What is the

N.R.E.

...?

NETFLIX

Recommandation Engine

Filters over **3,000 titles** at a time for **231 million subscribers**.

More than **1,300 recommendation clusters** based on user preferences.

80% of Netflix viewer activity is driven by personalized recommendations from the engine.

“Its estimated that the NRE saves Netflix over $1 billion per year in customer acquisition as of 2016.”

https://techjury.net/blog/netflix-statistics/

bip. xTech

Customer Segmentation with **Clustering**

Theory and practice of model building

March 2023

BIP CONSULTING

RCS

Giacomo Tanzi

HERE TO DARE

bip. xTech

“We use exponential technologies to power end-to-end digital solutions, supporting data-driven transformation and scaling up responsiveness to business evolution.”

xTech at a glance

Our centre of excellence in numbers

bip.xTech

The **largest** professional **DATA SCIENTIST** community in **Italy**!*

We use exponential technologies to power end-to-end digital solutions, supporting data-driven transformation and scaling up responsiveness to business evolution

Data & AI

Data Scientists, Data Visualizers, Data Strategists, Data Governance Experts

Cloud

Cloud Data Architects, Data Engineers, Microservice Experts

Tech Consulting

Network, IOT, Blockchain AR/VR Experts / Architects

Hyperautomation

RPA Experts, Low Code Developers

SW Solutions

UI Developers, Full Stack Developers, DevOps Engineer, Test Engineer

90%
Loyal customers – b2b service renewal

120+
Open **collaborations** with clients (**+10%** YoY)

500+
Projects successfully delivered in the last 3 years, across **20+ countries**

600+
Professionals

90%
Certified professionals

45+
Alliances with **tech partners**

Italy, US, UK, Spain & Brazil
Operations Hubs + ODC in LATAM and Eastern EU

10%
Invested **in R&D/y**

bip.xTech

Our Services

Data & AI

Data Strategy

Designing and bringing up-to-scale complex, data-driven organizations

Data Governance

Defining and setting up organizational/ operative models and data management practices

Data Science

Conceiving, designing and implementing Business Intelligence and AI-powered solutions at-scale

Cloud

Strategy and Governance

Creating the foundation of the multi-cloud strategy; setting up frameworks to optimize Cloud costs

Architecture and Optimization

Designing and building scalable and future-fit business applications; defining frameworks to optimize performance, quality, security, reliability and cost

Migration

Migrating applications and data platforms to the cloud & preparing the organization for the transition

Hyperautomation

Hyper-automation

Analyzing business processes and automating them by blending Robotic Process Automation and AI

Low-Code Platforms

Analyzing supporting clients on Low Code Platform adoption and implementation

SW Solutions

Software Architecture

Designing software architecture at application level and platform level to meet Client requirements

Solution DevOps

Designing, developing and deploying highly-scalable bespoke full-stack applications and replatforming legacy applications in a continuous integration and continuous testing approach

Tech Consulting

ICT Strategy and Advisory

Leading edge technology advisory; impact evaluation of innovative solutions; definition of roadmaps and migration plans

Architecture and Engineering

High- and low-level design of Digital Platforms (IT, Network, Blockchain, AR/VR,…)

Sourcing and Roll-Out Support

Analyzing and scouting the market, managing RFI/RFP and tendering, technical project management

Our Team

In our team we have

Data & AI

240+

Data Scientists

Data Governance Experts

Data Strategy Experts

Data Analysts and Visualizers

Cloud

160+

Cloud (Data) Architects

Cloud (Data) Engineers

Hyperautomation

60+

RPA Specialists

Low-code Developer

SW Solutions

60+

UI Developers

Full Stack Developers

DevOps Engineer

Test Engineer

Tech Consulting

80+

IT Experts/Architects

Network Experts

IOT Experts

Blockchain Experts

bip.xTech

Our Partnerships

# digital transformation # specialized solutions # best-of-breed technologies

"Talent wins games, but teamwork and intelligence win championships"

Steve Jobs

GLOBAL CLOUD SERVICE PROVIDERS

aws	Google Cloud	Microsoft
Advanced Partner	Premier Partner	Gold Partner

HIGHLIGHTS

- Google Cloud & IDC Webinar – Cognitive Suite
- Strategic partner of Microsoft AI HUB program
- Data Platform Modernization certified practice
- Low Code Strategy and Implementation certified practice

ACHIEVEMENTS

45+	Technology alliances
100+	2021 projects with Partner Technologies
300+	Certifications held

DATA ANALYTICS AND BUSINESS INTELLIGENCE

sas, Qlik, tableau

ARTIFICIAL INTELLIGENCE

iGenius, neurala

HYPERAUTOMATION AND RPA

AUTOMATION ANYWHERE, NICE, UiPath, appian

PROCESS MINING

IBM, celonis

BLOCKCHAIN

Algorand

LOW CODE

outsystems

DATA GOVERNANCE

Collibra, Informatica

NETWORKING & COMMUNICATIONS

cisco, Microsoft

IIOT

software AG

Certifications

Data & AI

Data Science

- sas
- Microsoft GOLD CERTIFIED Partner
- Google Cloud Certified
- CCA cloudera SPARK & HADOOP DEVELOPER
- databricks Academy
- Microsoft Certified AZURE AI ENGINEER ASSOCIATE
- aws machine learning
- ArcGIS
- TensorFlow
- Microsoft Certified AZURE DATA SCIENTIST ASSOCIATE
- lookML
- Google Analytics
- neo4j
- CERTIFIED ASSOCIATE Data Science v2.0 DELL Technologies
- Google Marketing Platform
- coursera
- iGenius
- neurala
- TEALIUM

Data Governance

- DAMA International
- Informatica
- collibra
- bip.xTech

Cloud

Cloud Platforms

- Google Cloud Certified
- Microsoft Azure
- amazon web services

Data Platforms

- Google Cloud Certified
- Microsoft Azure
- amazon web services
- Google Cloud Certified
- CCA cloudera SPARK & HADOOP DEVELOPER
- TERADATA

Data Visualization

- Qlik Sense
- Power BI
- tableau
- Looker

Hyperautomation

Robotic Process Automation

- UiPath Robotic Process Automation
- NICE
- appian
- blueprism
- AUTOMATION ANYWHERE

