



Arab Region Employees Salary Prediction

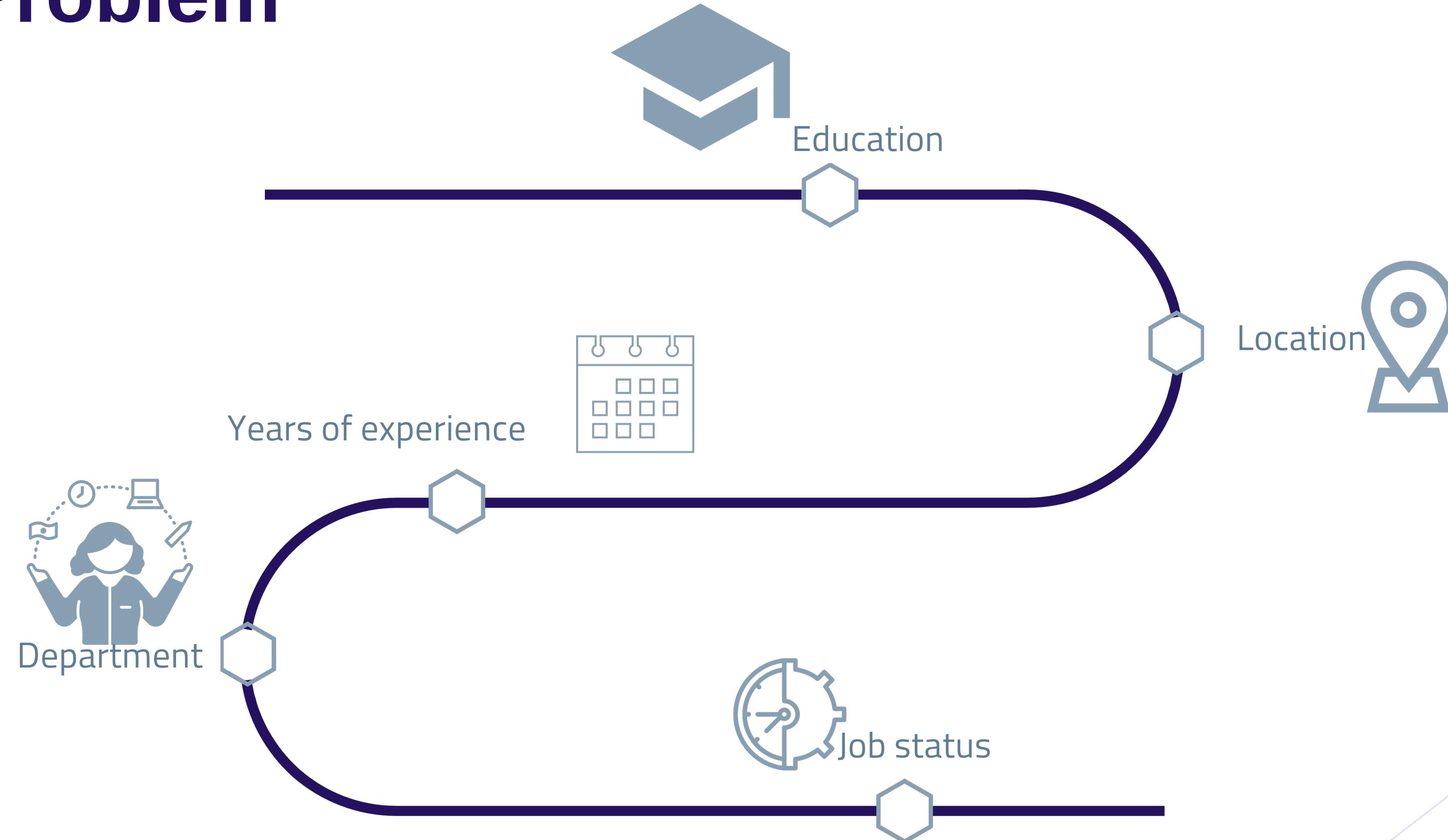
IT326 - 1st semester
Dr. Hanan Altamimi

Wissayif AlKherb | 443200459
Dimah Alsuwailem | 443200826
Jumanah Almuhsin | 443200866
Albandry Albakr | 443201085

Outline

01	Problem
02	Data Summarization: Data Information, Graphs
03	Data Preprocessing: Data Cleaning, Data Transformation, Feature selection
04	Data Mining Techniques: Classification and findings , Clustering and findings
05	Results

Problem



Data Summarization

- Data Information

our attributes are:

