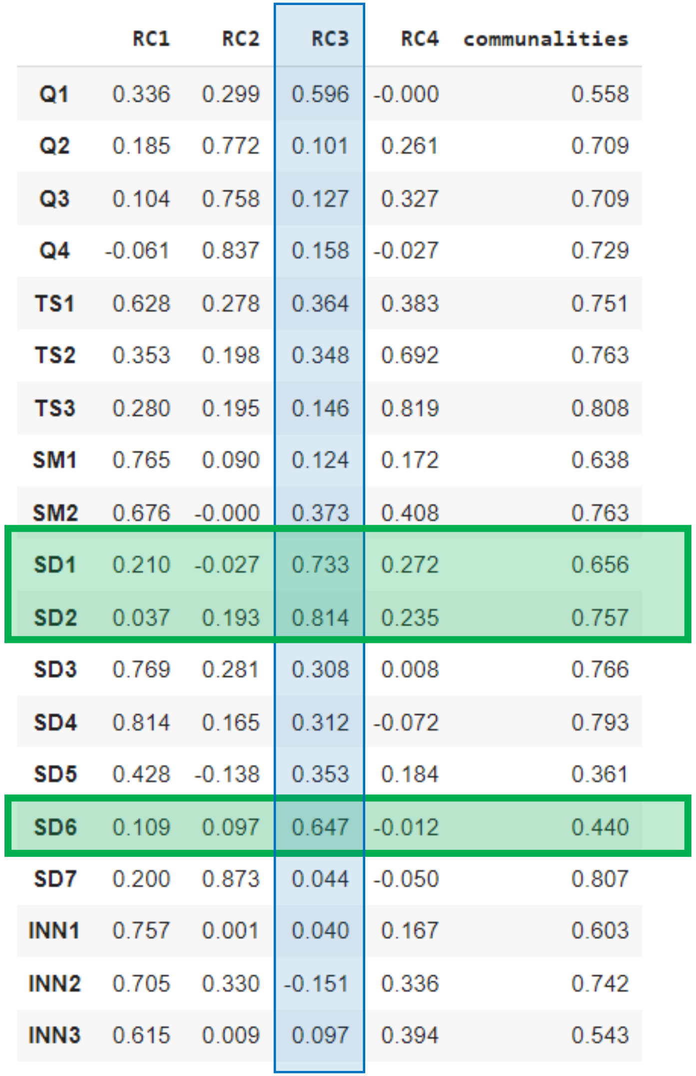
	RC1	RC2	RC3	RC4	communalities
Q1	0.336	0.299	0.596	-0.000	0.558
Q2	0.185	0.772	0.101	0.261	0.709
Q3	0.104	0.758	0.127	0.327	0.709
Q4	-0.061	0.837	0.158	-0.027	0.729
TS1	0.628	0.278	0.364	0.383	0.751
TS2	0.353	0.198	0.348	0.692	0.763
TS3	0.280	0.195	0.146	0.819	0.808
SM1	0.765	0.090	0.124	0.172	0.638
SM2	0.676	-0.000	0.373	0.408	0.763
SD1	0.210	-0.027	0.733	0.272	0.656
SD2	0.037	0.193	0.814	0.235	0.757
SD3	0.769	0.281	0.308	0.008	0.766
SD4	0.814	0.165	0.312	-0.072	0.793
SD5	0.428	-0.138	0.353	0.184	0.361
SD6	0.109	0.097	0.647	-0.012	0.440
SD7	0.200	0.873	0.044	-0.050	0.807
INN1	0.757	0.001	0.040	0.167	0.603
INN2	0.705	0.330	-0.151	0.336	0.742
INN3	0.615	0.009	0.097	0.394	0.543

Four Factor Solution

Interpreting the factors:

1. Combination of:
    - sales and marketing support
    - complaints handling
    - innovativeness

   Basically: a “general competence factor”
2. Quality
3. **Delivery**



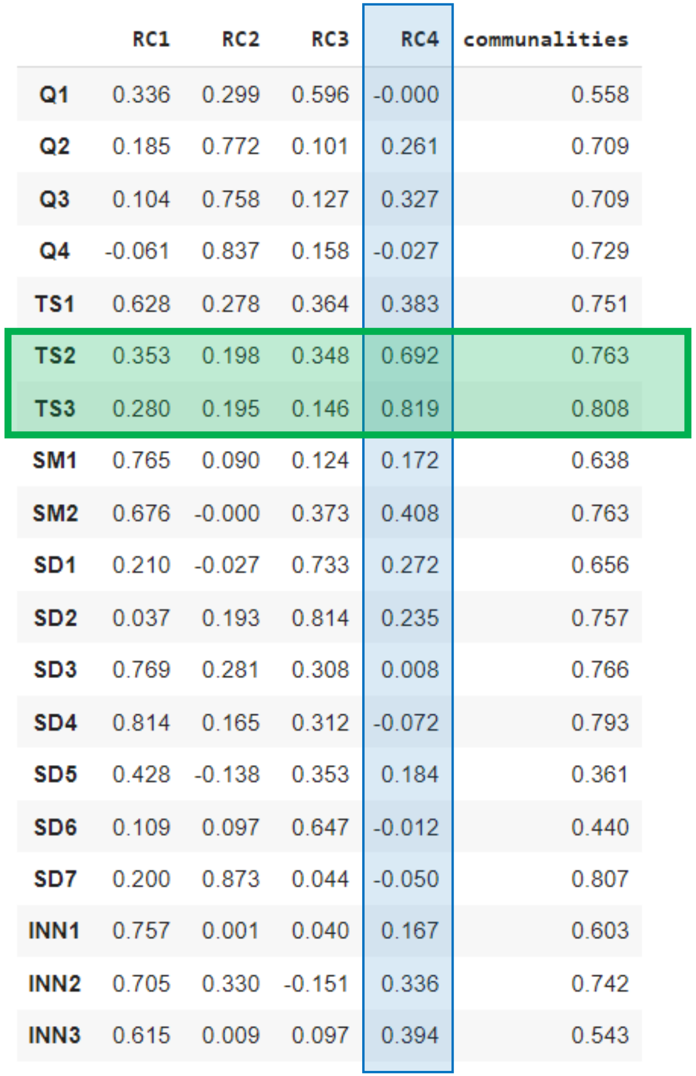
	RC1	RC2	RC3	RC4	communalities
Q1	0.336	0.299	0.596	-0.000	0.558
Q2	0.185	0.772	0.101	0.261	0.709
Q3	0.104	0.758	0.127	0.327	0.709
Q4	-0.061	0.837	0.158	-0.027	0.729
TS1	0.628	0.278	0.364	0.383	0.751
TS2	0.353	0.198	0.348	0.692	0.763
TS3	0.280	0.195	0.146	0.819	0.808
SM1	0.765	0.090	0.124	0.172	0.638
SM2	0.676	-0.000	0.373	0.408	0.763
SD1	0.210	-0.027	0.733	0.272	0.656
SD2	0.037	0.193	0.814	0.235	0.757
SD3	0.769	0.281	0.308	0.008	0.766
SD4	0.814	0.165	0.312	-0.072	0.793
SD5	0.428	-0.138	0.353	0.184	0.361
SD6	0.109	0.097	0.647	-0.012	0.440
SD7	0.200	0.873	0.044	-0.050	0.807
INN1	0.757	0.001	0.040	0.167	0.603
INN2	0.705	0.330	-0.151	0.336	0.742
INN3	0.615	0.009	0.097	0.394	0.543

Four Factor Solution

Interpreting the factors:

1. Combination of:
    - sales and marketing support
    - complaints handling
    - innovativeness

   Basically: a "general competence factor"
2. Quality
3. Delivery
4. **Technical expertise**



	RC1	RC2	RC3	RC4	communalities
Q1	0.336	0.299	0.596	-0.000	0.558
Q2	0.185	0.772	0.101	0.261	0.709
Q3	0.104	0.758	0.127	0.327	0.709
Q4	-0.061	0.837	0.158	-0.027	0.729
TS1	0.628	0.278	0.364	0.383	0.751
TS2	0.353	0.198	0.348	0.692	0.763
TS3	0.280	0.195	0.146	0.819	0.808
SM1	0.765	0.090	0.124	0.172	0.638
SM2	0.676	-0.000	0.373	0.408	0.763
SD1	0.210	-0.027	0.733	0.272	0.656
SD2	0.037	0.193	0.814	0.235	0.757
SD3	0.769	0.281	0.308	0.008	0.766
SD4	0.814	0.165	0.312	-0.072	0.793
SD5	0.428	-0.138	0.353	0.184	0.361
SD6	0.109	0.097	0.647	-0.012	0.440
SD7	0.200	0.873	0.044	-0.050	0.807
INN1	0.757	0.001	0.040	0.167	0.603
INN2	0.705	0.330	-0.151	0.336	0.742
INN3	0.615	0.009	0.097	0.394	0.543

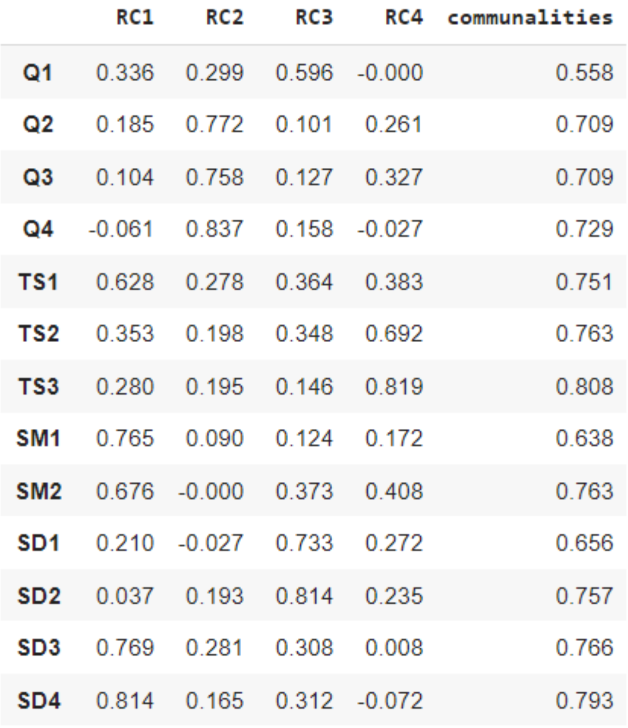
Four Factor Solution

Interpreting the factors:

1. Combination of:
    - sales and marketing support
    - complaints handling
    - innovativeness

   Basically: a "general competence factor"
2. Quality
3. Delivery
4. Technical expertise

Not what the survey designer expected!

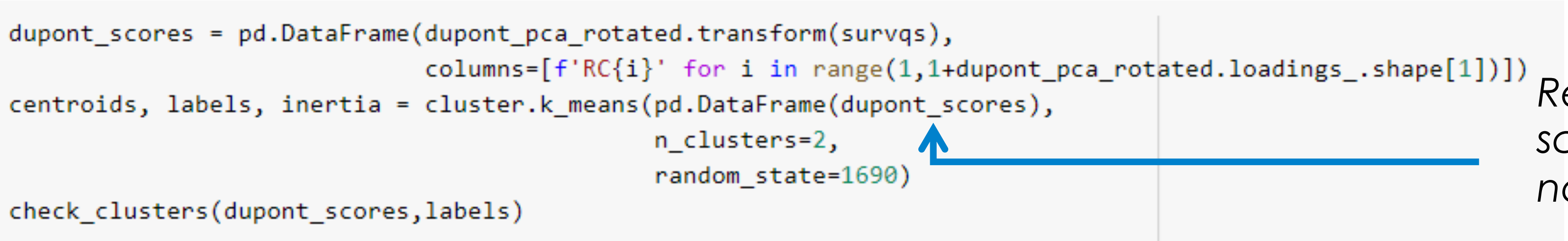


	RC1	RC2	RC3	RC4	communalities
Q1	0.336	0.299	0.596	-0.000	0.558
Q2	0.185	0.772	0.101	0.261	0.709
Q3	0.104	0.758	0.127	0.327	0.709
Q4	-0.061	0.837	0.158	-0.027	0.729
TS1	0.628	0.278	0.364	0.383	0.751
TS2	0.353	0.198	0.348	0.692	0.763
TS3	0.280	0.195	0.146	0.819	0.808
SM1	0.765	0.090	0.124	0.172	0.638
SM2	0.676	-0.000	0.373	0.408	0.763
SD1	0.210	-0.027	0.733	0.272	0.656
SD2	0.037	0.193	0.814	0.235	0.757
SD3	0.769	0.281	0.308	0.008	0.766
SD4	0.814	0.165	0.312	-0.072	0.793

Q1-Q4 = questions about quality
TS1-TS3 = questions about tech support
SM1-SM2 = questions about sales and marketing support
SD1-SD7 = questions about supply and delivery
INN1-INN3 = questions about innovation

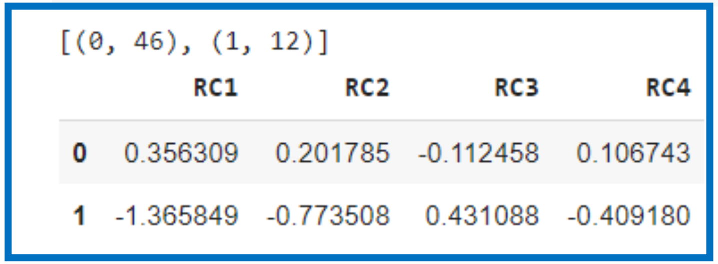
Interpretable Segmentation

Idea: run cluster analysis on the *factor scores*



```
dupont_scores = pd.DataFrame(dupont_pca_rotated.transform(survqs),
                             columns=[f'RC{i}' for i in range(1,1+dupont_pca_rotated.loadings_.shape[1])])
centroids, labels, inertia = cluster.k_means(pd.DataFrame(dupont_scores),
                                             n_clusters=2,
                                             random_state=1690)
check_clusters(dupont_scores,labels)
```

Remember: factor scores are already normalized



```
[(0, 46), (1, 12)]
```

	RC1	RC2	RC3	RC4
0	0.356309	0.201785	-0.112458	0.106743
1	-1.365849	-0.773508	0.431088	-0.409180

Interpretation: segments differ in their ratings along the four factors!

Segments

1. The Mainstream
2. The Haters

Summary: Segmentation

- Factor analysis can be used to:
  - Uncover structure in survey questions
  - Reduce dimensionality for more interpretable segmentation
  - Focus attention on key factors
- Factor and cluster (and regression) analysis can all be linked!
- Real data is messy: very rarely is the story clean cut

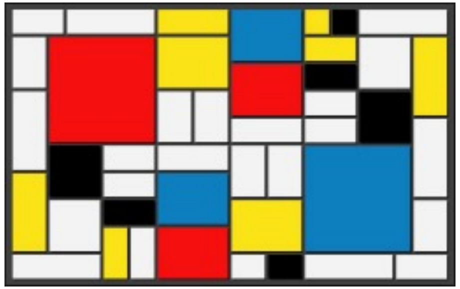
Application to Positioning

Where will we play?

Segmentation

S

Discovering and profiling groups of customers with similar needs and preferences



Targeting

T

Evaluating segment attractiveness and targeting most attractive ones

How will we win?

Positioning

P

Defining value proposition for target segments and developing a marketing plan

What is Positioning?

- Placing the product/service with respect to alternatives in the mind of the customer (Reis and Trout)

- Positioning statement
  - **Who** is the product for?
  - **What** does the product have to offer?
  - **How** is the product different?

- In other words: "How does my company deliver value to my (target) customer better than the competition"

wayfair ALLMODERN

Product Differentiation & Positioning

- “There is no such thing as a commodity”

Theodore Levitt

- “No matter how commonplace a product may appear, it does not have to be a commodity. Every product, every service can be differentiated”

Dermot Dunphy, CEO, Sealed Air

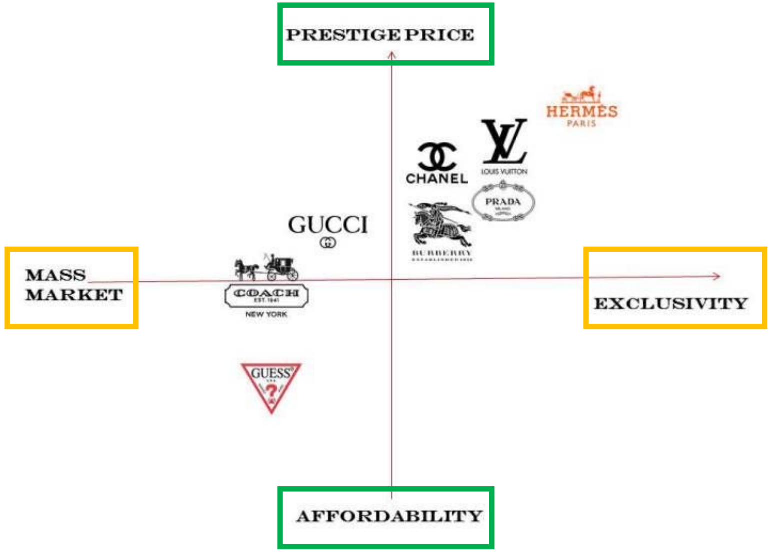
- **Differentiation** can be achieved on
  - Product benefits
  - Quality of customer service
  - Psychological associations of brand, …

- **Positioning**: the image created in the minds of target consumers relative to other brands in the category

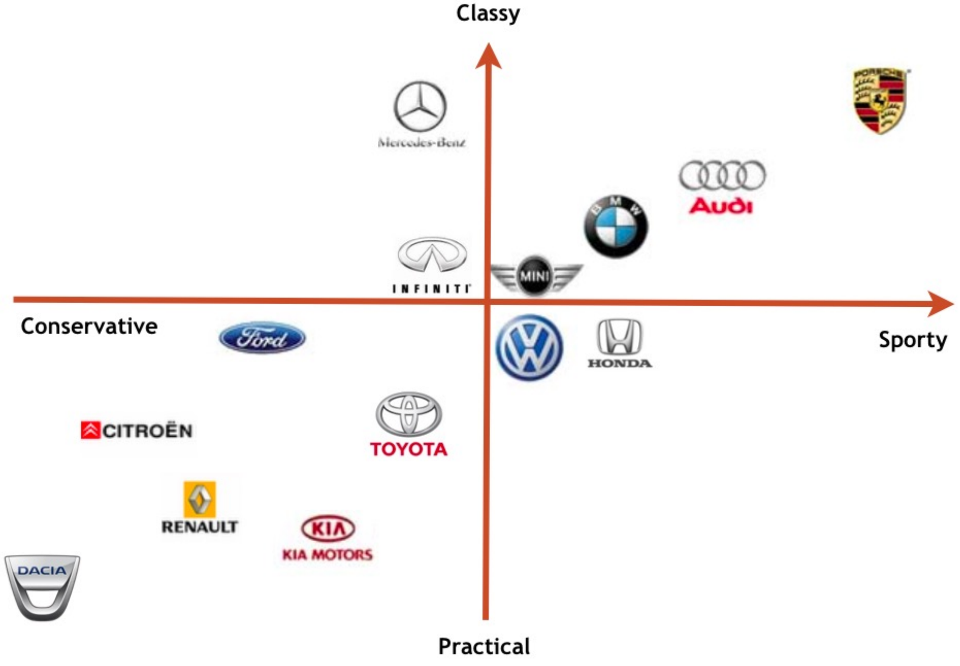
Factor Analysis: Relevance to Positioning

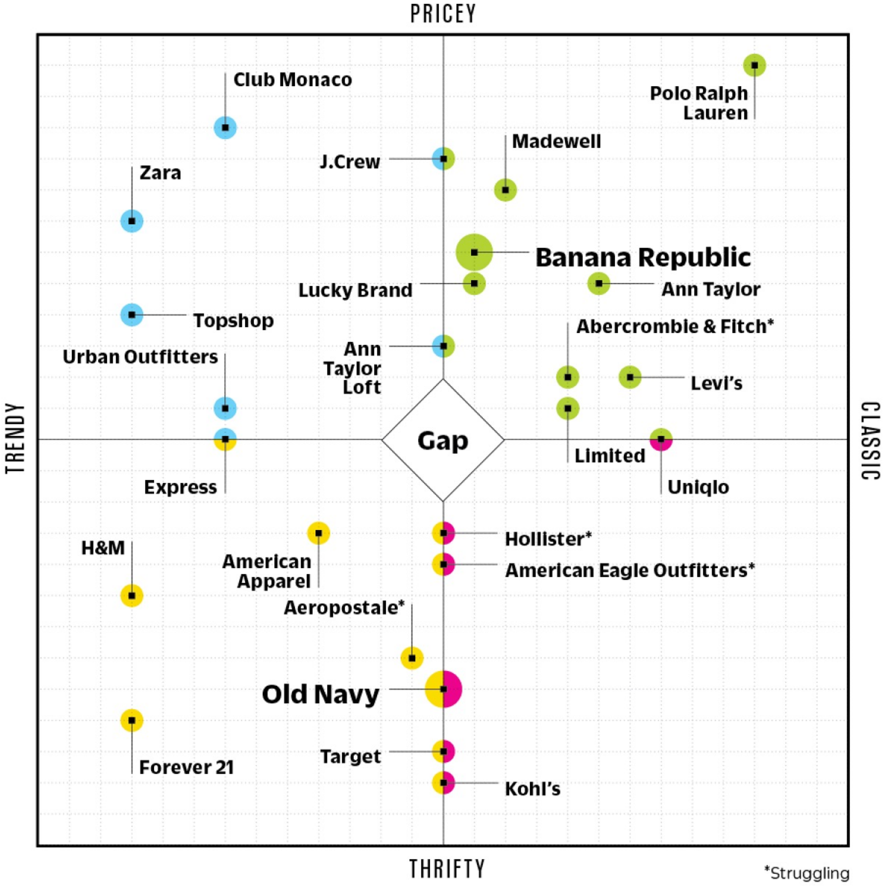
Understanding how *many* variables in a dataset capture a *few* unique constructs can be helpful for brand positioning

What factors are relevant in determining a brand’s position in the marketplace?



Tool: Perceptual Maps for Brands





What Actually is a Perceptual Map?

- Visual representation of how target customers view competing alternatives

- Dimensionality reduction:
  - Attitudes, opinions, survey questions, … → two dimensions

- Characteristics:
  - **Axes**: underlying dimensions characterize how customers differentiate among alternatives
  - **Distance**: pairwise distances between alternatives directly indicate how close or far apart the products are in the minds of customers

Uses of Perceptual Maps

- Understanding **market structure** in the minds of consumers
  - How do my **customers** perceive my brand?
  - Who are my key **competitors**?
  - How can I **communicate** my brand positioning in a way that is consistent with my customer's views, or how I want my brand to be seen?
- **Problem detection**: do people see us like we see us?
- Differentiated positioning and/or new product development

Look for a "hole" in the map!

(But also ask: why is there a hole?)

Perceptual Maps – Beer Brands

Example: Rate 20 different beers on 6 dimensions: Taste, Refreshing, Quality, Alcohol Content, High Class, Expensive

Average scores by beer:

Beer	Taste	Refreshing	Quality	Alcohol	Class	Expensive
Budweiser	1.6	2.4	1.4	2.5	1.1	1.4
Bud Light	1.1	2.8	1.1	1.4	1.5	1.1
Miller Light	1.4	2.5	1.1	1.1	1.4	1.5
…						
Stella Artois	4.1	3.4	2.8	2.6	4.6	4.4
Victory Lager	4.9	4.1	4.9	3.8	3.6	4.6
Chimay	4.4	2.5	4.9	4.9	4.8	4.6

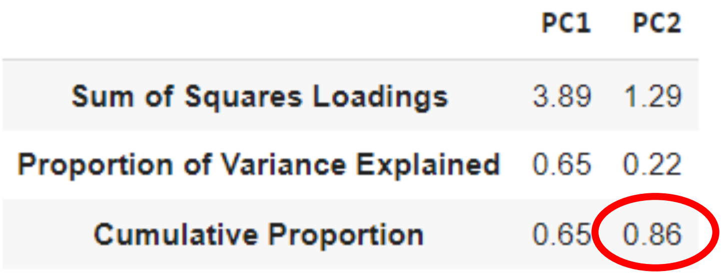
What dimensions underlie consumers' judgements?

Steps: Factor Analysis for Perceptual Maps

1. Use factor analysis to convert many judgments into 2+ underlying dimensions
    1. Two factors is ideal: results in a single perceptual map (two dimensions)
    2. More than two factors requires one map per pair of factors
2. Name and interpret the dimensions → axes of the map
3. Plot the factor scores → positions on the map

Let’s make a map using Python!

Beer: Factor Analysis



	PC1	PC2
Sum of Squares Loadings	3.89	1.29
Proportion of Variance Explained	0.65	0.22
Cumulative Proportion	0.65	0.86

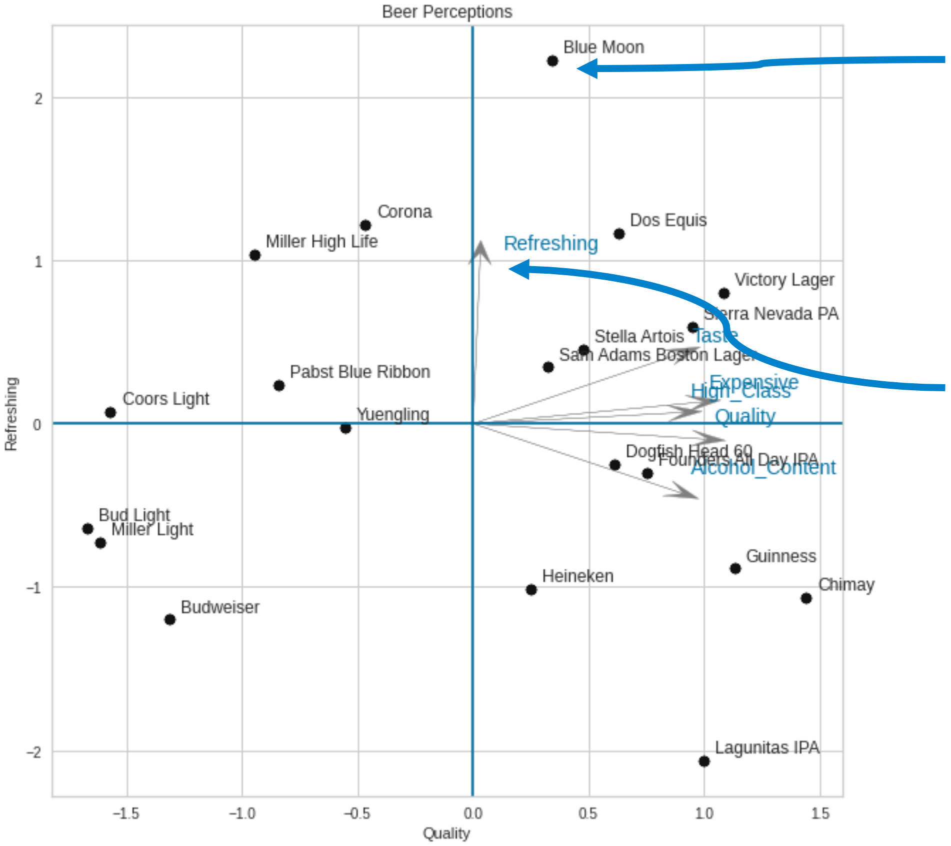
Force two factors

Two factors explain a high proportion of the variance

	RC1	RC2
Taste	0.848	0.406
Refreshing	0.029	0.973
Quality	0.944	-0.089
Alcohol_Content	0.841	-0.396
High_Class	0.842	0.064
Expensive	0.924	0.124

Factor 1: Quality
Factor 2: Refreshing

Perceptual Map: Plot the Factor Scores

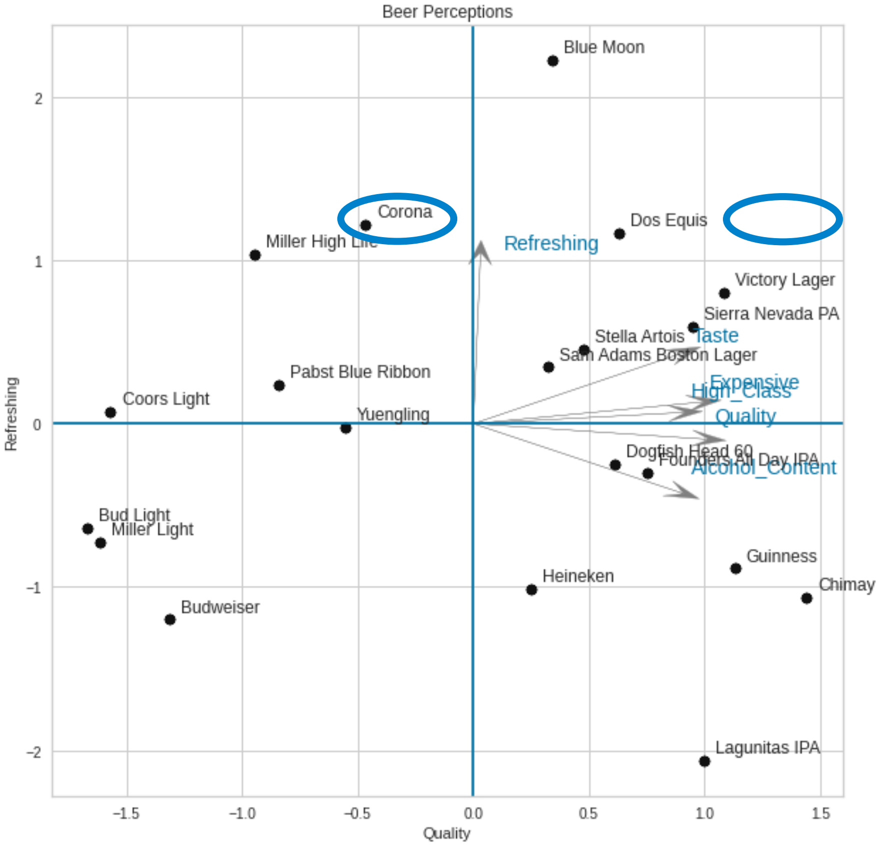


Blue Moon’s factor scores

Projection of variables onto 2D space of first 2 components

- Angle: closer angle = higher positive association
- Direction: association of variables with components

Perceptual Map: Plot the Factor Scores



If Corona wants to be viewed as higher quality, what are some marketing strategies to doing so?