AI Engine for Advaced BOTs

- Microsoft Cognitive Services

Process Mining

- celonis
- my invenio
- Fortress IQ

SW Solutions

Testing

- hp Loadrunner

Programming

- ORACLE Certified Professional Java SE 7 Programmer
- LPIC-1

Devops

- Microsoft Azure
- amazon web services

Low-Code

- Microsoft
- outsystems

Tech Consulting

Project Management

- PSM
- PMP
- PRINCE2

Service Management

- ISO 20000 Certified
- ITIL FOUNDATION

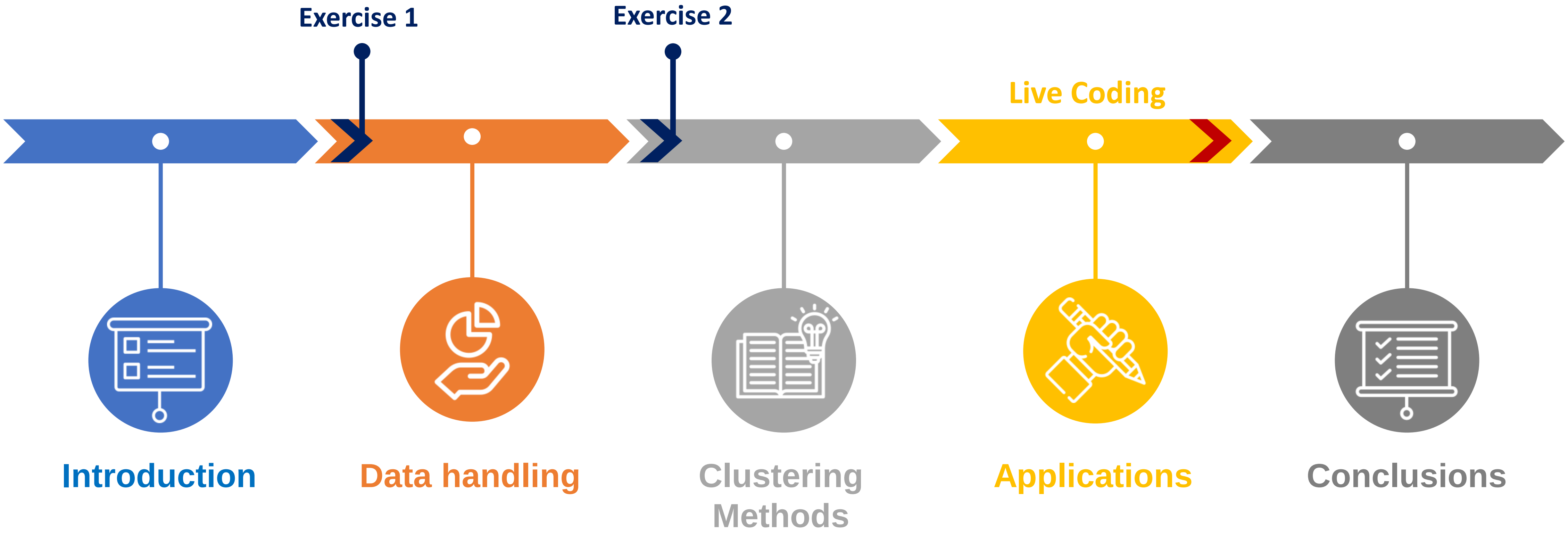
IT Architectures

- COBIT Foundation
- TOGAF

Networking

- CISCO CERTIFIED Architect
- CISCO CERTIFIED CCNA
- CISCO CERTIFIED CCNP
- tmforum CAREER CERTIFIED
- ONF CERTIFIED SDN ASSOCIATE
- ZABBIX

Overview



Exercise 1

Exercise 2

Live Coding

Introduction

Data handling

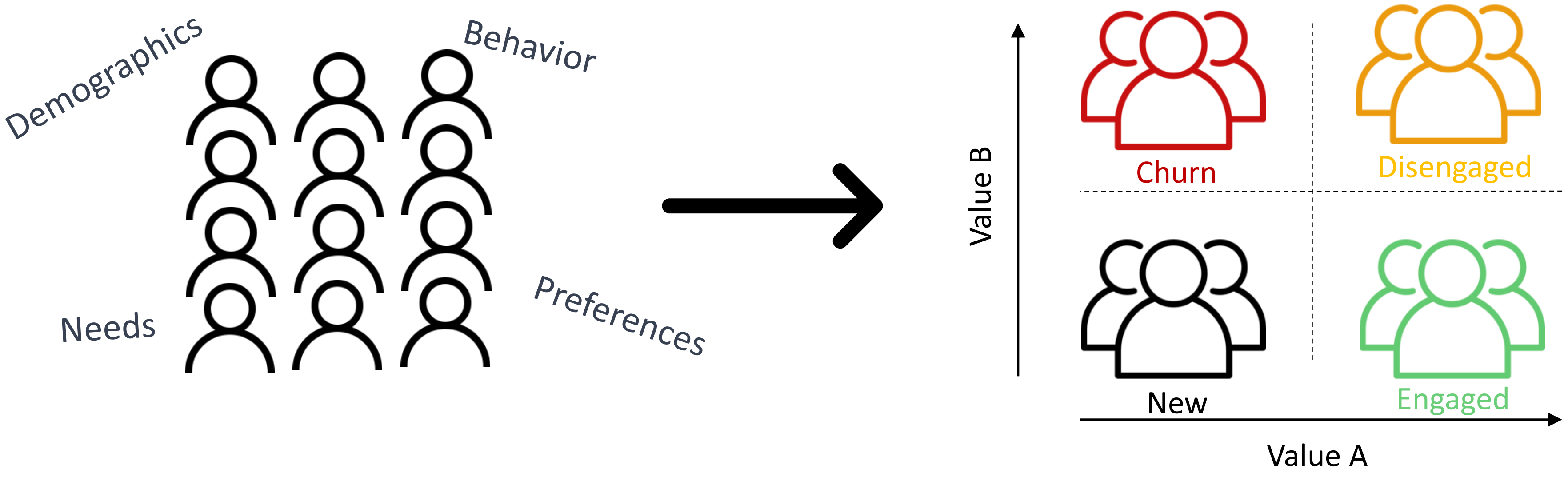
Clustering Methods

Applications

Conclusions

Customer Segmentation

Customer segmentation is a procedure **grouping similar customers together**, to better understand them.



Customer Segmentation steps

Criteria identification

Segmentation criteria depend on the problem in consideration.
Demographics, behavior, needs, and preferences are typical features adopted to segment customers.

Data collection and analysis

Data need to be collected via surveys, focus groups, customer feedback, digital interactions, etc.
Data are then analyzed to look for **clusters, trends and patterns.**

Costumer-segment creation

Businesses create customer **segments** identifying groups **with similar characteristics**.
Personas or profiles describing the typical customer in each segment are often created and adopted.

Target strategy development

Businesses develop **targeted strategies tailored to each segment**.
Personalized marketing messages, offers, products, promotions, etc. are often used.

Measurement and refinement

Effectiveness measures, employing tracking key metrics (customer engagement, retention, revenue, etc.), are **useful to modify** the segmentation and marketing **strategies if needed**.

Iterative update

Continuous and iterative approach improves results.
Steps repetition and strategy tuning to ensure business quality.

What is clustering?

Clustering is a **machine learning and data analysis** technique that **collects objects** (data points) **together into clusters**, based on **the similarity** among their characteristics or **features**.