ID

Education

Department

Job status

Location

Start date

Years

Salary

Job rate

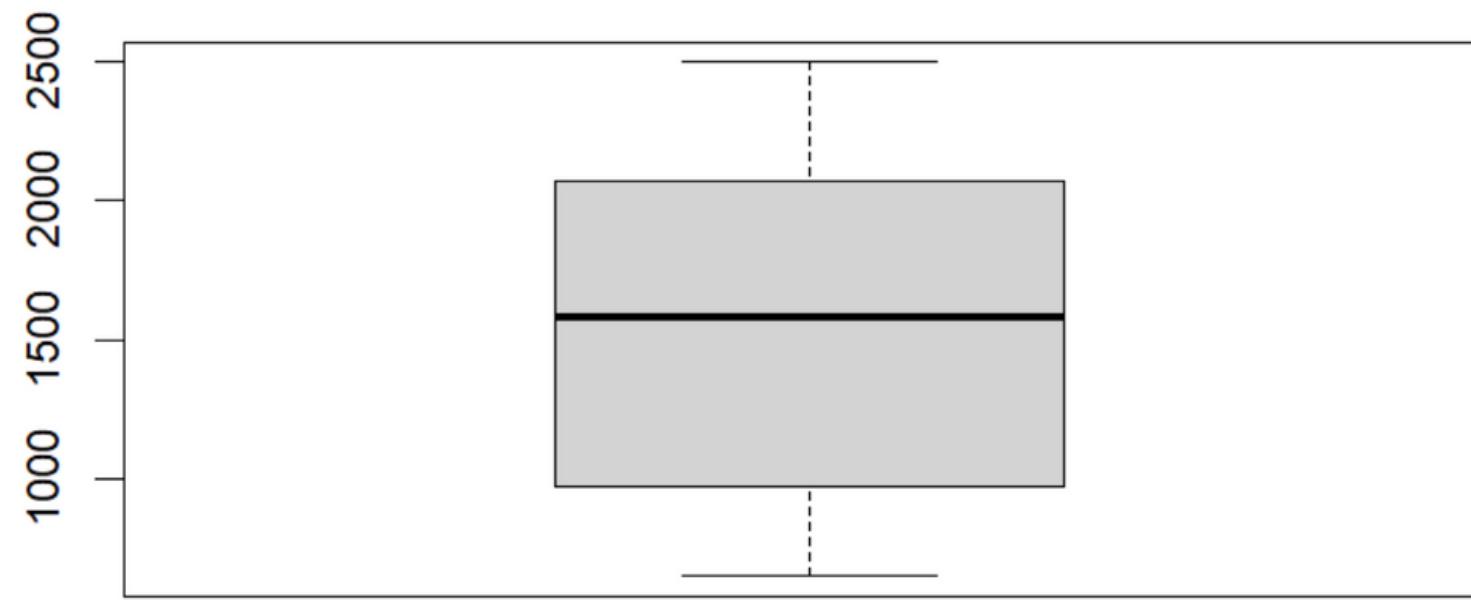
Permission

link of the dataset: <https://www.kaggle.com/datasets/qusaybtoush1990/employees>

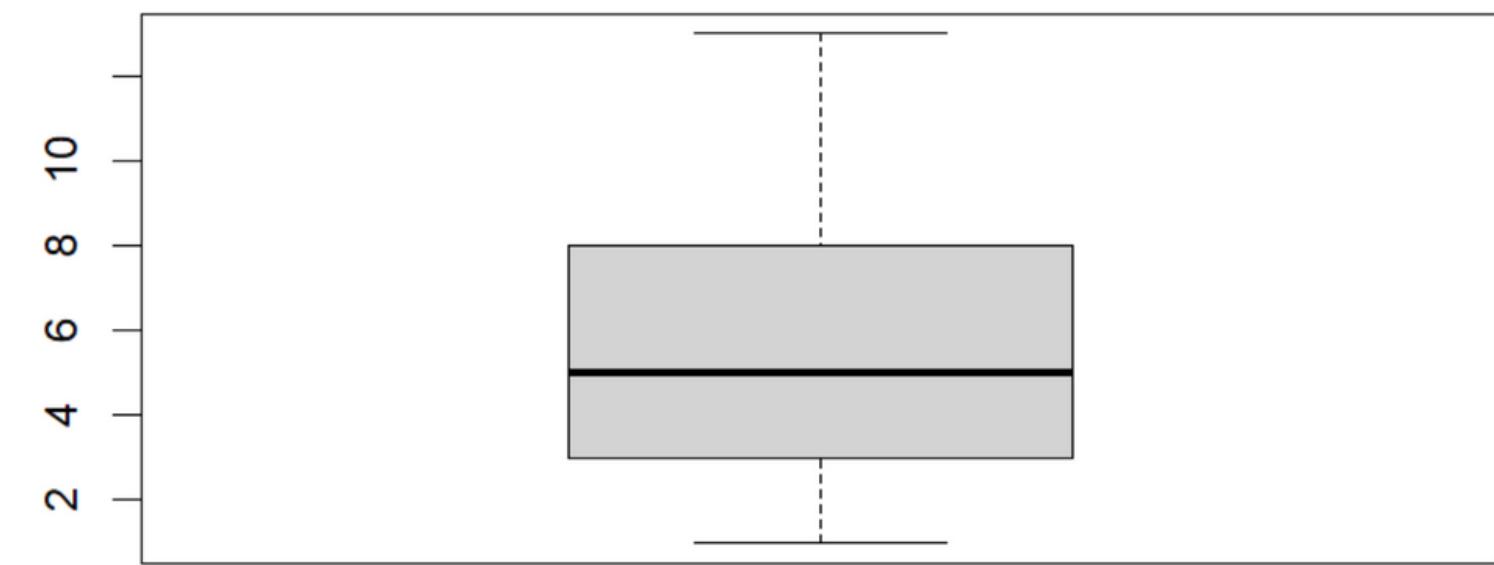
Data Summarization

- Data Graphs

Box plot for salary



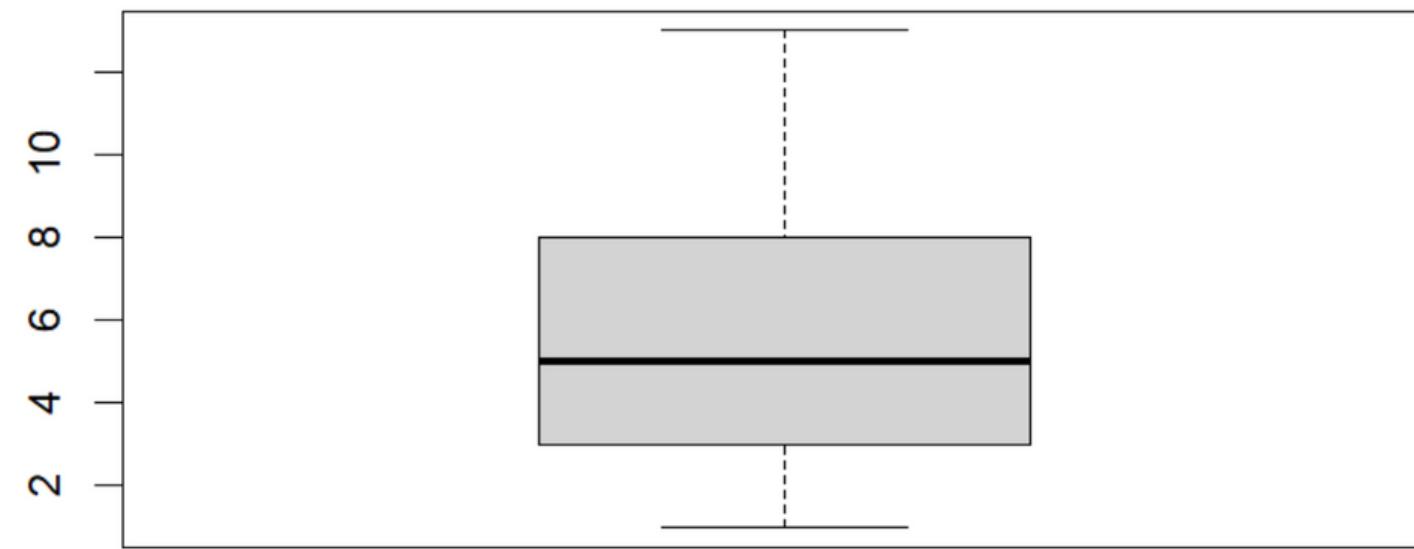
Box plot for years



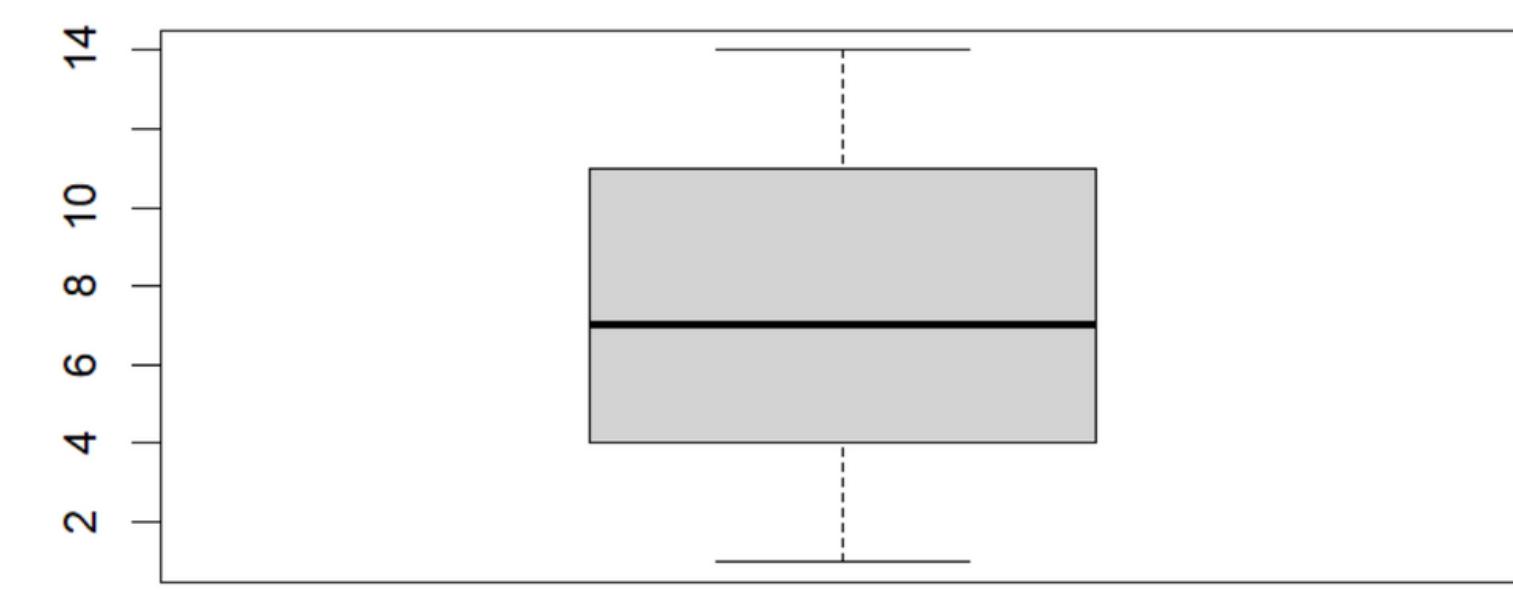
