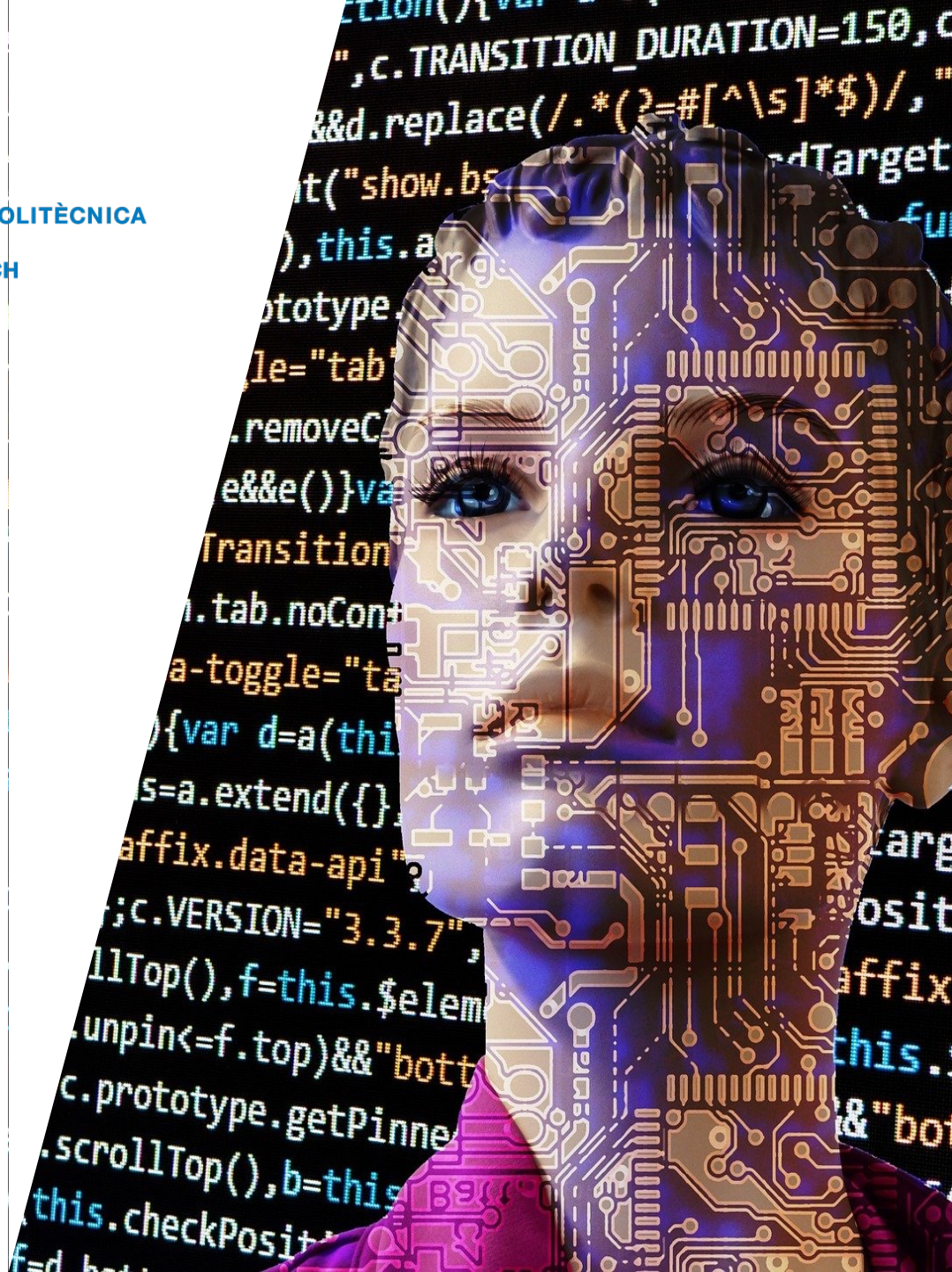
TechTalentLab

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Customer Segmentation

Customer Analytics

Dr. Vicenc Fernandez



UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

TechTalentLab

Customer Segmentation

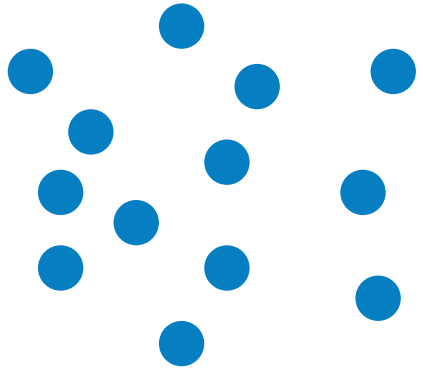
Why, why, why?

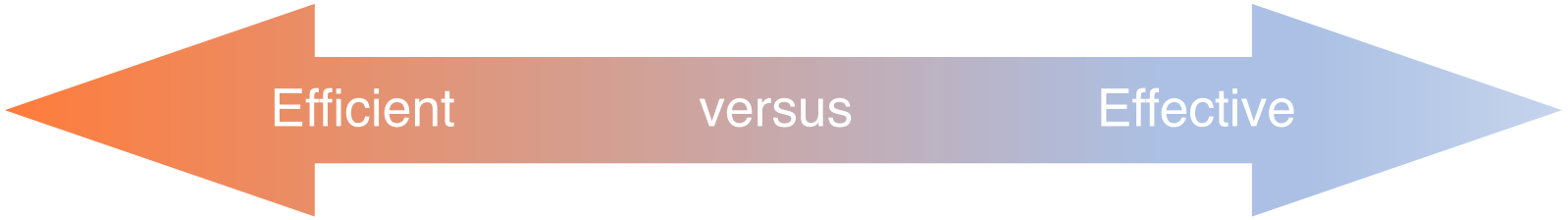
Customer Analytics, why?

Descriptive analytics (PAST), why?

Customer Segmentation, why?

Introduction





Market Matching Strategy

Segmentation

- Act of dissecting the marketplace into submarkets that require different marketing mixes