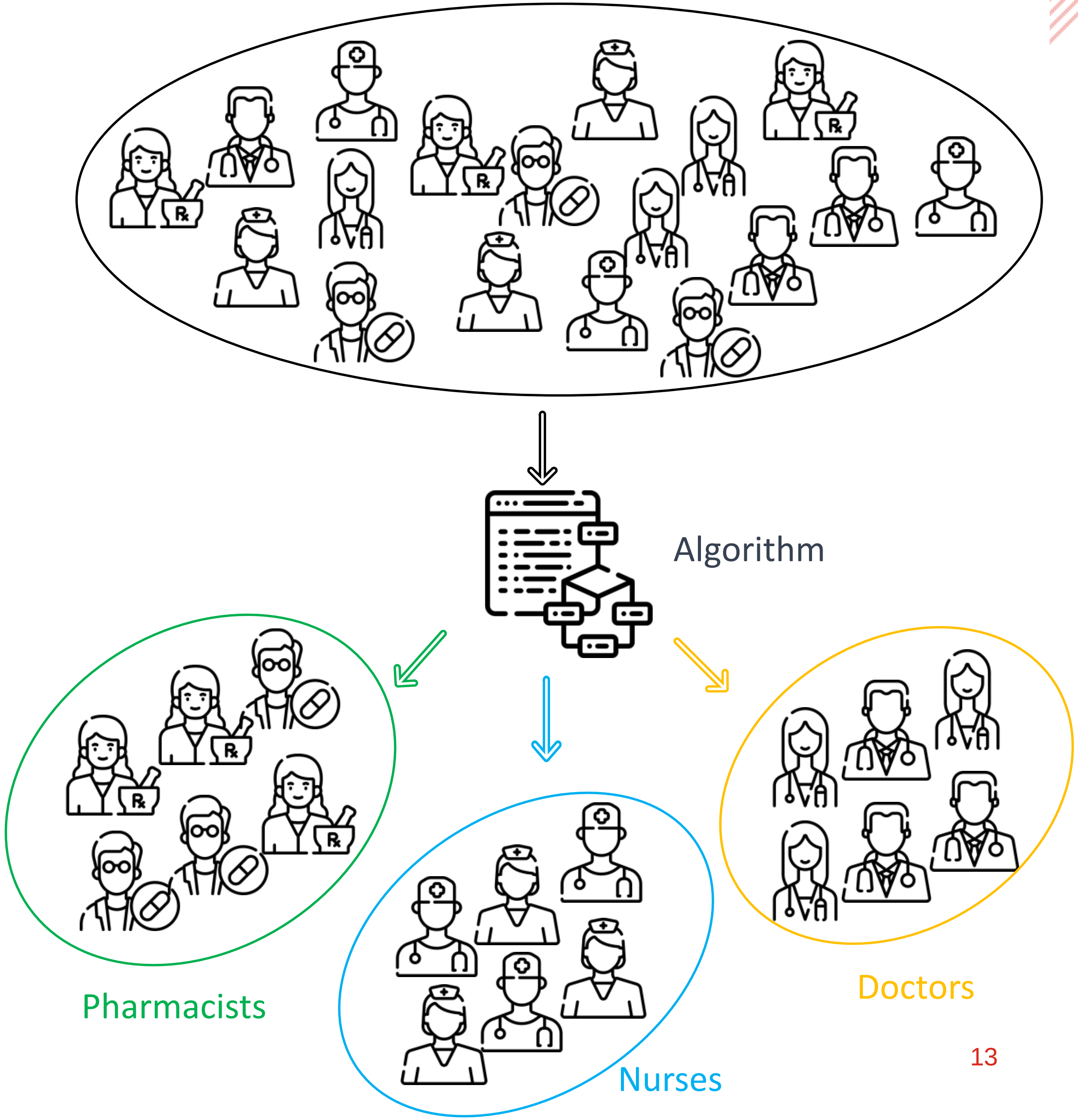
ML techniques are needed for large dimensionality domains, when the number of features is huge or when there is not a clear distinction between points.

Features represent all the available information we have of a particular object (age, gender, job,…)

Similar ≠ Equal

Given a group of doctors, nurses and pharmacists, it is natural to create 3 clusters separating the 3 given jobs.
However, if other different jobs are added, we may prefer to group together all healthcare workers.

When are two data points *similar*?



Similarity

Similarity between objects **can be defined in different ways**, depending on the need.

Given a group of doctors, nurses and pharmacists, what are the similarities?

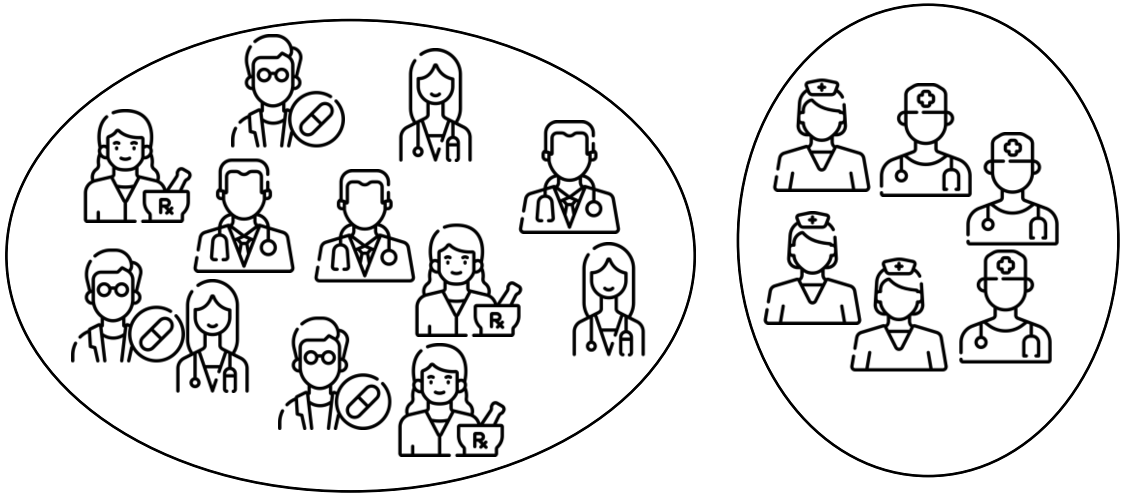
- If the interest is dress color, doctors and pharmacists are grouped together because of the common mandatory white color for their jackets, while nurses are allowed to wear jackets of different colors.

- If we are interested in hospital roles, we can say that nurses and doctors have more interactions with patients, forming a joint group, whereas pharmacists are to be treated separately since showcasing a low patient interaction.