Summary: Positioning

- Perceptual maps are useful tools for understanding positioning and developing brand and product strategy
  - Understanding the competitive environment in the minds of consumers
  - Look for “holes” (but always ask why!)

- Data-driven perceptual maps can be created through factor analysis:
  - Dimensions = factors
  - Map positions = factor scores

Takeaways: Factor Analysis

- Factor analysis: turn complex data into intuitive and meaningful factors
  - Lots of jargon: loadings, variance explained, scores, …
  - How it works: principal components analysis

- Many applications:
  - Simplifying data for easy description
  - Factor + cluster analysis for segmentation
  - Positioning through perceptual maps

Next Class

- Positioning Concept Check due before next class
- We will review Ford Ka and study CLV

Last Time

- Factor analysis/PCA:
  - Tool to turn complex data into intuitive and meaningful factors
  - Many applications:
    - Simplifying data for easy description
    - Factor + cluster analysis for segmentation
    - Positioning through perceptual maps
- Positioning:
  - "How does my company deliver value to my (target) customer better than the competition"
  - **Who** is the product for? **What** does the product have to offer? **How** is the product different?
- Today: Perceptual Maps in Python + Ford Ka + CLV

Today

Part 1: Perceptual Maps

1. Implementation in Python

Part 2: Ford Ka

1. Case Discussion

Part 3: Customer Lifetime Value

1. Definition
2. Problem: how to increase CLV?

Today’s Goals

Understand:

- The difficulties in establishing a marketing strategy
- What is customer lifetime value
- What are some strategies for increasing CLV

Be able to:

- Build a perceptual map in Python with survey data
- Conduct an end-to-end marketing strategy and be aware of pitfalls
- Measure CLV
- Quantify what your company must do to increase CLV

Course Roadmap

STP Analytics (Identify Value)	Customer Analytics (Deliver Value)	4P Analytics (Capture Value)
Module 1	**Module 2**	**Module 3**
What datasets can we use? How can we segment and target our customers? How should we position our products/services?	How much are our customers worth? Are our customers leaving? How do our customers make choices?	How do we build a new product? How should we price our products? How do we distribute them? How do we quantify the impact of our promotions?

Positioning

Perceptual Maps

What Actually is a Perceptual Map?

- Visual representation of how target customers view competing alternatives

- Dimensionality reduction:
  - Attitudes, opinions, survey questions, … → two dimensions

- Characteristics:
  - **Axes**: underlying dimensions characterize how customers differentiate among alternatives
  - **Distance**: pairwise distances between alternatives directly indicate how close or far apart the products are in the minds of customers

Steps: Factor Analysis for Perceptual Maps

1. Use factor analysis to convert many judgments into 2+ underlying dimensions
    1. Two factors is ideal: results in a single perceptual map (two dimensions)
    2. More than two factors requires one map per pair of factors
2. Name and interpret the dimensions → axes of the map
3. Plot the factor scores → positions on the map

Let’s make a map using Python!

Create a Perceptual Map of MBA Perceptions

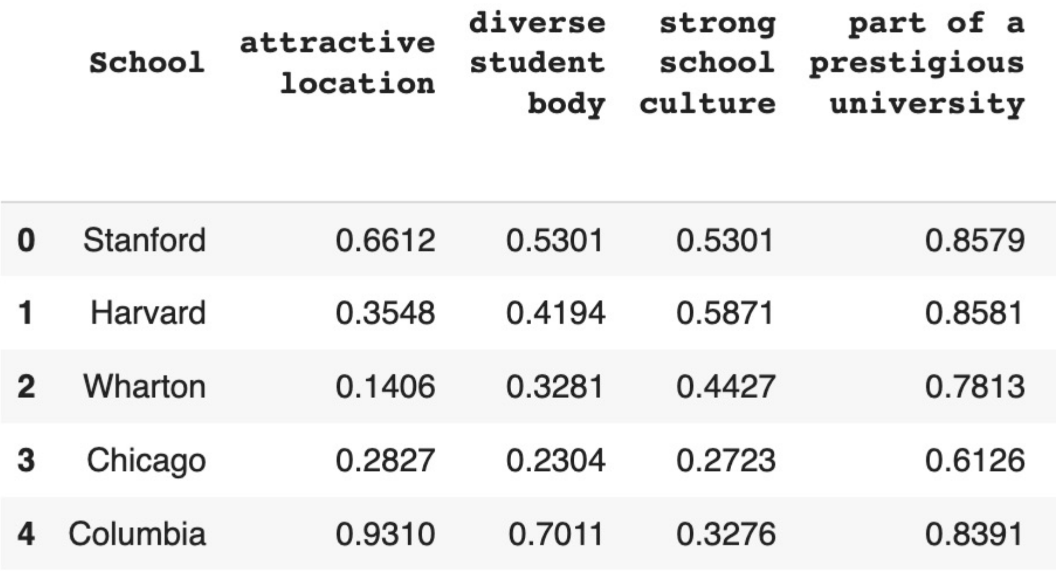
- Apply PCA to the results from our business school survey
- What factors emerge?
- How are various schools positioned along these factors?

In groups, create a perceptual map (or perhaps maps) of MBA perceptions.

15 minutes

Notice the Structure of the Data

- Different from structure of data for clustering or factor analysis of attitudinal surveys

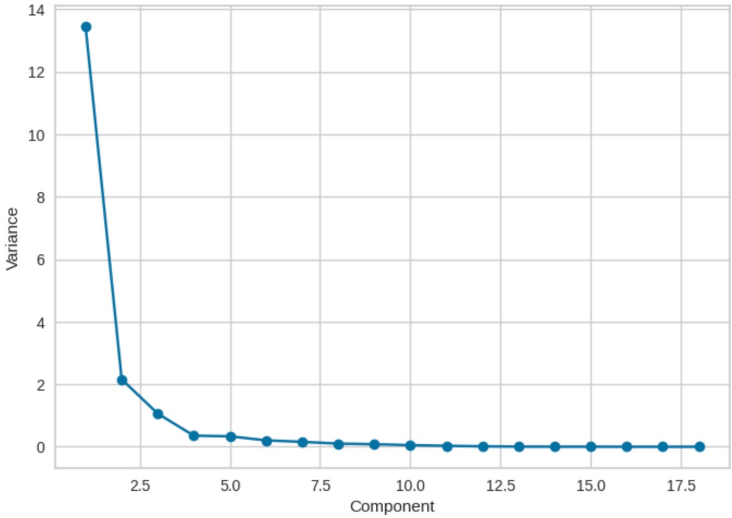


	School	attractive location	diverse student body	strong school culture	part of a prestigious university
0	Stanford	0.6612	0.5301	0.5301	0.8579
1	Harvard	0.3548	0.4194	0.5871	0.8581
2	Wharton	0.1406	0.3281	0.4427	0.7813
3	Chicago	0.2827	0.2304	0.2723	0.6126
4	Columbia	0.9310	0.7011	0.3276	0.8391

Rows are not at individual-level but aggregated across individuals

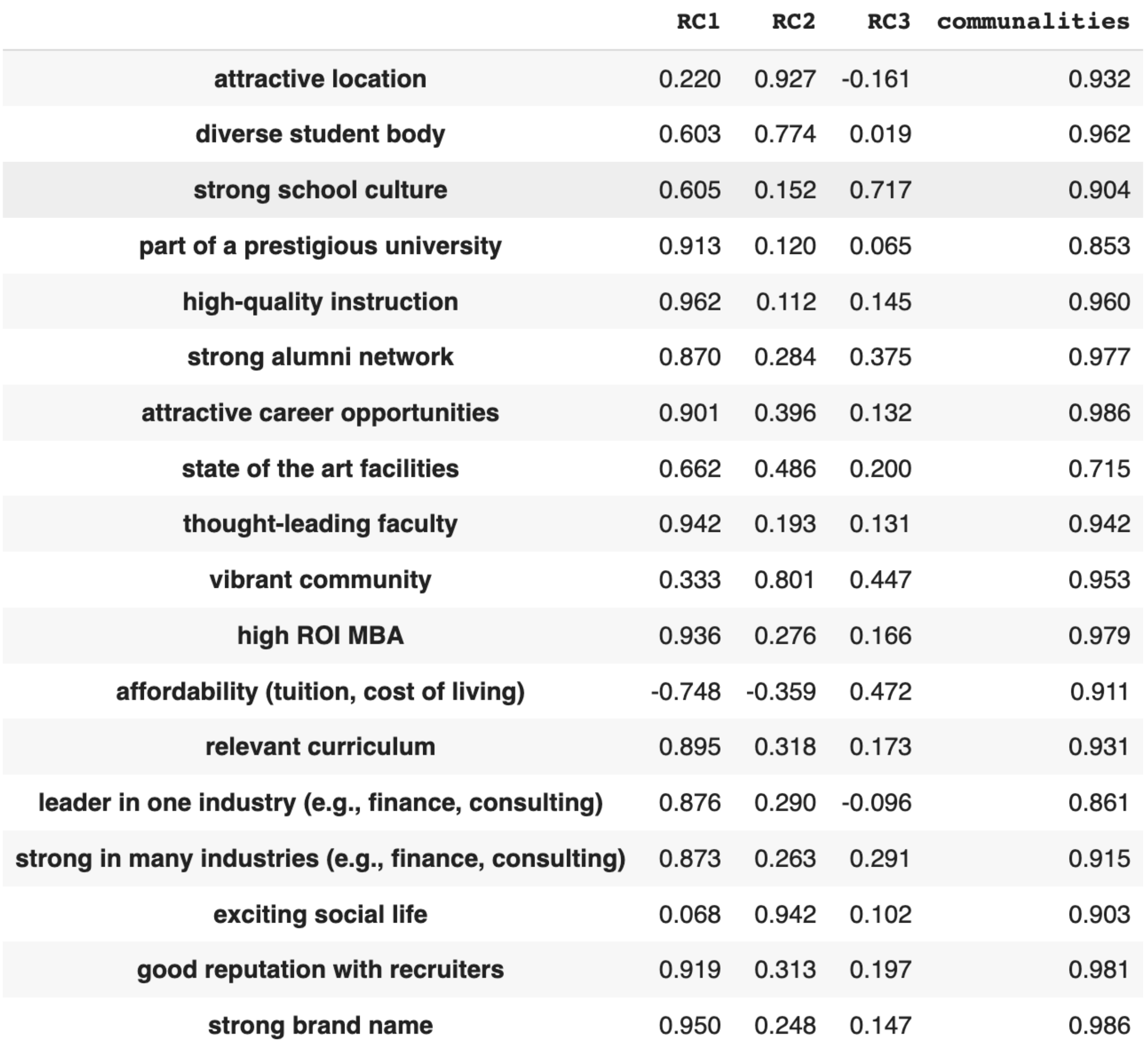
How Many Factors Are There?

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Sum of Squared Loadings	13.43	2.16	1.06	0.36	0.33	0.20	0.16	0.10	0.08	0.05
Proportion of Variance Explained	0.75	0.12	0.06	0.02	0.02	0.01	0.01	0.01	0.00	0.00
Cumulative Proportion	0.75	0.87	0.93	0.94	0.96	0.97	0.98	0.99	0.99	1.00



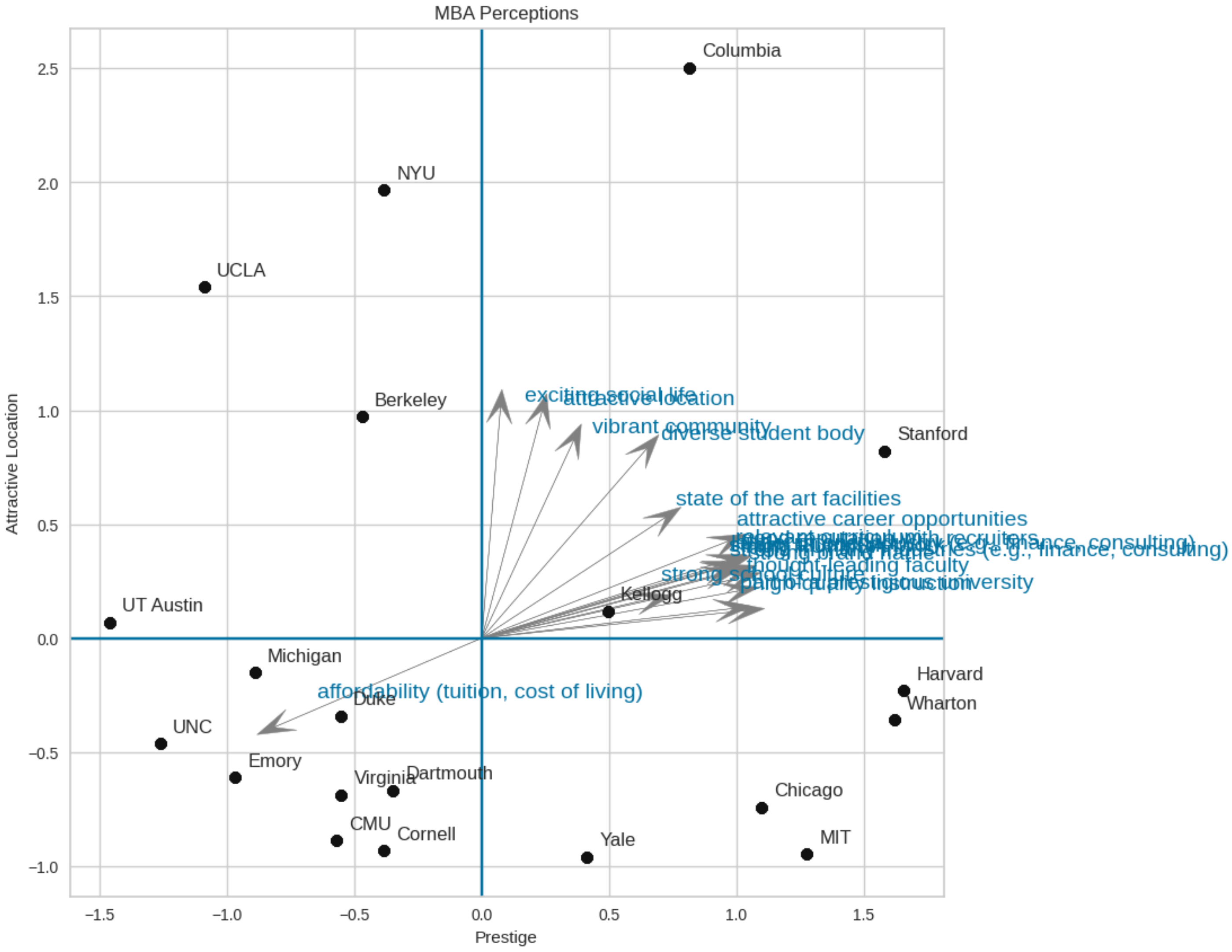
Name and Interpret the Factors

- RC1: Prestige
- RC2: Attractive Location
- RC3: Enriching Experience

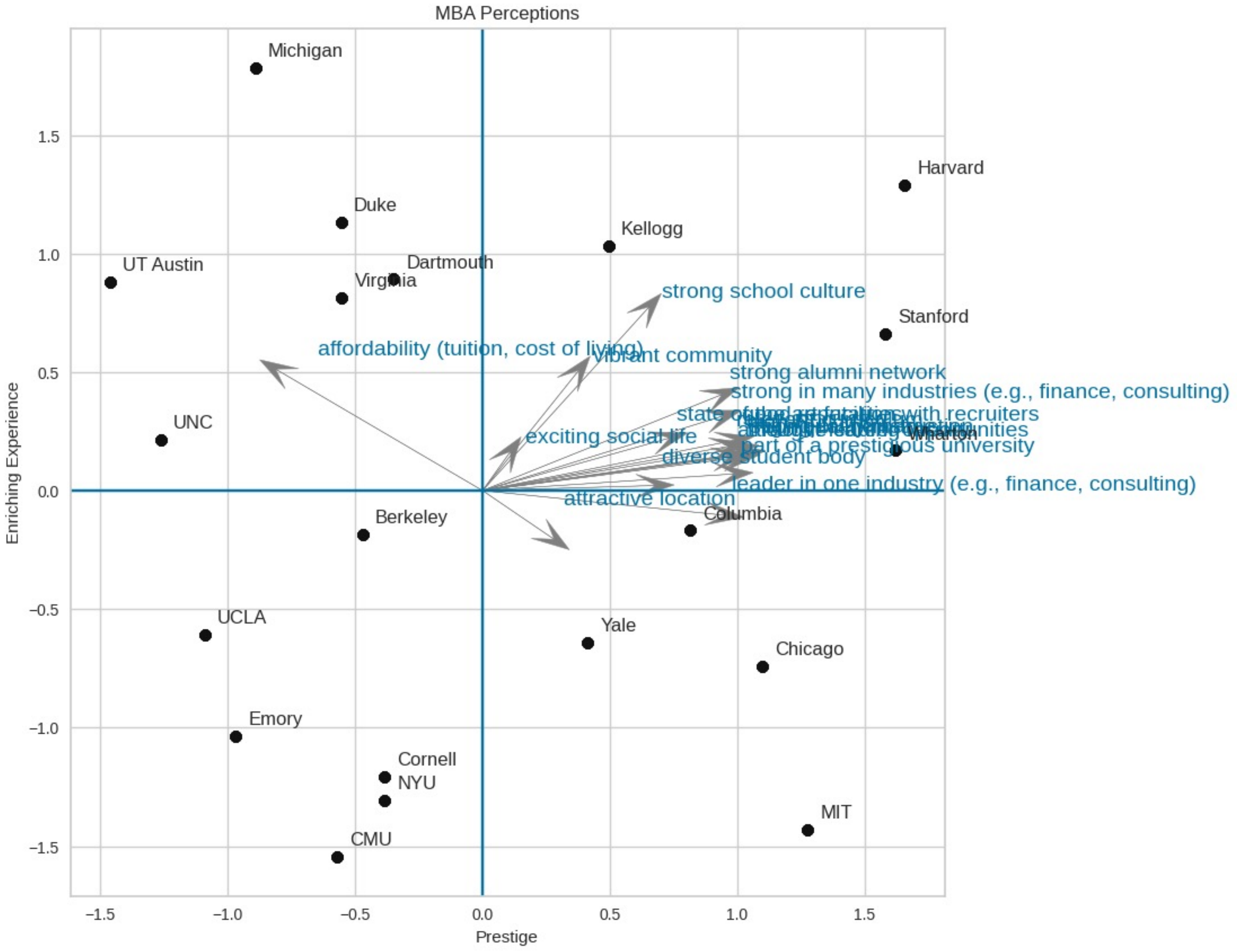


	RC1	RC2	RC3	communalities
attractive location	0.220	0.927	-0.161	0.932
diverse student body	0.603	0.774	0.019	0.962
strong school culture	0.605	0.152	0.717	0.904
part of a prestigious university	0.913	0.120	0.065	0.853
high-quality instruction	0.962	0.112	0.145	0.960
strong alumni network	0.870	0.284	0.375	0.977
attractive career opportunities	0.901	0.396	0.132	0.986
state of the art facilities	0.662	0.486	0.200	0.715
thought-leading faculty	0.942	0.193	0.131	0.942
vibrant community	0.333	0.801	0.447	0.953
high ROI MBA	0.936	0.276	0.166	0.979
affordability (tuition, cost of living)	-0.748	-0.359	0.472	0.911
relevant curriculum	0.895	0.318	0.173	0.931
leader in one industry (e.g., finance, consulting)	0.876	0.290	-0.096	0.861
strong in many industries (e.g., finance, consulting)	0.873	0.263	0.291	0.915
exciting social life	0.068	0.942	0.102	0.903
good reputation with recruiters	0.919	0.313	0.197	0.981
strong brand name	0.950	0.248	0.147	0.986

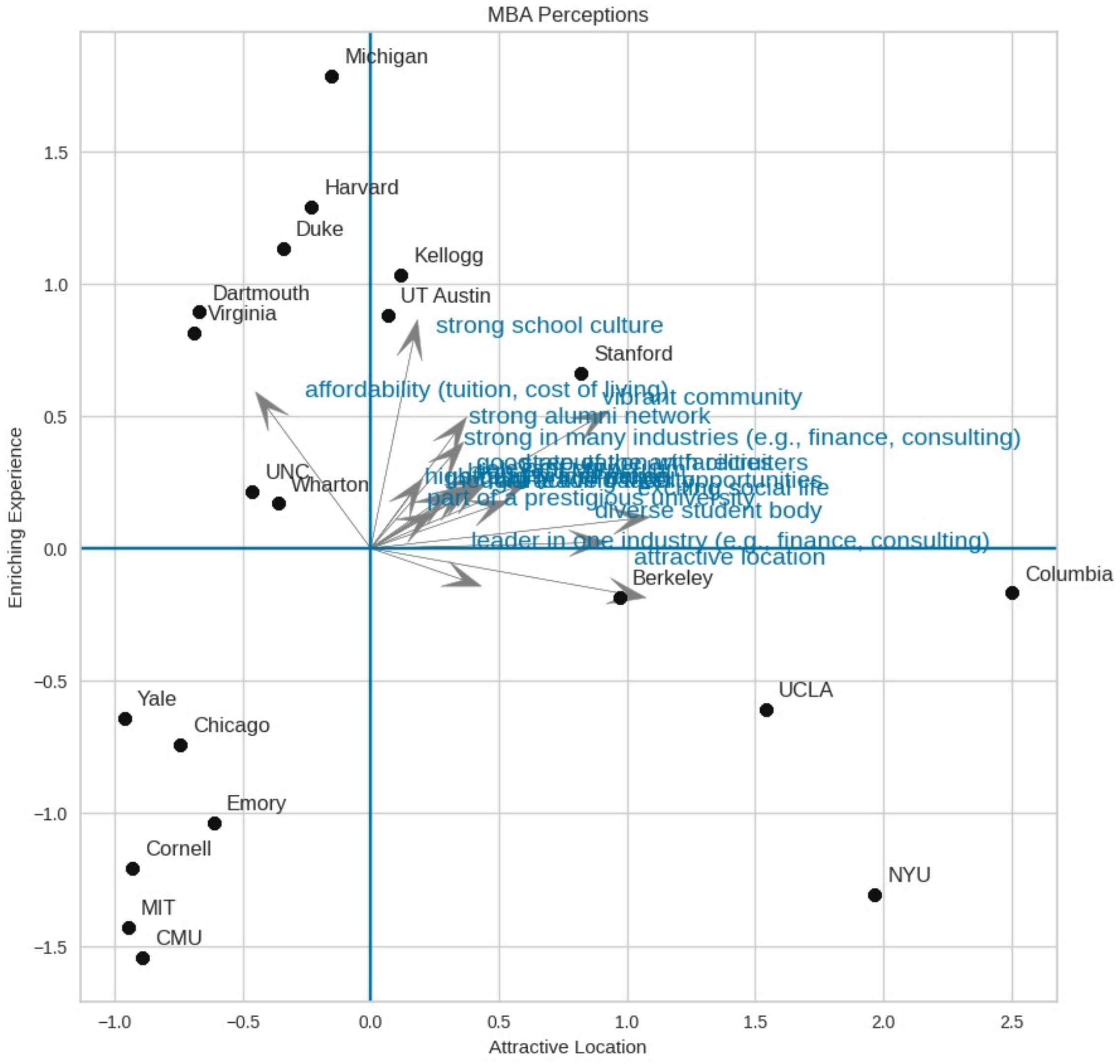
Factor 1 vs. Factor 2



Factor 1 vs. Factor 3



Factor 2 vs. Factor 3



Ford Ka

Ford Ka

How did Ford (and other car makers) segment the overall car market? What was the typical small car marketing strategy in the past?

- **Size-tiers** based segmentation (A, B, …; small, midsize, large, luxury)
  - Is that segmentation? No: product categorization
  - But **Price** correlated with size, so car companies targeted based on demographics:
    - small cars: young , lower income buyers
    - large cars: wealthier, older buyers and families

- How about competition?
  - Aligned to these segments:
    - small car companies focus on **cost-differentiation**
    - large/luxury cars companies adopt a **product differentiation** strategy

How did the small car market change?

- Becoming attractive and growing:
  - Environmental changes – high taxes on fuel, lower taxes on small cars, environmental-consciousness, traffic, easier parking
  - Demographic changes – more women in the labor force, smaller household size
- Competition:
  - Started innovating on style and design features
  - Luxury brands entered the market (BMW, Mercedes)
- Twingo's successful launch in 1992
  - Confirmed the need for urban, stylish car

Intense competitive pressure

How did Ford react?

- Launched new Fiesta for Trend-B market (better performance and more features in a small car):
  - Couldn't compete with Twingo → Ford Ka

- How did Ford design Ka?
  - Purpose was a quick and cheap response to Twingo
  - Built on Fiesta chassis (frame)
  - Difference with Renault: Renault started with market research before design

For Ford, market research started after they designed Ka

Is the existing segmentation approach still applicable for Ford Ka?

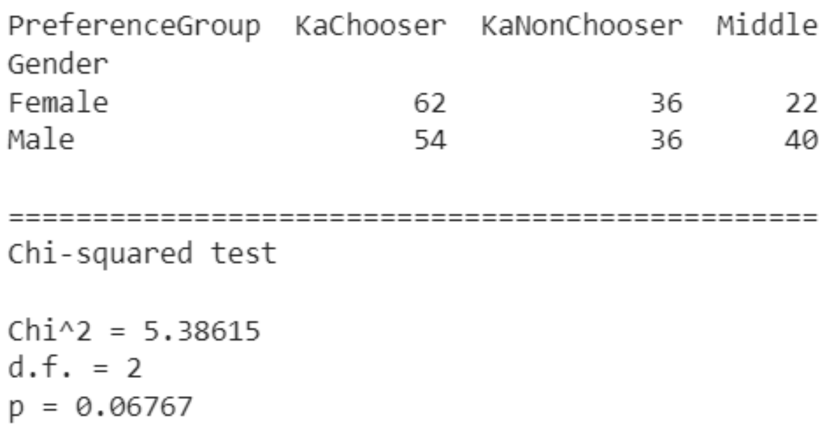
- **No!** Why?
- Small car market is changing rapidly and deviating from demographic segmentation
- Change in consumer preferences:
  - Price is no longer the driving factor
  - People are looking for luxury features in their small cars
- **Need-based segmentation**: small B category is now split into Basic, Trend, and Other (luxury)

Overall, shift from consumer demographics to consumer needs

Market Research Objectives

- Main objective: 250,000 units per year in Europe
- Decisions:
  - How to segment the small car market?
  - Which segment(s) to target?
- Exploratory research
  - 30 focus groups + interviews to assess consumer reaction to the KA
  - What are the key learnings?
    - Polarized responses (young found it risky, older thought it was for young people)
    - Among top 3 choices: Also Twingo + Tigra → style is major appeal for Ka
    - Among bottom 3: did not include Twingo + Tigra → less broad appeal
- Descriptive research
  - Attitudinal survey of 250 customers about preference/perception

Can we use the "old" segmentation?
Can gender separate Ka choosers and non-choosers?