Data Summarization

- Data Graphs

Box plot for job rate



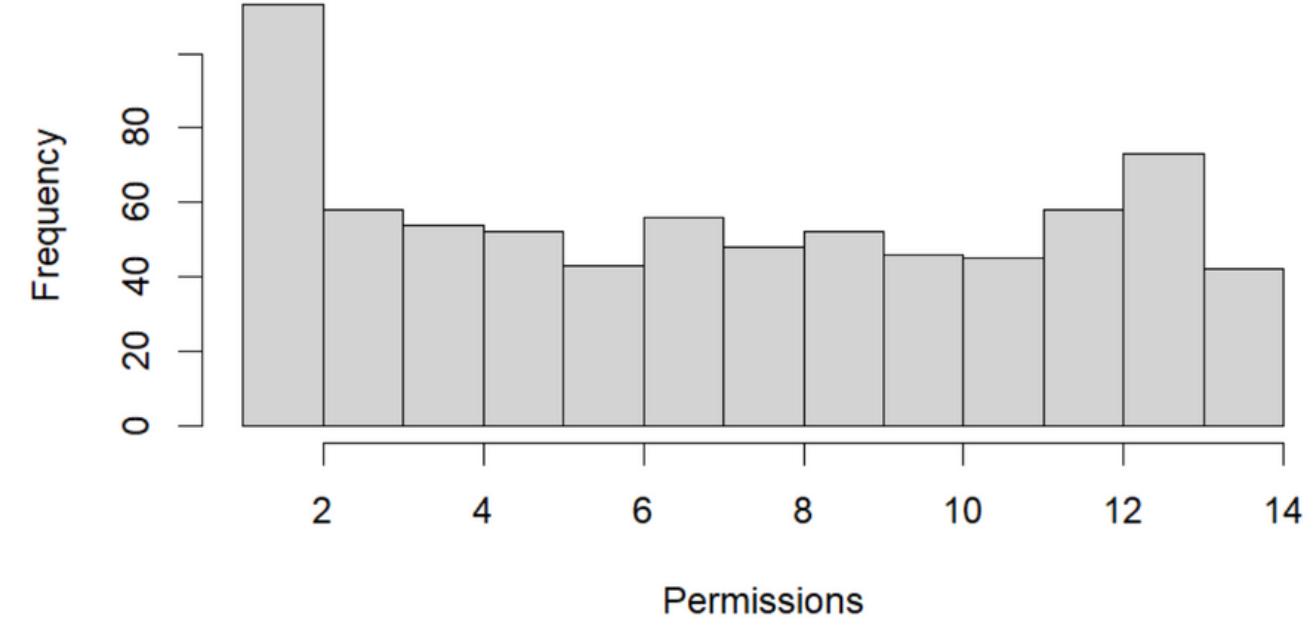
Box plot for permissions



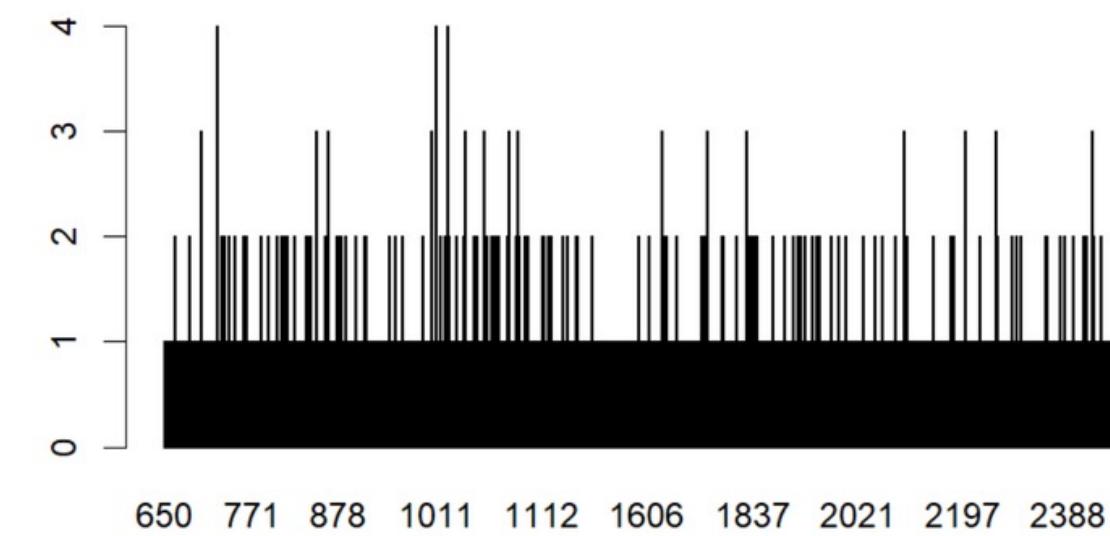
Data Summarization

- Data Graphs

Histogram of permissions



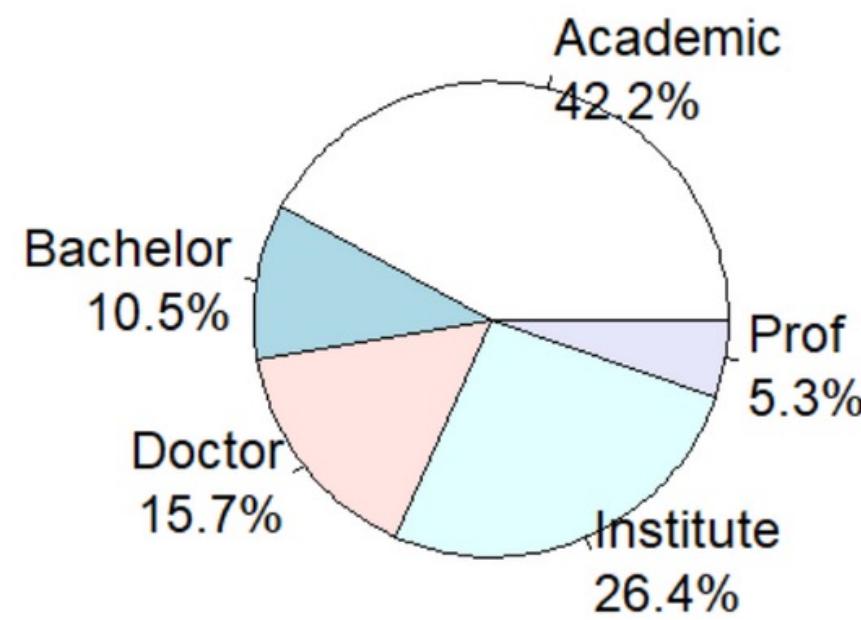
Bar Chart for salary



Data Summarization

- Data Graphs

Pie chart for Education



Data Preprocessing

We used three Preprocessing techniques:

- Data cleaning
- Data transformation
- Feature selection



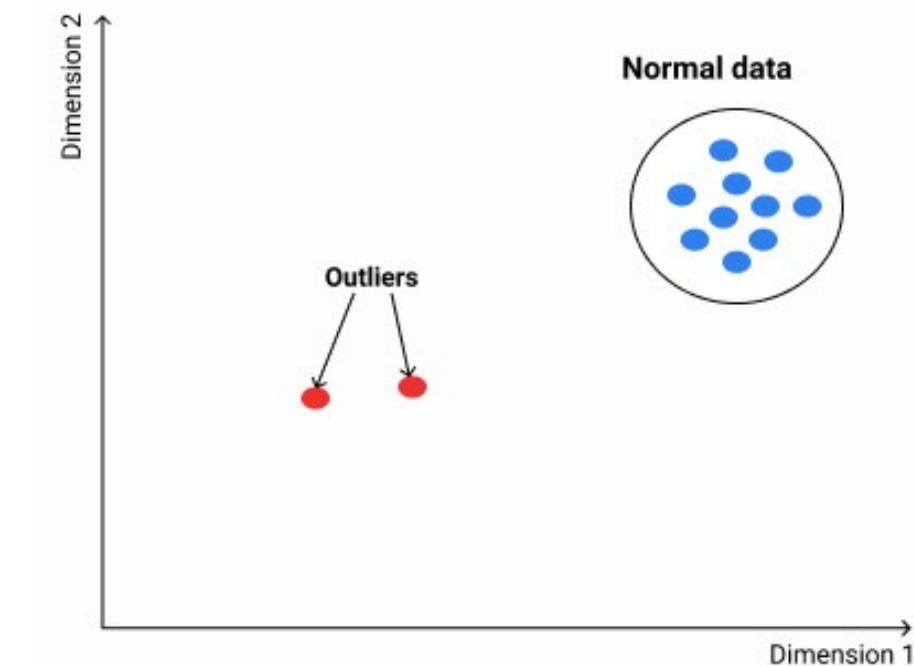
Data Preprocessing

- Data cleaning

missing or null
values

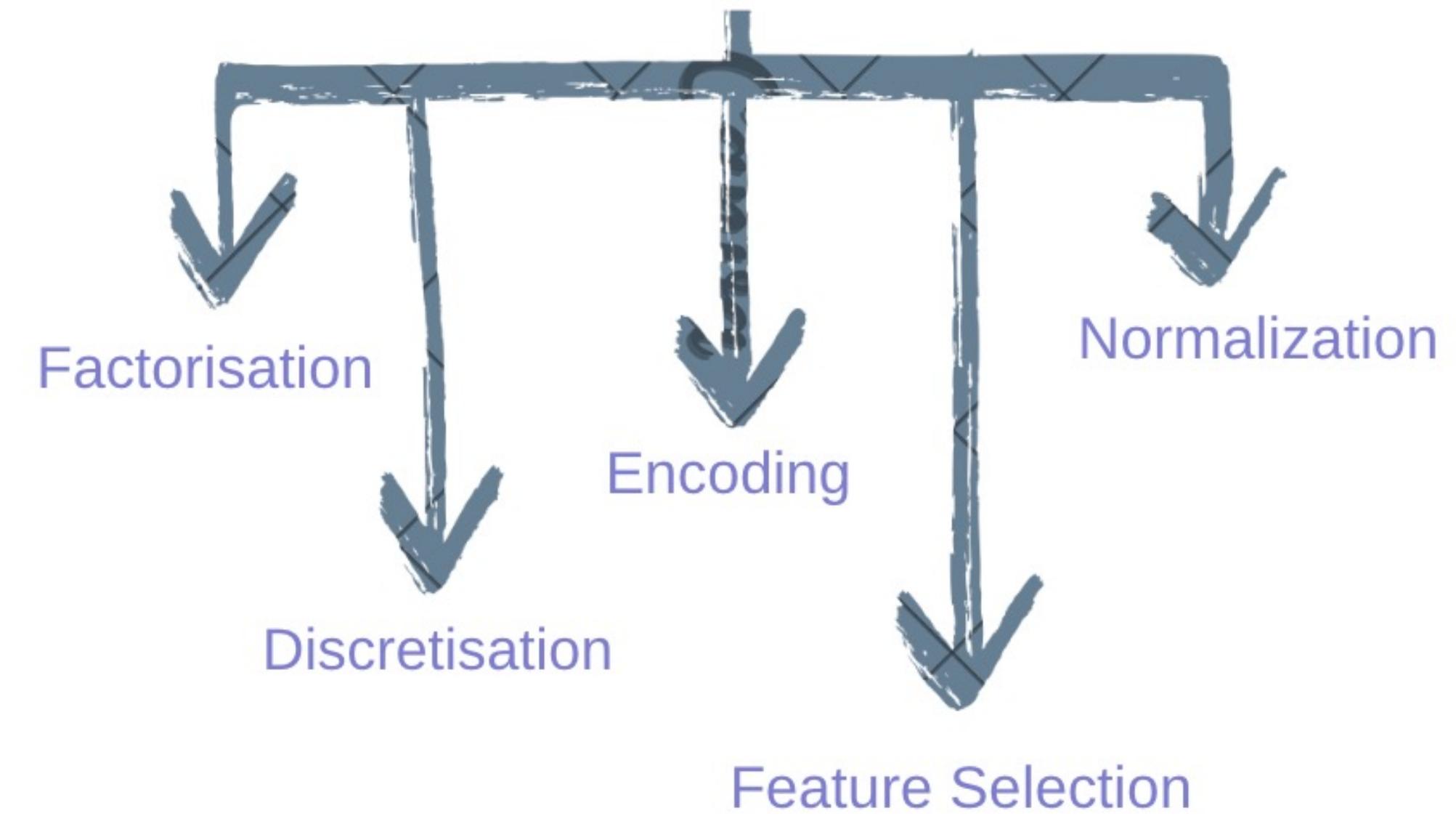


outliers



Data Preprocessing

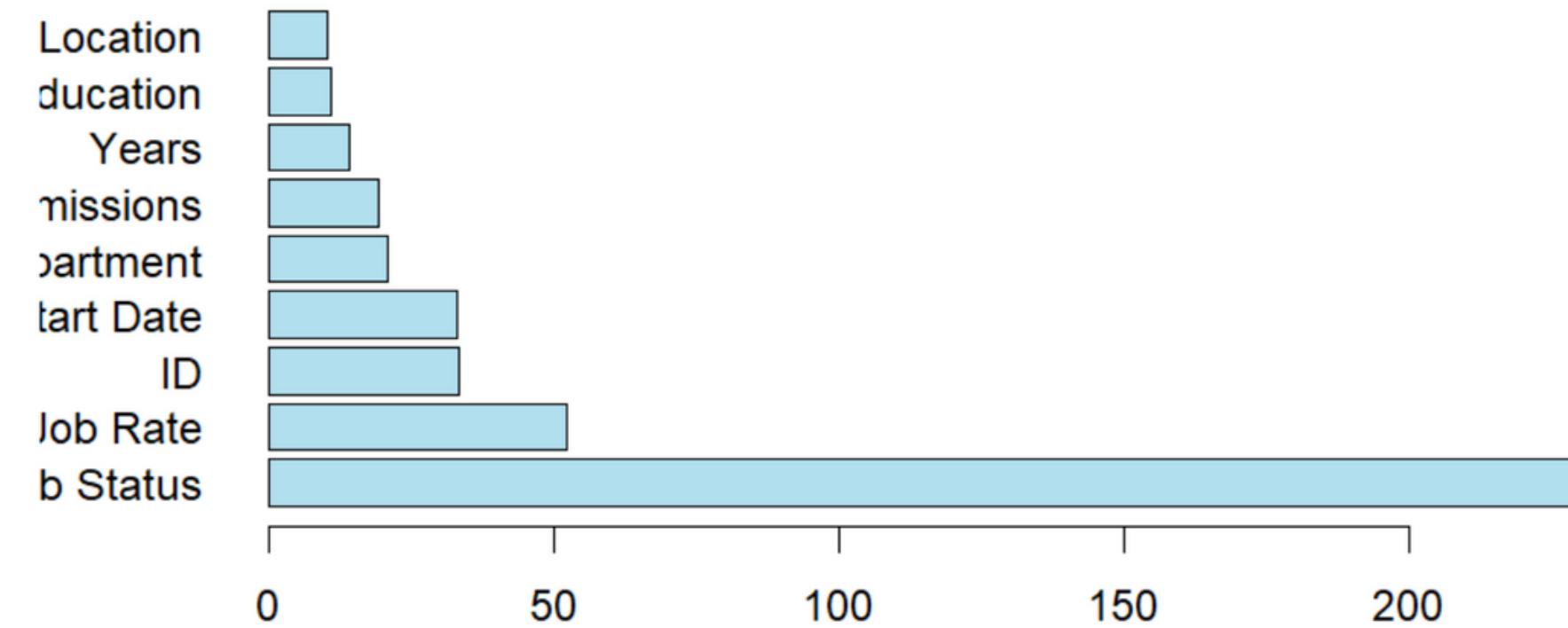
Data transformation



Data Preprocessing

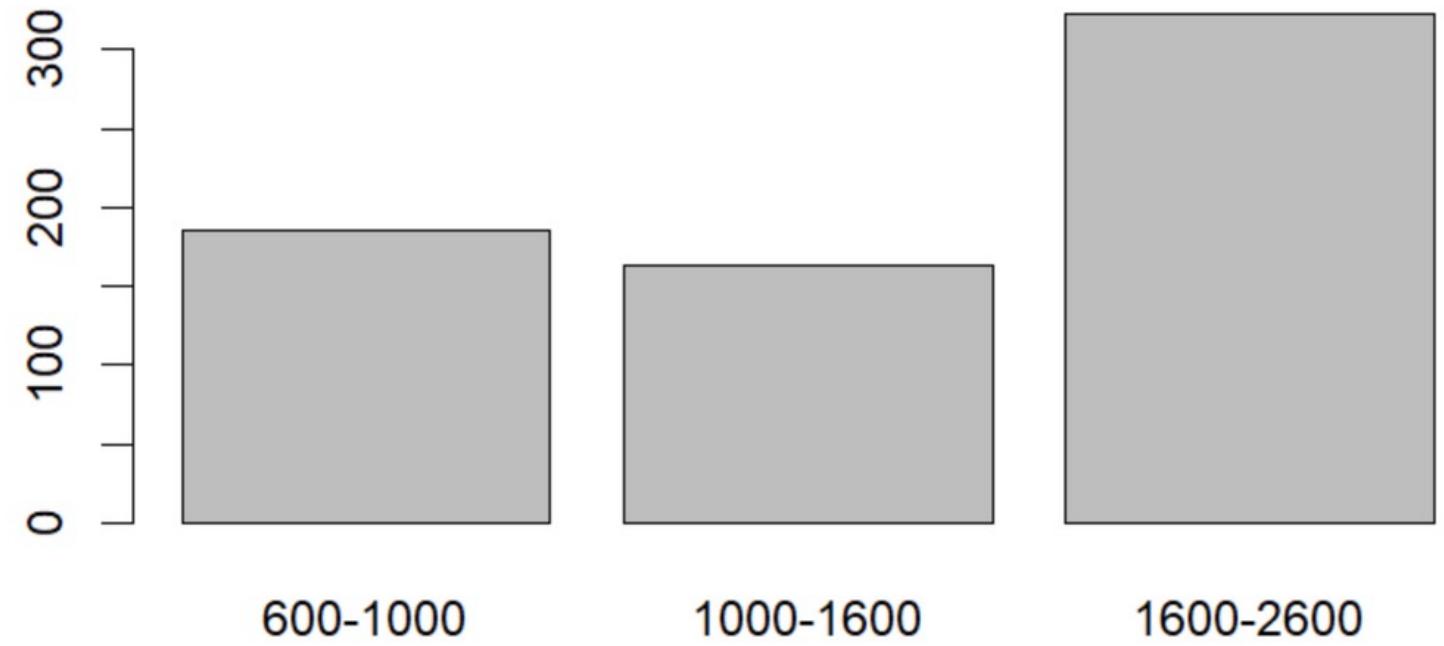
- Feature selection

assist in showing the most relevant features, leading to simpler and more interpretable models.