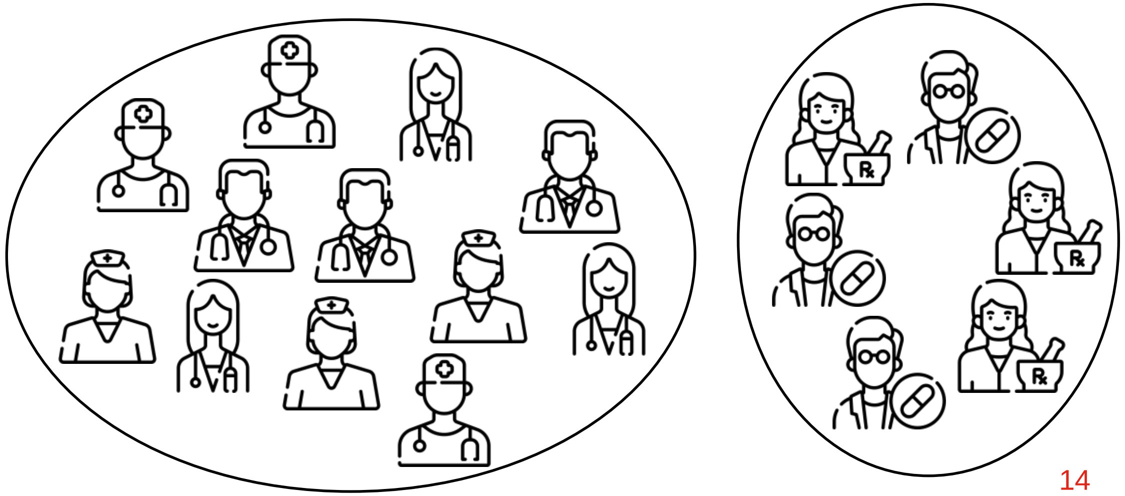
The feature of interest influences the definition of similarity.

How can we compute quantitatively similarity?

- Similarity by dress color



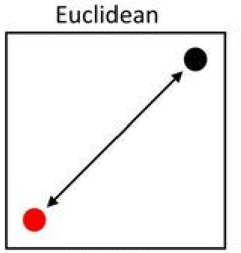
- Similarity by hospital roles



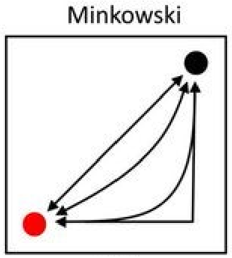
Distance as similarity measure

Distance: numerical measure describing "space" between two objects → **Smaller distance = More similarity**

Different distances can be adopted.

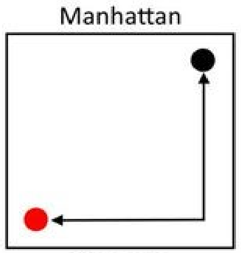


The Euclidean distance between two points in physical space is the length of a straight line between them, identifying the shortest possible connecting path.

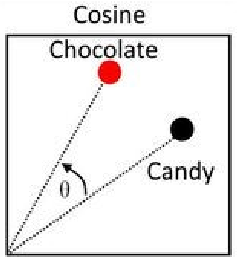


The Minkowski distance is a metric generalizing both the Euclidean distance and the Manhattan distance.

Used in Physics and e.g. General Relativity.



In a grid plan, the travel distance between street corners is given by the Manhattan distance: the number of east–west and north–south blocks one must traverse to get between those two points.



Cosine distance is the cosine of the angle between the vectors of two data points. Maximal for superposed vectors, and it decreases when vectors point in different directions.

Exercise 1

The concept of Similarity

How would you separate these 10 famous people into clusters?

- You can use only 3 clusters labeled A,B or C
- You can use whatever reasoning
- The most imaginative method wins!

Person	Cluster ?
Barack Obama	
Beyoncé	
Elon Musk	
Lionel Messi	
Bill Gates	
Leonardo DiCaprio	
Cristiano Ronaldo	
Kim Kardashian	
Tom Hanks	
Stephen Hawking	

Exercise 1

The concept of Similarity

How would you separate these 10 famous people into clusters?

- You can use only 3 clusters labeled A,B or C
- You can use whatever reasoning
- The most imaginative method wins!

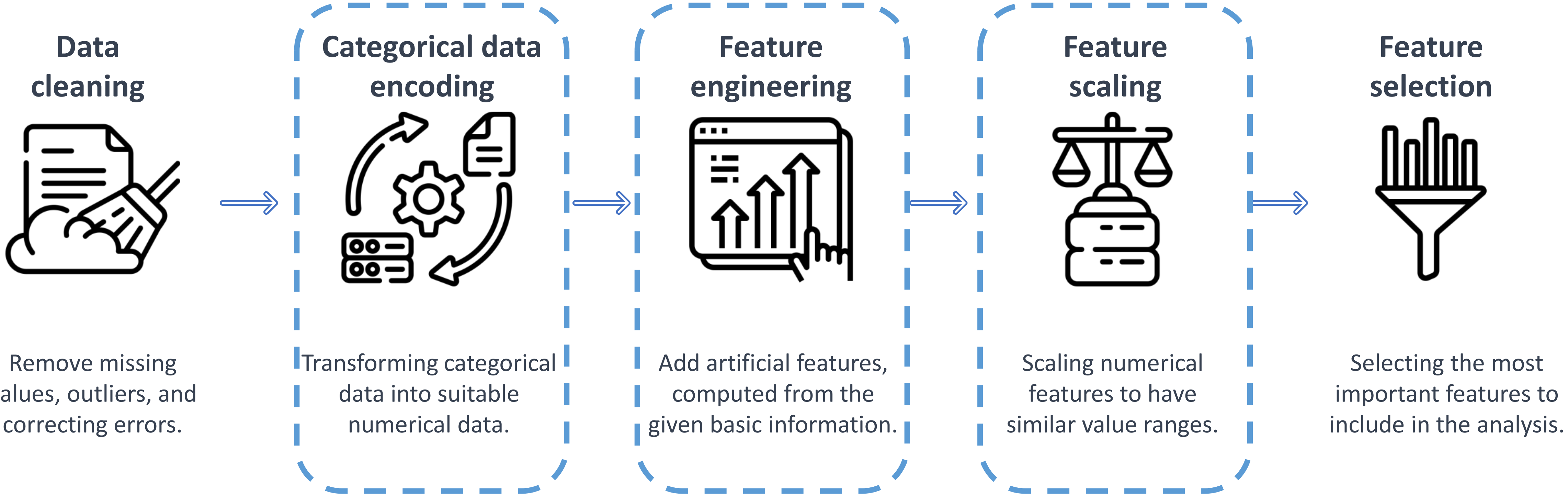
In which cluster would you add **me?**

Person	Cluster ?
Barack Obama	
Beyoncé	
Elon Musk	
Lionel Messi	
Bill Gates	
Leonardo DiCaprio	
Cristiano Ronaldo	
Kim Kardashian	
Tom Hanks	
Stephen Hawking	

Data Handling

Preprocessing steps

To compute distance and capture similarities **between data**, a **preprocessing** procedure **is needed** to properly prepare available data.



Data cleaning

Remove missing values, outliers, and correcting errors.

Categorical data encoding

Transforming categorical data into suitable numerical data.

Feature engineering

Add artificial features, computed from the given basic information.

Feature scaling

Scaling numerical features to have similar value ranges.

Feature selection

Selecting the most important features to include in the analysis.

Encoding categorical data (1/2)

Encoding (transforming): assigning one or more numeric values to a categorical feature.