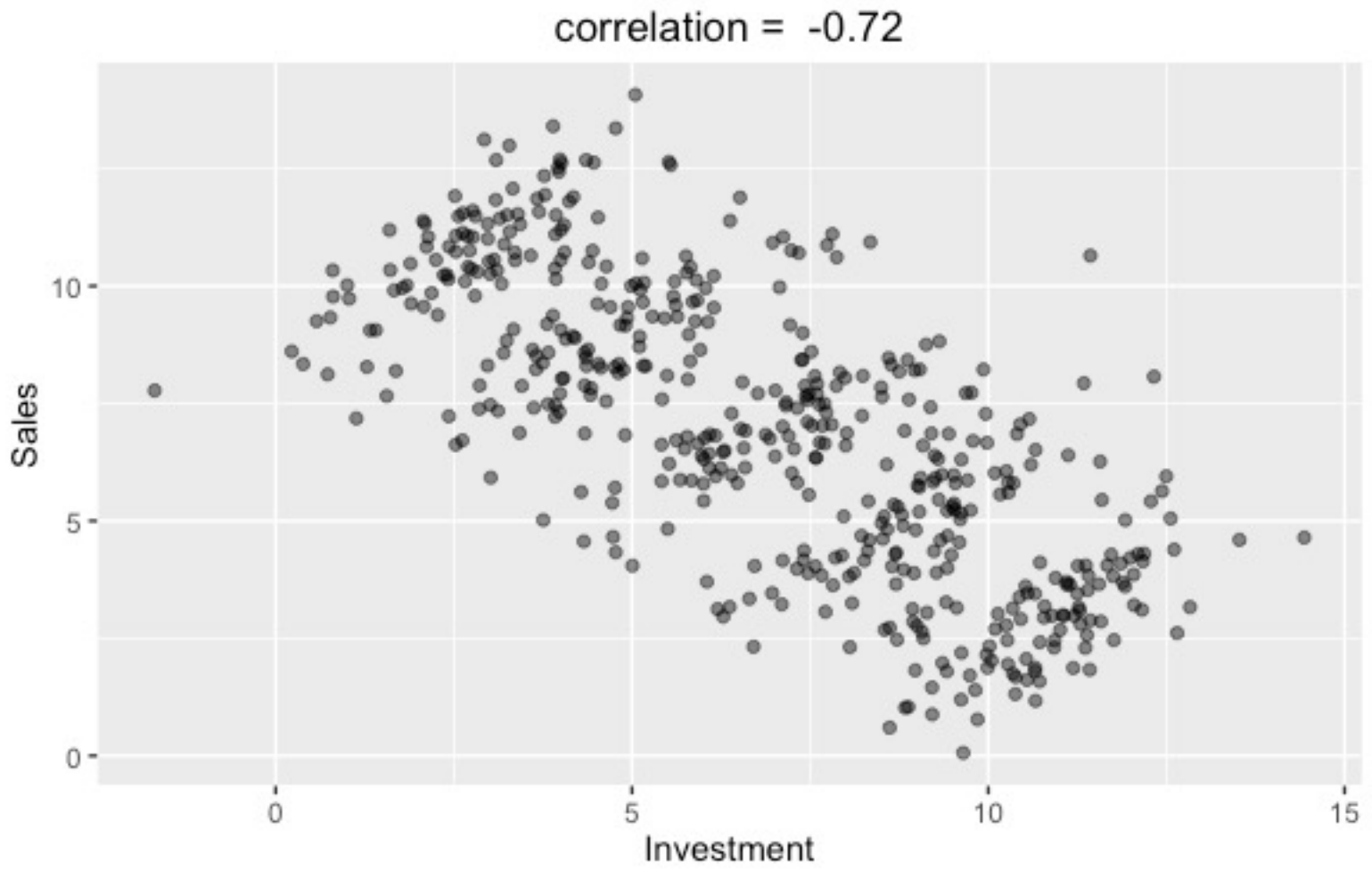
Targeting

- Process of reviewing market segments and deciding which one(s) to pursue

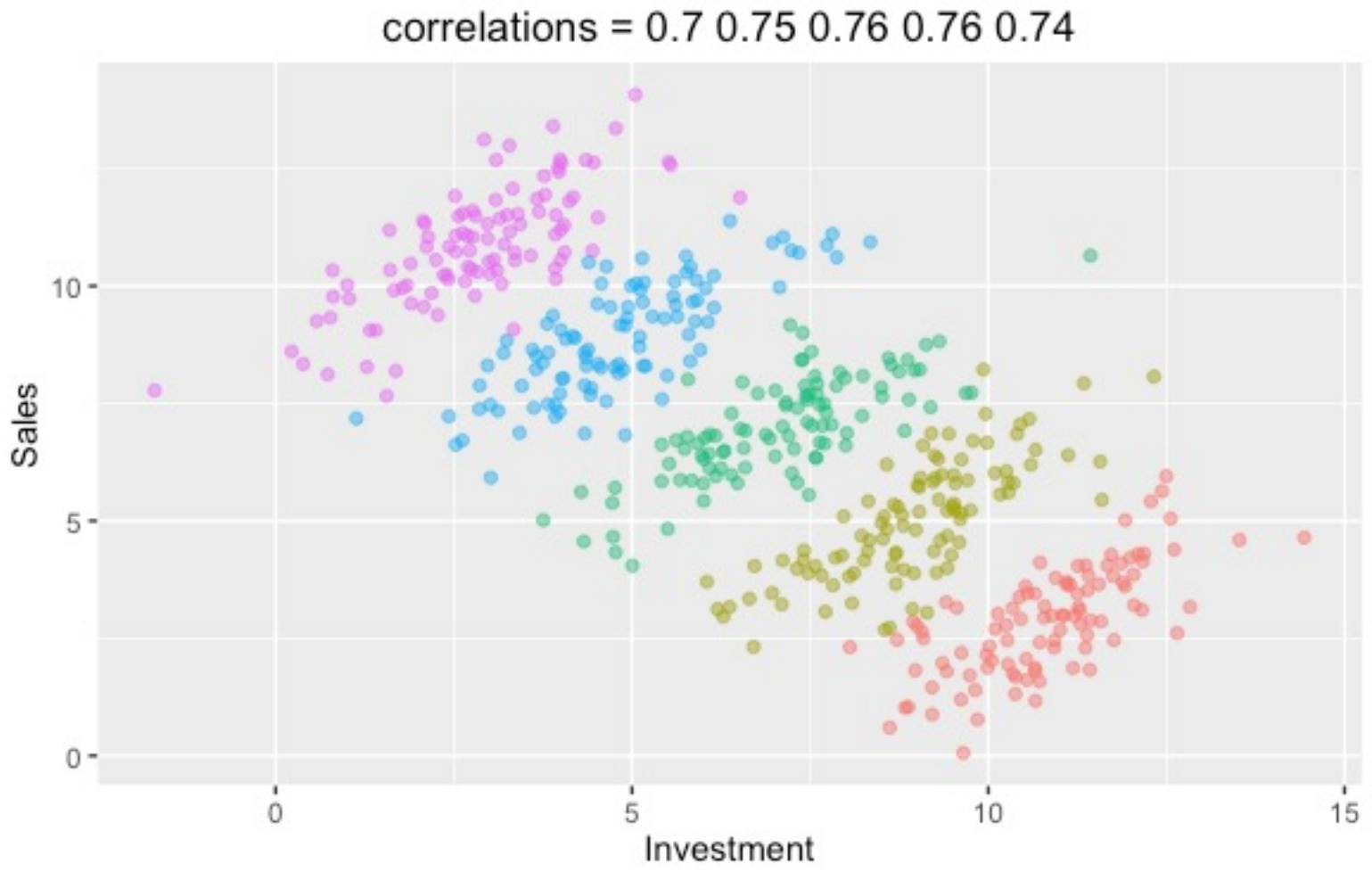
Positioning

- Establishing a differentiating image for a product or service in relation to its competition

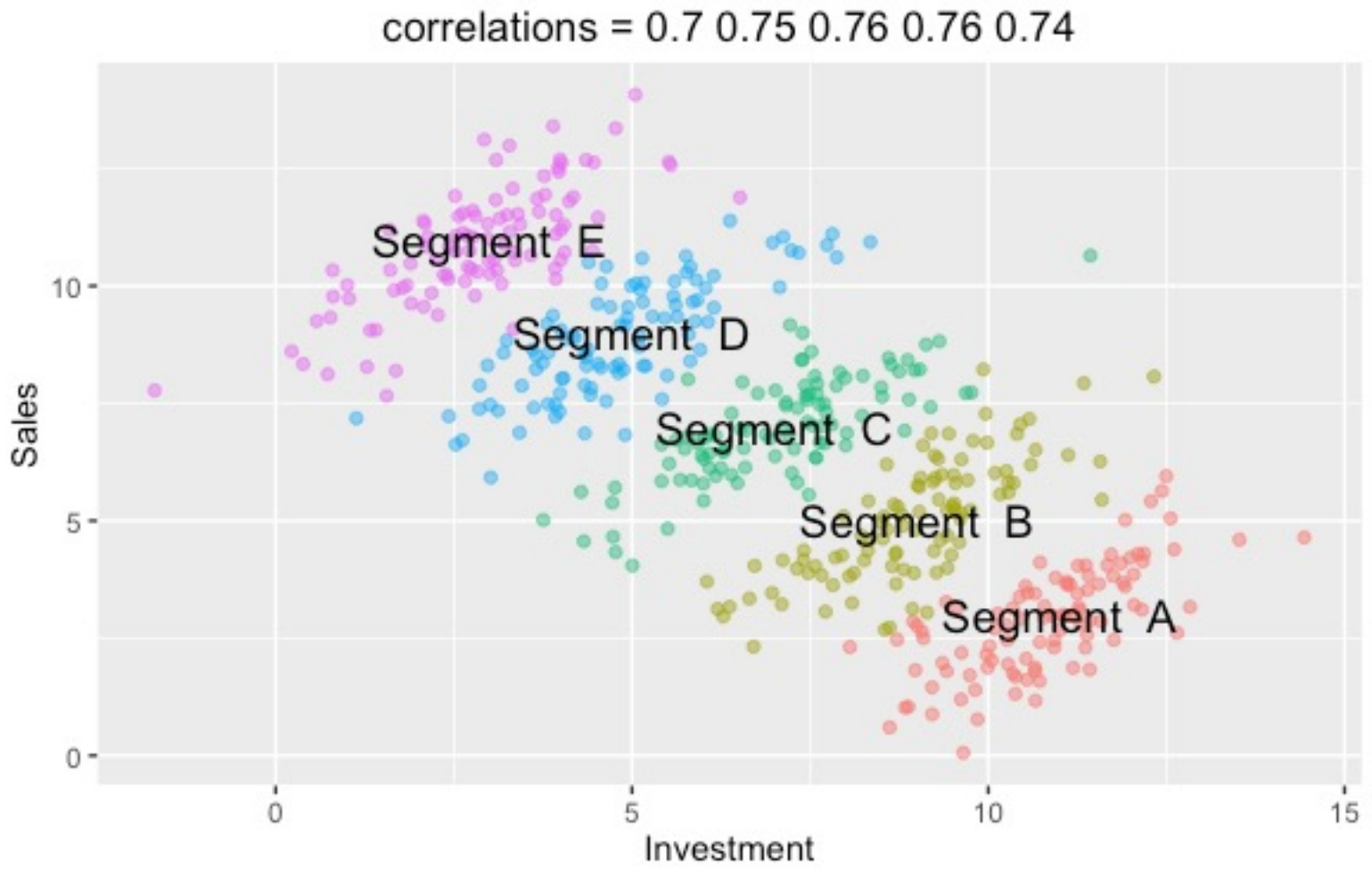
What is happening here?



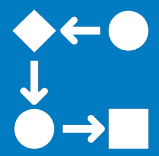
What is happening here?



What is happening here?



Steps for Customer Segmentation

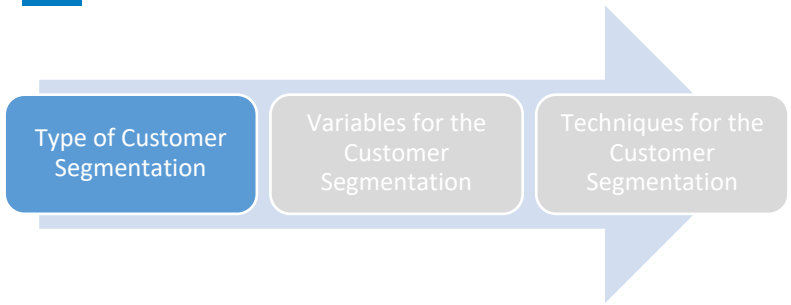


Types of Customer Segmentation

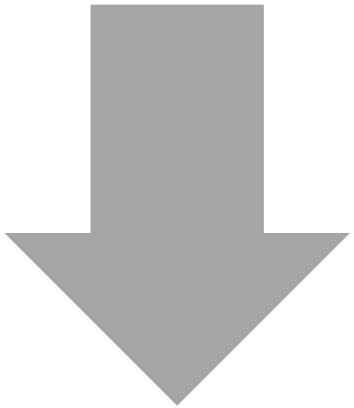
Variables for the Customer Segmentation

Techniques for the Customer Segmentation

Types of Customer Segmentation



Theoretical-driven segmentation



Data-driven segmentation

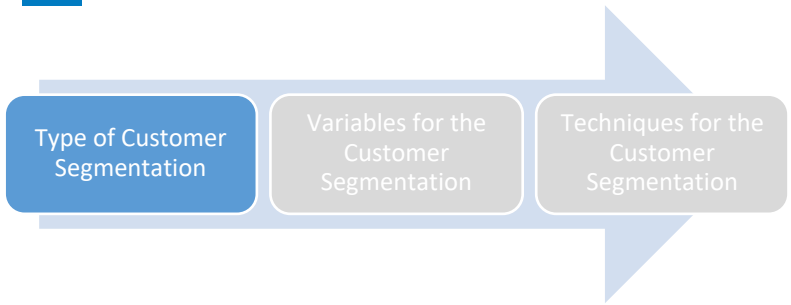
Types of Customer Segmentation

Type of Customer Segmentation

Variables for the Customer Segmentation

Techniques for the Customer Segmentation

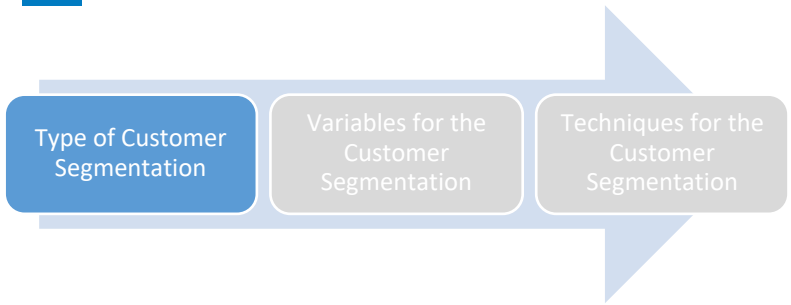
Types of Customer Segmentation



We are working in an online company that sells fruit and vegetable

Who are our customers?
How can we segment them?