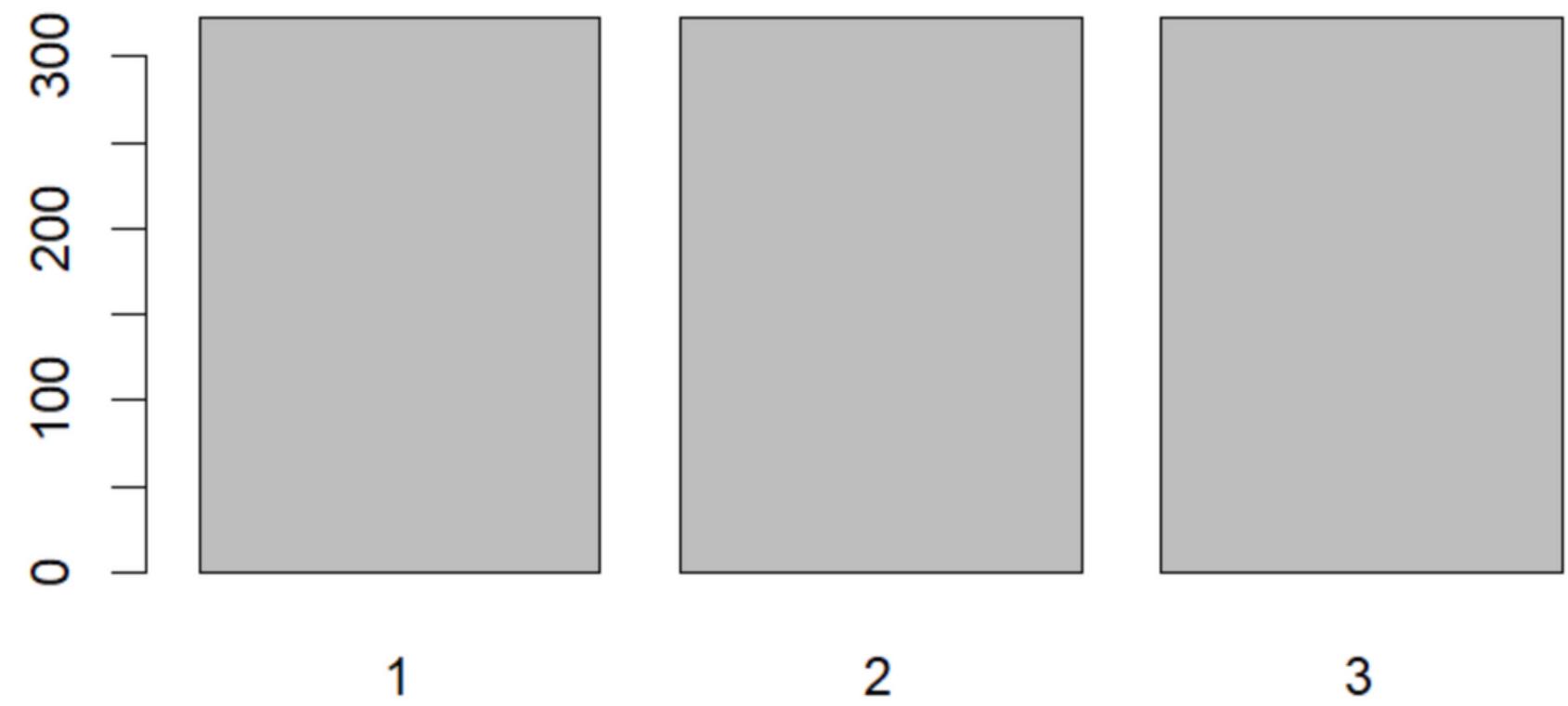


Imbalance

Before



After



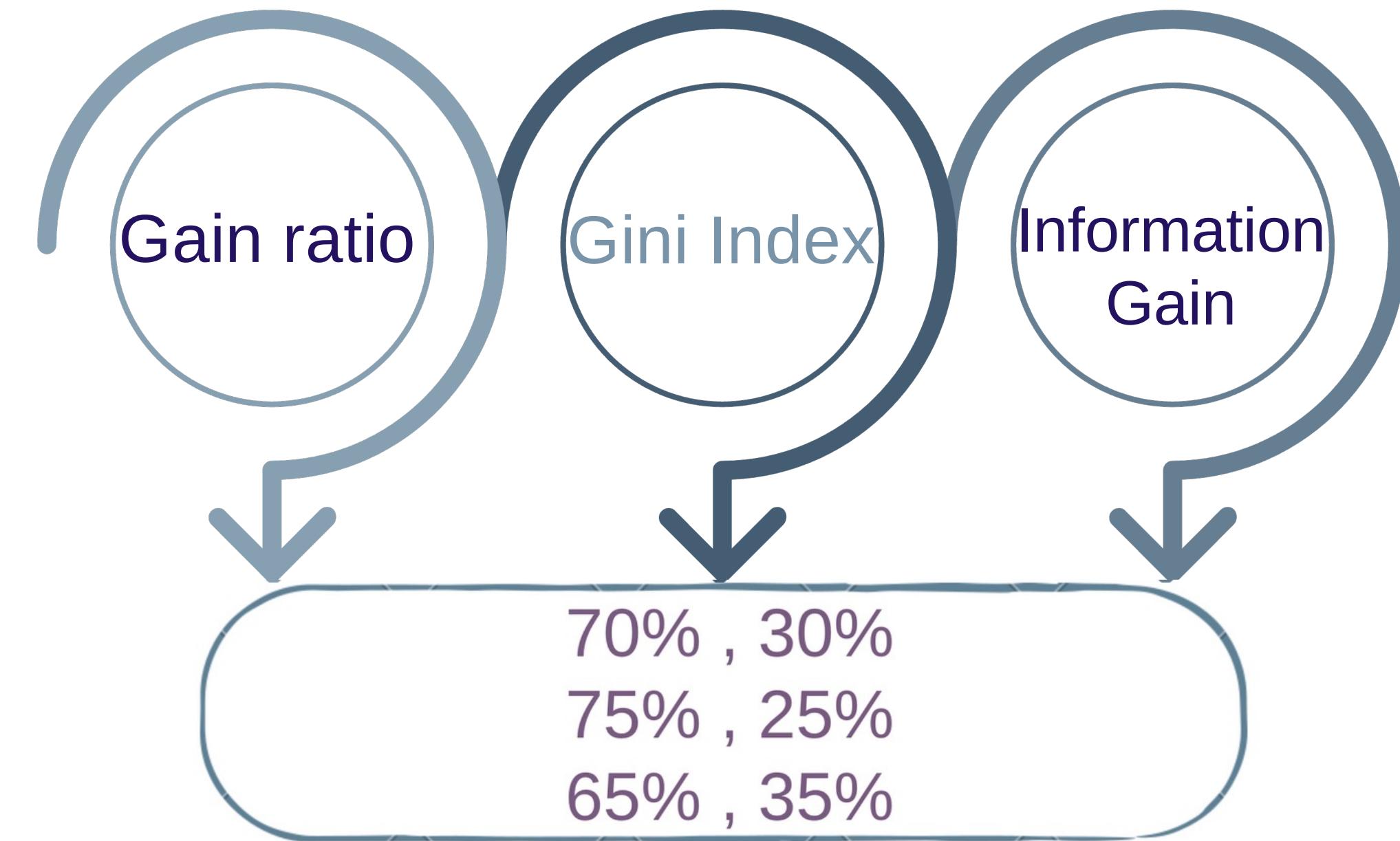
Data Mining Techniques

Classification