- **Ordinal encoding:** assigns a numerical value to each category based on their rank or order. Useful when a **clear hierarchy** or scale in data is available.

ID	Education
1	Bachelors Degree
2	Ph. D.
3	Masters Degree
4	Bachelors Degree

→

ID	Size
1	1
2	3
3	2
4	1

- **One-hot encoding**: creates a binary (0/1) variable for each category in the original feature. Easy comparison of different categories but create **many variables**.

ID	Role
1	Doctor
2	Nurse
3	Doctor
4	Pharmacists

→

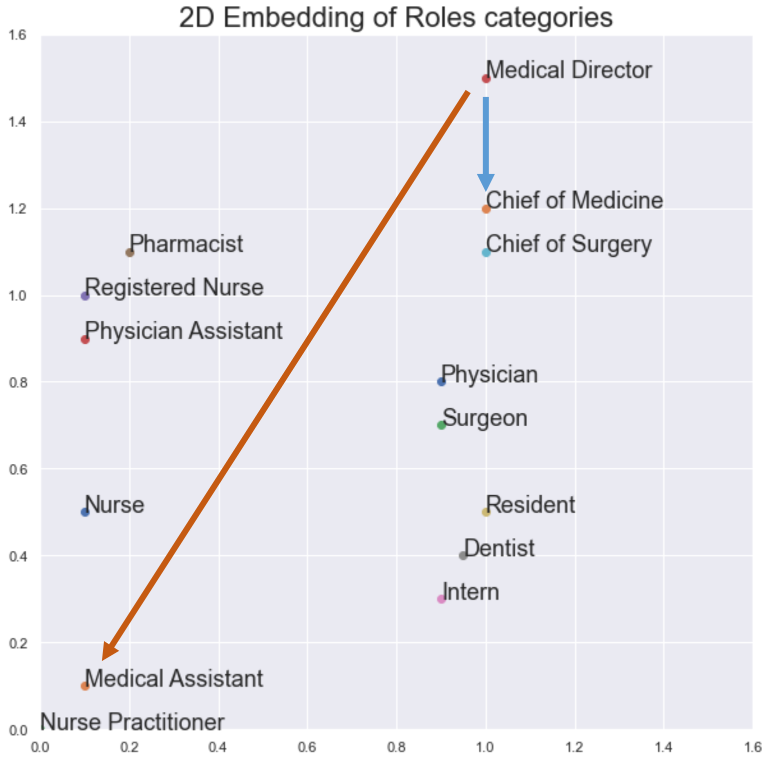
ID	is_doctor	is_nurse	is_pharmacist
1	1	0	0
2	0	1	0
3	1	0	0
4	0	0	1

Encoding categorical data (2/2)

Encoding (transforming): assigning one or more numeric values to a categorical feature.

- **Neural network encoding:** generates vectors from categorical data.
  Useful for **large number of categories** (e.g., *word2vec* recast all English dictionary words!)

ID	Role	x	y
1	Nurse	0.1	0.5
2	Surgeon	0.9	0.7
3	Resident	1	0.5
4	Medical Director	1	1.5



Feature engineering

Creation of new features:

- **add new features** derived from the original ones
- **transform existing** features to improve the usefulness

→

- **improve** a models predictive **performance**
- **reduce** computational effort and data **needs**
- **improve interpretability** of the results

Example: to **evaluate hospital performance**, a useful feature might the ratio:

$$\boldsymbol{patients\ per\ doctor} = \frac{number\ of\ patients}{number\ of\ doctors}$$

Feature engineering requires a deep understanding of the problem and domain knowledge

Feature Scaling

Transforming **numerical values** in a dataset **to a common scale** (usually between 0 and 1 or -1 and 1)

helps feature **comparability**

balances input feature **differences**

Example: using **Euclidean distance**, features with **higher magnitude** are **dominant with respect to** those with **lower magnitudes**.

Employee	Age	Salary
1	44	73000
2	27	47000
3	27	53000
4	38	62000
5	40	57000
6	35	53000
7	48	78000

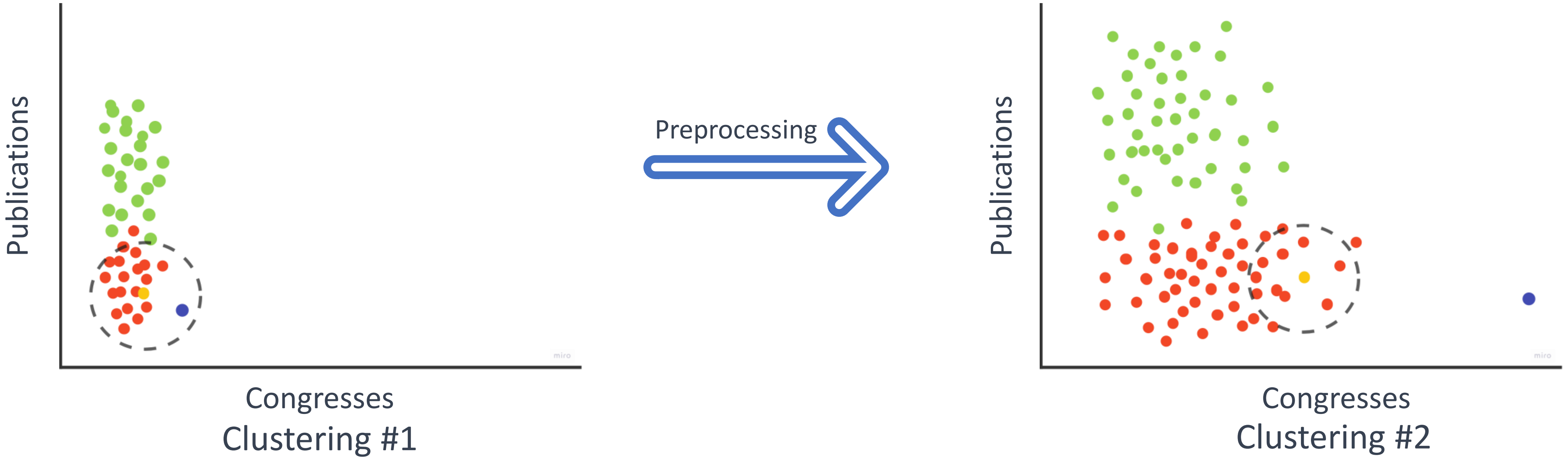
Employee	Age norm	Salary norm
1	0.809	0.838
2	0	0
3	0	0.193
4	0.523	0.483
5	0.619	0.322
6	0.380	0.193
7	1	1

Feature scaling improves performance of machine learning models **reducing** possible **biases** stemming from differences in feature scales

Preprocessing impact on Clusters

Example: cluster doctors (each point below) by their behavior towards scientific community looking publications and congresses attendances.

Note: without preprocessing data, the points yellow and blue belong to the cluster of "red" doctors and seem similar to each other. Indeed, their spatial distance in the plot is not very large. However, after the preprocessing (feature scaling), they appear further away from each other, leading to another conclusion.