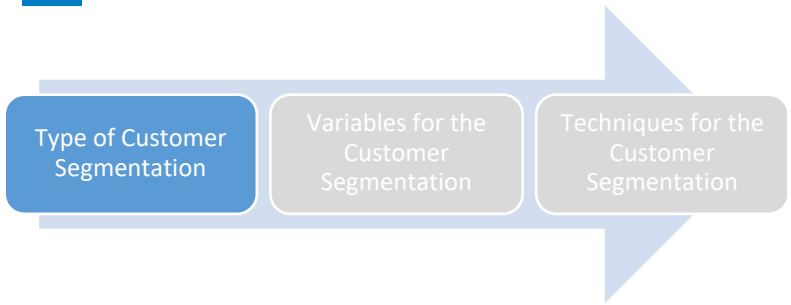
Types of Customer Segmentation



We are working in Wallapop

Who are our customers?
How can we segment them?

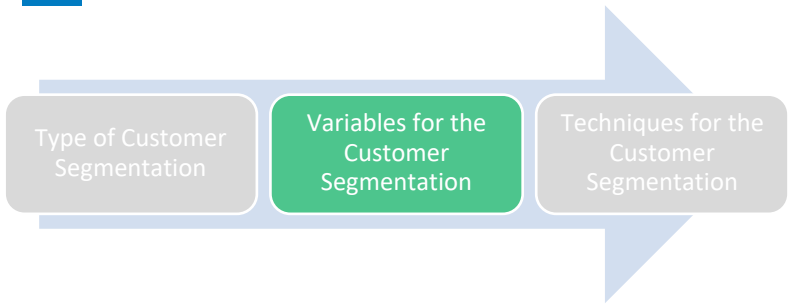
Types of Customer Segmentation



We are working in an airline.

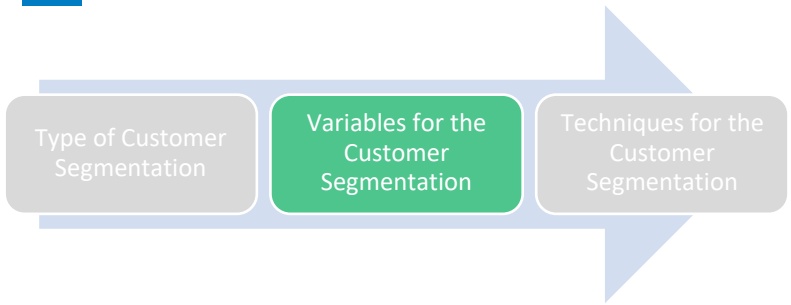
Who are our customers?
How can we segment them?

Variables for the Customer Segmentation

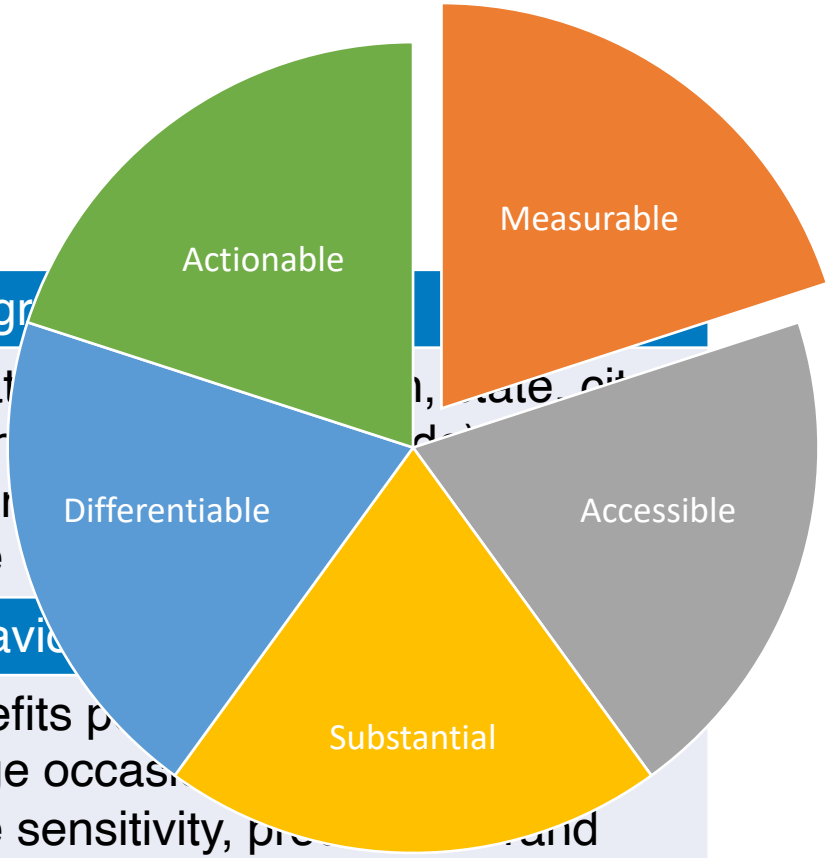


Demographic	Geographic
Age, gender, household size, family status, income, race, occupation, religion, class, nationality, education	Location (country, region, state, city, neighborhood, postal code), distance, climate, population density, taste
Psychographic	**Behavioral/Attitudinal**
Lifestyle, activities, interests	Benefits perceived/expected, loyalty, usage occasion/rate, user status, price sensitivity, product or brand attitude

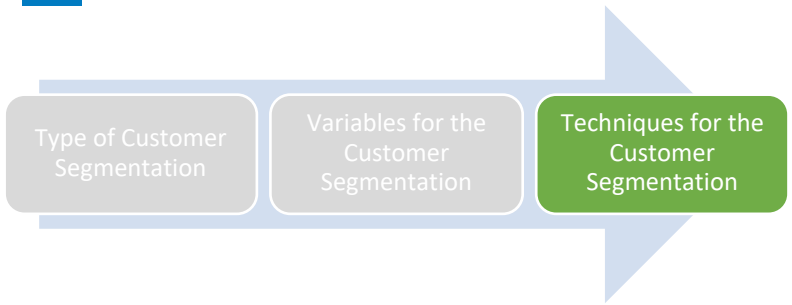
Variables for the Customer Segmentation



Demographic	Geogr
Age, gender, household size, family status, income, race, occupation, religion, class, nationality, education	Locat… neigh… distan… taste
Psychographic	**Behavi…**
Lifestyle, activities, interests	Benefits p… usage occas… price sensitivity, pre… …and attitude



Techniques for the Customer Segmentation



- K-means clustering
- **Hierarchica clustering**
- **DBSCAN - Density-Based Spatial Clustering of Applications with Noise**
- **AGNES Agglomerative Nesting**

Group Activity

We have two datasets:

- transactions.csv, where we have all our customers' purchases
- customers.csv, where you have demographic information about our customers

You need to build a customer segmentation with the available information

- You cannot search information on the web
- You need to prepare 1 slide for presenting your results.