Clustering

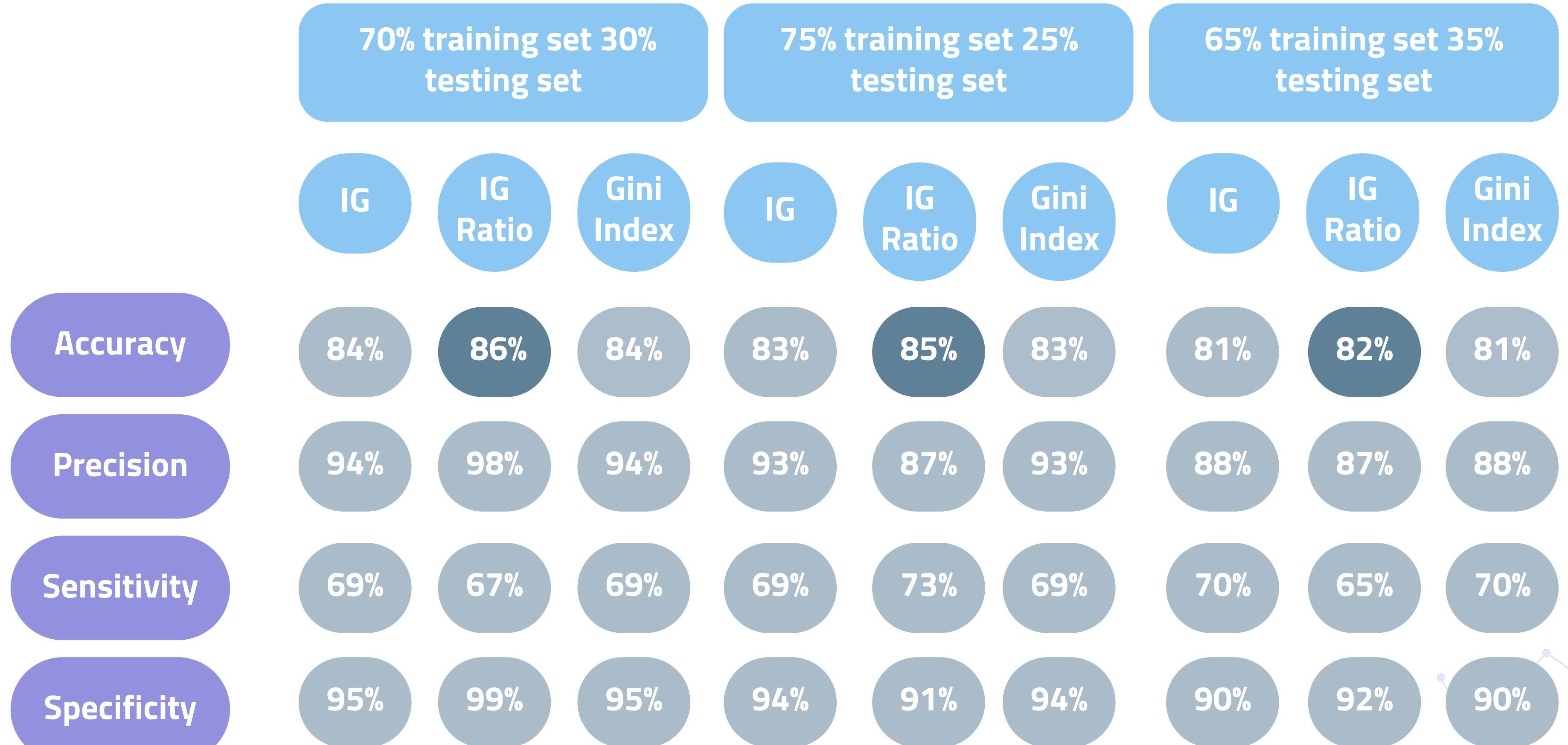
Classification

Algorithms:



Findings

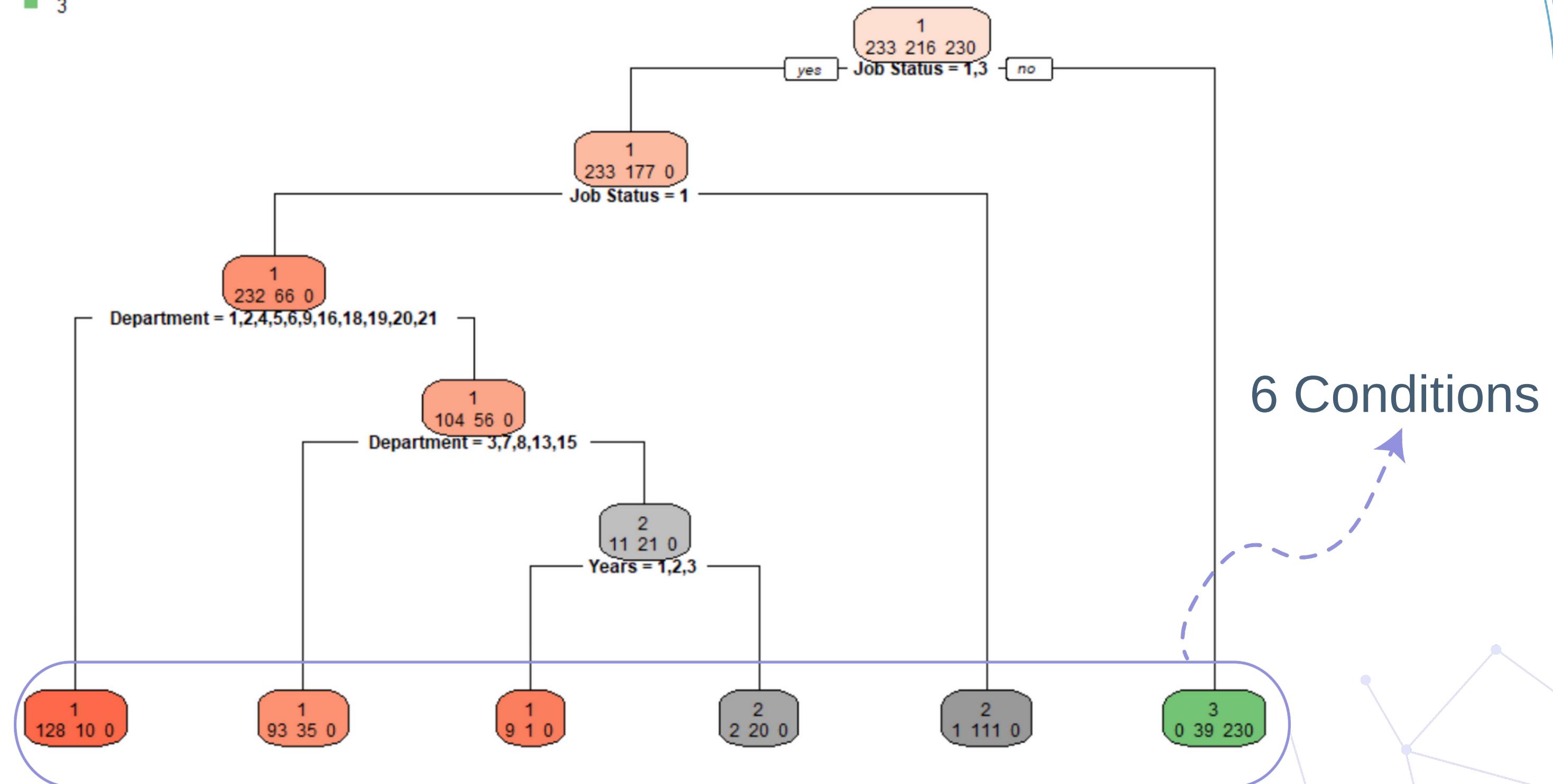
Classification:



Findings

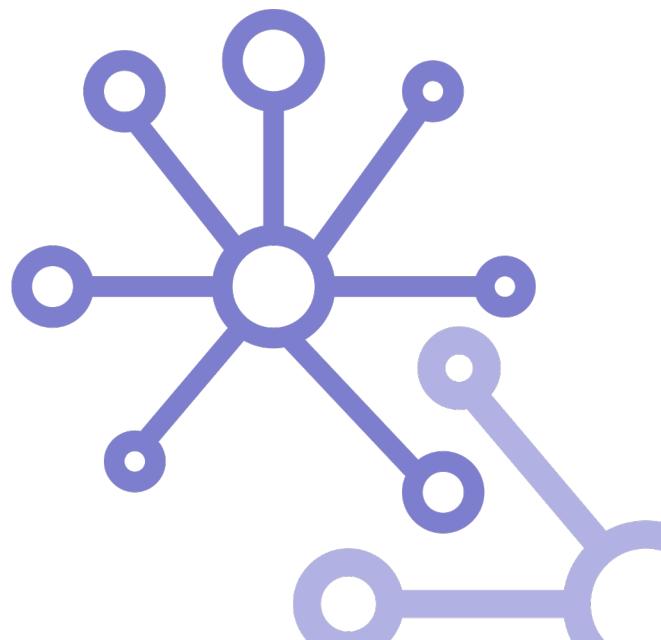
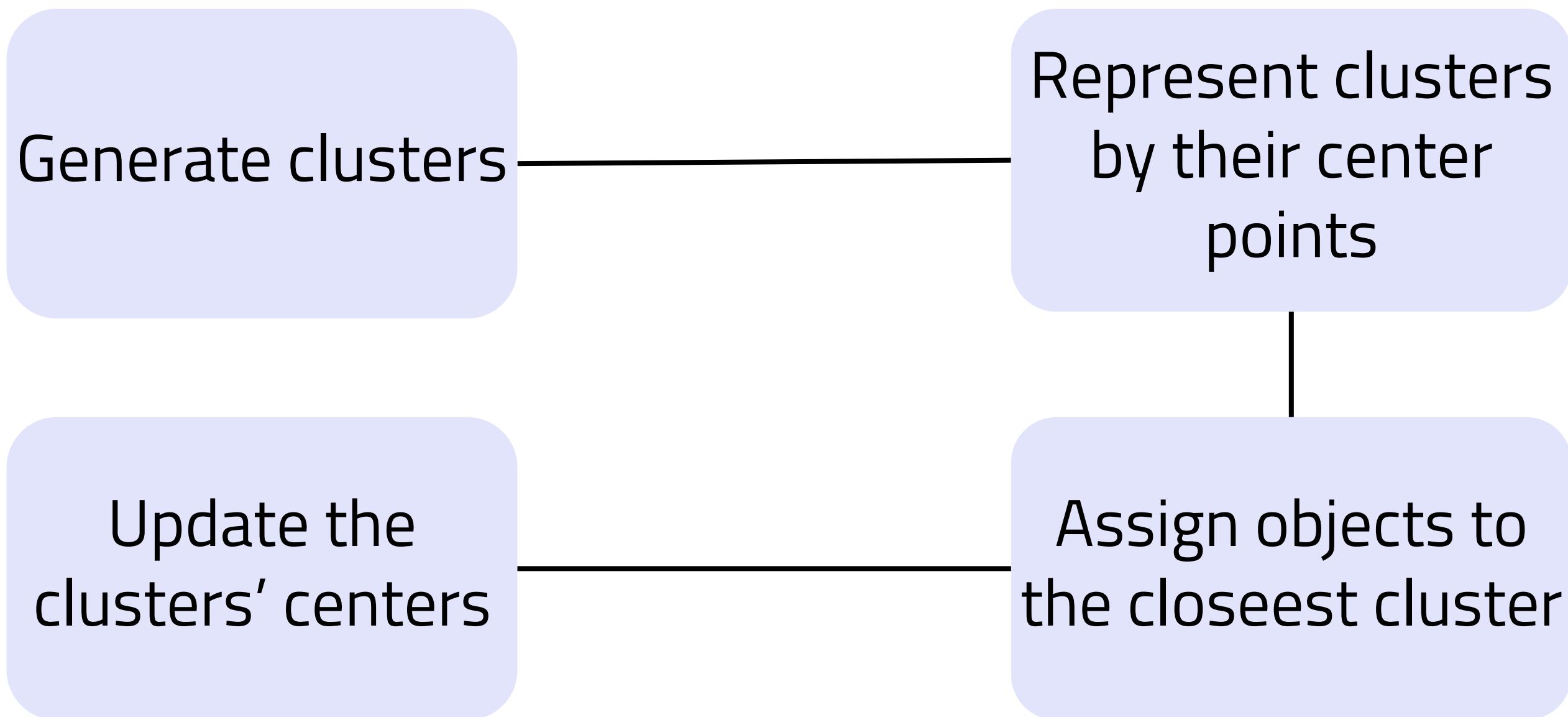
Classification:

■ 1
■ 2
■ 3



Clustering

How K-means algorithm is done?



Clustering

Evaluation and Comparison

	K=2	K=3	K=4
Average Silhouette width	0.3	0.25	0.22
total within-cluster sum of square	40750.1	34529.21	29822.08
BCubed precision	0.3349891	0.3352973	0.3362805
BCubed recall	0.512721	0.3487142	0.2555154

Clustering

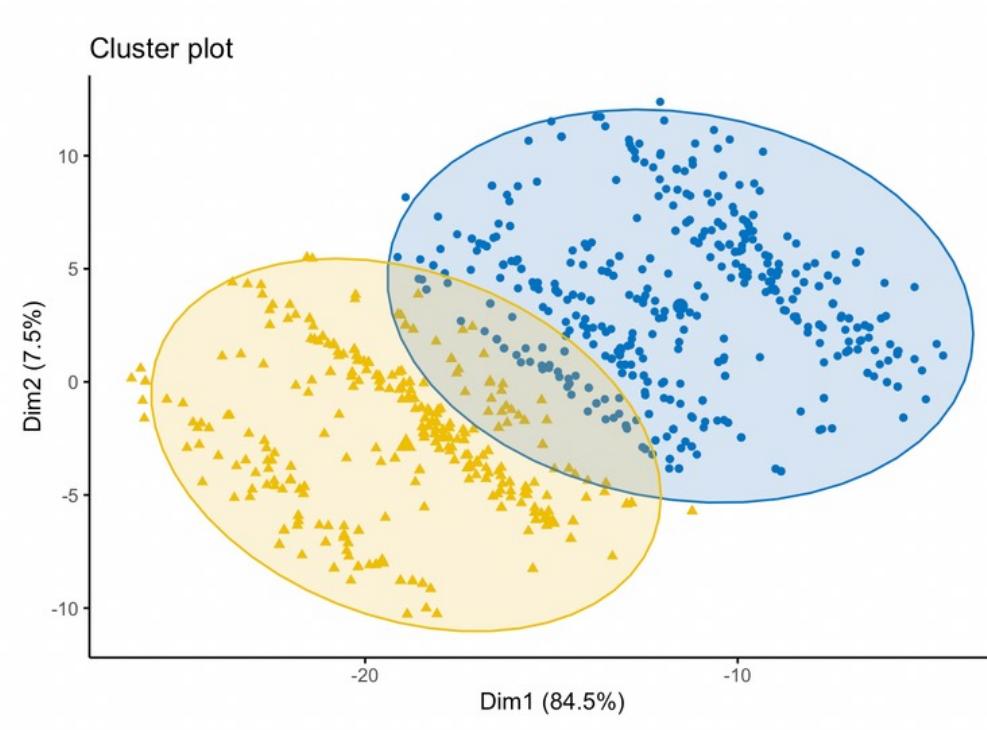
- **Partition our data using k-mean algorithm:**
- We tried three different k-mean values (2,3 and 4).

2. Cluster evaluation:

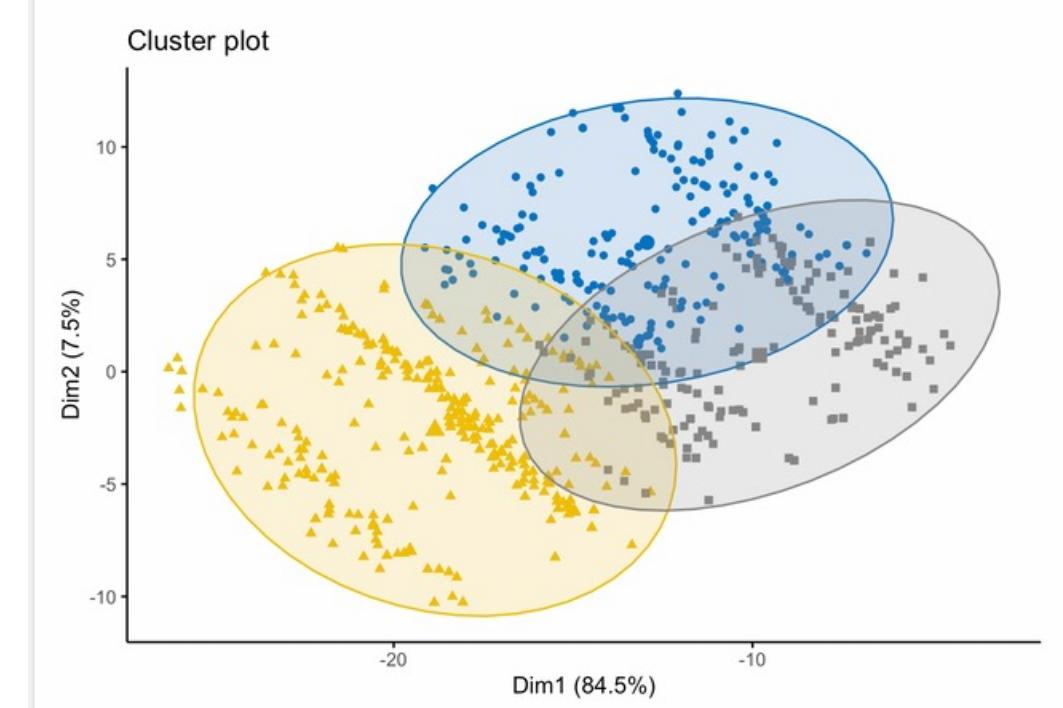
- We calculated the average silhouette width, total within cluster sum of square and BCubed (precision and recall).