Data preprocessing can change the resulting distance score (and the clustering).
It must be carefully applied based on the available data and the scope of our analysis.

Preprocessing

How would you encode the following features to one or more numbers?

- **Collaboration Attitude**:
  - POSITIVE
  - NEUTRAL
  - NEGATIVE
  - WILLING
  - UNWILLING

- **Gender**:
  - MALE
  - FEMALE

- **Sport:**
  - Basketbal_player
  - Horse_rider
  - Olympic jumper
  - Teenager football player

Exercise 2

Preprocessing

How would you encode the following features to one or more numbers?

- **Collaboration Attitude**:
    - POSITIVE = 1
    - NEUTRAL = 0
    - NEGATIVE = -1
    - WILLING = 0.5
    - UNWILLING= -0.5

- **Gender**:
    - MALE -> IS_MALE = 1 , IS_FEMALE = 0
    - FEMALE -> IS_MALE =0 , IS_FEMALE = 1

- **Sport:**
    - Basketbal_player-> (1,0)
    - Horse_rider>(0,1)
    - Olympic jumper >(1,1)
    - Teenager football player>(0,1)

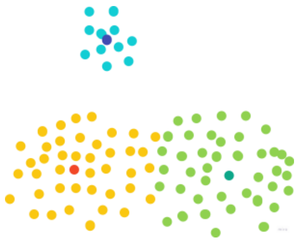
Clustering Methods

Top Clustering approaches

Different clustering methods are available and **suited for different scenarios**.

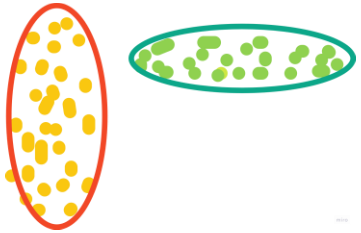
Centroid models

Points are clustered **minimizing distance from centroids** whose position are iteratively updated, until convergence.



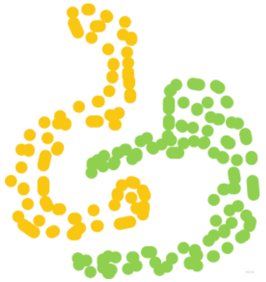
Distribution models

Data points are segmented to fit **independent distributions.**



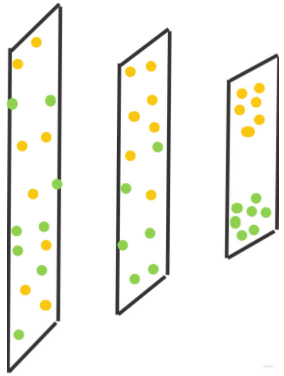
Density models

Points are clustered if are close to other points. **Clusters are high density regions**, separated by low density areas.



Neural models

Data points are embedded in a **lower dimensional space** using a non-linear transformation that preserves information.



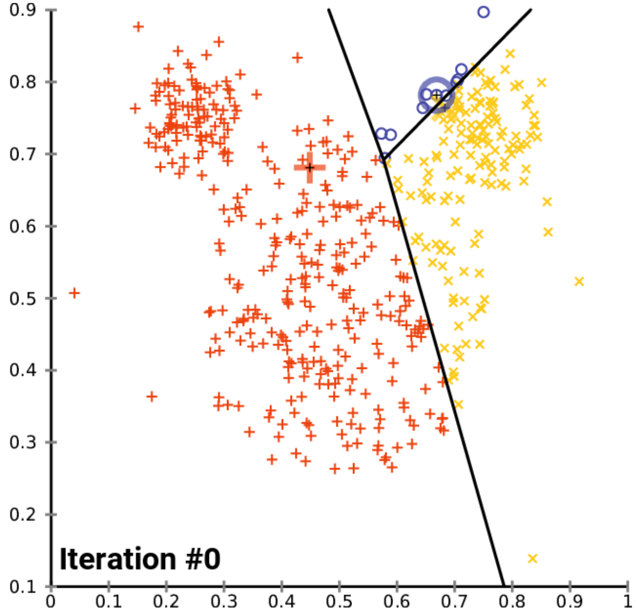
K-Means

The steps of the algorithm are:

1. Select *k* cluster centroids at random.
2. Assign each data point to the closest centroid cluster.
3. Update centroid of each cluster, using the mean of all the points in that cluster.
4. Repeat point 2 and 3 until the cluster assignments no longer changes

At the end of the algorithm, each data point is assigned to one of the k clusters based on its proximity to the centroids.

Pros	Cons
Relatively simple to implement	Choice of *k* in advance
Powerful tool for EDA	Dependent on the initial values of centroids
Identification of pattern in large data set	Clustering outliers
Computationally efficient	Cluster must be spherical and have equal variance



DBSCAN (Density-Based Clustering)

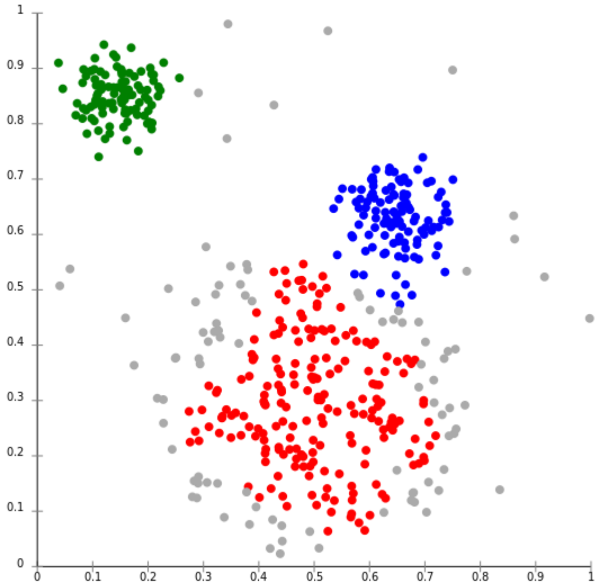
Clusters are based only on the **density** of the data points.

The steps of the algorithm are:

1. Randomly select a data point.
2. If the neighborhood contains a minimum number of points, the data point is labeled a **core point**, and all other points within the radius are assigned to the same cluster. Otherwise, the point is considered **noise**.
3. Repeat until all the data points are considered.

Two points will belong to the same cluster if it is possible to **connect them passing through high-density regions**.