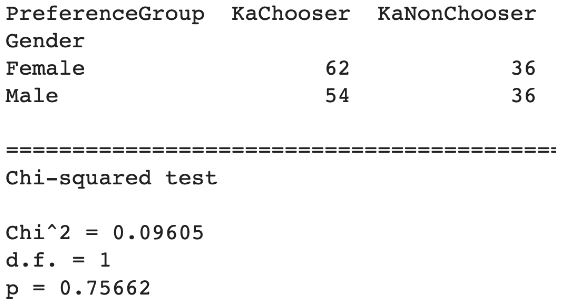


```
PreferenceGroup  KaChooser  KaNonChooser  Middle
Gender
Female                  62            36      22
Male                    54            36      40

================================================
Chi-squared test

Chi^2 = 5.38615
d.f. = 2
p = 0.06767
```

Ka choosers are more likely to be female

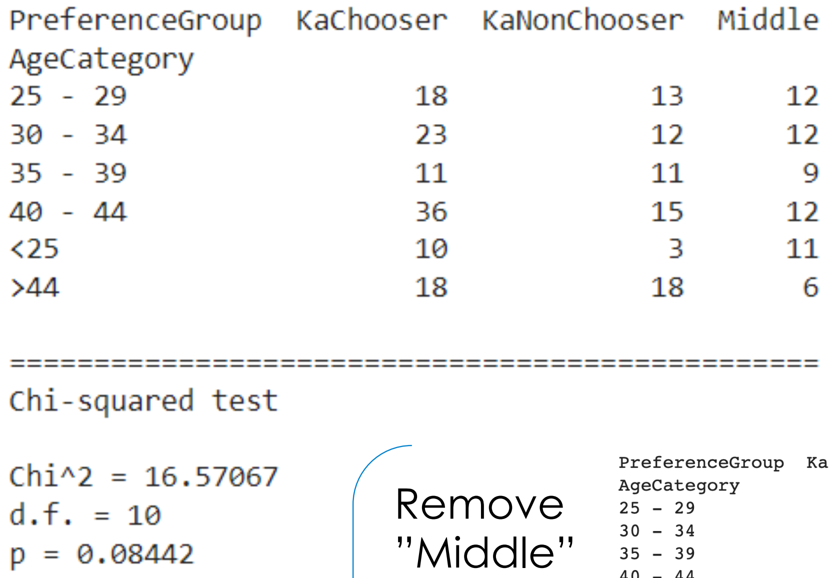
P-value=0.068 → Variables are independent (can't reject the null hypothesis of independence)

Remove "Middle"



```
PreferenceGroup  KaChooser  KaNonChooser
Gender
Female                  62            36
Male                    54            36

==========================================
Chi-squared test

Chi^2 = 0.09605
d.f. = 1
p = 0.75662
```

Can we use the “old” segmentation?
Can age separate Ka choosers and non-choosers?

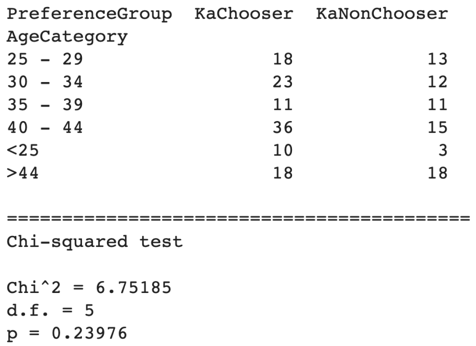


```
PreferenceGroup  KaChooser  KaNonChooser  Middle
AgeCategory
25 - 29                 18            13      12
30 - 34                 23            12      12
35 - 39                 11            11       9
40 - 44                 36            15      12
<25                     10             3      11
>44                     18            18       6

================================================
Chi-squared test

Chi^2 = 16.57067
d.f. = 10
p = 0.08442
```

Ka choosers are more likely to be 40-44 years old

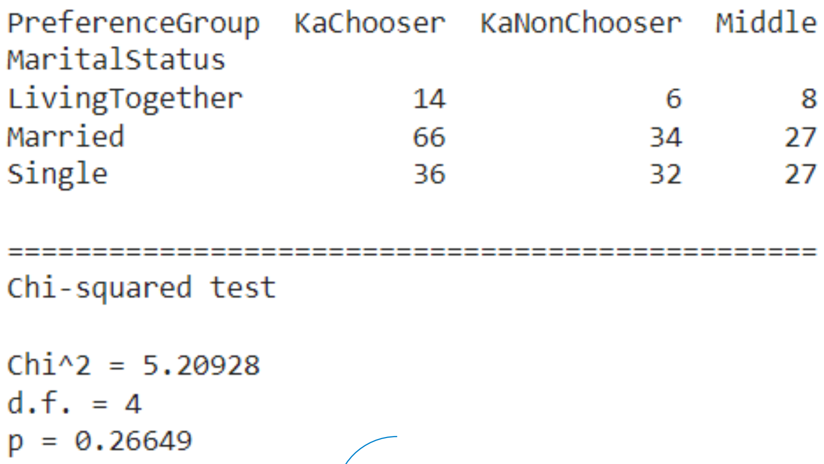
P-value=0.084 → Variables are independent (can’t reject the null hypothesis of independence)

Remove "Middle"



```
PreferenceGroup  KaChooser  KaNonChooser
AgeCategory
25 - 29                 18            13
30 - 34                 23            12
35 - 39                 11            11
40 - 44                 36            15
<25                     10             3
>44                     18            18

==========================================
Chi-squared test

Chi^2 = 6.75185
d.f. = 5
p = 0.23976
```

Can we use the “old” segmentation?
Can marital status separate Ka choosers and non-choosers?

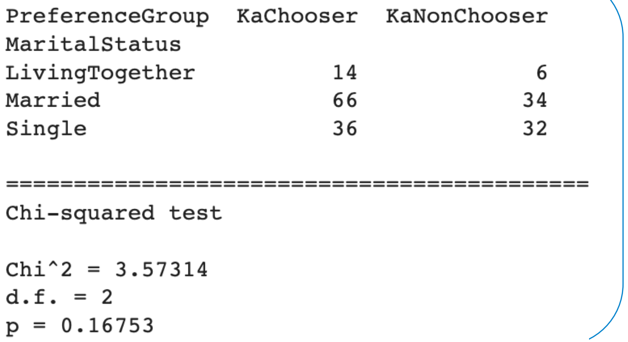


```
PreferenceGroup  KaChooser  KaNonChooser  Middle
MaritalStatus
LivingTogether          14             6       8
Married                 66            34      27
Single                  36            32      27

==================================================
Chi-squared test

Chi^2 = 5.20928
d.f. = 4
p = 0.26649
```

Ka choosers are more likely to be married

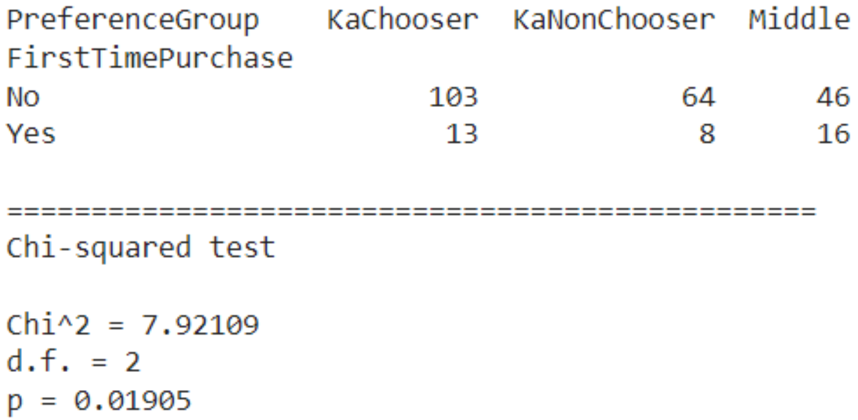
P-value=0.266 → Variables are independent (can’t reject the null hypothesis of independence)

Remove "Middle"



```
PreferenceGroup KaChooser  KaNonChooser
MaritalStatus
LivingTogether         14             6
Married                66            34
Single                 36            32

==========================================
Chi-squared test

Chi^2 = 3.57314
d.f. = 2
p = 0.16753
```

Can we use the “old” segmentation?
Can first car separate Ka choosers and non-choosers?

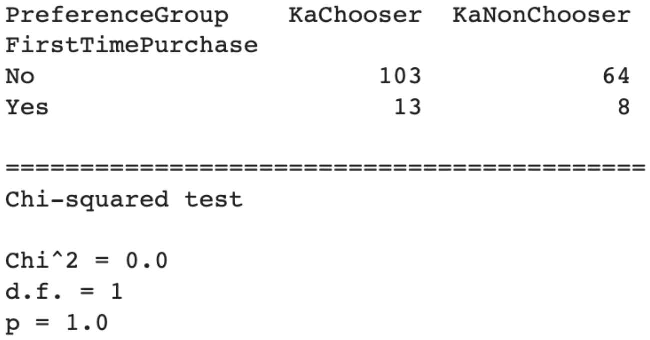


```
PreferenceGroup     KaChooser  KaNonChooser  Middle
FirstTimePurchase
No                        103            64      46
Yes                        13             8      16

================================================
Chi-squared test

Chi^2 = 7.92109
d.f. = 2
p = 0.01905
```

Ka choosers are less likely to be first time buyers

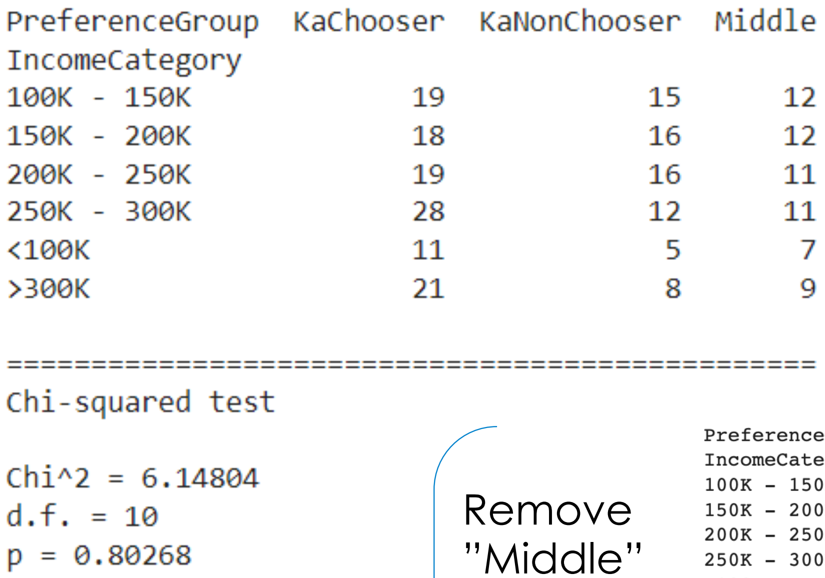
P-value=0.019 → Variables are not independent

Remove "Middle"



```
PreferenceGroup    KaChooser  KaNonChooser
FirstTimePurchase
No                       103            64
Yes                       13             8

==========================================
Chi-squared test

Chi^2 = 0.0
d.f. = 1
p = 1.0
```

Can we use the "old" segmentation?
Can income separate Ka choosers and non-choosers?

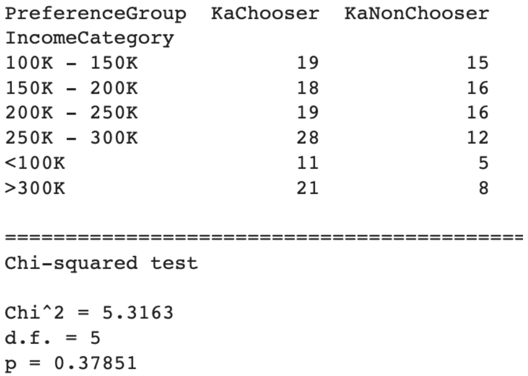


```
PreferenceGroup  KaChooser  KaNonChooser  Middle
IncomeCategory
100K - 150K             19            15      12
150K - 200K             18            16      12
200K - 250K             19            16      11
250K - 300K             28            12      11
<100K                   11             5       7
>300K                   21             8       9

================================================
Chi-squared test

Chi^2 = 6.14804
d.f. = 10
p = 0.80268
```

Ka choosers are more likely to have income greater than 250k

P-value=0.802 → Variables are independent (can't reject the null hypothesis of independence)

Remove "Middle"



```
PreferenceGroup  KaChooser  KaNonChooser
IncomeCategory
100K - 150K             19            15
150K - 200K             18            16
200K - 250K             19            16
250K - 300K             28            12
<100K                   11             5
>300K                   21             8

==========================================
Chi-squared test

Chi^2 = 5.3163
d.f. = 5
p = 0.37851
```

Old Segmentation Approach

Simply impossible!

- What is the alternative?
    - Needs-based segmentation
    - Attitudinal/psychographics data

Attitudinal Segmentation

- Cluster analysis on factors derived from the attitudinal variables

- Two-step solution
  - Factor Analysis on the attitudinal questions
  - Use the resulting factors to cluster

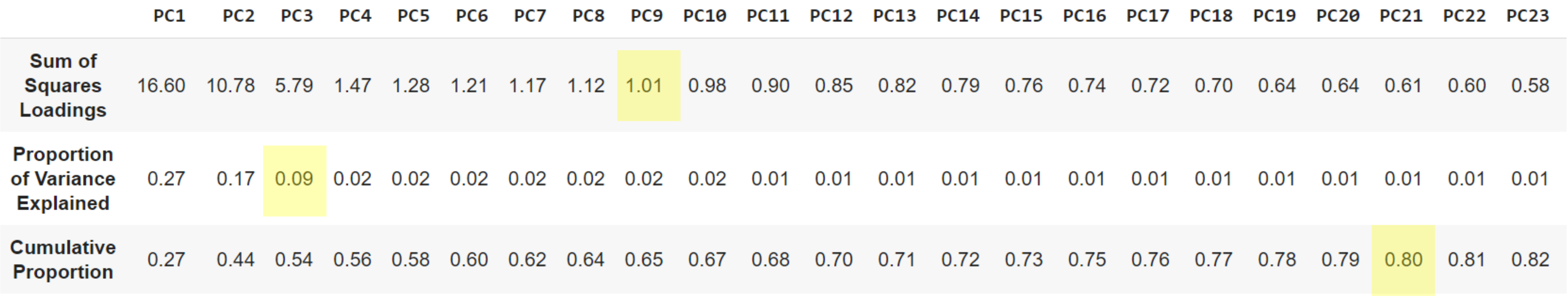
- Same as Dupont example

Steps to Factor Analysis with PCA

1. Estimate all the principal components (**without rotation**)
2. Determine the number of components (factors) to keep:
   - Eigenvalues>1, Cumulative Variance>80%, scree plot, managerial relevance
3. Compute the **rotated** factor loading matrix to understand (and name!) the underlying factors
4. Compute the factor scores:
   - For each observation, what are the values of the factors?

Dimension Reduction – Steps 1 and 2

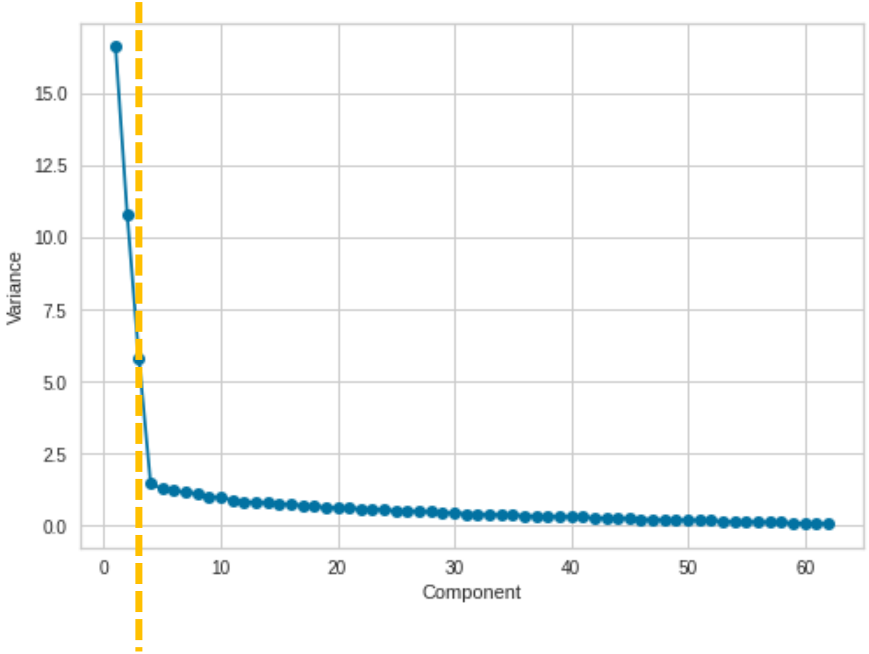
- Note: The table below is a truncated output
- How many factors do we retain?



	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23
Sum of Squares Loadings	16.60	10.78	5.79	1.47	1.28	1.21	1.17	1.12	1.01	0.98	0.90	0.85	0.82	0.79	0.76	0.74	0.72	0.70	0.64	0.64	0.61	0.60	0.58
Proportion of Variance Explained	0.27	0.17	0.09	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Cumulative Proportion	0.27	0.44	0.54	0.56	0.58	0.60	0.62	0.64	0.65	0.67	0.68	0.70	0.71	0.72	0.73	0.75	0.76	0.77	0.78	0.79	0.80	0.81	0.82

Dimension Reduction – Step 1 and 2

- But big drop after three factors (from 5.79 to 1.47)
  - Every additional factor only explains 2% of variance



Let's use three factors.

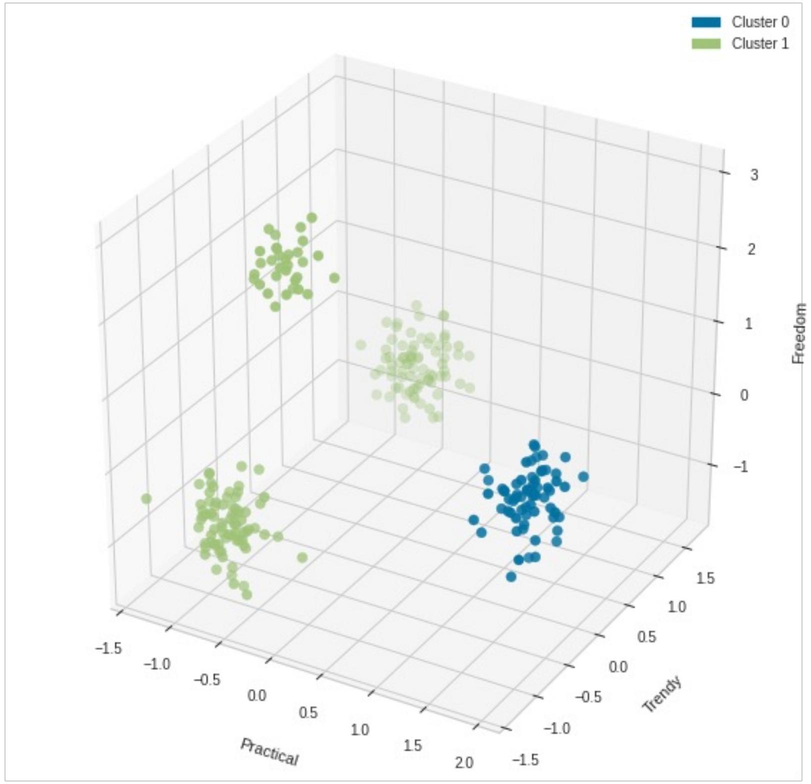
Dimension Reduction – Step 3 with 3 Factors

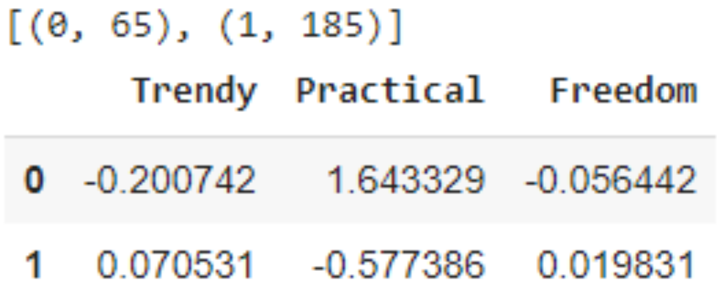
Statement	RC1	RC2	RC3	h2	Statement	RC1	RC2	RC3	h2
I am fashion conscious.	0.91	0.02	0.07	0.834	Fuel economy comes at the expense of performance.	-0.03	0.67	-0.02	0.445
Buying a car on a lower interest rate does not interest me.	0.9	0.04	0.13	0.822	I love to drive.	-0.03	0.09	-0.05	0.012
I always want the latest style and design in a vehicle.	0.77	-0.3	-0.02	0.678	I want a car that is fuel economic.	-0.05	0.03	-0.12	0.017
When it comes to cars my heart rules my head.	0.77	-0.13	-0.16	0.633	I have always been fascinated by cars which have a cult following.	-0.06	0.71	0.03	0.512
A car is an extension of oneself.	0.74	-0.26	-0.11	0.632	I am interested in car maintenance.	-0.08	-0.22	0.63	0.458
People ought to buy domestic products for the good of the country.	0.73	-0.27	-0.09	0.616	I do not believe that a Swatch branded car will be successful.	-0.08	0.68	0.06	0.475
My car must have a very individual interior.	0.73	-0.29	-0.14	0.638	For me a car is a symbol of freedom and independence.	-0.09	-0.11	0.66	0.458
Nowadays smart cars are mainly foreign brands.	0.71	-0.28	-0.11	0.596	I want a car that is easy to handle.	-0.1	0.13	-0.04	0.029
I want to buy a car that makes a statement about me.	0.65	-0.63	0	0.822	I like to believe that the car I drive will one day become a cult car.	-0.1	0.72	-0.05	0.533
In today's world it is anti-social to drive big cars.	0.65	-0.49	-0.46	0.873	City driving is my main concern.	-0.1	0.67	-0.08	0.466
I want a car that is trendy.	0.63	-0.32	0.41	0.668	I want a practical car.	-0.11	0.69	-0.02	0.482
Domestic made is best made.	0.59	0.48	-0.2	0.611	I do not have the time to worry about car maintenance.	-0.12	0.69	-0.1	0.497
Small cars are for kids.	0.59	0.42	-0.25	0.588	I want a car that has character.	-0.13	-0.2	0.61	0.425
I consider myself an authority on cars.	0.56	0.53	-0.25	0.651	I prefer cars with high performance.	-0.23	0.5	0.57	0.633
Having a masculine car is important to me.	0.5	0.32	0.44	0.542	I want a car that drives well on country roads.	-0.33	0.62	0.47	0.708
Small cars are for women.	0.46	0.41	0.4	0.538	I want a car that is nippy and zippy.	-0.37	-0.81	0.18	0.82
A car is a fashion accessory to me.	0.45	0.38	0.34	0.462	I want a vehicle that is environmentally friendly.	-0.5	-0.33	-0.39	0.51
Small cars are not prestigious.	0.34	0.75	0.24	0.737	I wish there were stricter exhaust regulations.	-0.53	-0.3	-0.3	0.461
The government should implement policies that favor public transportation.	0.26	0.12	-0.67	0.537	I prefer buying my next car from the same car manufacturer.	-0.53	-0.3	-0.37	0.509
The government is right to tax large cars more heavily than small cars.	0.16	0.14	-0.59	0.396	Small cars are much safer nowadays.	-0.58	-0.45	0.24	0.6
Many manufacturers do not really care about their customers needs.	0.16	-0.49	-0.59	0.611	I want a comfortable car.	-0.62	0.51	0.48	0.87
Small cars take up less room in today's traffic.	0.11	-0.68	0.08	0.476	I want the most equipment I can get for my money.	-0.63	-0.41	0.25	0.627
I am looking for a car which delivers a smooth ride.	0.09	-0.08	-0.06	0.019	Good aerodynamics help fuel economy.	-0.64	-0.38	0.26	0.625
My car must function with total reliability.	0.09	-0.04	-0.08	0.015	I have a relationship with my car.	-0.72	0.08	0.52	0.796
When buying a car I only consider a national make.	0.09	0.05	-0.69	0.484	Quality and reliability of products are my main concerns.	-0.72	0.06	0.51	0.781
I would rather deal with a manufacturer's rep than a salesperson.	0.06	-0.69	0.11	0.491	Most environmentally friendly products do not perform as well as those they replaced	-0.74	0.25	0.18	0.644
I prefer small cars.	0.06	-0.63	0.01	0.397	I want a car equipped with the latest features and technology.	-0.77	0.21	0.18	0.675
Today's cars are more efficient than yesterday's.	0.06	-0.05	-0.02	0.006	Image is not important to me in a car.	-0.77	0.16	0.11	0.634
My car must function with total reliability.	0.06	-0.03	-0.17	0.034	Cars all look the same these days.	-0.79	0.19	0.04	0.67
Basic transportation is all I need.	0.01	0.64	-0.54	0.708	The car I buy must be able to handle long motorway journeys.	-0.89	-0.14	0.25	0.877
A car is an extension of oneself.	0	-0.02	0.11	0.013	One should not spend beyond ones means.	-0.89	-0.04	-0.09	0.799

Dimension Reduction – Step 3 Naming

- Positive loading means statement is positively correlated with factor
- Negative loading means that opposite statement is positively correlated with factor
- Factor 1: Trendy
  - High score: "I am fashion conscious"; "I always want the latest style and design in a vehicle"
  - Low score (neg): "Cars all look the same these days"
- Factor 2: Practical
  - High score: "Small cars are not prestigious"; "I want a practical car"
  - Low score: "I want a car that is nippy and zippy"; "I prefer small cars"
- Factor 3: Freedom
  - High score: "For me a car is a symbol of freedom and independence"
  - Low score: "The government should implement policies that favor public transportation"

Cluster Analysis – 2 segments



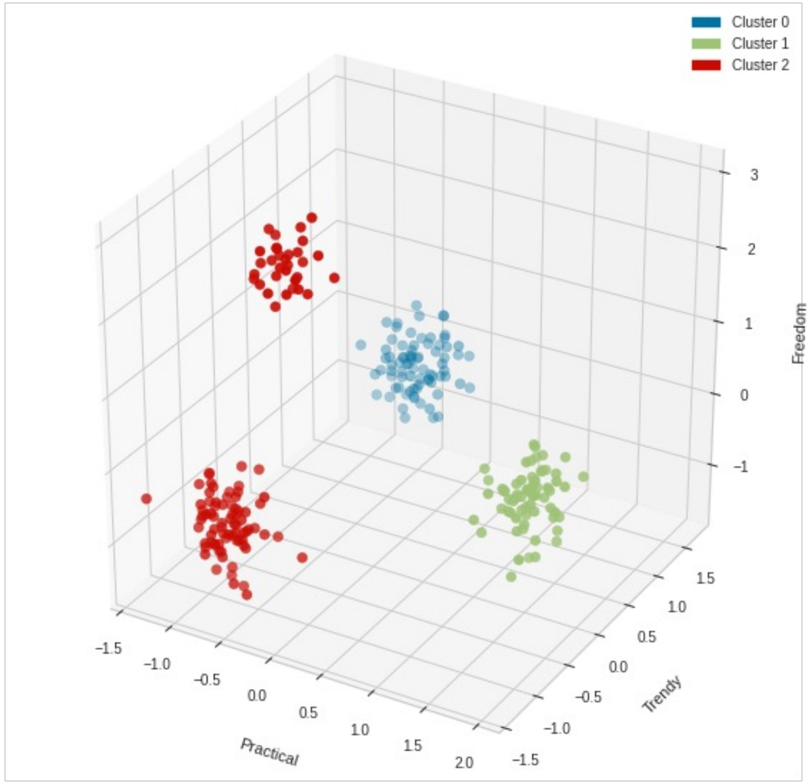


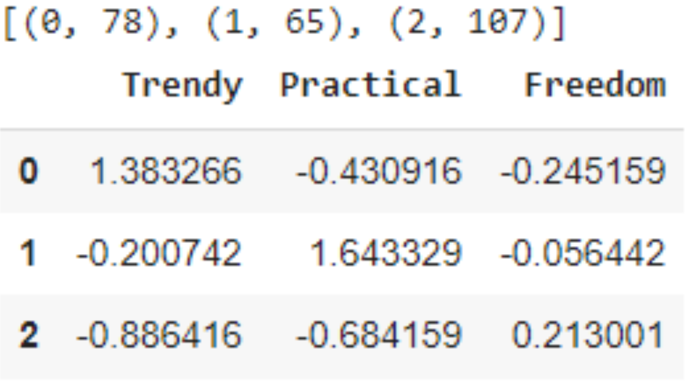
```
[(0, 65), (1, 185)]
```

	Trendy	Practical	Freedom
0	-0.200742	1.643329	-0.056442
1	0.070531	-0.577386	0.019831