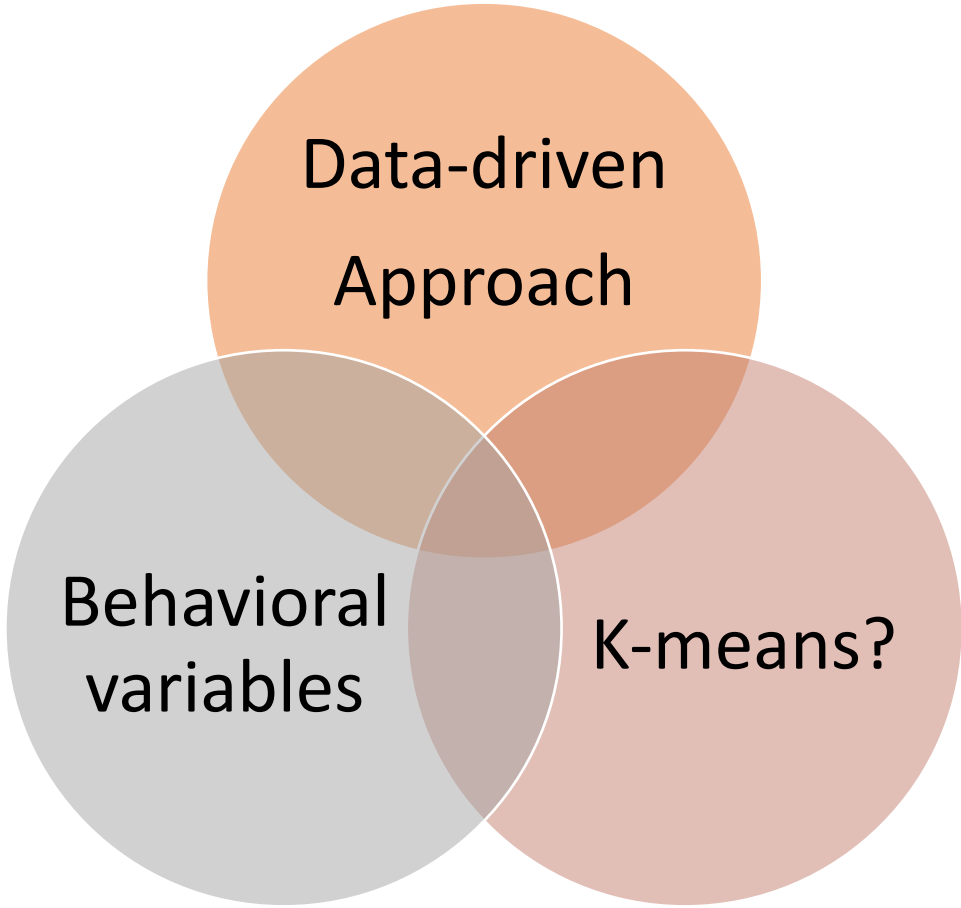
Groups of 3 people **30 minutes**

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

TechTalentLab

RFM Model

RFM Model



Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
0001	1	0	0	0	0	0	0	?	?	?
0002	1	0	0	1	0	0	0	?	?	?
0003	1	0	1	0	0	0	0	?	?	?
0004	1	0	1	0	1	1	1	?	?	?
0005	1	0	1	1	1	0	1	?	?	?
0006	1	1	1	1	0	1	0	?	?	?
0007	1	1	0	1	0	1	0	?	?	?
0008	1	1	1	1	1	1	1	?	?	?
0009	1	1	1	1	1	1	0	?	?	?
0010	1	0	1	1	0	1	1	?	?	?
…										

Who is the ***best*** customer?
Who is the person who has the highest probability to donate again?

Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
0001	1	0	0	0	0	0	0	?	?	?
0002	1	0	0	1	0	0	0	?	?	?
0003	1	0	1	0	0	0	0	?	?	?
0004	1	0	1	0	1	1	1	?	?	?
0005	1	0	1	1	1	0	1	?	?	?
0006	1	1	1	1	0	1	0	?	?	?
0007	1	1	0	1	0	1	0	?	?	?
Bob	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***?***	***?***	***?***
0009	1	1	1	1	1	1	0	?	?	?
0010	1	0	1	1	0	1	1	?	?	?
…										

Who is the ***best*** customer?
Who is the person who has the highest probability to donate again?

FREQUENCY

How many times did the donor give money in the last 7 years?

Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
0001	1	0	0	0	0	0	0	?	?	?
0002	1	0	0	1	0	0	0	?	?	?
0003	1	0	1	0	0	0	0	?	?	?
0004	1	0	1	0	1	1	1	?	?	?
0005	1	0	1	1	1	0	1	?	?	?
0006	1	1	1	1	0	1	0	?	?	?
0007	1	1	0	1	0	1	0	?	?	?
Bob	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***?***	***?***	***?***
0009	1	1	1	1	1	1	0	?	?	?
0010	1	0	1	1	0	1	1	?	?	?
…										

Who is the ***worst*** customer?
Who is the person who has the lowest probability to donate again?

Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Sara	1	0	0	0	0	0	0	?	?	?
0002	1	0	0	1	0	0	0	?	?	?
0003	1	0	1	0	0	0	0	?	?	?
0004	1	0	1	0	1	1	1	?	?	?
0005	1	0	1	1	1	0	1	?	?	?
0006	1	1	1	1	0	1	0	?	?	?
0007	1	1	0	1	0	1	0	?	?	?
Bob	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***?***	***?***	***?***
0009	1	1	1	1	1	1	0	?	?	?
0010	1	0	1	1	0	1	1	?	?	?
…										

Who is the ***worst*** customer?
Who is the person who has the lowest probability to donate again?

FREQUENCY

How many times did the donor give money in the last 7 years?

Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Sara	1	0	0	0	0	0	0	?	?	?
0002	1	0	0	1	0	0	0	?	?	?
0003	1	0	1	0	0	0	0	?	?	?
0004	1	0	1	0	1	1	1	?	?	?
0005	1	0	1	1	1	0	1	?	?	?
0006	1	1	1	1	0	1	0	?	?	?
0007	1	1	0	1	0	1	0	?	?	?
Bob	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***?***	***?***	***?***
0009	1	1	1	1	1	1	0	?	?	?
0010	1	0	1	1	0	1	1	?	?	?
…										

After Bob,
who is the person who has the highest probability to donate again?

Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Sara	1	0	0	0	0	0	0	?	?	?
0002	1	0	0	1	0	0	0	?	?	?
0003	1	0	1	0	0	0	0	?	?	?
Mary	1	0	1	0	1	1	1	?	?	?
0005	1	0	1	1	1	0	1	?	?	?
0006	1	1	1	1	0	1	0	?	?	?
0007	1	1	0	1	0	1	0	?	?	?
Bob	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***?***	***?***	***?***
John	1	1	1	1	1	1	0	?	?	?
0010	1	0	1	1	0	1	1	?	?	?
…										

After Bob,
who is the person who has the highest probability to donate again?

Why did you decide Mary or John?

RECENCY: How recently did the donor give money? When was the last time the donor give money?

FREQUENCY: How many times did the donor give money in the last 7 years?

Why did you decide Mary or John?

RECENCY: How recently did the donor give money? When was the last time the donor give money?

Which is more important?

FREQUENCY: How many times did the donor give money in the last 7 years?

Donor Types???

ALIVE: An active donor → This person gives frequently and gave recently

DORMANT: This person has not given recently, but is likely to give again with the right development prompts

LAPSED: This person has not given recently and is not likely to give again

Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Sara	1	0	0	0	0	0	0	?	?	?
0002	1	0	0	1	0	0	0	?	?	?
0003	1	0	1	0	0	0	0	?	?	?
Mary	1	0	1	0	1	1	1	?	?	?
0005	1	0	1	1	1	0	1	?	?	?
0006	1	1	1	1	0	1	0	?	?	?
0007	1	1	0	1	0	1	0	?	?	?
Bob	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***?***	***?***	***?***
John	1	1	1	1	1	1	0	?	?	?
0010	1	0	1	1	0	1	1	?	?	?
…										

What type of Donor are Mary and John? Alive, Dormant, or Lapsed

Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Sara	1	0	0	0	0	0	0	?	?	?
0002	1	0	0	1	0	0	0	?	?	?
0003	1	0	1	0	0	0	0	?	?	?
Mary	1	0	1	0	1	1	1	?	?	?
0005	1	0	1	1	1	0	1	?	?	?
0006	1	1	1	1	0	1	0	?	?	?
0007	1	1	0	1	0	1	0	?	?	?
Bob	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***?***	***?***	***?***
John	1	1	1	1	1	1	0	?	?	?
Cris	1	0	1	1	0	1	1	?	?	?
…										

Who is more valuable: Mary or Cris? Both have 5 donations

Why did you decide Mary or John?

RECENCY: How ***recently*** did the donor give money? When was the ***last time*** the donor give money?

FREQUENCY: How many times did the donor give money in the last 7 years?

Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Sara	1	0	0	0	0	0	0	?	?	?
0002	1	0	0	1	0	0	0	?	?	?
0003	1	0	1	0	0	0	0	?	?	?
Mary	1	0	1	0	1	1	1	?	?	?
0005	1	0	1	1	1	0	1	?	?	?
0006	1	1	1	1	0	1	0	?	?	?
0007	1	1	0	1	0	1	0	?	?	?
Bob	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***?***	***?***	***?***
John	1	1	1	1	1	1	0	?	?	?
Cris	1	0	1	1	0	1	1	?	?	?
…										

Could you predict what is going to happen with Cris in the following 3 years?

Let's see an example: donations

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Cris	1	0	1	1	0	1	1	?	?	?

ID	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Cris	1	0	1	1	0	1	1	0	1	1

But, do you think that people are doing this kind of things??

Could you predict what is going to happen with Cris in the following 3 years?

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

TechTalentLab

RFM Segmentation

The Monetary Score

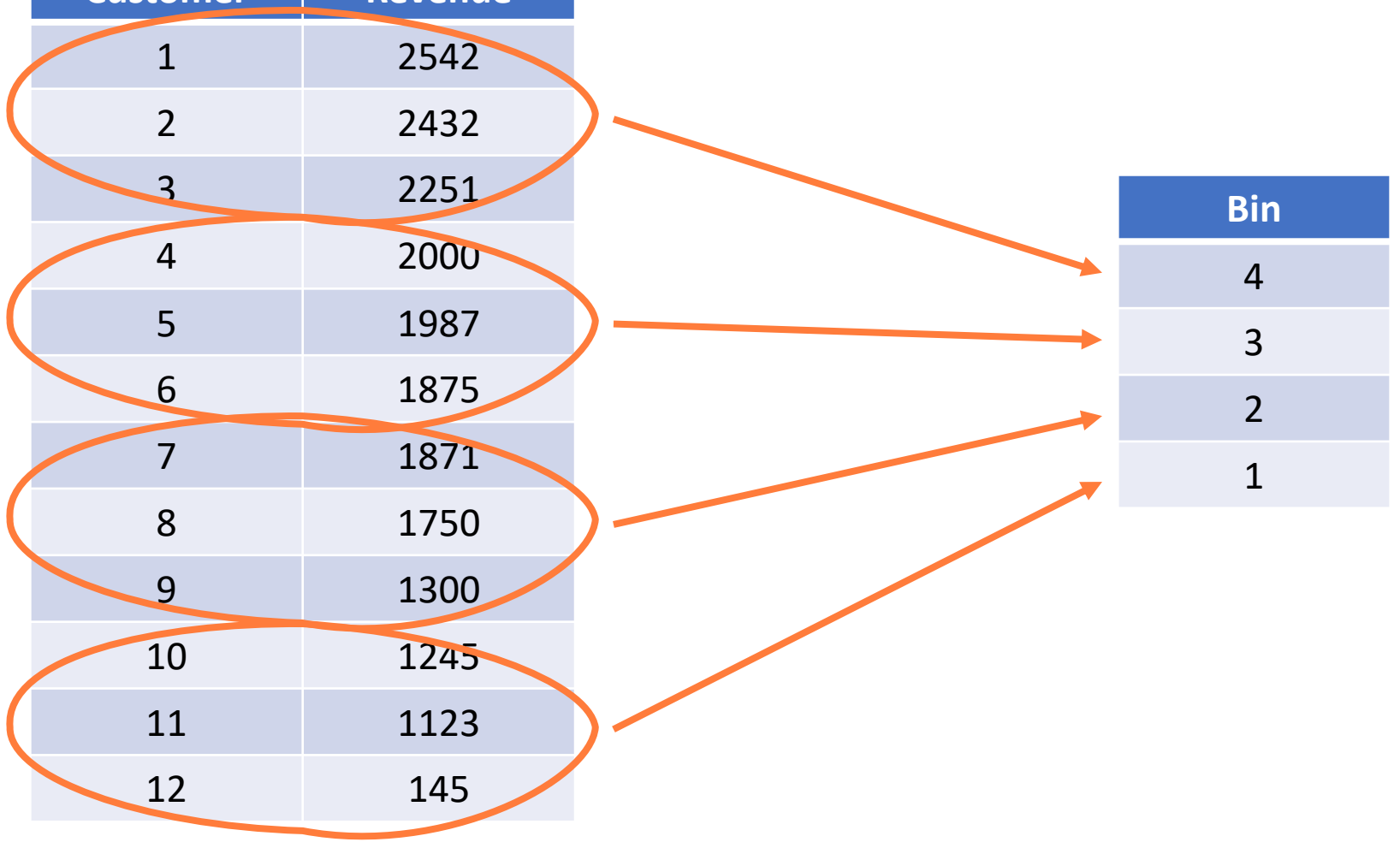
Calculate the total revenue by each customer