Pros	Cons
Handles clusters with arbitrary shapes and sizes	Sensitive to the choice of initial parameters
Identifies outliers and noise	Not memory efficient
No need to specify the number of clusters	Does not provide centroid information



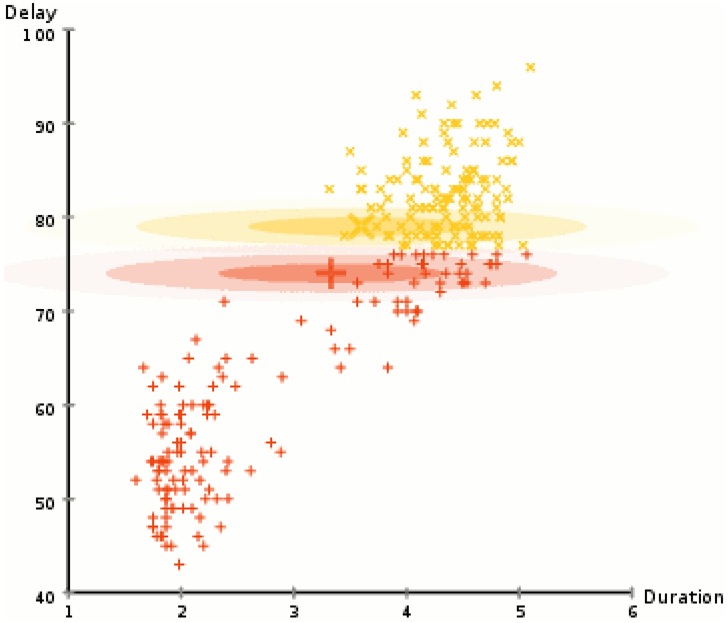
https://commons.wikimedia.org/wiki/File:DBSCAN-Gaussian-data.svg

Distribution Models

Distribution-based clustering, also known as Gaussian Mixture Model (GMM) clustering, is another popular unsupervised learning algorithm used for clustering data points.

Unlike K-means and DBSCAN, which assign data points to a single cluster, GMM clustering works by modeling the data as a combination of several Gaussian distributions. Each Gaussian distribution represents a cluster, and the algorithm estimates the parameters of the distributions to fit the data, together with the probability that a data point belongs to each of the k clusters.

Pros	Cons
Captures complex data structure	Computationally intensive
Allows for overlapping clusters	Real data may not be well described by Gaussian distribution
Allows for data sampling	



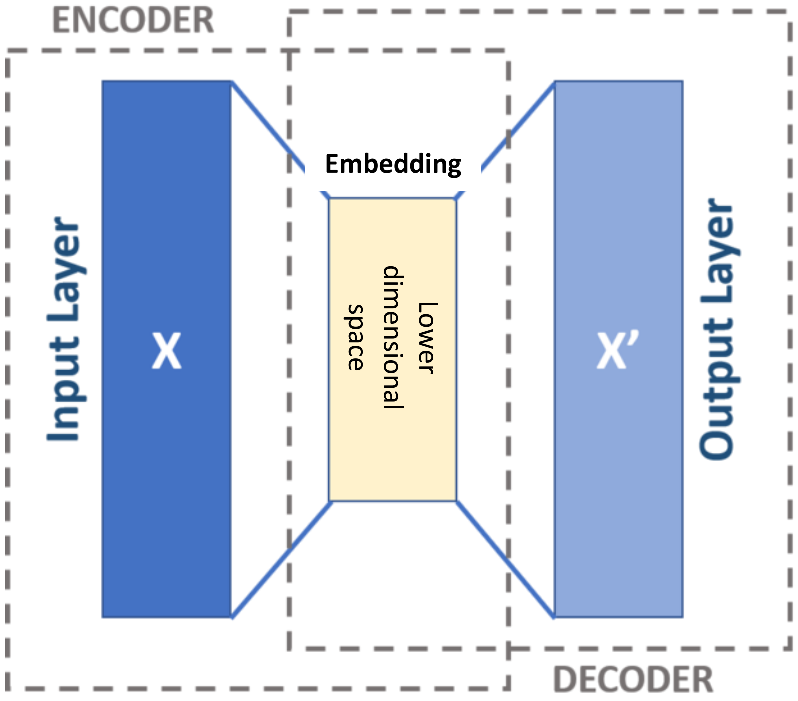
https://commons.wikimedia.org/wiki/File:EM_Clustering_of_Old_Faithful_data.gif

Autoencoders

Autoencoders are neural networks that learn to **encode data into a lower-dimensional representation**, and then decode it back into its original form. In clustering, this means that an autoencoder can learn to group similar data points together by representing them with similar encoded vectors. The process of training an autoencoder involves minimizing the difference between the original data and its decoded representation, which encourages the network to learn a compressed representation that captures the most important features of the data.

Usually, **a standard clustering method** like k-Means **is used** to find centroids in the encoded low-dimensional space.

Pros	Cons
Handles high-dimensional data	Computationally expensive
Learns non-linear relationships	More difficult to train



Recap

Recap comparison

Algorithm	Pros	Cons
K-Means	Relatively simple to implement	Choice of *k* in advance
	Powerful tool for EDA	Dependent on the initial values of centroids
	Identification of pattern in large data set	Clustering outliers
	Computationally efficient	Cluster must be spherical and have equal variance
DBSCAN	Handles clusters with arbitrary shapes and sizes	Sensitive to the choice of initial parameters
	Identifies outliers and noise	Not memory efficient
GMM	Captures complex data structure	Computationally intensive
	Allows for overlapping clusters	Estimation of the parameters of Gaussian distribution
Autoencoders	Handles high-dimensional data	Computationally expensive
	Learns non-linear relationships	More difficult to train

Applications

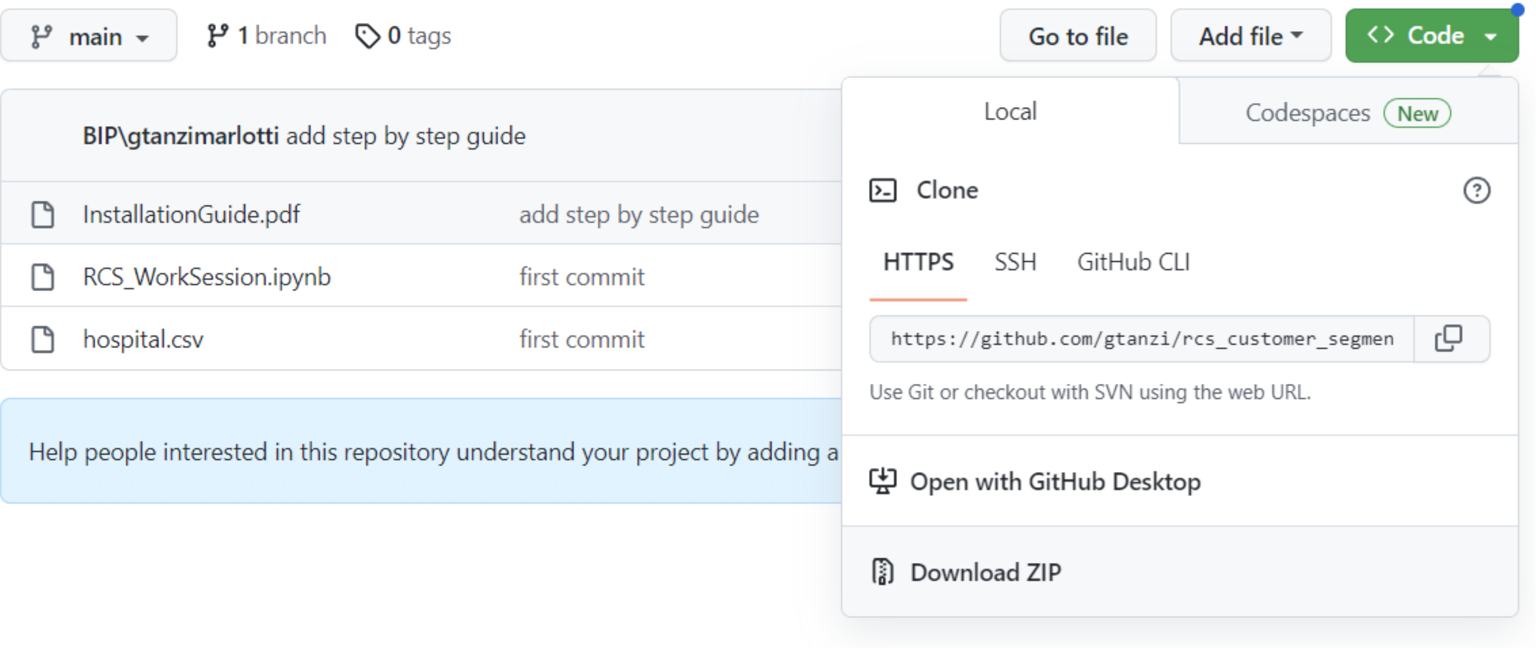
Use Case: Customer Segmentation using K-Means