Between SS / Total SS: 32.0%

Cluster Analysis – 3 segments



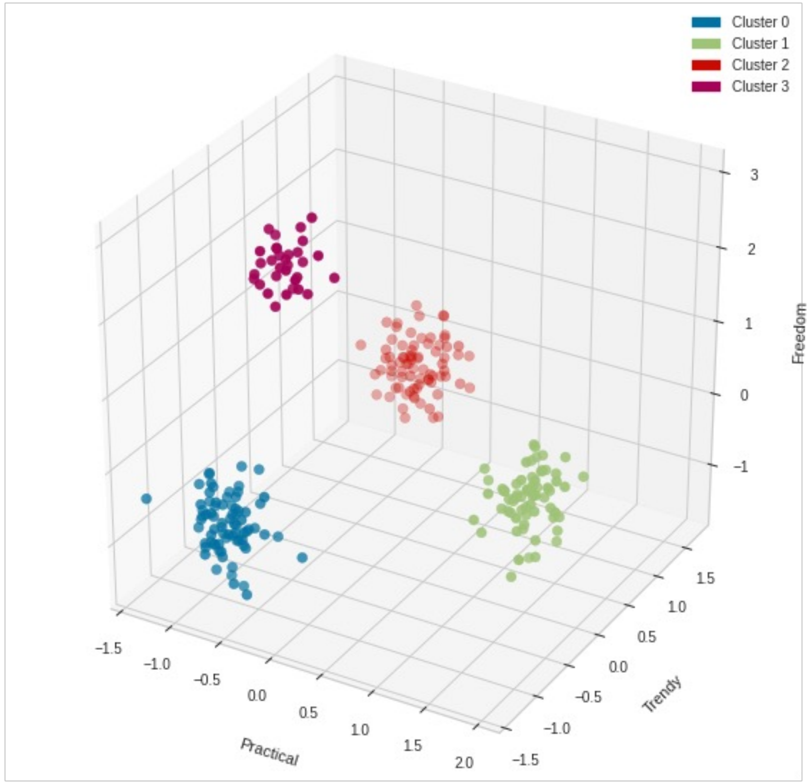


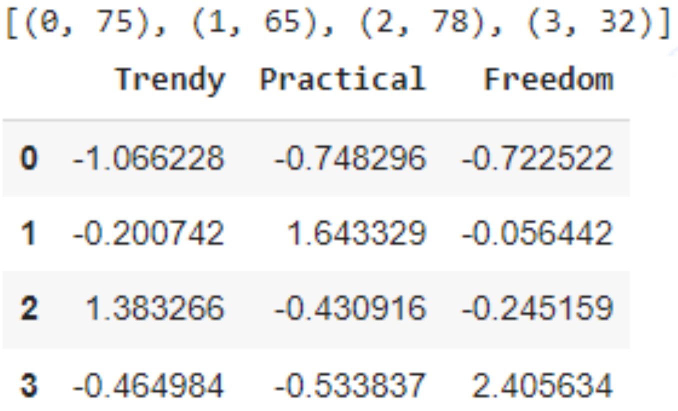
```
[(0, 78), (1, 65), (2, 107)]
```

	Trendy	Practical	Freedom
0	1.383266	-0.430916	-0.245159
1	-0.200742	1.643329	-0.056442
2	-0.886416	-0.684159	0.213001

Between SS / Total SS: 65.0%

Cluster Analysis – 4 segments



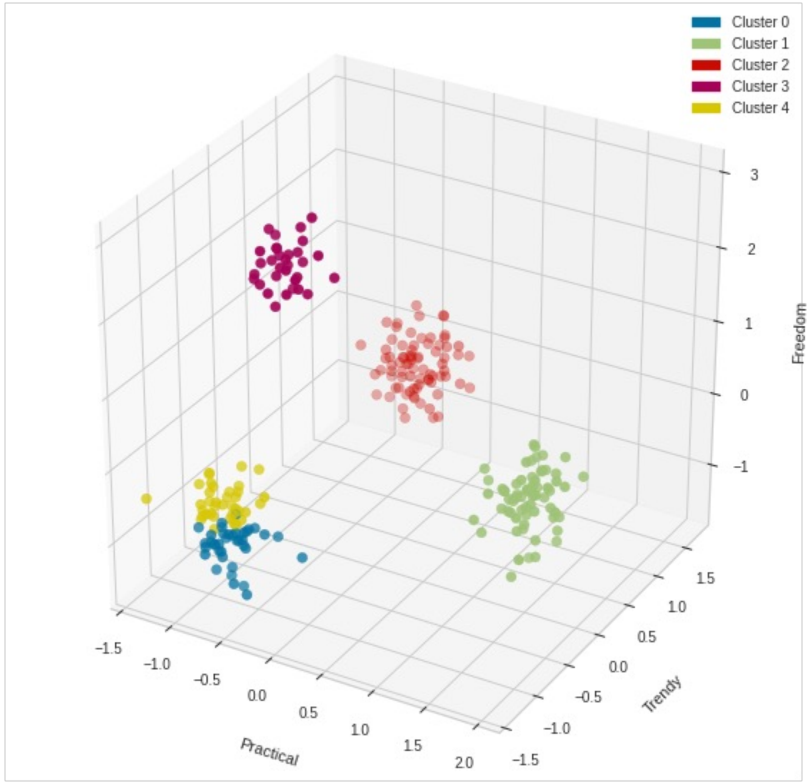


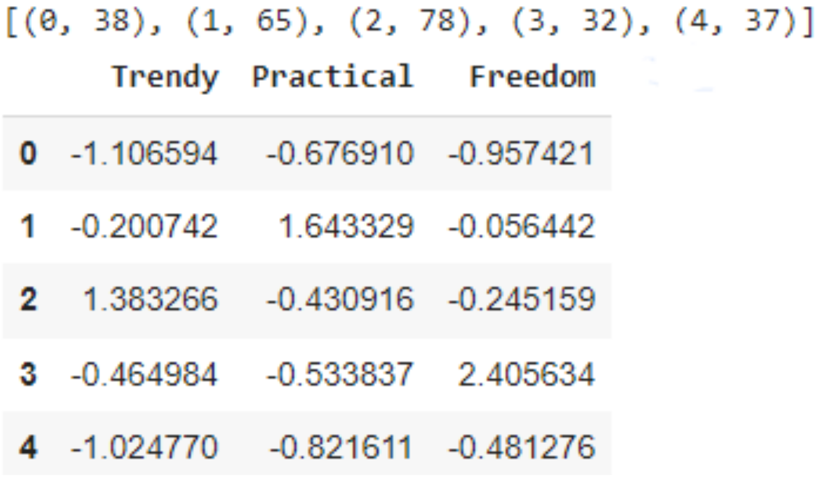
```
[(0, 75), (1, 65), (2, 78), (3, 32)]
```

	Trendy	Practical	Freedom
0	-1.066228	-0.748296	-0.722522
1	-0.200742	1.643329	-0.056442
2	1.383266	-0.430916	-0.245159
3	-0.464984	-0.533837	2.405634

Between SS / Total SS: 95.0%

Cluster Analysis – 5 segments



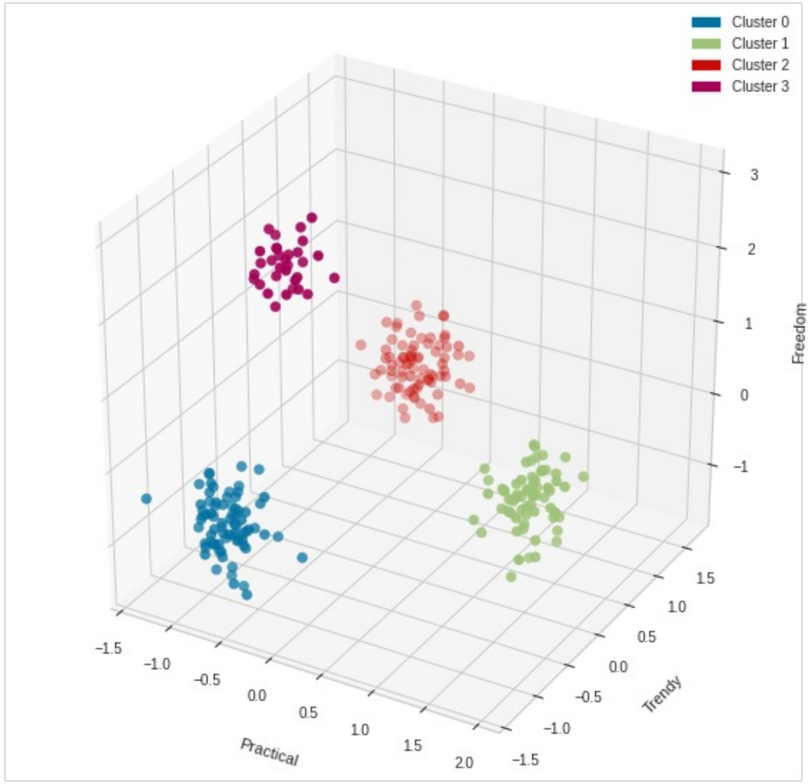


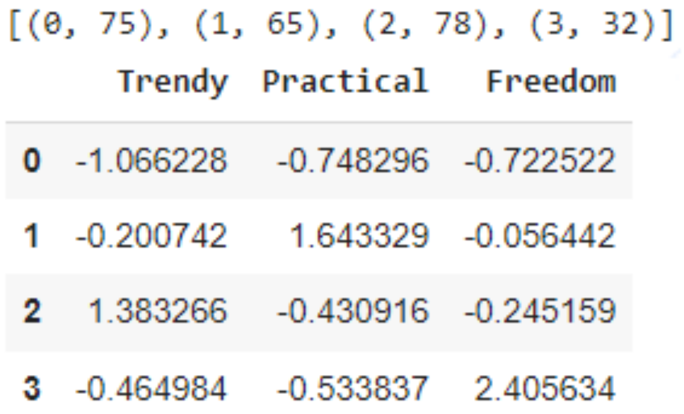
```
[(0, 38), (1, 65), (2, 78), (3, 32), (4, 37)]
```

	Trendy	Practical	Freedom
0	-1.106594	-0.676910	-0.957421
1	-0.200742	1.643329	-0.056442
2	1.383266	-0.430916	-0.245159
3	-0.464984	-0.533837	2.405634
4	-1.024770	-0.821611	-0.481276

Between SS / Total SS: 96.0%

Cluster Analysis – 4 segments



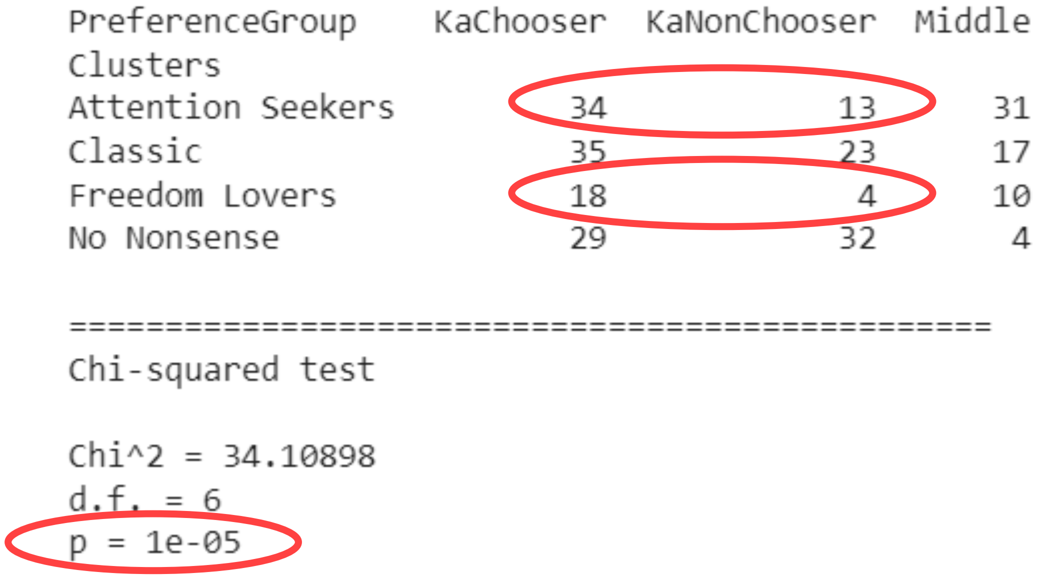


```
[(0, 75), (1, 65), (2, 78), (3, 32)]
      Trendy  Practical   Freedom
0  -1.066228  -0.748296  -0.722522
1  -0.200742   1.643329  -0.056442
2   1.383266  -0.430916  -0.245159
3  -0.464984  -0.533837   2.405634
```

Between SS / Total SS: 95.0%

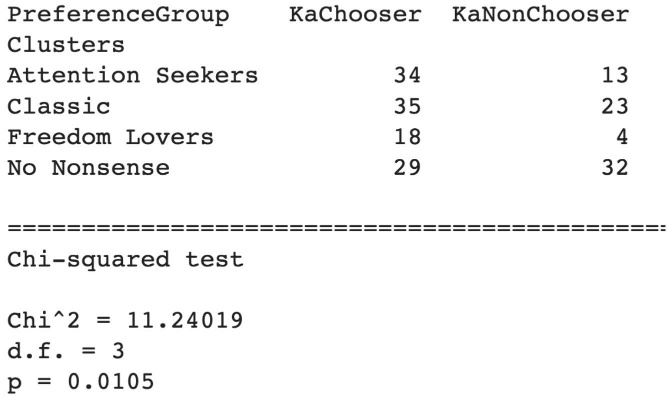
- Segment 1: Classics (75)
- Segment 2: No-nonsense (65)
- Segment 3: Attention-Seekers (78)
- Segment 4: Freedom Lovers/Car Lovers (32)

What segmentation approach do you recommend and who is your target buyer? Why?



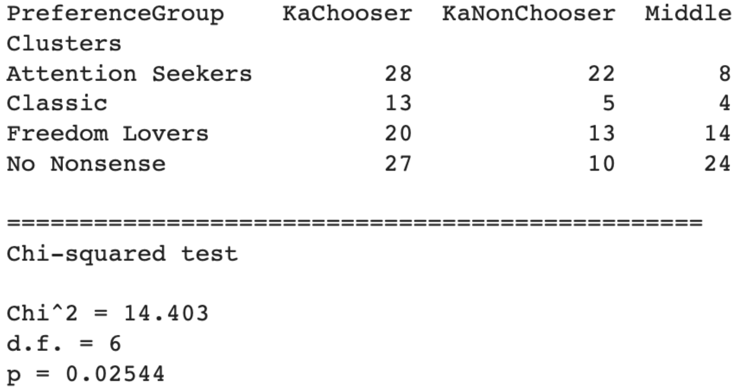
```
PreferenceGroup    KaChooser  KaNonChooser  Middle
Clusters
Attention Seekers         34            13      31
Classic                   35            23      17
Freedom Lovers            18             4      10
No Nonsense               29            32       4

================================================
Chi-squared test

Chi^2 = 34.10898
d.f. = 6
p = 1e-05
```

Remove "Middle" After



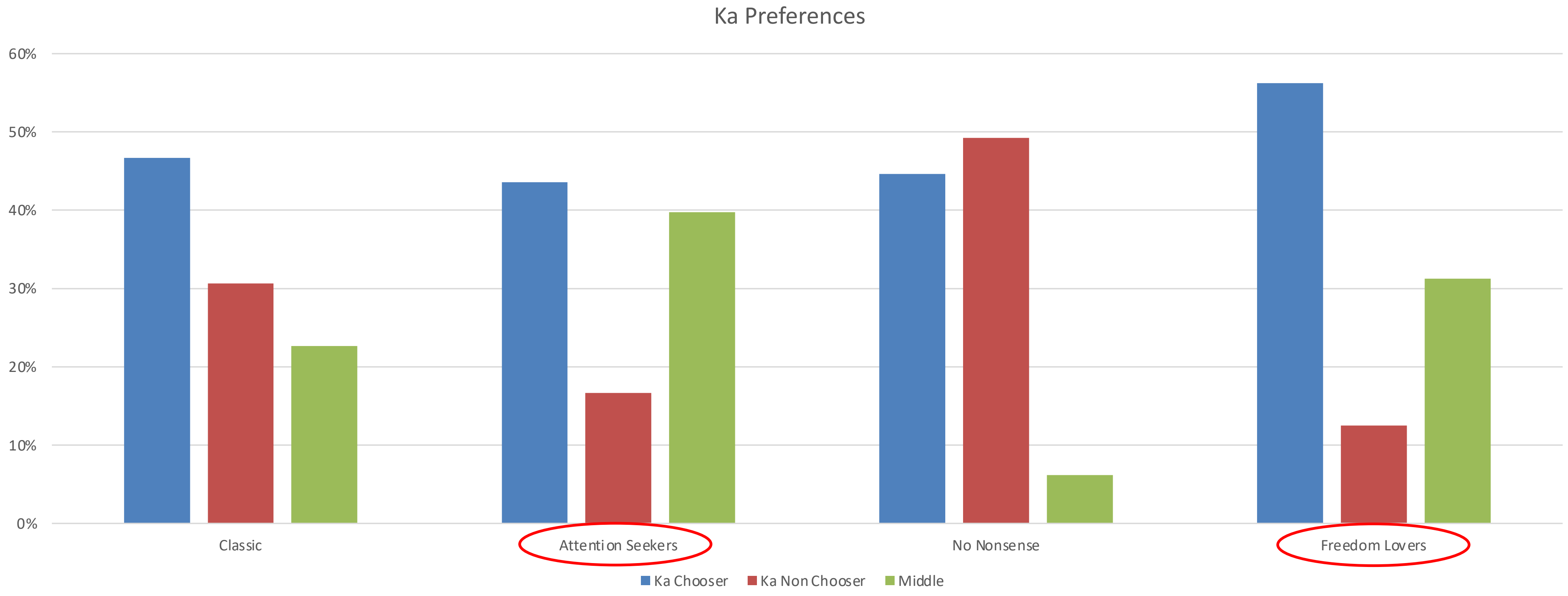
```
PreferenceGroup    KaChooser  KaNonChooser
Clusters
Attention Seekers         34            13
Classic                   35            23
Freedom Lovers            18             4
No Nonsense               29            32

============================================
Chi-squared test

Chi^2 = 11.24019
d.f. = 3
p = 0.0105
```

Remove "Middle" Before



```
PreferenceGroup    KaChooser  KaNonChooser  Middle
Clusters
Attention Seekers         28            22       8
Classic                   13             5       4
Freedom Lovers            20            13      14
No Nonsense               27            10      24

================================================
Chi-squared test

Chi^2 = 14.403
d.f. = 6
p = 0.02544
```

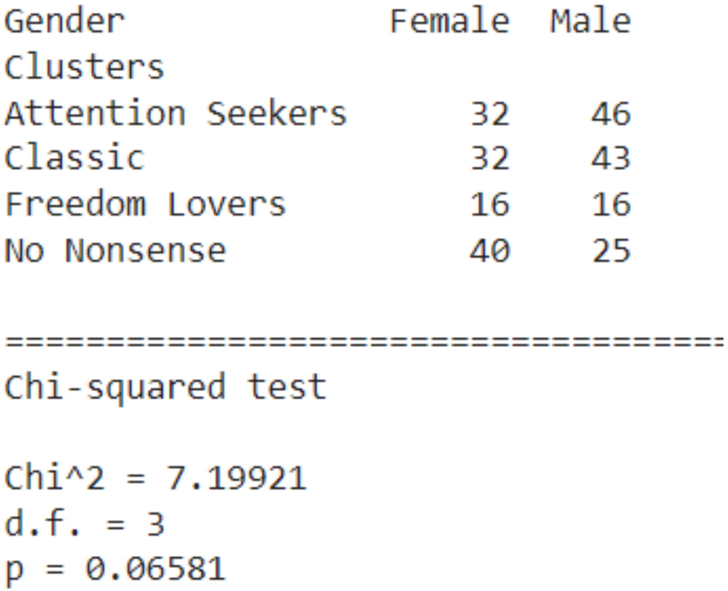
Cluster Results and Ka Preference



N=78

N=32

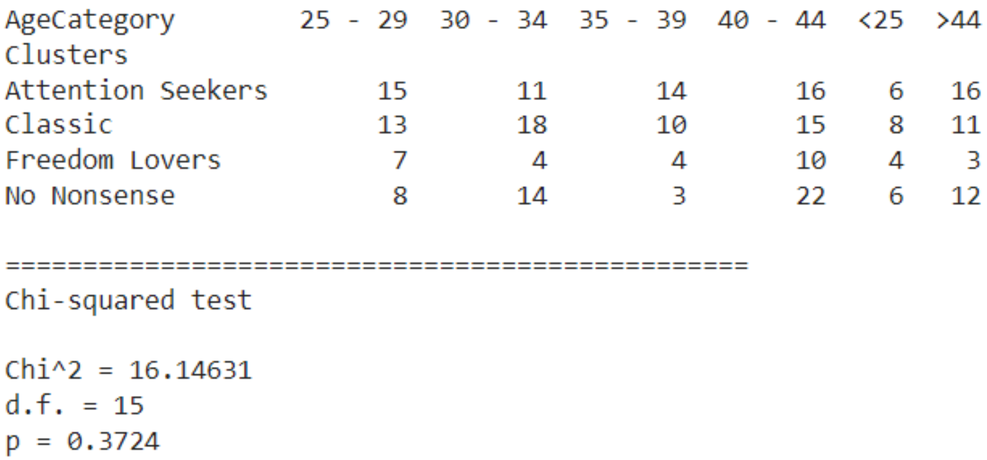
Relating the Clusters with Demographics
Gender



```
Gender              Female  Male
Clusters
Attention Seekers       32    46
Classic                 32    43
Freedom Lovers          16    16
No Nonsense             40    25

====================================
Chi-squared test

Chi^2 = 7.19921
d.f. = 3
p = 0.06581
```

- Attention-seekers are more likely to be male
- P-value=0.066: relationship marginally significant

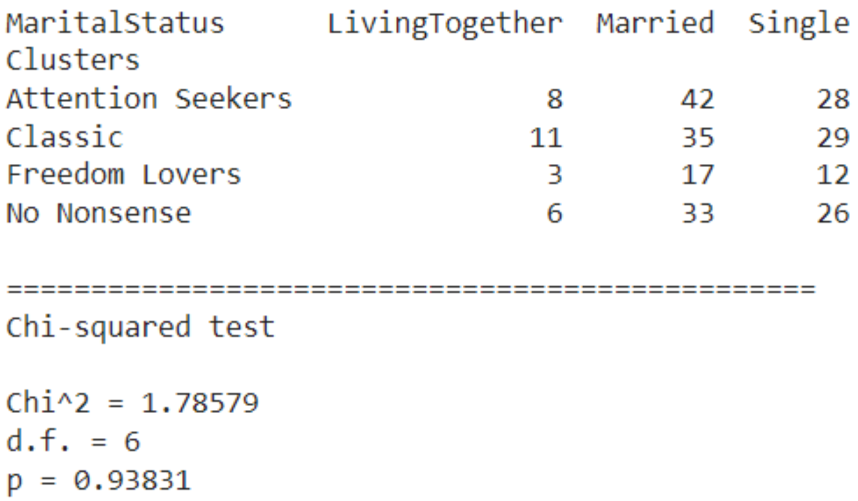
Relating the Clusters with Demographics
Age



```
AgeCategory        25 - 29  30 - 34  35 - 39  40 - 44   <25  >44
Clusters
Attention Seekers       15       11       14       16     6   16
Classic                 13       18       10       15     8   11
Freedom Lovers           7        4        4       10     4    3
No Nonsense              8       14        3       22     6   12

================================================
Chi-squared test

Chi^2 = 16.14631
d.f. = 15
p = 0.3724
```

- Age is not a clear descriptor of the segments
- Relationship is not significant

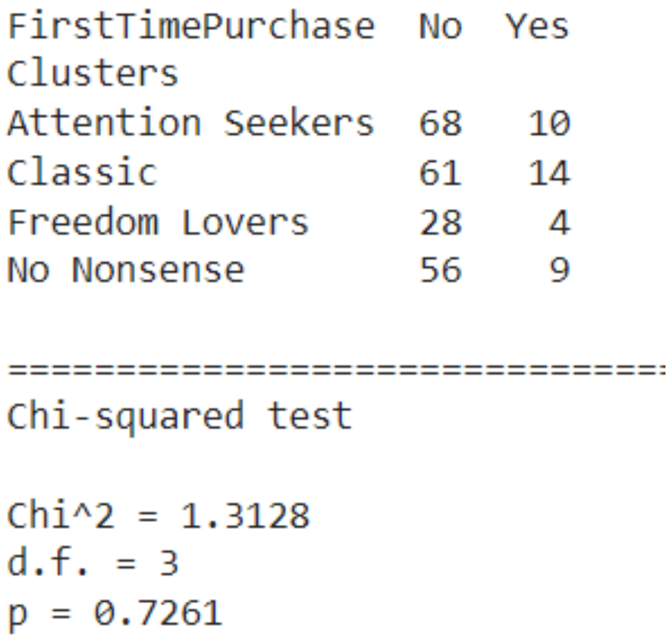
Relating the Clusters with Demographics Marital Status



```
MaritalStatus       LivingTogether  Married  Single
Clusters
Attention Seekers                8       42      28
Classic                         11       35      29
Freedom Lovers                   3       17      12
No Nonsense                      6       33      26

================================================
Chi-squared test

Chi^2 = 1.78579
d.f. = 6
p = 0.93831
```

- No significant difference across segments
- Relationship is not significant

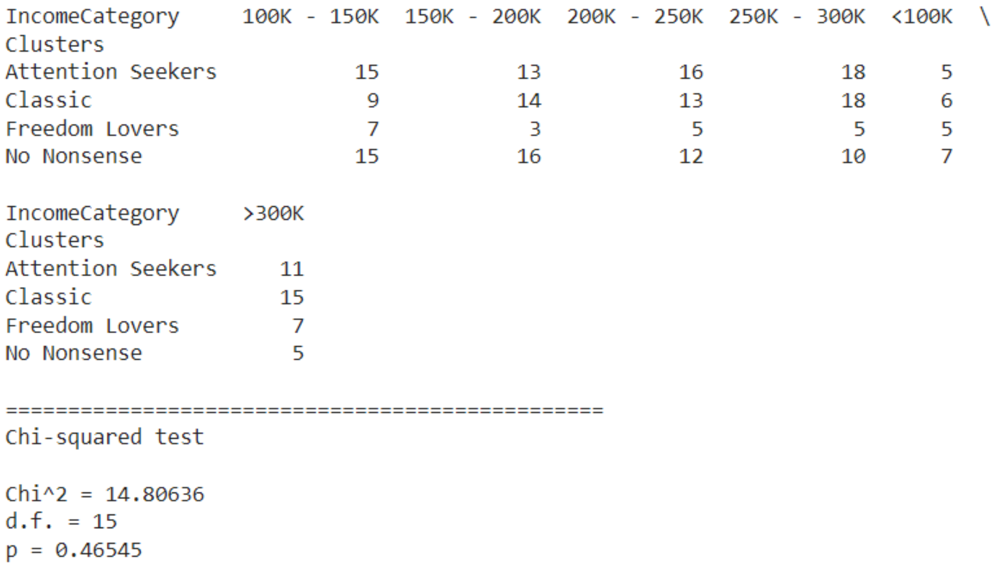
Relating the Clusters with Demographics
First Car



```
FirstTimePurchase  No  Yes
Clusters
Attention Seekers  68   10
Classic            61   14
Freedom Lovers     28    4
No Nonsense        56    9

==============================
Chi-squared test

Chi^2 = 1.3128
d.f. = 3
p = 0.7261
```

- Target segments are somewhat less likely to be first car buyers
- Relationship is not significant

Relating the Clusters with Demographics Income



```
IncomeCategory     100K - 150K  150K - 200K  200K - 250K  250K - 300K  <100K  \
Clusters
Attention Seekers           15           13           16           18      5
Classic                      9           14           13           18      6
Freedom Lovers               7            3            5            5      5
No Nonsense                 15           16           12           10      7

IncomeCategory     >300K
Clusters
Attention Seekers     11
Classic               15
Freedom Lovers         7
No Nonsense            5

================================================
Chi-squared test

Chi^2 = 14.80636
d.f. = 15
p = 0.46545
```

- Income does not discriminate well between the segments
- Relationship is not significant

Relating the Clusters with Demographics

- Almost no relationship between the clusters and demographics
- Could we have predicted it?
- Why is it a problem?