Define the number of bins for the score

Bin the revenue values: the highest value is the most highest revenue

The Monetary Score

Customer	Revenue
1	2542
2	2432
3	2251
4	2000
5	1987
6	1875
7	1871
8	1750
9	1300
10	1245
11	1123
12	145



Bin
4
3
2
1

Frequency Score

Calculate the total number of purchases

Define the number of bins for the score

Bin the frequencyvalues: the highest value is the most frequency

The Recency Score

Select the date of most recent purchase

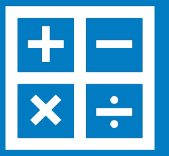
Define the number of bins for the score

Bin the recency values: the highest value is the most recency

RFM Score

RFM score is generated which is simply the three individual scores concatenated into a single value.

recency_score	frequency_score	monetary_score	rfm_score
2	2	2	222
4	5	5	455
5	1	1	511
5	4	4	544
5	4	5	545
5	4	3	543
4	5	5	455



RFM Segmentation

Segment	Description	R	F	M
Champions	Bought recently, buy often and spend the most	4 - 5	4 - 5	4 - 5
Loyal Customers	Spend good money. Responsive to promotions	2 - 5	3 - 5	3 - 5
Potential Loyalist	Recent customers, spent good amount, bought more than once	3 - 5	1 - 3	1 - 3
New Customers	Bought more recently, but not often	4 - 5	<= 1	<= 1
Promising	Recent shoppers, but haven't spent much	3 - 4	<= 1	<= 1
Need Attention	Above average recency, frequency & monetary values	2 - 3	2 - 3	2 - 3
About To Sleep	Below average recency, frequency & monetary values	2 - 3	<= 2	<= 2
At Risk	Spent big money, purchased often but long time ago	<= 2	2 - 5	2 - 5
Can't Lose Them	Made big purchases and often, but long time ago	<= 1	4 - 5	4 - 5
Hibernating	Low spenders, low frequency, purchased long time ago	1 - 2	1 - 2	1 - 2
Lost	Lowest recency, frequency & monetary scores	<= 2	<= 2	<= 2

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

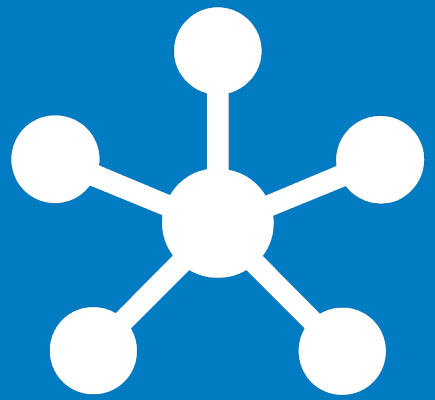
TechTalentLab

Clustering

K-Means

K-means clustering tries to find the centroids of a fixed number of clusters of points in a high-dimensional space.

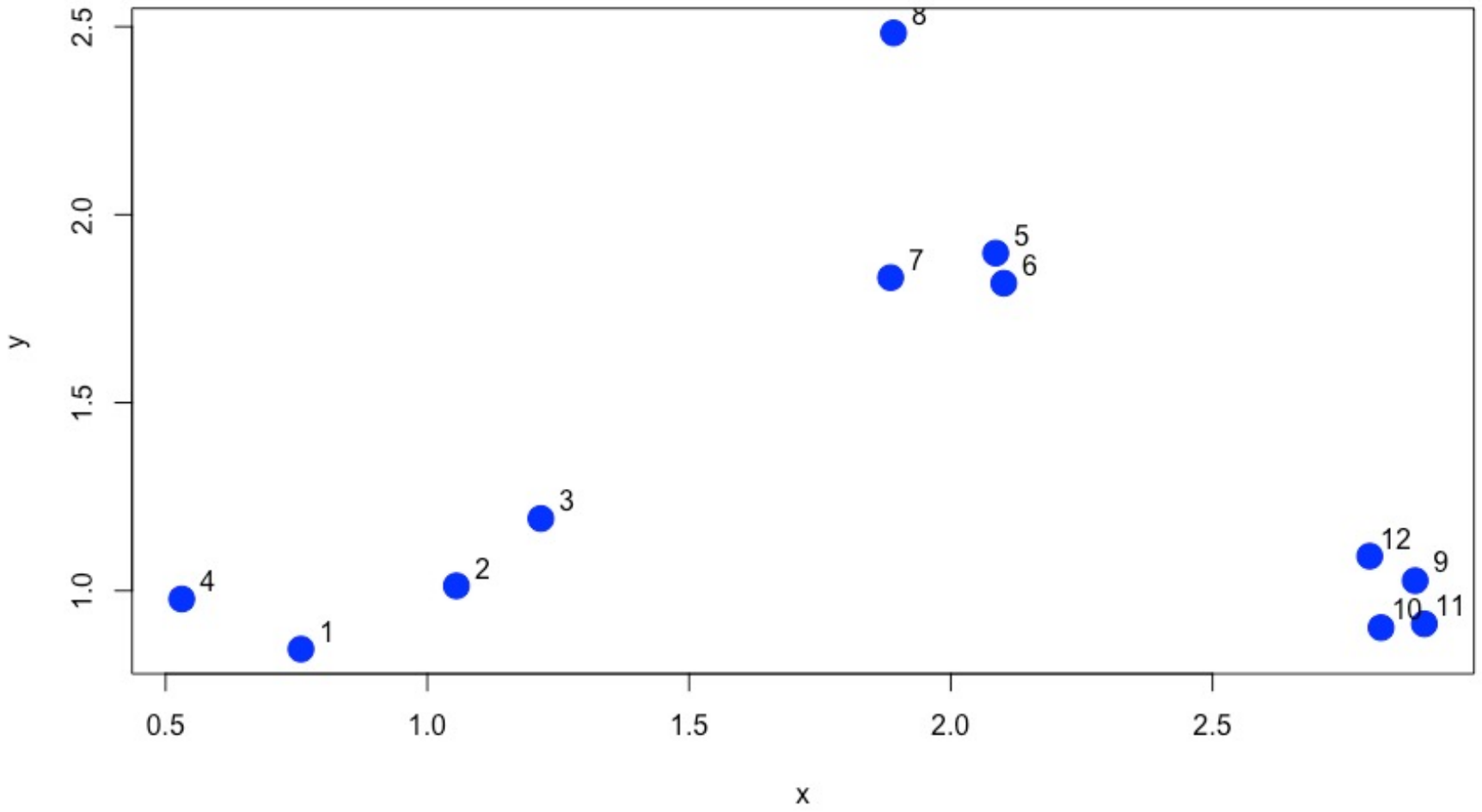
The idea is very simple: to cluster cases, in other words, to group cases.