We will analyze a dataset containing information about people working in hospitals such as:

Age,
Gender,
Education,
Role ,
Average Visit Duration,
Prescription Attitude,
Publications Number,
Congresses Attended,
Partnerships with Pharmaceutical Companies,
Collaboration Attitude with Pharmaceutical Companies,
Average Opened Email,
Visits to the company landing page during last year
Access to Online Services during last year,
Attended Events during last year ,
Network Usage,
ClickthroughRate,
OpenRate,
Engagement Driver

Use Case: Customer Segmentation using Clustering

https://github.com/gtanzi/rcs_customer_segmentation



main ▾ | 1 branch | 0 tags | Go to file | Add file ▾ | <> Code ▾

BIP\gtanzimarlotti add step by step guide

InstallationGuide.pdf	add step by step guide
RCS_WorkSession.ipynb	first commit
hospital.csv	first commit

Help people interested in this repository understand your project by adding a

Local | Codespaces New

Clone

HTTPS | SSH | GitHub CLI

```
https://github.com/gtanzi/rcs_customer_segmen
```

Use Git or checkout with SVN using the web URL.

Open with GitHub Desktop

Download ZIP

Conclusions

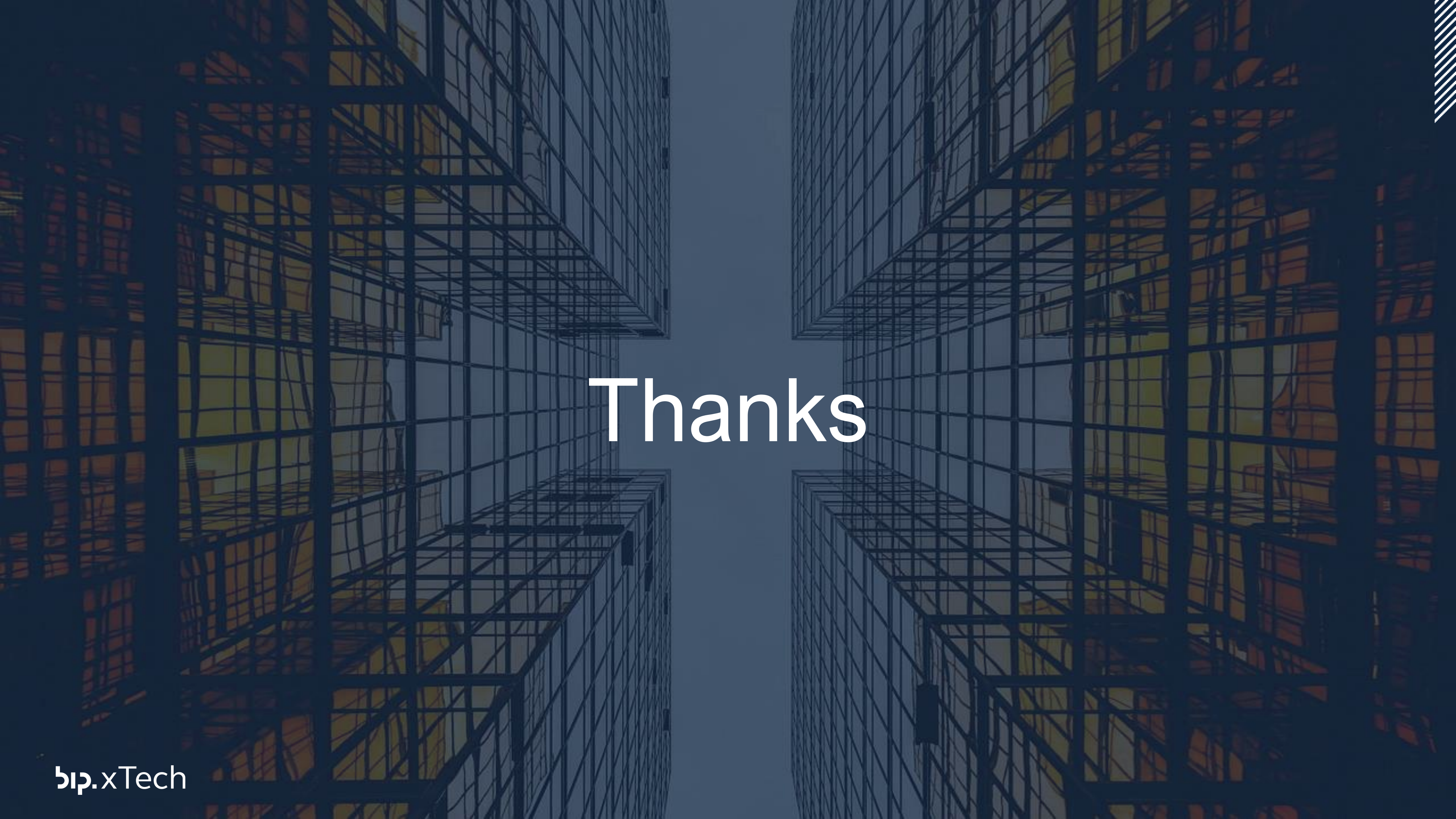
Clustering methods are **powerful tools** for customer segmentation

There are **various clustering methods** available, each with its own advantages and limitations.

The choice of clustering method depends on the data and the problem at hand, as well as the specific business objectives.

Preprocessing and feature engineering are important steps in clustering analysis to ensure the quality and relevance of the input data.

Visualization and interpretation of the clustering results are essential to gain insights and **inform business decisions**.



Thanks