Chosen Segmentation Approach
Potential implementation problems

- Attitudinal Segmentation: Attention-seekers and Freedom Lovers
  - Demographic segmentation is not possible
  - Needs-based segmentation found two clusters with positive opinions on Ka
- Potential Problems:
  - How do you reach those two targets? Which media?
  - Uncertainty on segment sizes (original survey wasn't random)
    - 43% of respondents are in those segments. Does this mean 43% of the population?
  - Compatibility with Ford's image
  - Resistance from upper management
  - New approach for the company which should be embraced by all (dealers)

What happened?

- First months were very slow
- In 1998 Ford sold 266,000 Ford Ka ☑
  - Sales target (250,000) achieved
  - Ford Ka became best selling car in Europe
    - Outsold the Renault Twingo
    - Did very well in U.K., Germany, and Scandinavia
    - Did less well in Eastern Europe
    - Sold in Australia and Brazil
- In 2002 sales dropped below 200,000 units
  - Lost lead to Twingo
- In 2005
  - Ford ka is launched in South Africa

What happened?

- 2006 and forward
  - In Jan. 2006 sales in UK reached 400,000 units
  - "Today's Ka driver still loves the funky look of Ka but also wants more technology and luxury" Ford Ka Marketing Director 01/06

- 2008 and forward:
  - Opened a joint plant with Fiat in Poland
  - The new Ford Ka appeared in Bond's "Quantum of Solace"

- The "Evil Twin" viral marketing campaign

"In order to differentiate the dashing SportsKa from its more sedate namesake, the new model is presented as the "Evil Twin" of the Ford Ka. In this viral movie, a pigeon that flies over the SportsKa is struck by its bonnet, and instantly killed. This Ka clearly has a devilish streak."

Agency: Ogilvy, London, The Viral Factory, London

To Summarize: Marketing Strategy

1. Needs-Based/Benefits/Behavioral Segmentation
    - Group customers into segments based on similar needs/benefits (if necessary, with factor analysis)
2. Segment Identification
    - For each segment, determine which demographics or behaviors make the segments distinct and identifiable
3. Segment Attractiveness
    - Based on profitability, ease of reach, risks, determine which segments should be pursued
4. Positioning Strategy
    - For each segment, create a "value proposition" (what can you offer them that is different from the competition?)
5. Implementation
    - Beware of the dangers when implementing

The Role of Analytics

Statistical tools are no replacement for managerial judgement.

But

These tools can provide insights about consumers unavailable from simple analyses.

Customer Lifetime Value

Course Roadmap

STP Analytics (Identify Value)	Customer Analytics (Deliver Value)	4P Analytics (Capture Value)
Module 1	**Module 2**	**Module 3**
What datasets can we use?	How much are our customers worth?	How do we build a new product?
How can we segment and target our customers?	Are our customers leaving?	How should we price our products? How do we distribute them?
How should we position our products/services?	How do our customers make choices?	How do we quantify the impact of our promotions?

What is CLV?

Some Inspirational Words…

- "Success is getting the right customers … and keeping them."

  Charles Cawley, Founder MBNA
- "The most important single thing is to focus obsessively on the customer. Our goal is to be earth's most customer-centric company."

  Jeff Bezos
- "There is only one boss—the customer. And he can fire everybody in the company from the chairman on down, simply by spending his money somewhere else."

  Sam Walton
- Customer is everywhere → customer is an asset

Marketing is about creating customers and keeping them

Modern Marketing Context: New Business Models

- Subscription Pricing
- Sharing Economy
- Direct to Consumer

moviepass™
Blue Apron
BIRCHBOX◆
Adobe
DOLLAR SHAVE CLUB
SHAVE TIME. SHAVE MONEY.
NETFLIX
amazon.com Prime
CLASSPASS

DOLLAR SHAVE CLUB
SHAVE TIME. SHAVE MONEY.
allbirds
soylent
Casper

airbnb
Uber
TURO

Tech Trends: Internet of Things (IoT)

- A network of physical devices that can transfer data to one another without human intervention

Tesla Connectivity

Google Nest

Amazon Echo

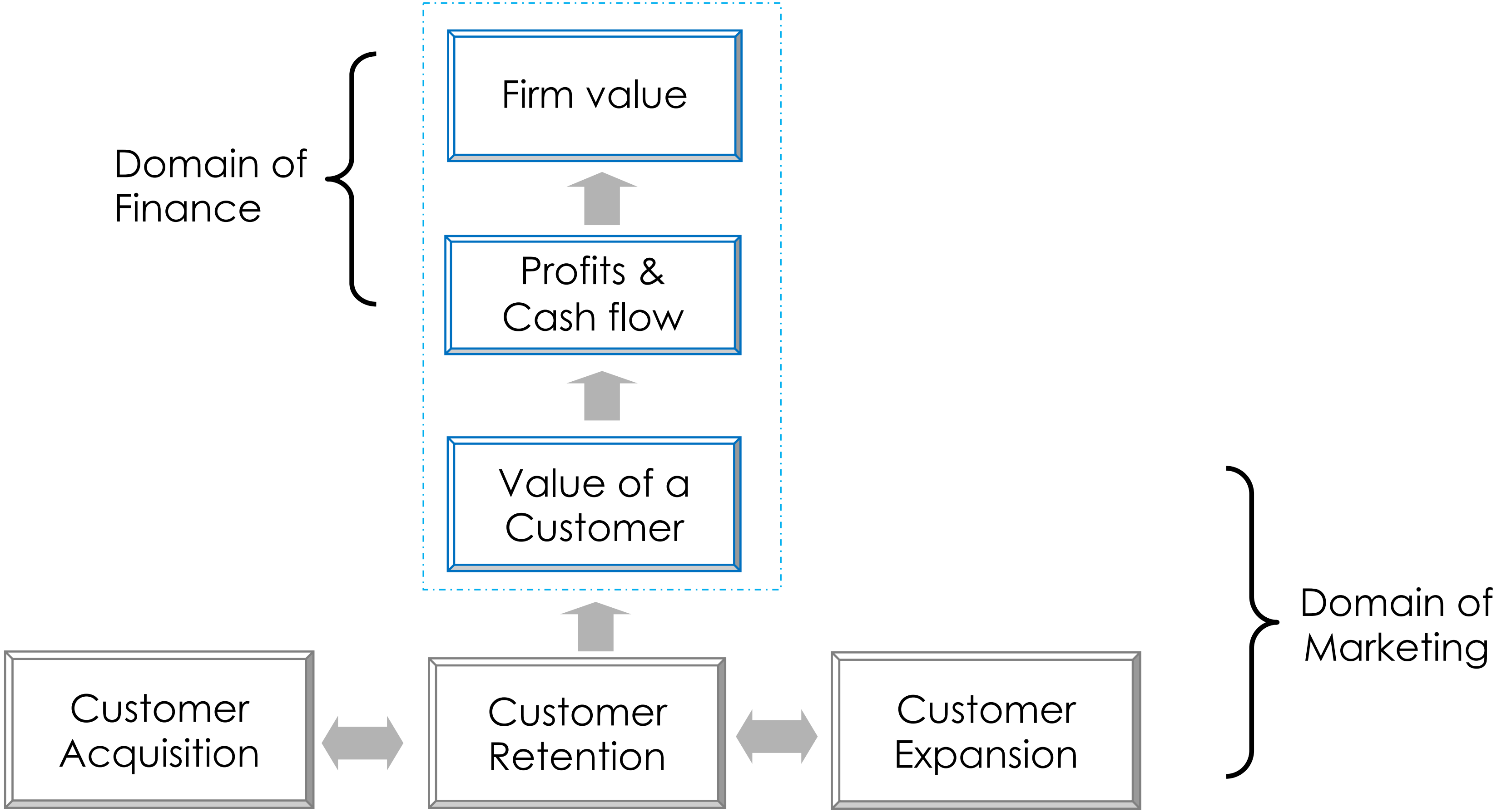
Modern Marketing Context: Digital Transformation

- Technological trends impacting marketing
  - Internet of Things
  - Retail analytics

- Data and analytics drive marketing activities
  - More connected with customers
  - Personalization

Technology has transformed all areas of marketing

Customer as an Asset and Firm Value



Domain of Finance

Firm value

Profits & Cash flow

Value of a Customer

Domain of Marketing

Customer Acquisition

Customer Retention

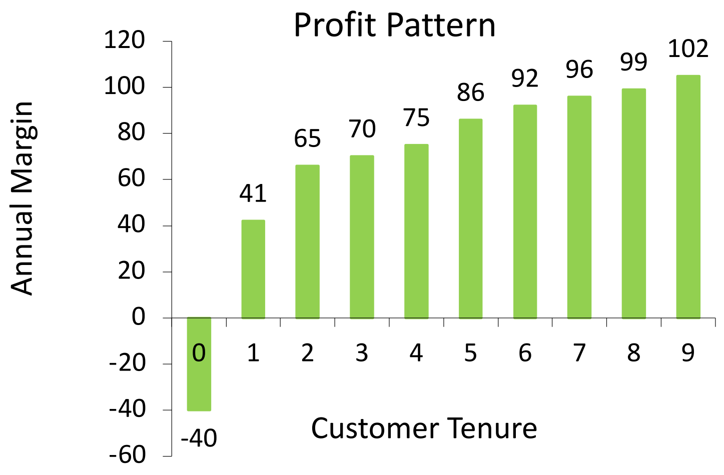
Customer Expansion

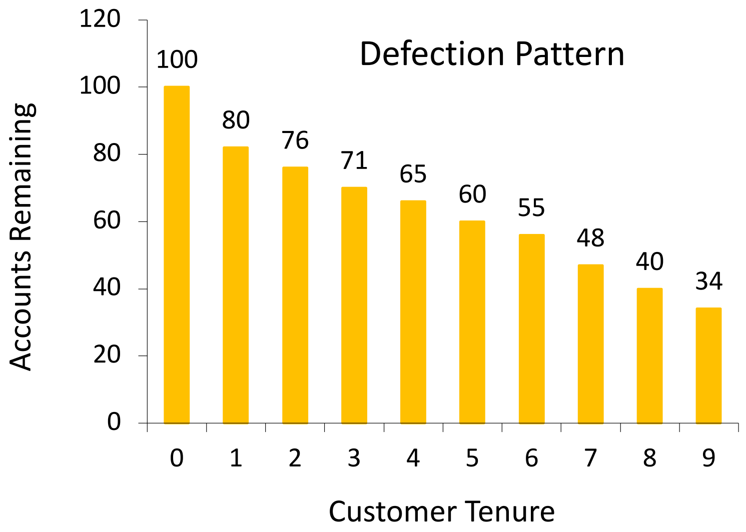
What is Customer Lifetime Value (CLV)?

Customer Lifetime Value is the net present value of all future streams of profits that a customer generates over the life of the relationship with the firm

- Net Present Value:
  - Sum of all future cash flow discounted back to today's value
  - How much money we make from a customer over many periods but discounted
  - $100 today is better than $100 in a year
- Profit: Revenue - Cost
- Relationship: Mutually beneficial interactions with customers

Measuring CLV - Intuition





Profit 1st year

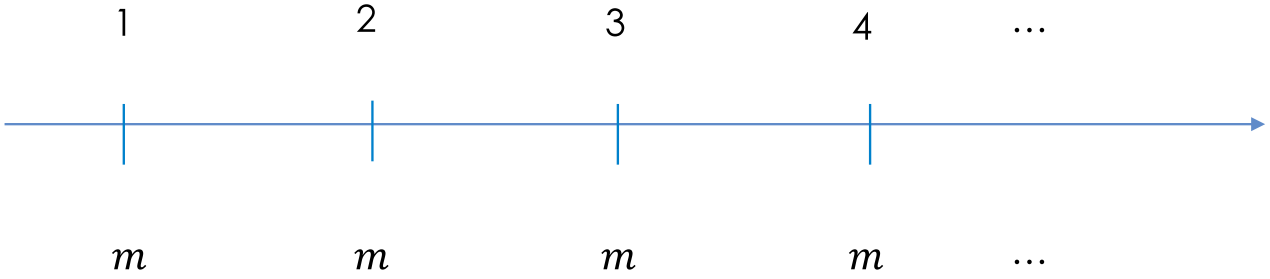
Acquisition Cost

Probability that customer is still here ~ Probability of not churning

$$CLV = -40 + \frac{(\$41)*(.80)}{(1+0.1)} + \frac{(\$65)*(.76)}{(1+0.1)^2} + \ldots$$

Value of Money – discounted at 10%

Measuring CLV – Modeling



Period	1	2	3	4	…
Margin	m	m	m	m	…

Measuring CLV – Modeling

Period	1	2	3	4	…
Margin	m	m	m	m	…
Prob (Alive)	r	r^2	r^3	r^4	…

Measuring CLV – Modeling – Retention rate

- Annual retention rate r
  - Can be computed for other periods (e.g., monthly)

- Example: at beginning of the year, 100 customers and at the end of the year, 80 remain:

$$r = \frac{80}{100} = 0.8$$

Measuring CLV – Modeling

Period	1	2	3	4	…
Margin	m	m	m	m	…
Prob (Alive)	r	r^2	r^3	r^4	…
Expected Margin	mr	mr^2	mr^3	mr^4	…
Discounted Expected Margin	$\frac{mr}{(1+i)}$	$\frac{mr^2}{(1+i)^2}$	$\frac{mr^3}{(1+i)^3}$	$\frac{mr^4}{(1+i)^4}$	…

Measuring CLV – Modeling

- Constant annual margin *m*
- Constant annual retention rate *r*
- Annual discount rate *i* and Acquisition cost *AC*

$$CLV = \frac{mr}{(1+i)} + \frac{mr^2}{(1+i)^2} + \frac{mr^3}{(1+i)^3} + \cdots - AC$$

Measuring CLV – Computation

$$CLV = \frac{mr}{(1+i)} + \frac{mr^2}{(1+i)^2} + \frac{mr^3}{(1+i)^3} + \cdots - AC$$

$$CLV = \frac{mr}{(1+i)}\left(1 + \frac{r}{(1+i)} + \frac{r^2}{(1+i)^2} + \cdots\right) - AC$$

Let $k = \frac{r}{1+i}$

$$CLV = mk(\underbrace{1 + k + k^2 + \cdots}) - AC$$

$$CLV = \frac{mk}{1-k} - AC$$

(Recall the formula for the sum of an infinite geometric series)

$$(1 + k + k^2 + \cdots) = \frac{1}{1-k}$$

Thus,

$$CLV = m\left(\frac{r}{1+i-r}\right) - AC$$

Measuring CLV – Modeling

- Constant annual margin *m*;
- Constant annual retention rate *r*
- Annual discount rate *i* and Acquisition cost *AC*

$$CLV = \frac{mr}{(1+i)} + \frac{mr^2}{(1+i)^2} + \frac{mr^3}{(1+i)^3} + \cdots - AC$$

$$= m\left(\frac{r}{1+i-r}\right) - AC$$

CLV and Margin Multiple

- Lifetime value of a customer

$$CLV = m\underbrace{\left(\frac{r}{1+i-r}\right)}_{\textbf{Margin multiple}} - AC$$

- m = margin
- i = discount rate
- r = retention rate
- AC = Acquisition Cost

Margin Multiple

$$\frac{r}{1+i-r}$$

Retention Rate	Discount Rate			
	10%	12%	14%	16%
60%	1.2	1.15	1.11	1.07
70%	1.75	1.67	1.59	1.52
80%	2.67	2.5	2.35	2.22
90%	4.5	4.09	3.75	3.46

Let's Put this in Practice

- 140 customers purchase 2,285 units per month
- Customers pay $12.50 per unit
- Variable cost per unit is $4.25
- Annual retention rate = 0.9, Annual discount rate = 12%

  i. What is the maximum the company should be willing to spend to acquire a new customer?

Question 1

- 140 customers purchase 2,285 units per month
- Customers pay $12.50 per unit
- Variable cost per unit is $4.25
- Annual retention rate = 0.9, Annual discount rate = 12%
  i. What is the **maximum** the company should be willing to spend to acquire a new customer?

- $CLV = m\left(\frac{r}{1+i-r}\right) - AC \geq 0$
- We need to find the AC such that the CLV equals 0 (breakeven).
  - $m = quantity * unit\ profit = \frac{2285}{140} * 12 \ * (12.5 - 4.25) = 1615.82$
    - Note the need to annualize
  - Margin multiple: $\left(\frac{r}{1+i-r}\right) = \left(\frac{0.9}{1+0.12-0.9}\right) = 4.09$
    - $i = 0.12, r = 0.9$
- $CLV \ = 1615.82(4.09) - AC = 6{,}610.18 \ - AC = 0$
- $AC = 6{,}610.18$

Let’s Put this in Practice

- 140 customers purchase 2,285 units per month
- Customers pay $12.50 per unit
- Variable cost per unit is $4.25
- Annual Retention rate = 0.9, Annual discount rate = 12%
    i. What is the maximum the company should be willing to spend to acquire a new customer?
    ii. What is the maximum the company should spend in total for this set of customers, once, now to increase retention rate to 0.95?

Question 2

- 140 customers purchase 2,285 units per month
- Customers pay $12.50 per unit
- Variable cost per unit is $4.25
- Annual Retention rate = 0.9, Annual discount rate = 12%
    i. What is the maximum the company should spend in total for this set of customers, **once**, now to increase retention rate to 0.95?

- Goal: increase CLV → $CLV_{new} - CLV_{now} \geq 0$
- $CLV_{now} = 6610.18 - AC$
- CLV if retention rate is 0.95:
    - Margin multiple: $\left(\frac{r}{1+i-r}\right) = \left(\frac{0.95}{1+0.12-0.95}\right) = 5.59$
    - Margin stays the same
    - $CLV_{new} = 1615(5.59) - AC - x = 9029.59 - AC - x$
- $9029.59 - AC - x - (6610.18 - AC) = 0$
- $x = 9029.59 - 6610.18 = 2{,}419$ per person
- Maximum willing to spend: $140(CLV_{new} - CLV_{now}) = 140x = 140(2{,}419) = 338{,}718$

Let's Put this in Practice

- 140 customers purchase 2,285 units per month
- Customers pay $12.50 per unit
- Variable cost per unit is $4.25
- Annual Retention rate = 0.9, Annual discount rate = 12%
  1. What is the maximum the company should be willing to spend to acquire a new customer?
  2. What is the maximum the company should spend in total for this set of customers, once, now to increase retention rate to 0.95?
  3. What is the maximum that the company should spend/customer, annually, to increase retention rate to 0.95?

Question 3

- 140 customers purchase 2,285 units per month
- Customers pay $12.50 per unit
- Variable cost per unit is $4.25
- Annual Retention rate = 0.9, Annual discount rate = 12%
    i. What is the maximum that the company should spend/customer, annually, to increase retention rate to 0.95?

- Goal: increase $CLV \rightarrow CLV_{new} - CLV_{now} \geq 0$
- In the worst case, we need the new CLV to be equal to the old CLV
- $CLV_{now} = 6610.18 - AC$
- CLV if retention rate is 0.95:
    - Margin multiple: $\left(\frac{r}{1+i-r}\right) = \left(\frac{0.95}{1+0.12-0.95}\right) = 5.59$
    - Annual cost per customer to increase from 4.09 to 5.59 → margin will go down
    - New margin $m' = 1615.82 - x$
    - $CLV_{new} = 5.59 * m' - AC = 5.59 * (1615.82 - x) - AC$
- $CLV_{new} = CLV_{now} \rightarrow 5.59 * (1615.82 - x) - AC = 6610.18 - AC$
- Maximum willing to pay:

$$x = 1615.82 - \frac{6610.18}{5.59} = 433$$

Q2 vs. Q3

Q2: What is the maximum the company should spend in total for this set of customers, once, now to increase retention rate to 0.95?

Paid one time at the beginning so part of acquisition cost

Q3: What is the maximum that the company should spend/customer, annually, to increase retention rate to 0.95?

Paid annually so part of margin

Managing CLV

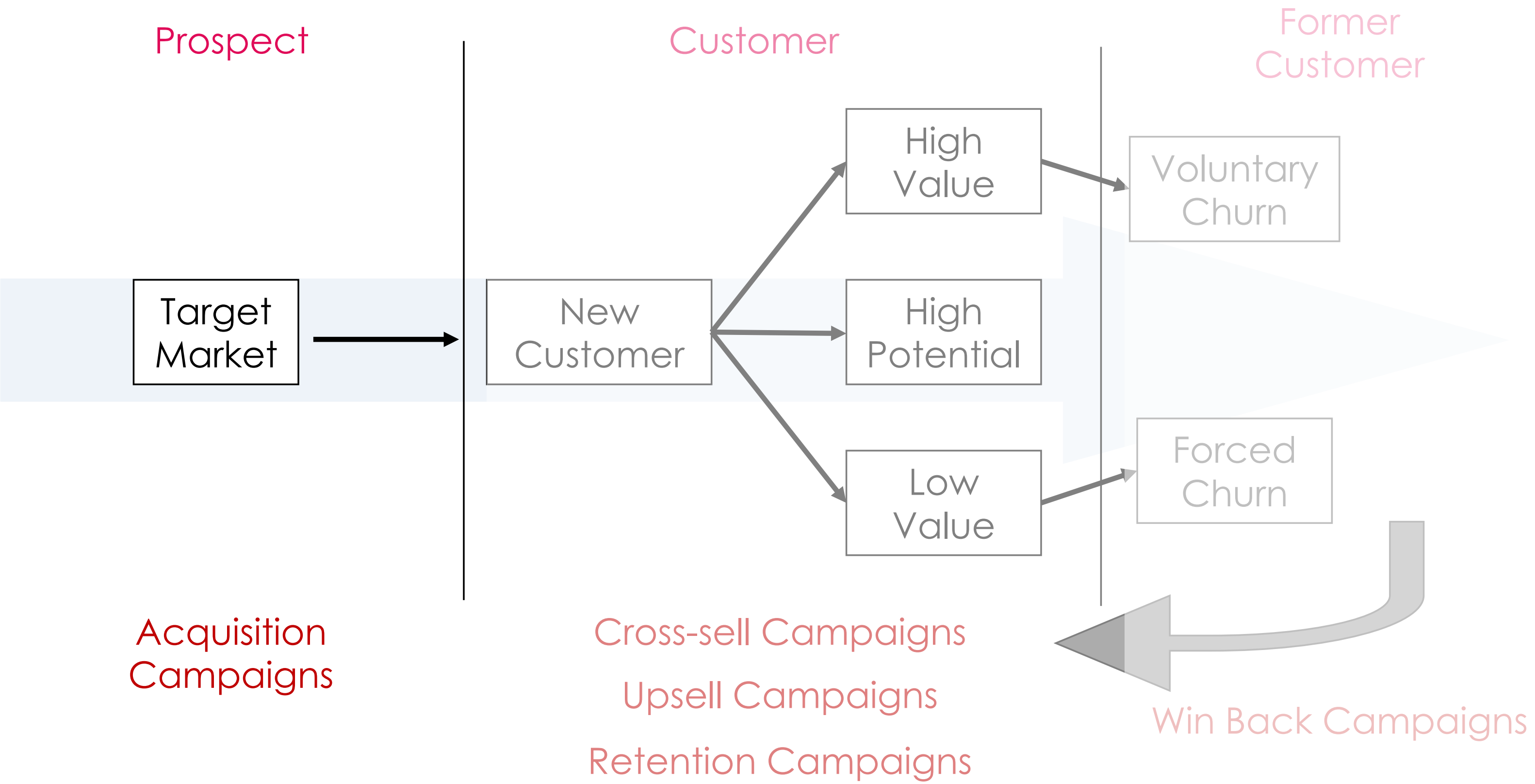
Increasing Customer Value: Strategies

$$CLV = m\left(\frac{r}{1+i-r}\right) - AC$$

- Three levers can be pulled:
  - Customer acquisition:
    - Reduce cost needed to get a customer → Decrease acquisition cost
  - Customer expansion
    - Increase profit from a customer → Increase margin
  - Customer retention:
    - Keep more of your customers → Increase retention rate

How can we achieve this (quantitatively and qualitatively)?

Customer Relationship Management



Prospect

Customer

Former Customer

Target Market

New Customer

High Value

High Potential

Low Value

Voluntary Churn

Forced Churn

Acquisition Campaigns

Cross-sell Campaigns

Upsell Campaigns

Retention Campaigns

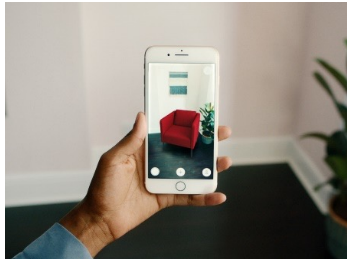
Win Back Campaigns

Managing CLV

Customer Acquisition

Customer Acquisition Strategies

- Advertising/Communications
  - E.g., free trial
- Affiliations
  - Amazon
  - EBay partner network
- Acquisitions
  - Meta,
    - Instagram (2012) $1B
    - WhatsApp (2014) $19B
    - Oculus VR (2014) $2B…
- New Technologies
  - Augmented reality



Ikea Place

Gucci, IOS app for shoes

Sephora Virtual Artist

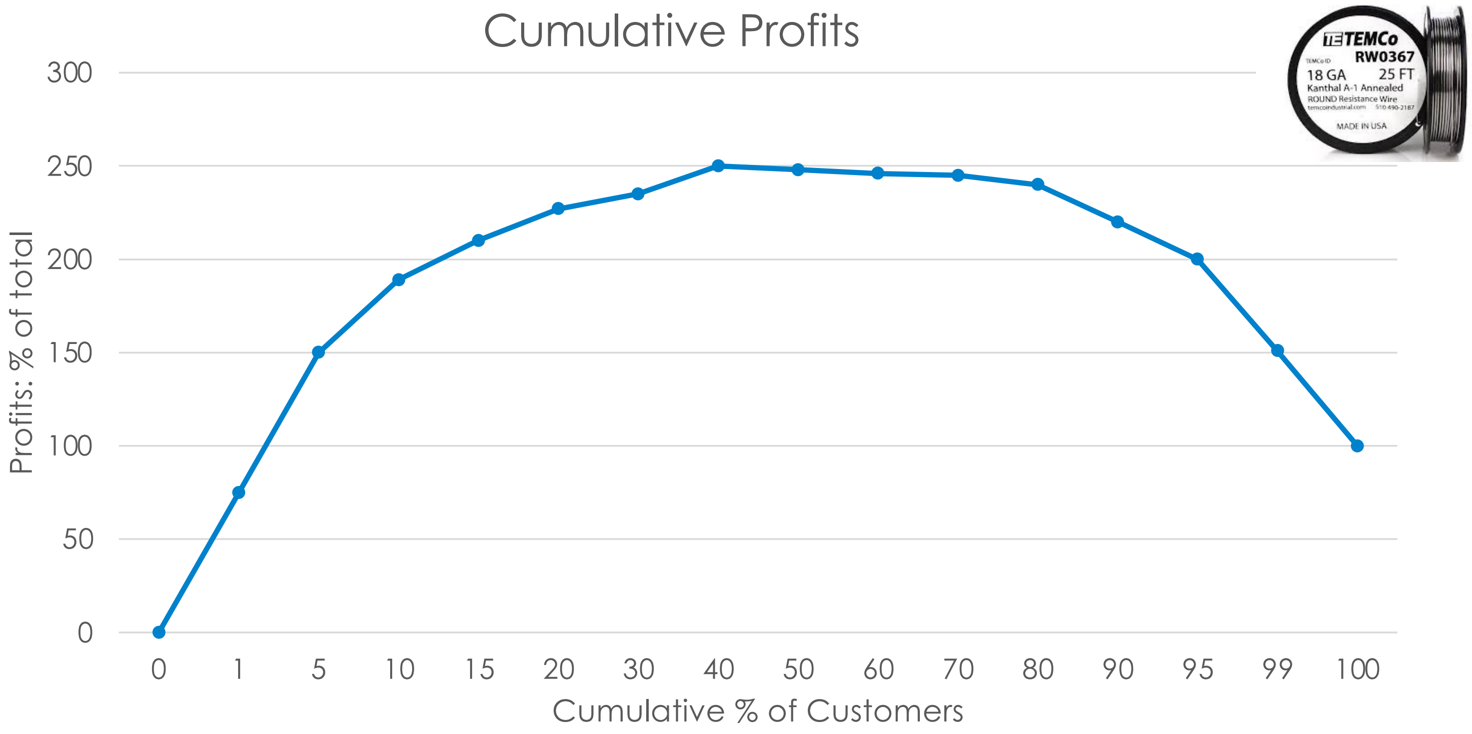
Customer Acquisition Costs

- Goal: decrease cost per acquisition
  - Note that the numbers are only illustrative

Activity	Cost/New Customer	Cost/Solicitation
Personal Selling	$500	$100
Direct Mail	$115	$1.50
Telemarketing	$95	$3.30
Website, e-mail	$30	$0.06

- Typically try to use the cheapest but:
  - Each channel has limited reach
  - Some products require a specific activity

All customers are important but…



Customers ordered by profitability

←Most profitable Least Profitable →

some customers are not worth it.