The algorithm is recursive and goes as follows

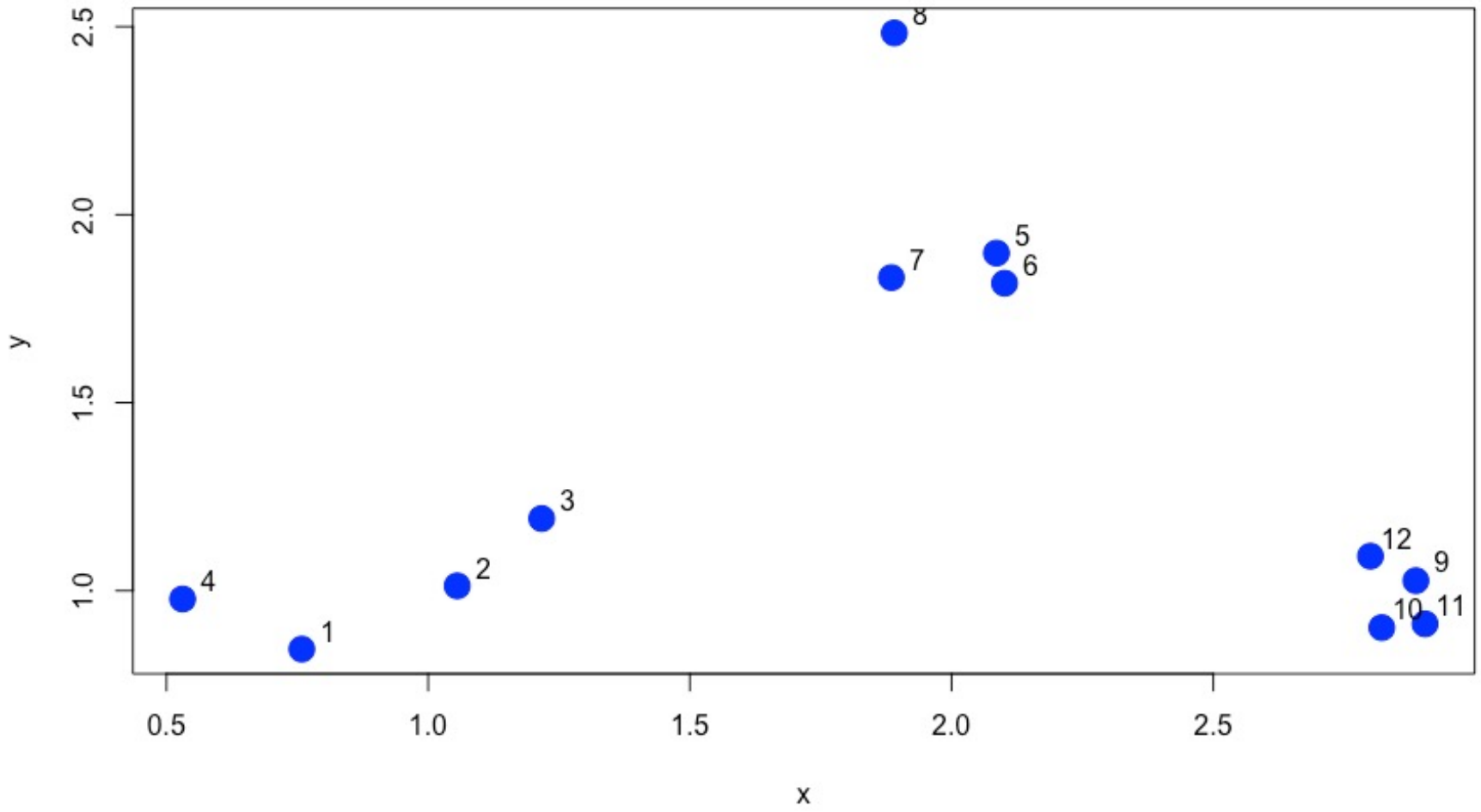
- Fix the number of clusters at some integer greater than or equal to 2
- Start with the “centroids” of each cluster
- Assign points to their closest centroid; cluster membership corresponds to the centroid assignment
- Recalculate centroid positions and repeat.