Clustering

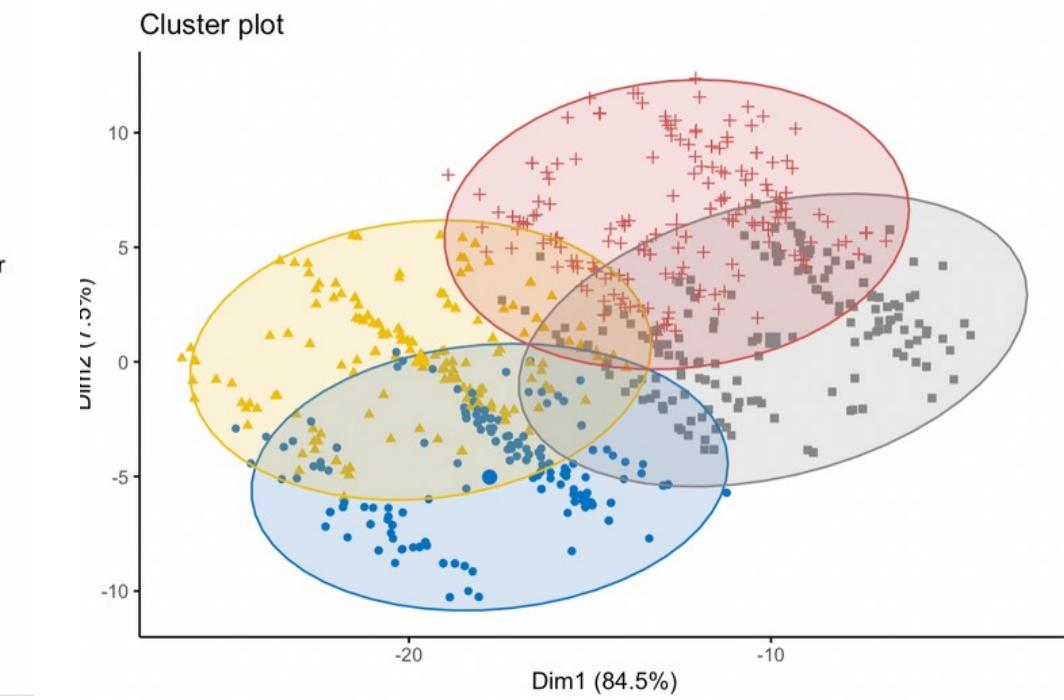
- Partition our data using k-mean algorithm:



k=2



k=3

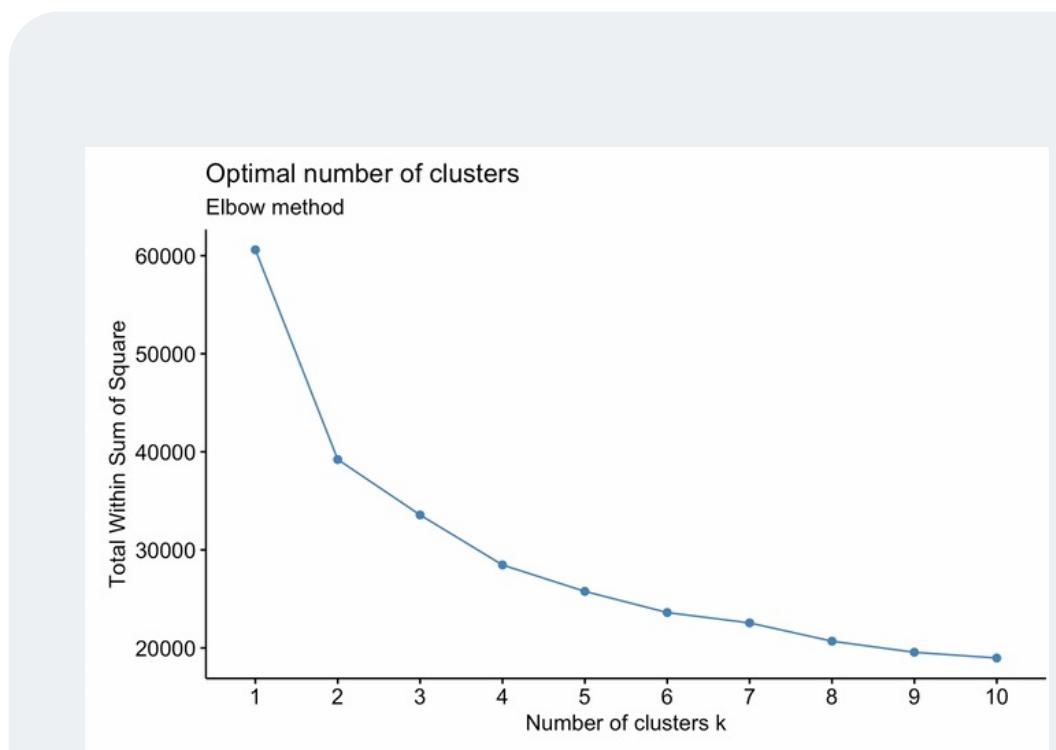
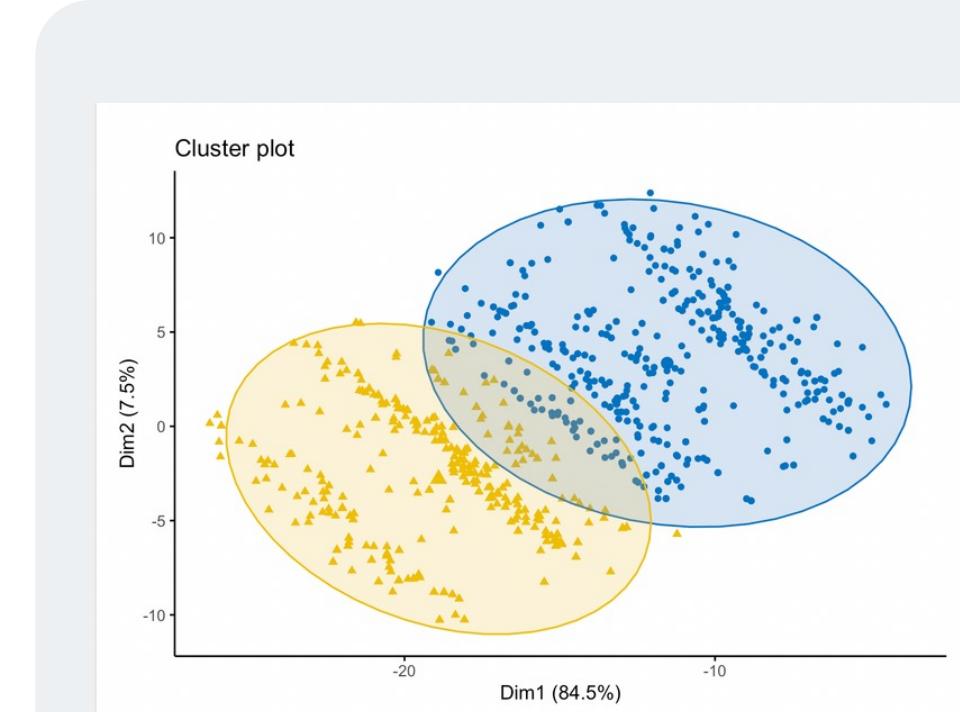


k=4

Findings

Clustering:

- Average silhouette width = 0.3 
- Total within cluster sum = 39732.43 
- BCubed (precision) = 0.3349891 
- BCubed (recall) = 0.512721 



Classification

**What is the
best to predict
Salary?**

Clustering

Thank you!

Do you have any question?