Next Class

- Recommender system – matrix factorization
- Customer retention

- Required reading: Matrix Factorization Techniques for Recommender Systems

B9651 – Marketing Analytics

Session 5: CRM + Churn

Professor Hortense Fong

Logistics

- CLV Concept Check due before next class
  - Graded for completion
- First Individual Assignment due Wed, Oct 16 at 8PM
- No class during CBS fall break (i.e., Oct 15 + 16)
- Midterm October 22 + 23
  - Material from Weeks 1-5, closed book, calculators allowed

Last Time

1. What is Customer Lifetime Value?
   - Definition: Customer Lifetime Value is the net present value of all future streams of profits that a customer generates over the life of the relationship with the firm

$$CLV = m\left(\frac{r}{1+i-r}\right) - AC$$

   - Problem: how to increase CLV?

2. Managing CLV
   - Customer Acquisition

Today Part 1: CRM

1. What is Customer Lifetime Value?
    1. Definition
    2. Problem: how to increase CLV?

2. Managing CLV
    1. Customer Acquisition
    2. Customer Expansion
        1. NMF + Implementation in Excel and Python
    3. Customer Retention

Course Roadmap

STP Analytics (Identify Value)	Customer Analytics (Deliver Value)	4P Analytics (Capture Value)
Module 1	**Module 2**	**Module 3**
What datasets can we use?	How much are our customers worth?	How do we build a new product?
How can we segment and target our customers?	Are our customers leaving?	How should we price our products? How do we distribute them?
How should we position our products/services?	How do our customers make choices?	How do we quantify the impact of our promotions?

Managing CLV

Customer Expansion

Customer Expansion

- How can we increase customer margin?
  - Increase usage
    - e.g., laundry detergent
  - Upsell
    - Switch customers to higher priced product or service
  - Bundling/cross-selling
    - Disney+/Hulu/ESPN
  - Reduce cost
    - Migration to online usage (banks, airlines)
  - Recommend products

Product Recommendations

- Netflix, Amazon, Whole Foods, and many others…

Recommended for You Based on Kindle Paperwhite, 6" High Resolution Display w...

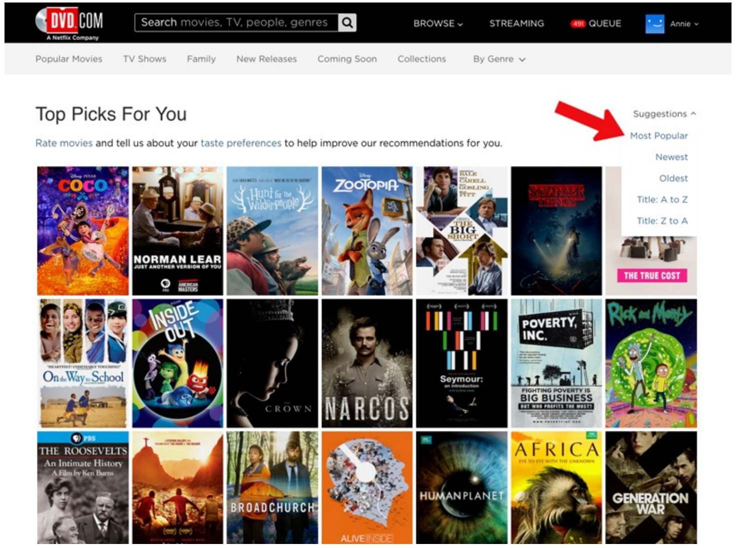
Page 1 of 5

MoKo Case for Kindle Paperwhite, Premium Thinnest and Lightest Leather Cover with… 898 $9.99 Prime

Swees Ultra Slim Leather Case Cover for Amazon All-New Kindle Paperwhite (Both 2012… 273 $3.99 Prime

Fintie SmartShell Case for Kindle Paperwhite - The Thinnest and Lightest Leather Cover for… 7,015 $14.99 Prime

Kindle Paperwhite, 6" High Resolution Display (212 ppi) with Built-in Light, Free 3G… 45,265 $159.99 Prime



Top Picks For You

Rate movies and tell us about your taste preferences to help improve our recommendations for you.

Recommendation System Algorithms

- Content-based filtering
  - Pre-define features of product

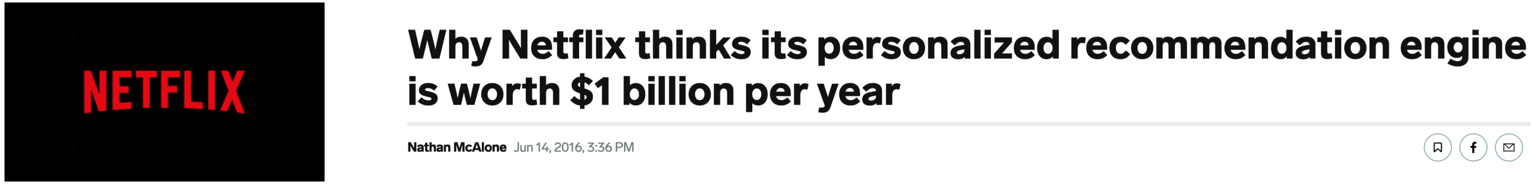
- Collaborative filtering (or nearest neighbors)
  - User based
  - Item based

- Matrix factorization methods
  - Use latent factors to represent users and items

The Netflix Prize

- Competition for best **collaborative filtering** algorithm to predict user ratings for films based on previous ratings
  - Goal: recommend content to users that they will like

- $1,000,000 prize



Why Netflix thinks its personalized recommendation engine is worth $1 billion per year

Nathan McAlone Jun 14, 2016, 3:36 PM

  - “the combined effect of personalization and recommendations save us more than $1B per year.”
  - keep subscribers from canceling

Source: https://www.businessinsider.com/netflix-recommendation-engine-worth-1-billion-per-year-2016-6

Customer Recommendation

User ratings of items

Users	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
User 1	5	.	5	2	.	3	4	5
User 2	1	2	.	.	4	3	1	5
User 3	3	1	5	.	4	.	.	.
User 4	1	3	4	.	1	3	1	4
User 5	1	.	2	.	4	3	2	2
User 6	.	.	1	.	1	2	4	1
User 7	5	.	.	2	2	3	4	5
User 8	4	3	3	5	.	.	.	3
User 9	4	3	5	.	.	1	5	.
User 10	1	2	.	5	.	.	3	4

User-Based Collaborative Filtering

Users	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
User 1	5	.	5	2	.	3	4	5
User 2	1	2	.	.	4	3	1	5
User 3	3	1	5	.	4	.	.	.
User 4	1	3	4	.	1	3	1	4
User 5	1	.	2	.	4	3	2	2
User 6	.	.	1	.	1	2	4	1
User 7	5	.	.	2	2	3	4	5
User 8	4	3	3	5	.	.	.	3
User 9	4	3	5	.	.	1	5	.
User 10	1	2	.	5	.	.	3	4

Customer Recommendation

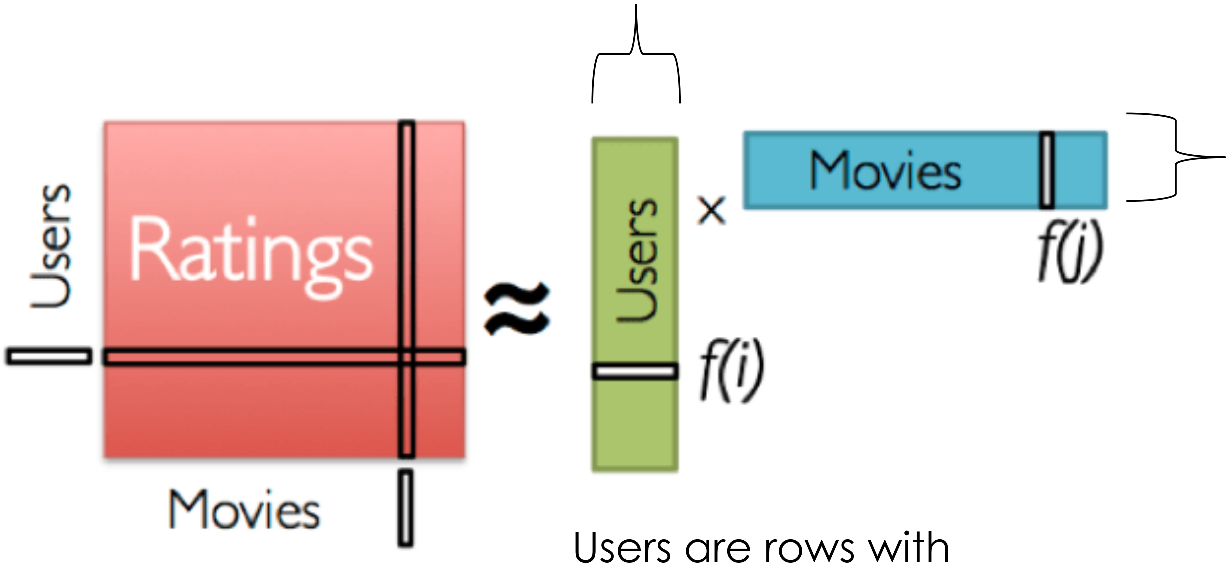
User ratings of items

Users	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
User 1	5	.	5	2	.	3	4	5
User 2	1	2	.	.	4	3	1	5
User 3	3	1	5	.	4	.	.	.
User 4	1	3	4	.	1	3	1	4
User 5	1	.	2	.	4	3	2	2
User 6	.	.	1	.	1	2	4	1
User 7	5	.	.	2	2	3	4	5
User 8	4	3	3	5	.	.	.	3
User 9	4	3	5	.	.	1	5	.
User 10	1	2	.	5	.	.	3	4

Item-Based Collaborative Filtering

Users	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
User 1	5	.	5	2	.	3	4	5
User 2	1	2	.	.	4	3	1	5
User 3	3	1	5	.	4	.	.	.
User 4	1	3	4	.	1	3	1	4
User 5	1	.	2	.	4	3	2	2
User 6	.	.	1	.	1	2	4	1
User 7	5	.	.	2	2	3	4	5
User 8	4	3	3	5	.	.	.	3
User 9	4	3	5	.	.	1	5	.
User 10	1	2	.	5	.	.	3	4

Matrix Factorization – Big Picture



User Latent Factors

Item Latent Factors

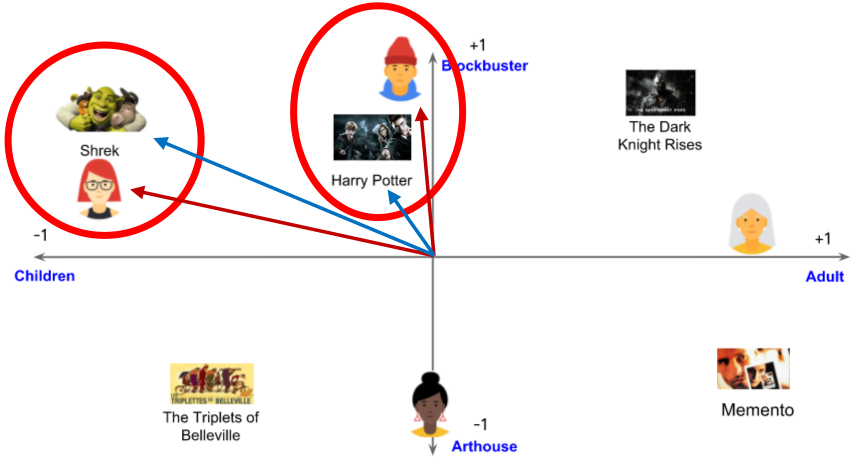
Movies are columns with values over the latent factors

Users are rows with values over the latent factors

Number of latent factors is a hyperparameter, it's something that we have to set

Matrix Factorization – Intuition

- The popularity of an item depends upon unobserved factors:
  - Children vs adults, arthouse vs blockbusters…
- Item latent vector q_i: extent to which the item i “contains” these factors
- User latent vector p_u: weights that user u assigns to each factor



Matrix Factorization – Model

- The rating is expressed in terms of the biases and latent vectors as

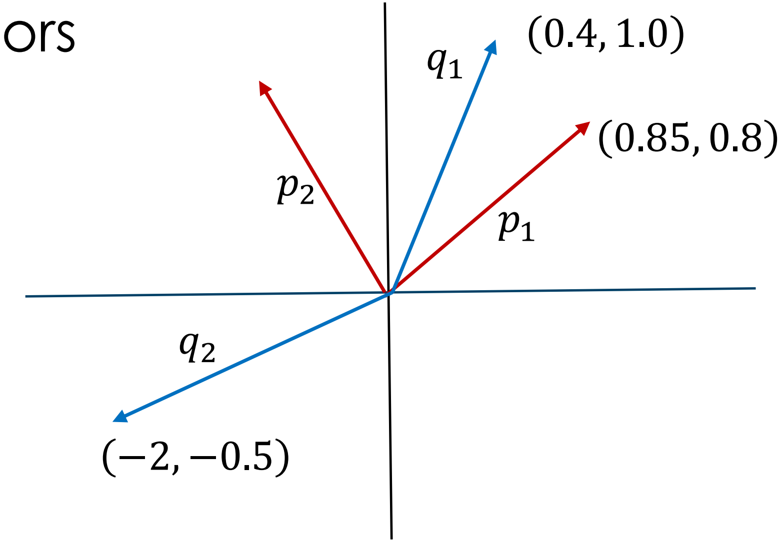
$$\underbrace{\hat{r}_{ui}}_{\text{Predicted rating for user u for item i}} = \underbrace{\mu}_{\text{Overall mean (intercept)}} + \boldsymbol{\alpha_u} + \boldsymbol{\beta_i} + \underbrace{p'_u q_i}_{\text{Captures match value}}$$

- Bias terms
  - User bias, $\boldsymbol{\alpha_u}$: some users tend to give higher scores than others
  - Item bias, $\boldsymbol{\beta_i}$: some items tend to be of higher quality than others
- Latent vectors
  - User latent factors – p_u
  - Item latent factors – q_i

Based on Koren, Bell and Volinsky, "Matrix Factorization Techniques for Recommender Systems, Computer, IEEE, August, 2009, pp 42-49

Matrix Factorization – Latent Space & Factors

- Latent factors are multidimensional vectors
  - User latent factors – p_u
  - Item latent factors – q_i
- Example in two dimensions
  - $p_1 = (0.85, 0.8)$
  - $q_1 = (0.4,1.0)$
  - $q_2 = (-2, -0.5)$



- Dot Product: user’s overall interest in item’s characteristics
  - $p_1' q_1 = 0.85 * 0.4 + 0.8 * 1.0 = 1.14$
  - $p_1' q_2 = 0.85 * -2 + 0.8 * -0.5 = -2.1$

Matrix Factorization – Objective Function

- The rating is expressed in terms of the biases and latent vectors as

$$\hat{r}_{ui} = \mu + \boldsymbol{\alpha_u} + \boldsymbol{\beta_i} + p'_u q_i$$

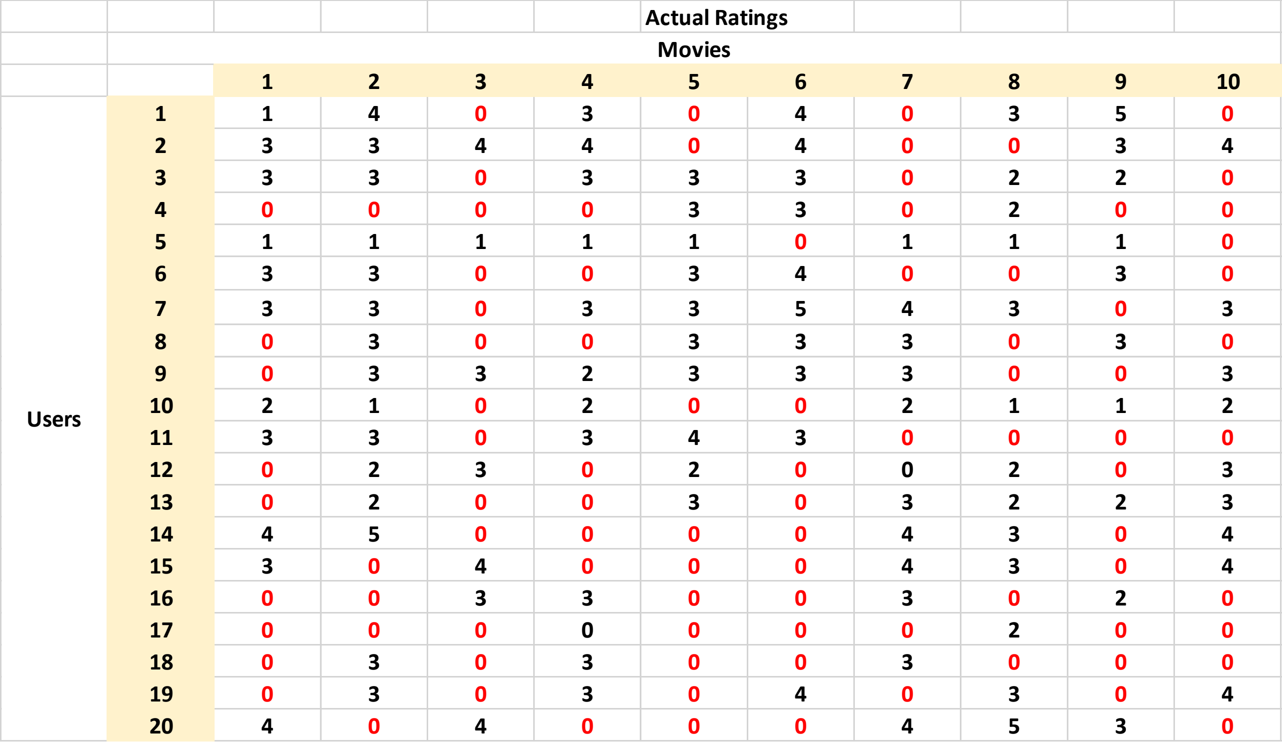
- We learn the model parameters $\theta = (\mu, \alpha_u, \beta_i, p_u, q_i)$ by minimizing the regularized squared error

$$\min_{\theta} \sum_{\forall u,i} \underbrace{\left(r_{u,i} - \mu - \alpha_u - \beta_i - p'_u q_i\right)^2}_{\text{Squared error}} + \underbrace{\lambda\left(\alpha_u^2 + \beta_i^2 + \|p_u\|^2 + \|q_i\|^2\right)}_{\text{Regularizer to prevent overfitting}}$$

- Only observed ratings in the data matrix are used in computing the squared error (first term in the objective function)

Movie Ratings Data: Example

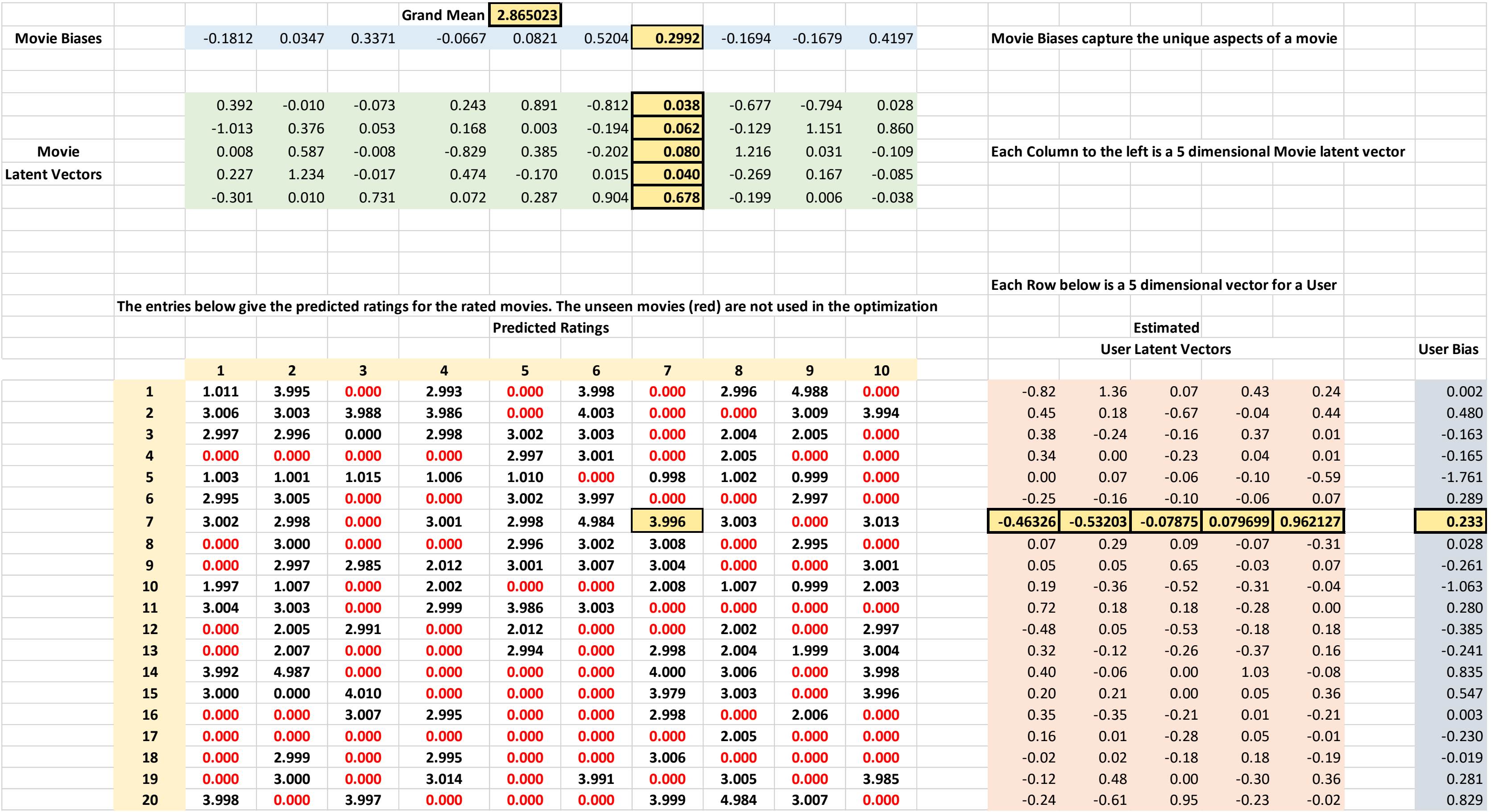
- Unseen movies are in red and are not used for obtaining parameters



						Actual Ratings					
		Movies									
		1	2	3	4	5	6	7	8	9	10
Users	1	1	4	0	3	0	4	0	3	5	0
	2	3	3	4	4	0	4	0	0	3	4
	3	3	3	0	3	3	3	0	2	2	0
	4	0	0	0	0	3	3	0	2	0	0
	5	1	1	1	1	1	0	1	1	1	0
	6	3	3	0	0	3	4	0	0	3	0
	7	3	3	0	3	3	5	4	3	0	3
	8	0	3	0	0	3	3	3	0	3	0
	9	0	3	3	2	3	3	3	0	0	3
	10	2	1	0	2	0	0	2	1	1	2
	11	3	3	0	3	4	3	0	0	0	0
	12	0	2	3	0	2	0	0	2	0	3
	13	0	2	0	0	3	0	3	2	2	3
	14	4	5	0	0	0	0	4	3	0	4
	15	3	0	4	0	0	0	4	3	0	4
	16	0	0	3	3	0	0	3	0	2	0
	17	0	0	0	0	0	0	0	2	0	0
	18	0	3	0	3	0	0	3	0	0	0
	19	0	3	0	3	0	4	0	3	0	4
	20	4	0	4	0	0	0	4	5	3	0

Matrix Factorization

$$\hat{r}_{ui} = \mu + \alpha_u + \beta_i + p'_u q_i$$



Grand Mean **2.865023**

Movie Biases	-0.1812	0.0347	0.3371	-0.0667	0.0821	0.5204	**0.2992**	-0.1694	-0.1679	0.4197
Movie Latent Vectors	0.392	-0.010	-0.073	0.243	0.891	-0.812	**0.038**	-0.677	-0.794	0.028
	-1.013	0.376	0.053	0.168	0.003	-0.194	**0.062**	-0.129	1.151	0.860
	0.008	0.587	-0.008	-0.829	0.385	-0.202	**0.080**	1.216	0.031	-0.109
	0.227	1.234	-0.017	0.474	-0.170	0.015	**0.040**	-0.269	0.167	-0.085
	-0.301	0.010	0.731	0.072	0.287	0.904	**0.678**	-0.199	0.006	-0.038

Movie Biases capture the unique aspects of a movie

Each Column to the left is a 5 dimensional Movie latent vector

Each Row below is a 5 dimensional vector for a User

The entries below give the predicted ratings for the rated movies. The unseen movies (red) are not used in the optimization

Predicted Ratings

	1	2	3	4	5	6	7	8	9	10	Estimated User Latent Vectors					User Bias
1	1.011	3.995	0.000	2.993	0.000	3.998	0.000	2.996	4.988	0.000	-0.82	1.36	0.07	0.43	0.24	0.002
2	3.006	3.003	3.988	3.986	0.000	4.003	0.000	0.000	3.009	3.994	0.45	0.18	-0.67	-0.04	0.44	0.480
3	2.997	2.996	0.000	2.998	3.002	3.003	0.000	2.004	2.005	0.000	0.38	-0.24	-0.16	0.37	0.01	-0.163
4	0.000	0.000	0.000	0.000	2.997	3.001	0.000	2.005	0.000	0.000	0.34	0.00	-0.23	0.04	0.01	-0.165
5	1.003	1.001	1.015	1.006	1.010	0.000	0.998	1.002	0.999	0.000	0.00	0.07	-0.06	-0.10	-0.59	-1.761
6	2.995	3.005	0.000	0.000	3.002	3.997	0.000	0.000	2.997	0.000	-0.25	-0.16	-0.10	-0.06	0.07	0.289
7	3.002	2.998	0.000	3.001	2.998	4.984	3.996	3.003	0.000	3.013	-0.46326	-0.53203	-0.07875	0.079699	0.962127	0.233
8	0.000	3.000	0.000	0.000	2.996	3.002	3.008	0.000	2.995	0.000	0.07	0.29	0.09	-0.07	-0.31	0.028
9	0.000	2.997	2.985	2.012	3.001	3.007	3.004	0.000	0.000	3.001	0.05	0.05	0.65	-0.03	0.07	-0.261
10	1.997	1.007	0.000	2.002	0.000	0.000	2.008	1.007	0.999	2.003	0.19	-0.36	-0.52	-0.31	-0.04	-1.063
11	3.004	3.003	0.000	2.999	3.986	3.003	0.000	0.000	0.000	0.000	0.72	0.18	0.18	-0.28	0.00	0.280
12	0.000	2.005	2.991	0.000	2.012	0.000	0.000	2.002	0.000	2.997	-0.48	0.05	-0.53	-0.18	0.18	-0.385
13	0.000	2.007	0.000	0.000	2.994	0.000	2.998	2.004	1.999	3.004	0.32	-0.12	-0.26	-0.37	0.16	-0.241
14	3.992	4.987	0.000	0.000	0.000	0.000	4.000	3.006	0.000	3.998	0.40	-0.06	0.00	1.03	-0.08	0.835
15	3.000	0.000	4.010	0.000	0.000	0.000	3.979	3.003	0.000	3.996	0.20	0.21	0.00	0.05	0.36	0.547
16	0.000	0.000	3.007	2.995	0.000	0.000	2.998	0.000	2.006	0.000	0.35	-0.35	-0.21	0.01	-0.21	0.003
17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.005	0.000	0.000	0.16	0.01	-0.28	0.05	-0.01	-0.230
18	0.000	2.999	0.000	2.995	0.000	0.000	3.006	0.000	0.000	0.000	-0.02	0.02	-0.18	0.18	-0.19	-0.019
19	0.000	3.000	0.000	3.014	0.000	3.991	0.000	3.005	0.000	3.985	-0.12	0.48	0.00	-0.30	0.36	0.281
20	3.998	0.000	3.997	0.000	0.000	0.000	3.999	4.984	3.007	0.000	-0.24	-0.61	0.95	-0.23	-0.02	0.829

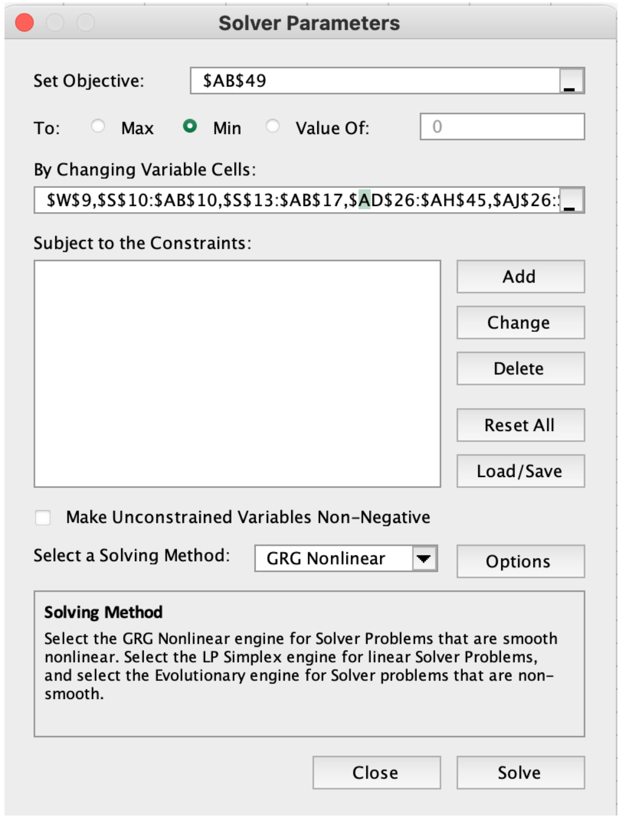
Let’s Go to Excel

Matrix Factorization

We choose 5 latent factors.