Example

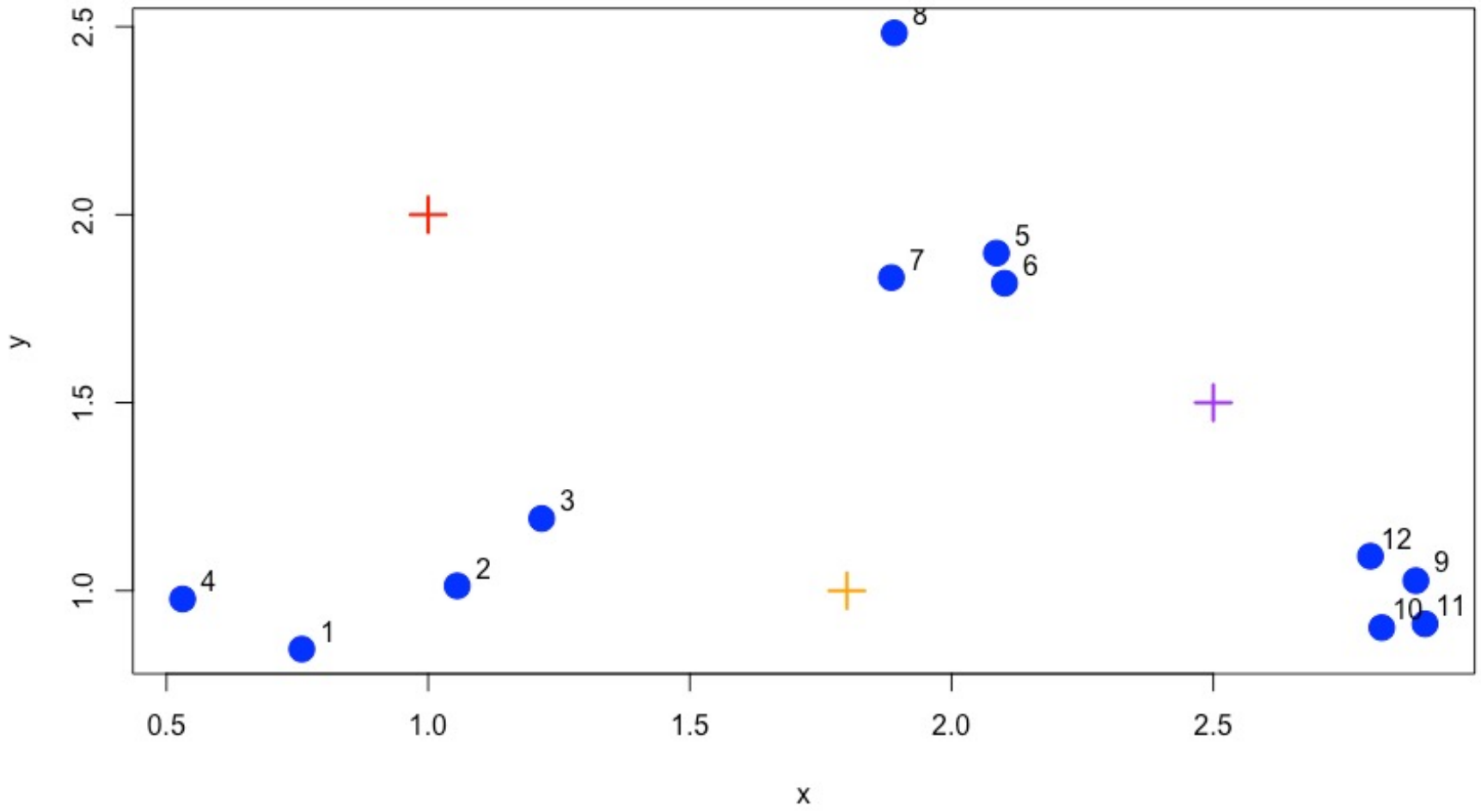


Example

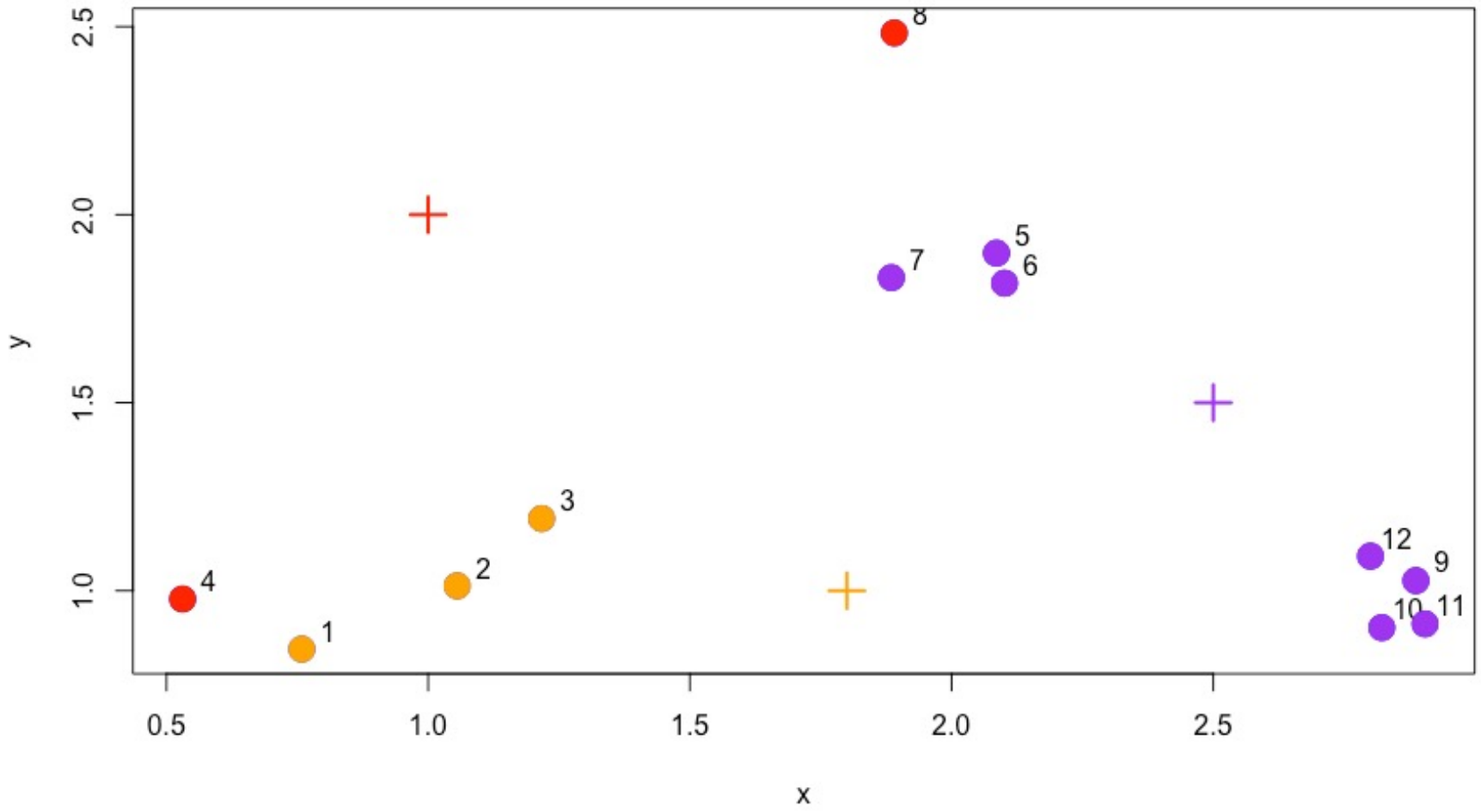
3 clusters

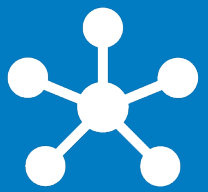


Example

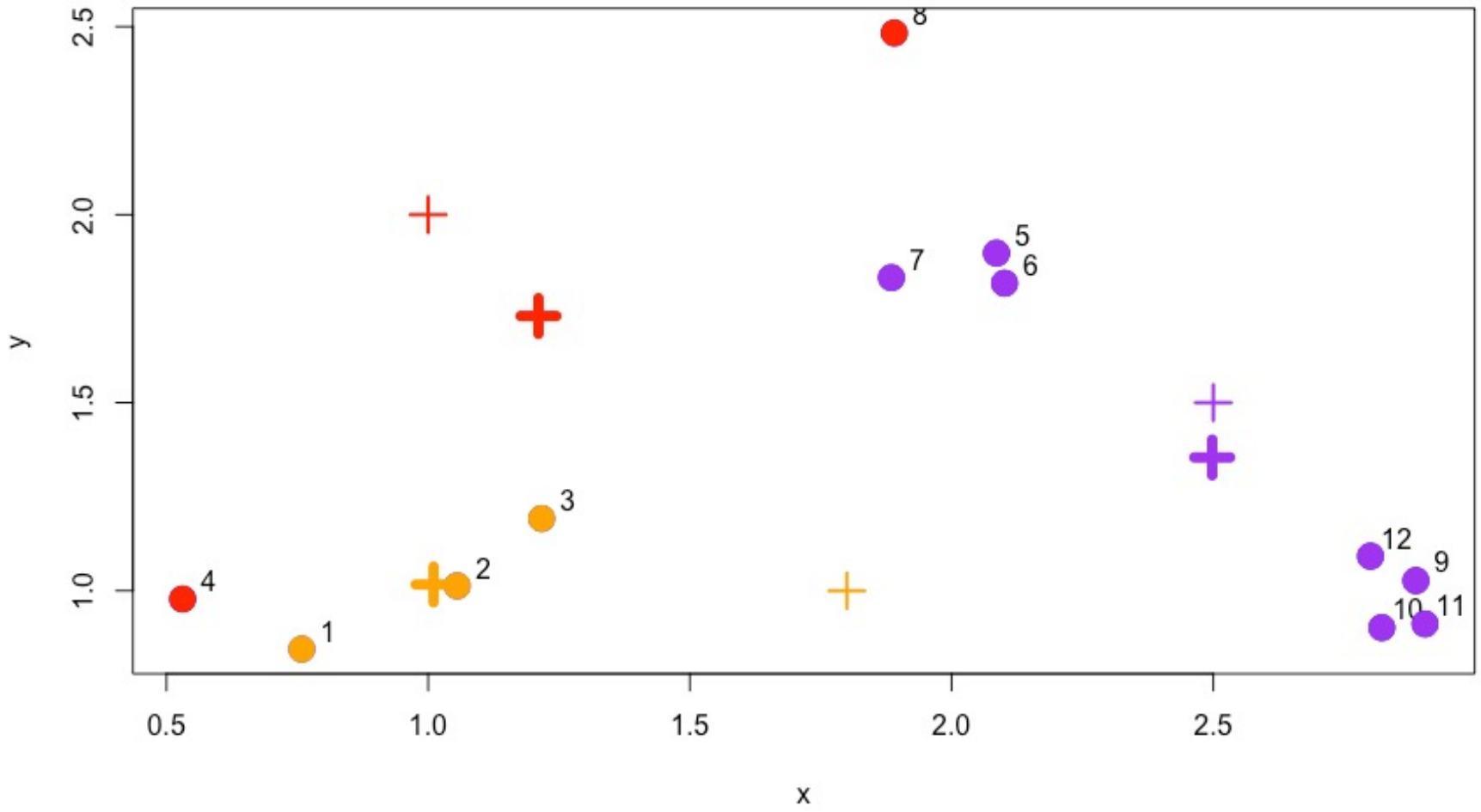


Example

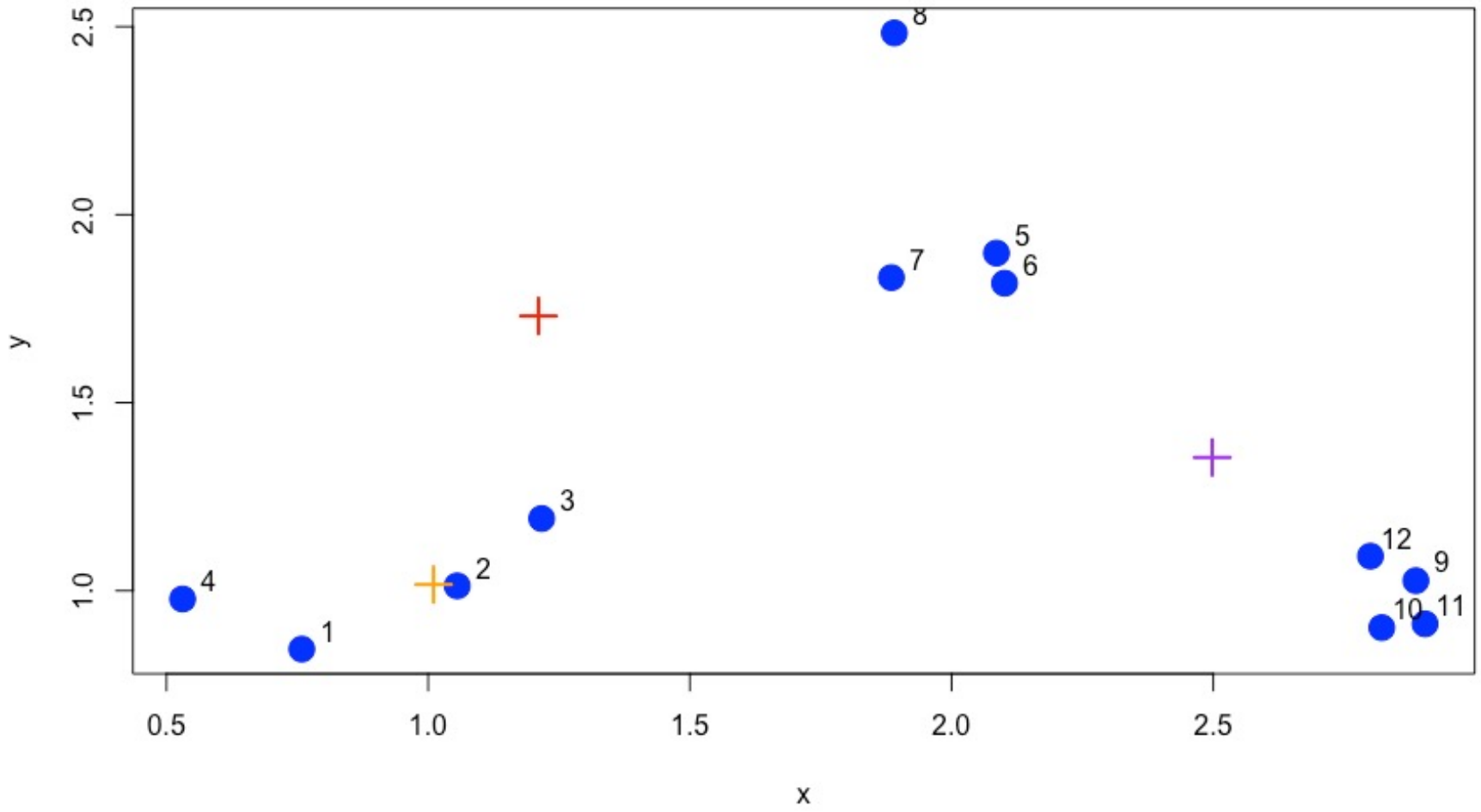




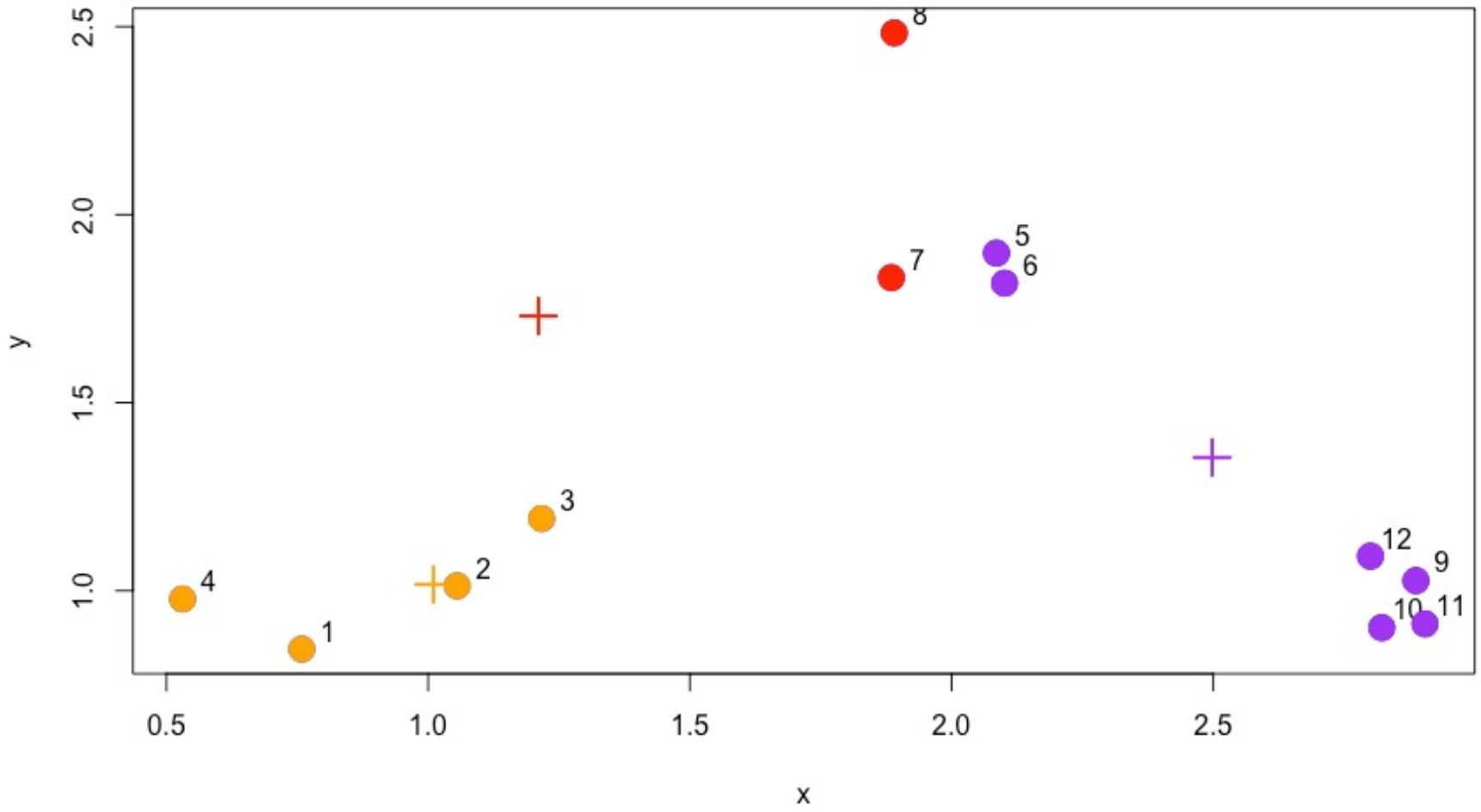
Example



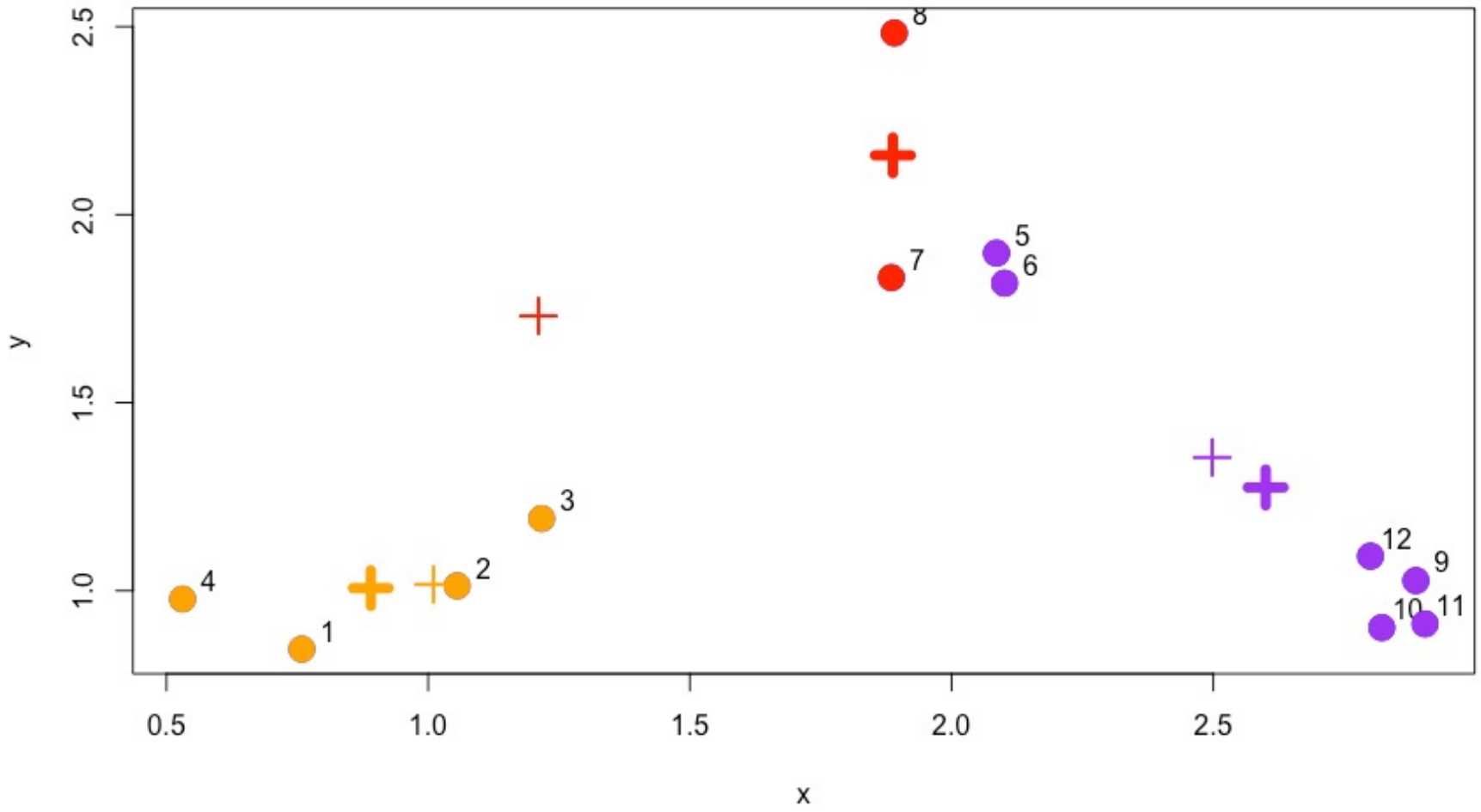
Example



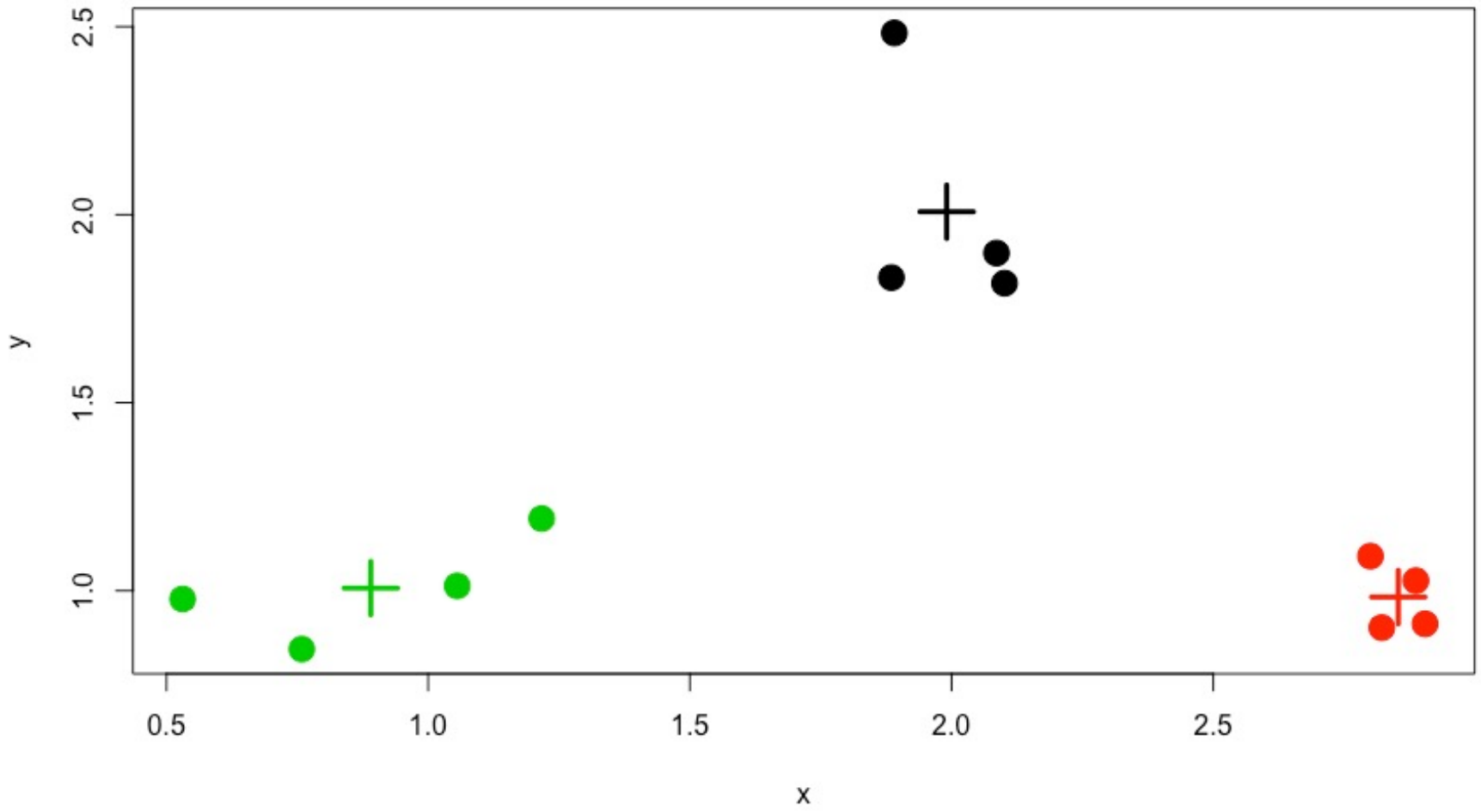
Example



Example



Example

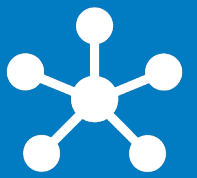


K-Means

```
> kmeansObj <- kmeans(dataFrame, centers = 3)
```

```
> kmeansObj$cluster
> kmeansObj$center
```

```
> plot(x, y, col=kmeansObj$cluster, pch=19, cex=2)
> points(kmeansObj$centers, col=1:3, pch=3, cex=3, lwd=3)
```



RFM Segmentation

Let’s see what we can do!!