Matrix Factorization Predictions

- Excel Solver



Solver Parameters

Set Objective: $AB$49

To: ○ Max ● Min ○ Value Of: 0

By Changing Variable Cells:

$W$9,$S$10:$AB$10,$S$13:$AB$17,$AD$26:$AH$45,$AJ$26:

Subject to the Constraints:

☐ Make Unconstrained Variables Non-Negative

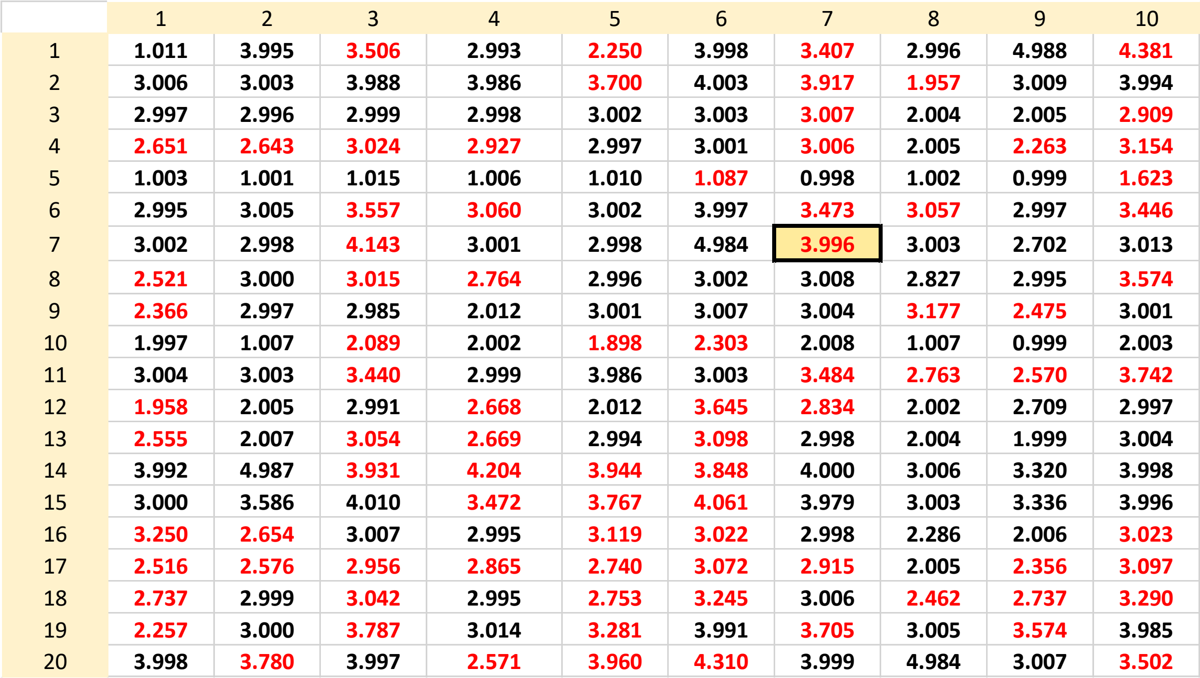
Select a Solving Method: GRG Nonlinear

Solving Method

Select the GRG Nonlinear engine for Solver Problems that are smooth nonlinear. Select the LP Simplex engine for linear Solver Problems, and select the Evolutionary engine for Solver problems that are non-smooth.

Matrix Factorization Predictions

- Given parameters, we can predict ratings for unseen movies (red)



	1	2	3	4	5	6	7	8	9	10
1	1.011	3.995	3.506	2.993	2.250	3.998	3.407	2.996	4.988	4.381
2	3.006	3.003	3.988	3.986	3.700	4.003	3.917	1.957	3.009	3.994
3	2.997	2.996	2.999	2.998	3.002	3.003	3.007	2.004	2.005	2.909
4	2.651	2.643	3.024	2.927	2.997	3.001	3.006	2.005	2.263	3.154
5	1.003	1.001	1.015	1.006	1.010	1.087	0.998	1.002	0.999	1.623
6	2.995	3.005	3.557	3.060	3.002	3.997	3.473	3.057	2.997	3.446
7	3.002	2.998	4.143	3.001	2.998	4.984	3.996	3.003	2.702	3.013
8	2.521	3.000	3.015	2.764	2.996	3.002	3.008	2.827	2.995	3.574
9	2.366	2.997	2.985	2.012	3.001	3.007	3.004	3.177	2.475	3.001
10	1.997	1.007	2.089	2.002	1.898	2.303	2.008	1.007	0.999	2.003
11	3.004	3.003	3.440	2.999	3.986	3.003	3.484	2.763	2.570	3.742
12	1.958	2.005	2.991	2.668	2.012	3.645	2.834	2.002	2.709	2.997
13	2.555	2.007	3.054	2.669	2.994	3.098	2.998	2.004	1.999	3.004
14	3.992	4.987	3.931	4.204	3.944	3.848	4.000	3.006	3.320	3.998
15	3.000	3.586	4.010	3.472	3.767	4.061	3.979	3.003	3.336	3.996
16	3.250	2.654	3.007	2.995	3.119	3.022	2.998	2.286	2.006	3.023
17	2.516	2.576	2.956	2.865	2.740	3.072	2.915	2.005	2.356	3.097
18	2.737	2.999	3.042	2.995	2.753	3.245	3.006	2.462	2.737	3.290
19	2.257	3.000	3.787	3.014	3.281	3.991	3.705	3.005	3.574	3.985
20	3.998	3.780	3.997	2.571	3.960	4.310	3.999	4.984	3.007	3.502

Let’s Go to Python

Matrix Factorization

Matrix Factorization – Pros & Cons

- Pros:
    - Requires only data on past user behavior (e.g., product ratings)
    - Domain free

- Cons:
    - Cold start problem

Managing CLV

Customer Retention

Customer Churn

- Churn: defection of customers

- Retention rates can be low in certain industries

Blue Apron	20%
huluPLUS	50%
CDNOW	60%
NETFLIX	75%
Spotify	81%
dish NETWORK	82%
verizon wireless	86%
amazon Prime	90%

Source: Various sources + [Telecommunications]()

Causes of Churn

- Company
  - Structural (poor ongoing customer experience)
  - Event based (a specific incident causes serious customer dissatisfaction)
- Competition
  - Promotion (switching)
  - Product/service (superior value proposition)
- Customer
  - Needs change, location changes…
  - http://www.youtube.com/watch?v=xmpDSBAh6RY&NR=1

Impact of Retention Rate

	Company 1	Company 2
Retention Rate	95%	90%
Churn Rate	5%	10%
Acquisition Rate	10%	10%

- How long will it take for each company to double its customer base?
  - Company 2
    - Not possible
  - Company 1 (starts with N customers)

Impact of Retention Rate

	Company 1	Company 2
Retention Rate	95%	90%
Churn Rate	5%	10%
Acquisition Rate	10%	10%

- How long does it take for each company to double its customer base?
  - Company 2
    - Not possible
  - Company 1 (starts with N customers)
    - What is the growth rate? How many customers after 2 years? x years?
      - $g = 5\%; N(1+g)^2; N(1+g)^x$
    - Suppose it takes x years to double:
      - $\mathrm{N}(1+\mathrm{g})^{\mathrm{x}} = 2N \rightarrow \mathrm{x} = \frac{\log(2)}{\log(1+g)} = 14.2\ yrs$

Impact of Retention Rate on Market Share

Case 1

$$\begin{array}{ccc} & \multicolumn{2}{c}{T = 2} \\ & \mathbf{A} & \mathbf{B} \\ T = 1 \quad \mathbf{A} & \mathbf{0.8} & 0.2 \\ \mathbf{B} & 0.2 & \mathbf{0.8} \end{array}$$

Case 2

$$\begin{array}{ccc} & \multicolumn{2}{c}{T = 2} \\ & \mathbf{A} & \mathbf{B} \\ \mathbf{A} & \mathbf{0.9} & 0.1 \\ \mathbf{B} & 0.2 & \mathbf{0.8} \end{array}$$

Case 3

$$\begin{array}{ccc} & \multicolumn{2}{c}{T = 2} \\ & \mathbf{A} & \mathbf{B} \\ \mathbf{A} & \mathbf{0.95} & 0.05 \\ \mathbf{B} & 0.2 & \mathbf{0.8} \end{array}$$

What is the long-run market share of A in the three cases?

Suppose A starts with 30% market share.

Impact of Retention Rate on Market Share

Case 1

$T = 2$

$$\begin{array}{cc|cc|} & & A & B \\ \hline & A & \mathbf{0.8} & 0.2 \\ T = 1 & & & \\ & B & 0.2 & \mathbf{0.8} \\ \hline \end{array}$$

Case 2

$T = 2$

$$\begin{array}{c|cc|} & A & B \\ \hline A & \mathbf{0.9} & 0.1 \\ B & 0.2 & \mathbf{0.8} \\ \hline \end{array}$$

Case 3

$T = 2$

$$\begin{array}{c|cc|} & A & B \\ \hline A & \mathbf{0.95} & 0.05 \\ B & 0.2 & \mathbf{0.8} \\ \hline \end{array}$$

What is the long-run market share of A in the three cases?

50%

67 %

80 %

Impact of Retention Rate on Market Share

- S_A = steady state market share of A
- S_B = steady state market share of B
- Transition matrix

$$\begin{array}{cccc} & & \multicolumn{2}{c}{T = 2} \\ & & \mathbf{A} & \mathbf{B} \\ T = 1 & \mathbf{A} & P_{AA} & P_{AB} \\ & \mathbf{B} & P_{BA} & P_{BB} \end{array}$$

- Since we're in a steady state:

$$S_A = P_{AA}S_A + P_{BA}S_B$$
$$S_B = P_{AB}S_A + P_{BB}S_B$$
$$S_A + S_B = 1$$

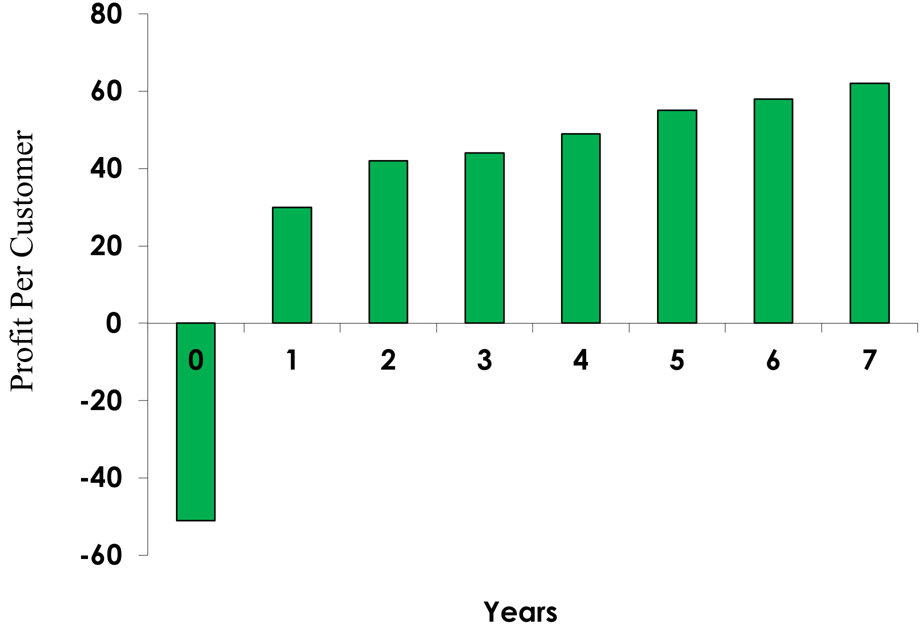
- We can solve for S_A and S_B in terms of the transition probabilities:

$$S_B = \frac{1 - P_{AA}}{1 - P_{AA} + P_{BA}}$$
$$(S_A = 1 - S_B)$$

Notice the starting market shares never show up!

Customer Lifecycle Profit Effect

- Annual profit per customer generally increases with customer tenure
  - Reduced price-sensitivity
  - Customer referrals
  - Reduced cost of serving customer
  - Increased purchases

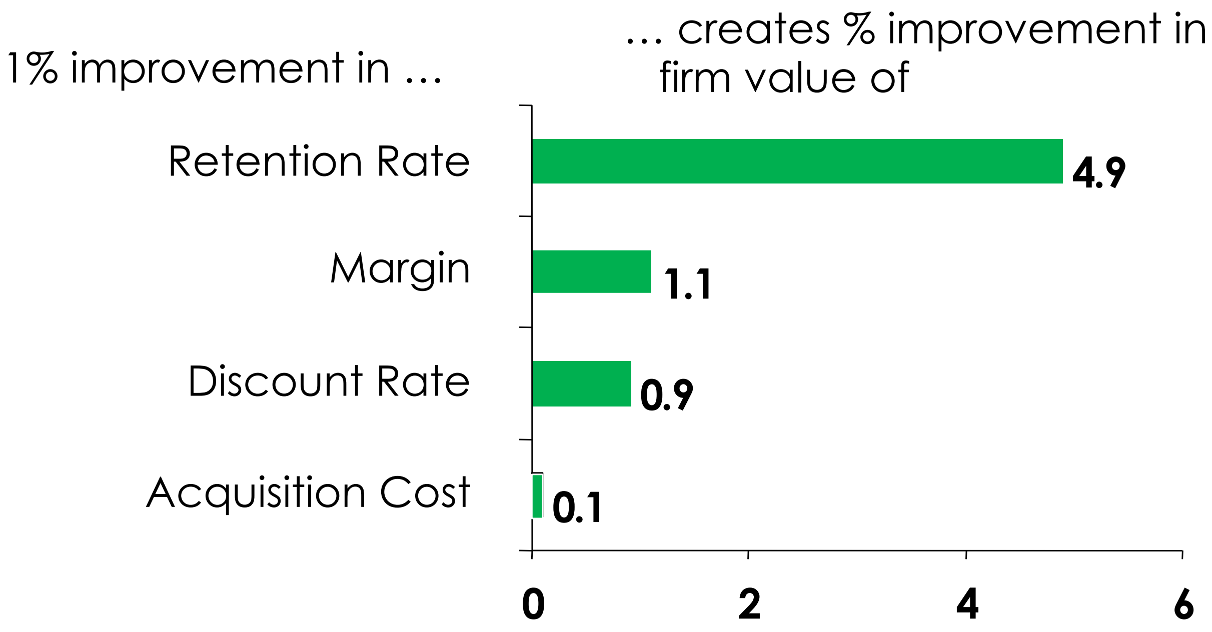


Source: Reichheld and Sasser (1990), “Zero Defections: Quality Comes to Service,” *HBR*, Sep-Oct.

Retention Rate & Firm Value

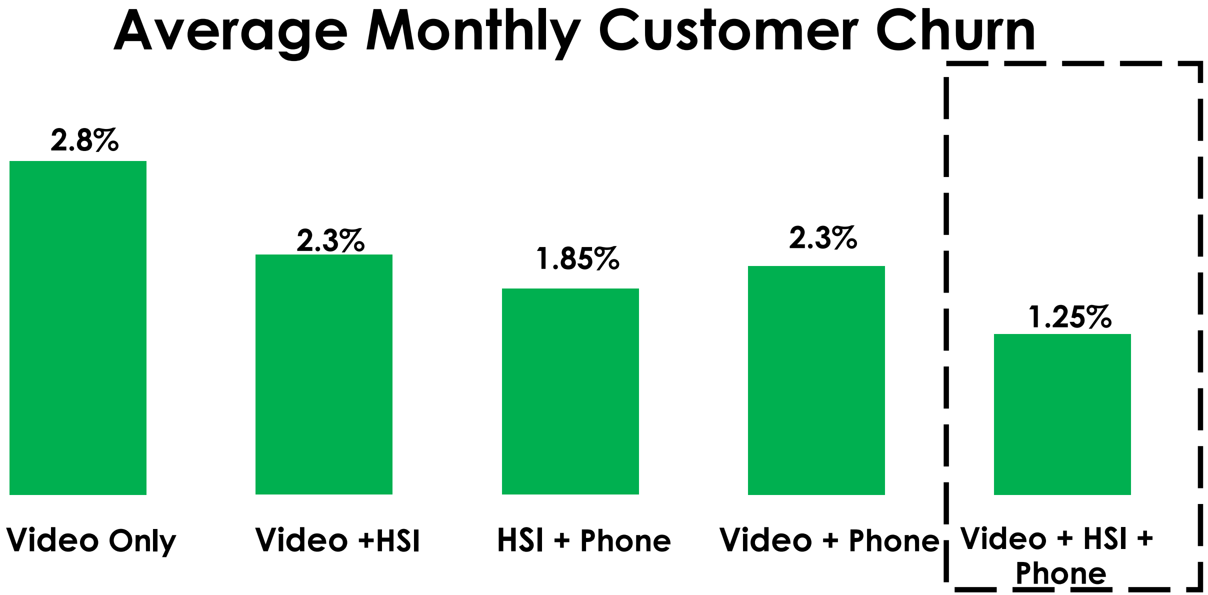
- Double benefit of higher retention rates
  - Inventory effect (# of customers)
  - Profit /customer effect (higher CLV)

- Largest impact on value



How to Increase Retention?

- Difficult question:
  - Depends on customers, competition, environment
- Some options
  - Quality customer service
  - Bundling
  - …



Takeaways

- Customers are assets and relationship management is important
- Lifetime value of a customer $CLV = m\left(\frac{r}{1+i-r}\right) - AC$

- CLV can be managed by optimizing
  1. Customer acquisition (AC)
  2. Customer expansion (m)
     - Recommendation systems + Implementation
  3. Customer retention (r)

Customer Churn

Course Roadmap

STP Analytics (Identify Value)	Customer Analytics (Deliver Value)	4P Analytics (Capture Value)
Module 1	**Module 2**	**Module 3**
What datasets can we use? How can we segment and target our customers? How should we position our products/services?	How much are our customers worth? Are our customers leaving? How do our customers make choices?	How do we build a new product? How should we price our products? How do we distribute them? How do we quantify the impact of our promotions?

Today Part 2: Modeling Churn

1. How can we forecast lifetime of customers?
2. How can we model customer survival using discrete-time customer base models?
   - Geometric Distribution
3. How can we incorporate customer heterogeneity?
   - Finite Mixture Models

Motivation

- Suppose you are Amazon
- You offer discounts to attract customers (e.g., student discount)
- You worry about customers churning
- How can you predict whether a customer will churn?

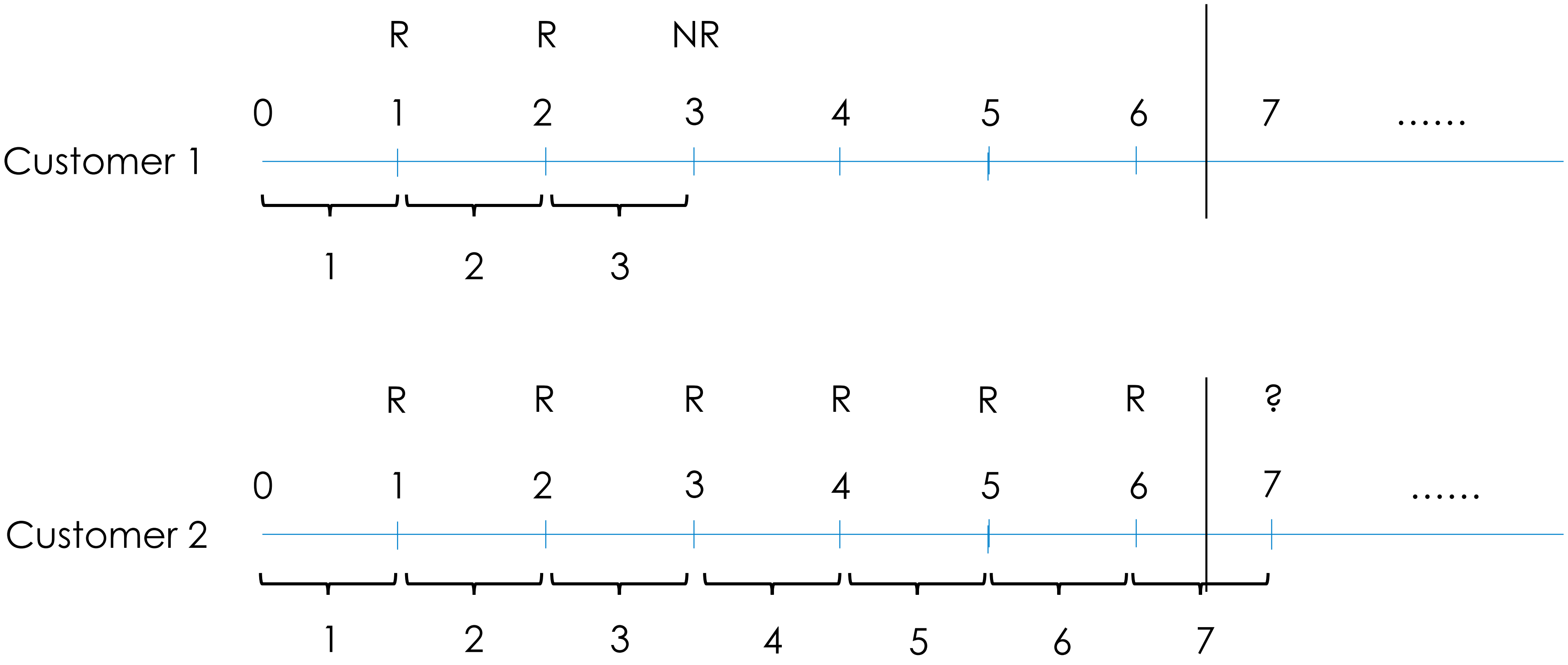
Customer Churn

- Customer retention is critical for improving the customer equity of the firm
- Customer churn (or survival) can be modeled probabilistically using survival models
- Survival models can be used for either continuous or discrete lifetime data
- Discrete survival model for forecasting customer retention

Subscription Context

- Consider a cohort of customers who joined at the beginning of Period 1
- Customers need to renew their subscription after the end of every period (e.g., pay the annual fee)
- Customers are observed till the end of Period T
  - The data is censored in that we do not observe what happens after Period T

Customer Journeys



R = Renewal, NR = No Renewal

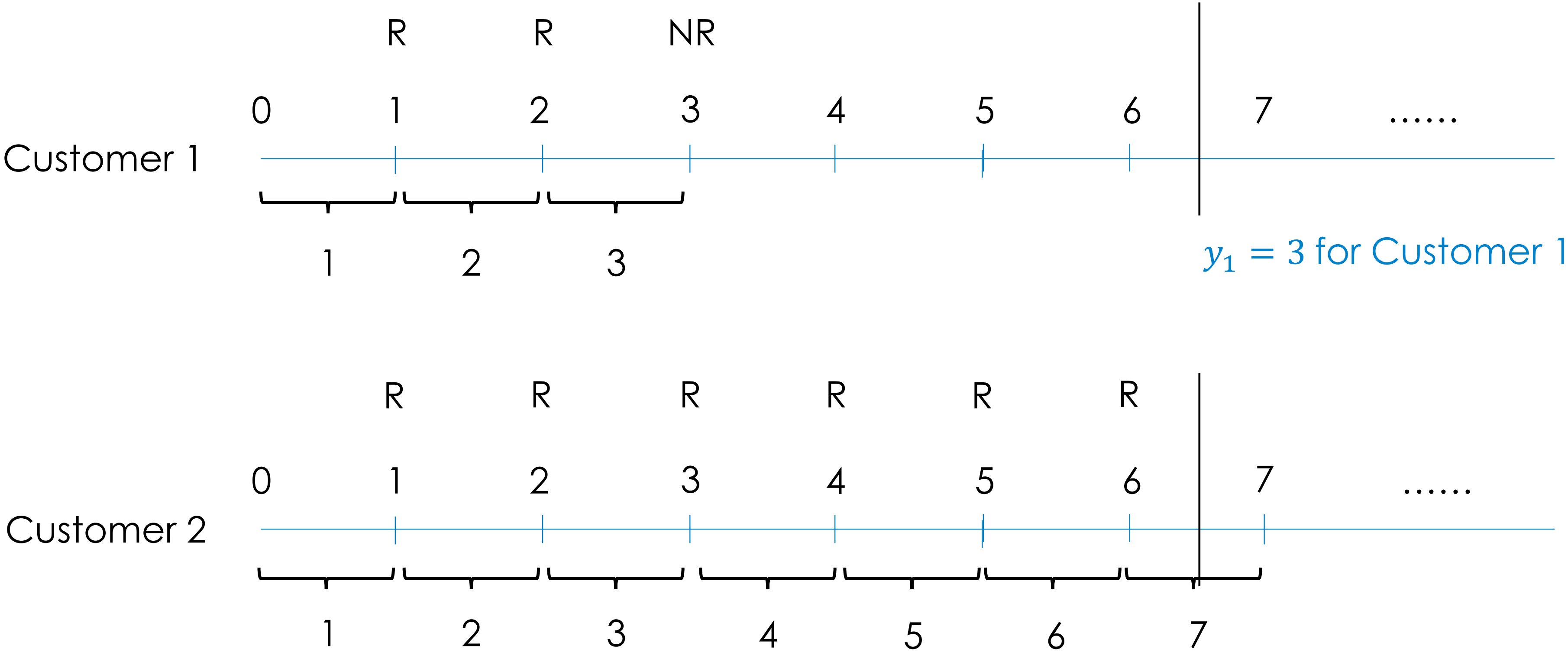
Customer Journey

- Let $X \geq 0$ be the number of renewals a customer makes before canceling
- Let $Y \geq 1$ be the number of periods the customer is with us after which he cancels

- We will focus on modeling Y
  - $Y = X + 1$
  - Lower case values represent actual values of Y
  - y_i: value for customer i

Customer Journeys

Let $Y \geq 1$ be the number of periods the customer is with us after which he cancels



$y_1 = 3$ for Customer 1

$y_2 > 6$ for Customer 2

Customer Journey

- Let $X \geq 0$ be the number of renewals a customer makes before canceling
- Let $Y \geq 1$ be the number of periods the customer is with us after which he cancels

- We will focus on modeling Y
  - $Y = X + 1$
  - Lower case values represent actual values of Y
  - y_i: value for customer i

- We can model the customer journeys probabilistically
- Assumption: lifetime comes from a Geometric distribution

Geometric Distribution

- θ: the probability that a customer does not renew the subscription at a renewal occasion

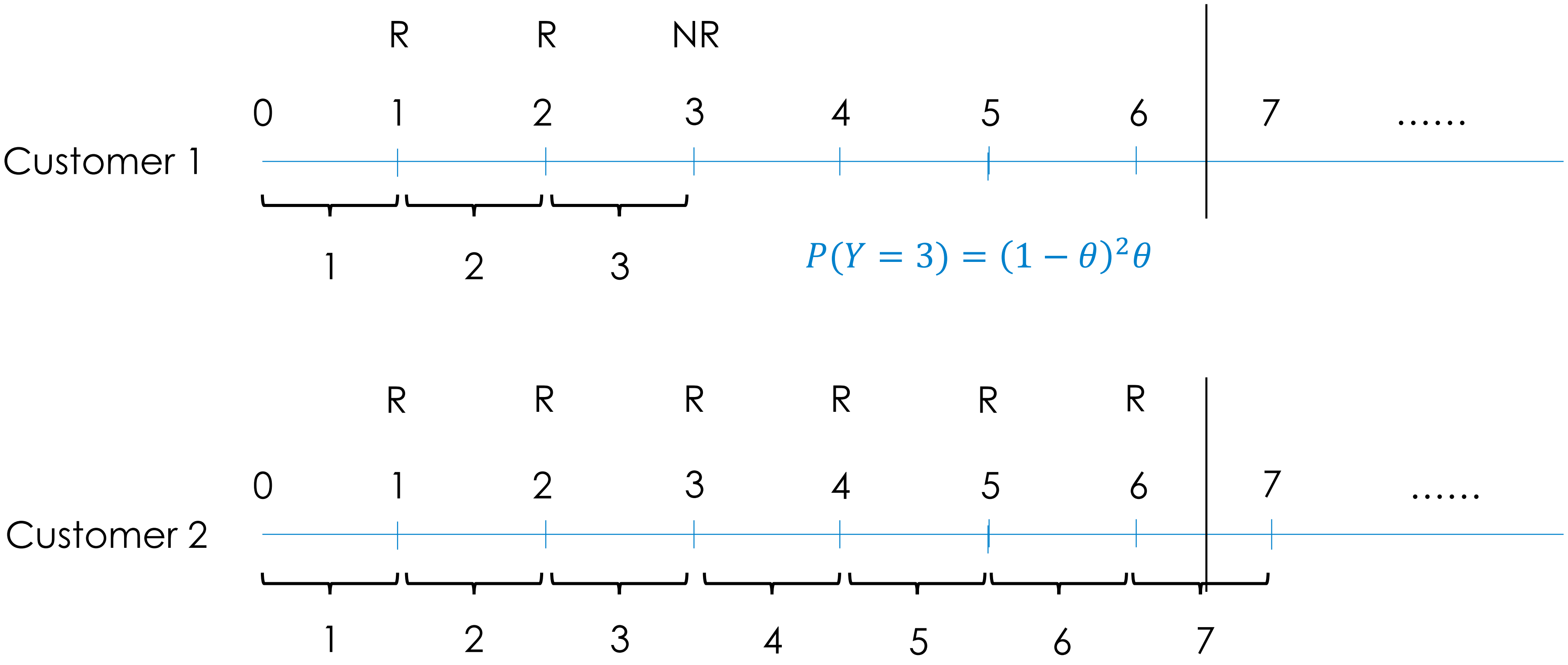
- θ is assumed to be the same for each customer and is time invariant

- Y (lifetime of the customer) is distributed geometric with probability θ
  - Number of “Bernoulli trials” until customer defects

$$\text{Prob}(Y=k)=(1-\theta)^{k-1}\theta$$

Goal: Find the value of θ that fits observed data the best

Customer Journeys – What are the Probabilities?



R R NR

0 1 2 3 4 5 6 7 ……

Customer 1

1 2 3

$P(Y = 3) = (1 - \theta)^2\theta$

R R R R R R

0 1 2 3 4 5 6 7 ……

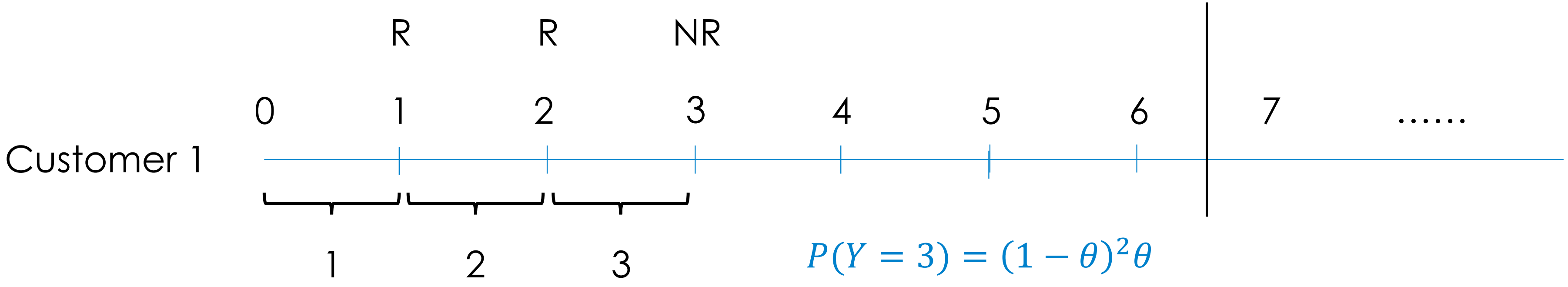
Customer 2

1 2 3 4 5 6 7

Geometric Distribution

- The p.m.f of Y is given by $f(k) = \text{Prob}(Y = k) = (1 - \theta)^{k-1}\theta$
- The c.d.f of Y is given by $F(k) = \text{Prob}(Y \leq k) = 1 - \text{Prob}(Y > k) = 1 - (1 - \theta)^k$
  - $\text{Prob}(Y > k)$ is the same as the event that the first k trials are not successes (i.e., $(1 - \theta)^k$)
- The survivor function is given by $S(k) = \text{Prob}(Y > k) = (1 - \theta)^k$

Customer Journeys – What are the Probabilities?



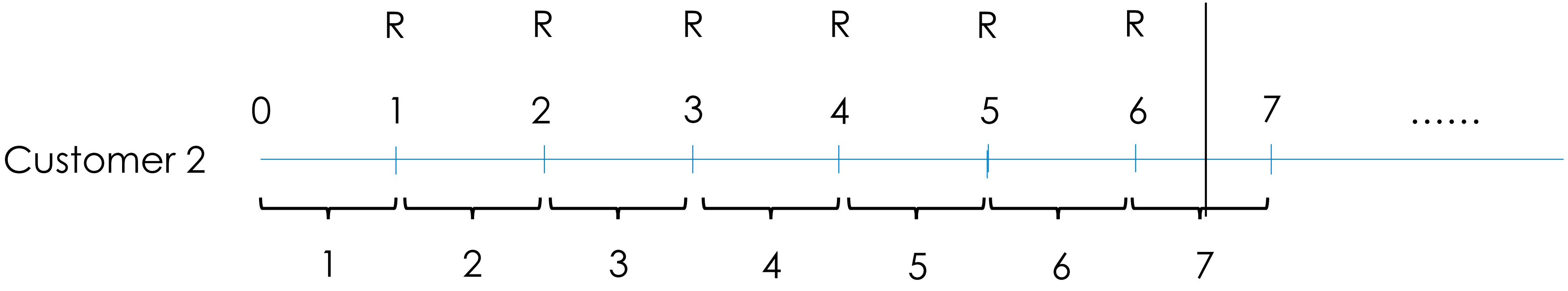
Customer 1

R R NR

0 1 2 3 4 5 6 7 ……

1 2 3

$P(Y = 3) = (1 - \theta)^2\theta$



Customer 2

R R R R R R

0 1 2 3 4 5 6 7 ……

1 2 3 4 5 6 7

$P(Y > 6) = (1 - \theta)^6$

Geometric Distribution

- The p.m.f of Y is given by $f(k) = \text{Prob}(Y = k) = (1-\theta)^{k-1}\theta$

- The c.d.f of Y is given by $F(k) = \text{Prob}(Y \leq k) = 1 - \text{Prob}(Y > k) = 1 - (1-\theta)^k$

- The survivor function is given by $S(k) = \text{Prob}(Y > k) = (1-\theta)^k$

- Why is it important?
  - We can use definitions to fit a geometric distribution to customer retention data
  - How? Maximum Likelihood

Maximum Likelihood Estimation (MLE)

- Maximum Likelihood Estimation
  - Use the data to find values of the model parameters (θ) that maximize the likelihood of observing the data that we have
- We estimate the model parameters by maximizing the likelihood function $L(\theta)$
- The resulting parameter estimates, θ_{ML} are called “maximum likelihood estimates"

Maximum Likelihood Estimation (MLE) Simple Example

- Imagine a coin lands Heads with prob p and Tails with prob $1-p$
- We toss the coin 10 i.i.d. times and obtain: TTTTTTTTTH
  - Is $p=0$ likely? $\Pr($TTTTTTTTTH | p=0)=0
  - Is $p=1$ likely? $\Pr($TTTTTTTTTH | p=1)=0
  - Is $p=0.5$ likely?
    - $\Pr($TTTTTTTTTH | p=0.5) $=\Pr(T|p=0.5)*\cdots*\Pr(H|p=0.5)=0.5^{10}=0.00097$
  - Is $p=0.1$ likely?
    - $\Pr($TTTTTTTTTH | p=0.1)$=\Pr(T|p=0.1)*\cdots*\Pr(H|p=0.1)=(0.9)^9(0.1)=0.039$

- We want to find the value of p that makes observing our data most **likely**

Maximum Likelihood Estimation (MLE)
Simple Example

- Likelihood corresponds to the probability of observing our data as a function of the parameters of a statistical model

$$\mathcal{L}(p|TTTTTTTTTH) = (1-p)^9 * p$$

- We need to find the value of p that maximizes $\mathcal{L}(p|TTTTTTTTTH)$. How?
  - Differentiate. But it is simpler and equivalent to maximizing log-likelihood.

$$\mathcal{LL}(p|TTTTTTTTTH) = 9\log(1-p) + \log(p)$$

$$\frac{d\mathcal{LL}(p|TTTTTTTTTH)}{dp} = -\frac{9}{1-p} + \frac{1}{p}$$

  - Maximized when $p = 0.1$

Likelihood

- Let y_i be the observed lifetime of customer i from period 1 to T
  - Lifetime can be complete (customer left) or censored (customer still here after T)
  - Our data doesn't go to the end of time
- Suppose T = 6

Customer (i)	Lifetime (y_i)
1	3
2	6
3	3
4	6
…	…

Number of periods customer is with us after which he/she cancels

Likelihood

- Let y_i be the observed lifetime of customer i from period 1 to T
  - Lifetime can be complete (customer left) or censored (customer still here after T)
- Let δ_i be the binary indicator that takes the value 1 if i is still alive after T

Customer (i)	**Lifetime (y_i)**	**Censored (δ_i)**
1	3	0
2	6	1
3	3	0
4	6	0
…	…	…

→ Censored, this person has not left at time 6

→ Not censored, this person left at time 6

Likelihood

- Let y_i be the observed lifetime of customer i from period 1 to T
    - Lifetime can be complete (customer left) or censored (customer still here after T)
- Let δ_i be the binary indicator that takes the value 1 if i is still alive after T
- What is the probability of observing the same journey as Customer 1?

Customer (i)	**Lifetime (y_i)**	**Censored (δ_i)**	**Likelihood ($L(\theta \mid y_i)$)**
1	3	0	?
2	6	1	?
3	3	0	?
4	6	0	?
…	…	…	…

Likelihood

- Let y_i be the observed lifetime of customer i from period 1 to T
    - Lifetime can be complete (customer left) or censored (customer still here after T)
- Let δ_i be the binary indicator that takes the value 1 if i is still alive after T
- What is the probability of observing the same journey as Customer 1?

Customer (i)	Lifetime (y_i)	Censored (δ_i)	Likelihood ($L(\theta \mid y_i)$)
1	3	0	$(1-\theta)^2\,\theta$
2	6	1	$(1-\theta)^6$
3	3	0	$(1-\theta)^2\,\theta$
4	6	0	$(1-\theta)^5\,\theta$
…	…	…	…

Likelihood

- Let y_i be the observed lifetime of customer i from period 1 to T
  - Complete or censored lifetime
- Let δ_i be the binary indicator that takes the value 1 if i is still alive after T
- What is the likelihood for observation i?

$$\mathcal{L}_i(\theta) = f(y_i|\theta)^{1-\delta_i} S(y_i|\theta)^{\delta_i} = ((1-\theta)^{y_i-1}\theta)^{1-\delta_i}((1-\theta)^{y_i})^{\delta_i}$$

  - If customer left ($\delta_i = 0$), likelihood is $\mathcal{L}_i(\theta) = f(y_i|\theta)$, the probability the customer left after y_i periods
  - If customer alive ($\delta_i = 1$), likelihood is $\mathcal{L}_i(\theta) = S(y_i|\theta)$, the probability the customer has not left after y_i periods
- The likelihood for the entire data is the product of the individual-level likelihoods

$$\mathcal{L}(\theta|\mathbf{D}) = \prod_{i=1}^{I} \mathcal{L}_i(\theta|y_i, \delta_i)$$

Log-Likelihood

- In practice, we maximize the log-likelihood

- The log-likelihood for customer i is given by

$$\mathcal{LL}_i(\theta|y_i,\delta_i) = (1-\delta_i)\log\big(f(y_i|\theta)\big) + \delta_i\log\big(S(y_i|\theta)\big)$$

$$= (1-\delta_i)\log((1-\theta)^{y_i-1}\theta) + \delta_i\log\big((1-\theta)^{y_i}\big)$$

- Overall log-likelihood
  - Sum of the individual-specific log-likelihoods

$$\mathcal{LL}(\theta|\mathcal{D}) = \sum_{i=1}^{I}\mathcal{LL}_i(\theta|y_i,\delta_i)$$

Let’s go to Excel

BetaGeometricDetailed_inclass.xlsx – Individual-Level

Estimate θ using MLE

10 minutes; feel free to work in groups

$$\mathcal{LL}_i(\theta|y_i,\delta_i) = (1-\delta_i)\log((1-\theta)^{y_i-1}\theta) + \delta_i \log\big((1-\theta)^{y_i}\big),$$
$$\mathcal{LL}(\theta|\mathcal{D}) = \textstyle\sum_{i=1}^{I} \mathcal{LL}_i(\theta|y_i,\delta_i)$$

Finite Mixture Model for Segmentation

- So far, all customer are the same
  - Why? Same θ (probability of churn)
  - How about heterogeneity?

- Expect different θs for different groups of customers

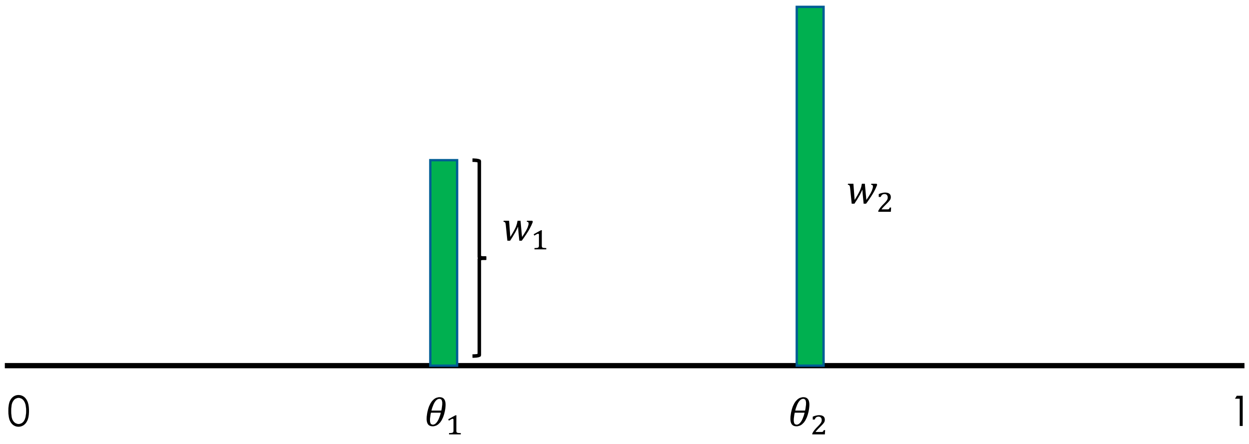
Finite Mixture Model for Segmentation

- So far, all customer are the same
  - Why? Same θ (probability of churn)
  - How about heterogeneity?

- We will assume that customers belong to a finite number of segments M
  - Customers differ in their θ values
    - θ_m: probability for all members in segment m
  - Segment sizes differ
    - w_m: size of segment m
  - $z_i \in \{1,2,\ldots,M\}$ indicate the latent segment for customer i
    - We don't observe which segment each customer is in
  - $\Pr(z_i = m) = w_m$, where $w_m \in [0,1]$, $\forall m$, and $\sum_{m=1}^{M} w_m = 1$

Finite Mixture Models: Point Mass Representation

Two segments



$$w_1 + w_2 = 1$$

Conditional Likelihood

- Within each segment, lifetimes distributed geometrically
- Conditional likelihood
  - Likelihood of the customer conditional on belonging to a particular segment
- For completed observation y_i, likelihood conditional on belonging to segment m is

$$\mathcal{L}_{im}(\theta_m|y_i) = f(y_i;\theta_m)$$

- For incomplete (censored) observation ($\delta_i = 1$)

$$\mathcal{L}_{im}(\theta_m|y_i) = S(y_i;\theta_m)$$

Finite Mixture Likelihood

- Segment membership is not observed
    - It must be integrated to the likelihood
- Size of segments also not observed
- Unconditional likelihood for an individual
    - Weighted average of the conditional likelihoods of that individual

- For two segments

Note: Likelihood vs log likelihood
Weighted average of logs different
from log of weighted average

$$\mathcal{L}_i(\theta_1, \theta_2, w_1, w_2 | y_i) = w_1 \mathcal{L}_{i1}(\theta_1 | y_i) + w_2 \mathcal{L}_{i2}(\theta_2 | y_i)$$

$$= w_1((1-\theta_1)^{y_i-1}\theta_1)^{1-\delta_i}((1-\theta_1)^{y_i})^{\delta_i} + w_2((1-\theta_2)^{y_i-1}\theta_2)^{1-\delta_i}((1-\theta_2)^{y_i})^{\delta_i}$$

Let’s go to Excel

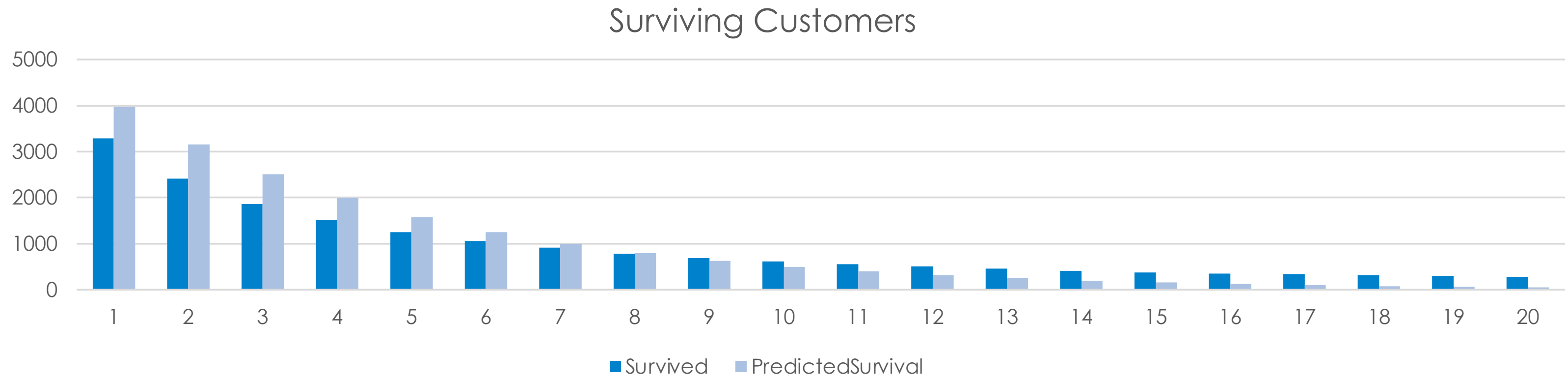
BetaGeometricDetailed_inclass.xlsx – FiniteMixtureIndividual

Estimate θs using MLE

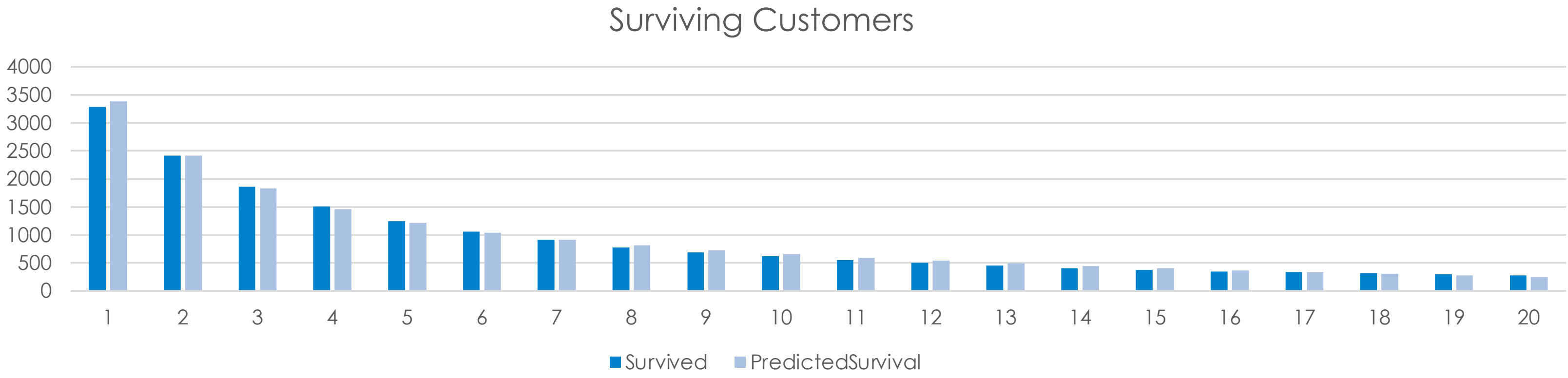
$$L_i(\theta_1, \theta_2, w|y_i) = w_1((1-\theta_1)^{y_i-1}\theta_1)^{1-\delta_i}((1-\theta_1)^{y_i})^{\delta_i}+w_2((1-\theta_2)^{y_i-1}\theta_2)^{1-\delta_i}((1-\theta_2)^{y_i})^{\delta_i}$$

One Segment vs. Two Segments

One Segment



Two Segments



Continuous Mixture Model - Beta Geometric Model

- So far:
  - Geometric Distribution: no heterogeneity
  - Finite Mixture: heterogeneity through segments
  - A more flexible model would have many segments but not parsimonious!
  - Instead, we assume that each customer has a unique value of θ

- Let θ_i be the probability associated with customer i
  - We assume that θ_i vary across customers according to a population distribution
  - θ_i is itself a random variable

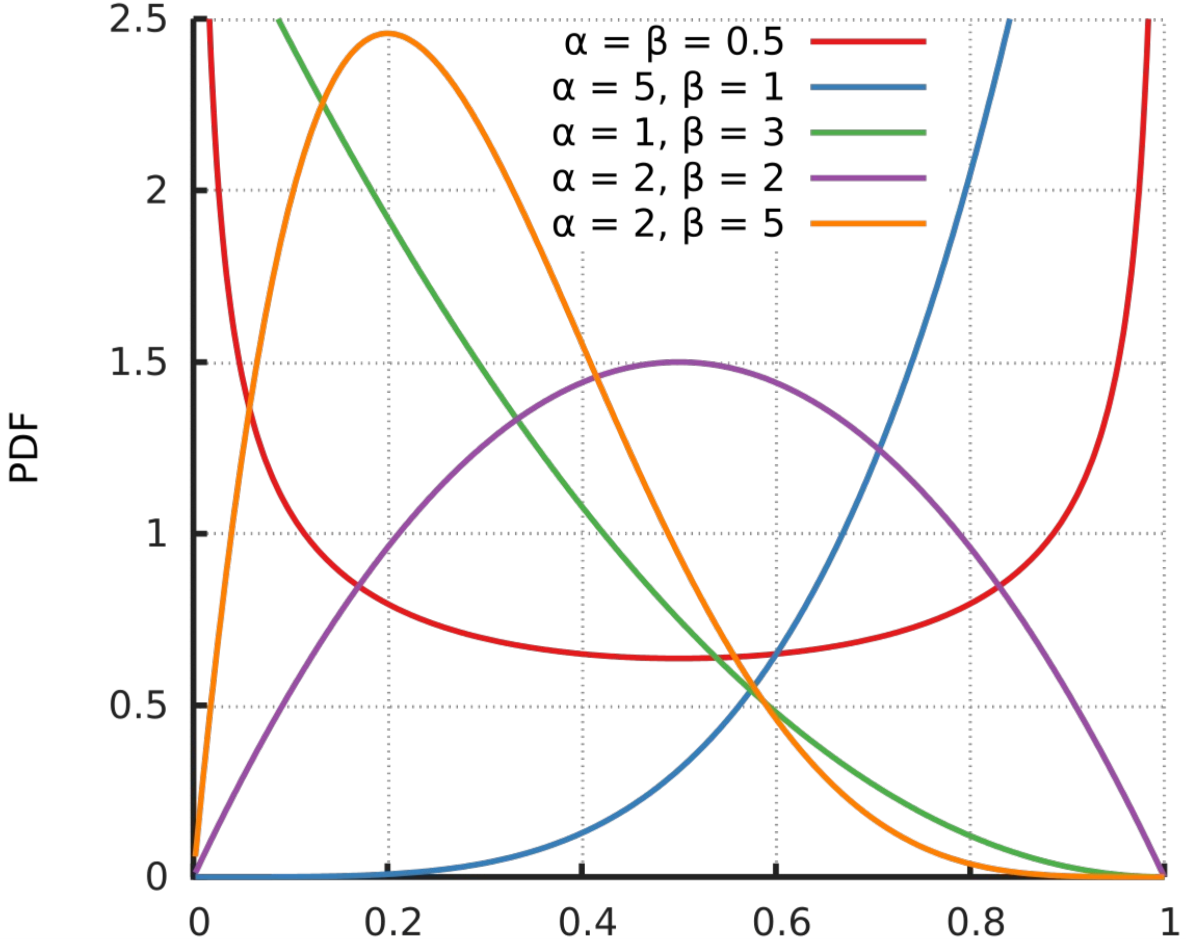
Population Distribution

- $\theta_i \sim Beta(a,b)$: Beta distribution
- The PDF of Beta distribution is given by

$$p(\theta|a,b) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}, a > 0\,, b > 0$$

- $B(a,b)$ is the beta function and the normalizing constant
- The gamma function $\Gamma(z)$ extends the factorial function to complex numbers
  - $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\ dx,\ \ Re(z) > 0$
  - For positive integers we have, $\Gamma(n) = (n-1)!$
  - $\Gamma(z+1) = \ z\,\Gamma(z)$

Why Beta distribution?



Beta Geometric Model

- Individual Level Model

$$f(y_i|\theta_i) = \text{Geometric}(y_i|\theta_i)$$

- Population Distribution

$$g(\theta_i|a,b) = \text{Beta}(\theta_i|a,b)$$

- Mixture Distribution: Beta Geometric

$$p(y_i|a,b) = \int_0^1 f(y_i|\theta_i)g(\theta_i|a,b)\,d\theta_i$$

- Why is it better?
  - Conjugacy – buys us algebraic convenience

Beta Geometric Model

- Complete Data

$$P(y_i = k|a, b) = \frac{B(a+1, b+k+1)}{B(a, b)}$$

- Censored Data

$$S(y_i = k|a, b) = \frac{B(a, b+k)}{B(a, b)}$$

- Implementation in Excel

Other Extensions

- Continuous time survival models
  - The time of churn is a continuous variable
- Non-contractual settings
  - Blood donations

Takeaways

- Modeling churn is related to the retention problem faced by companies
- Customer churn can be modeled using discrete-time customer base models
    - Geometric Distribution
- Models can be more complex and incorporate customer heterogeneity
    - Finite Mixture Models
- Possible extensions
    - Beta-Geometric Model
    - Continuous Survival Models