Labor Turnover Prediction

Data Mining

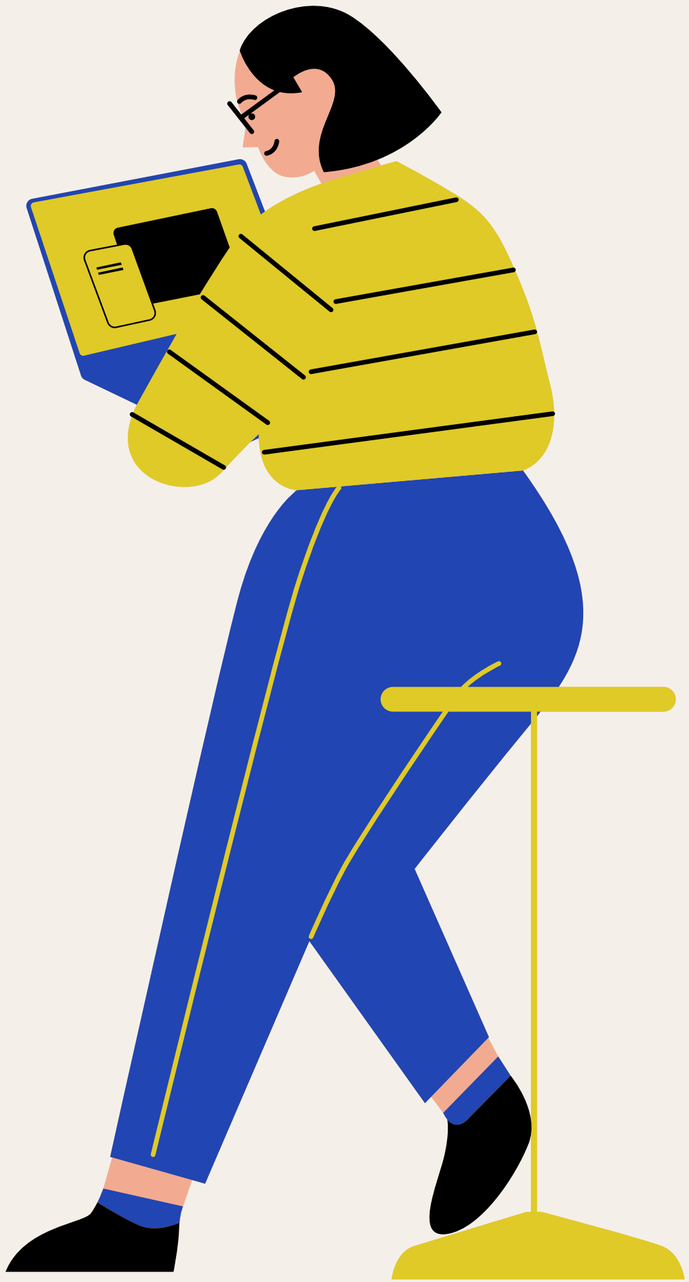
Fatima Rauf &
Menahil Baig

01 - Introduction

02 - Background

03 - Methodology

04 - Conclusions



Data

Mining

01 - Introduction

Analyzing data enables informed decision-making

Hypothesis 1: The relationship between turnover and satisfaction level is negative.

Hypothesis 2: The relationship between employee turnover and the last evaluation is negative.

Hypothesis 3: There is a negative relationship between time spent at the company and turnover.

Hypothesis 4: There is a positive relationship between working hours and turnover.

Hypothesis 5: There is a negative relationship between the level of salary and turnover.

Data

Visualization

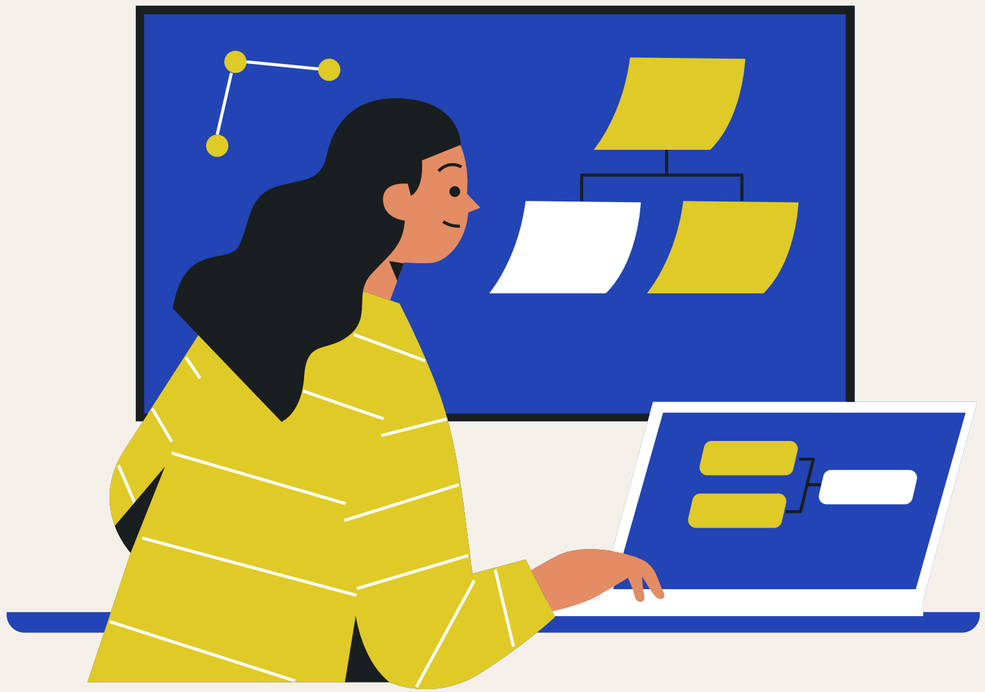
02 - Related Work

Hancock et al.

Meta-analytic review of employee turnover as a predictor of firm performance

- Hausknecht and Trevor
- dysfunctional turnover
- Employee turnover and organizational performance change
- Meta-analytic methods: effect sizes, moderator analyses, and curvilinearity
- Occupation, job level, and industry negatively affected the relationship as was seen for managerial employees

02 - Related Work



Zhao et al.

Employee turnover prediction with machine learning: A reliable approach

SUPERVISED LEARNING ALGORITHMS

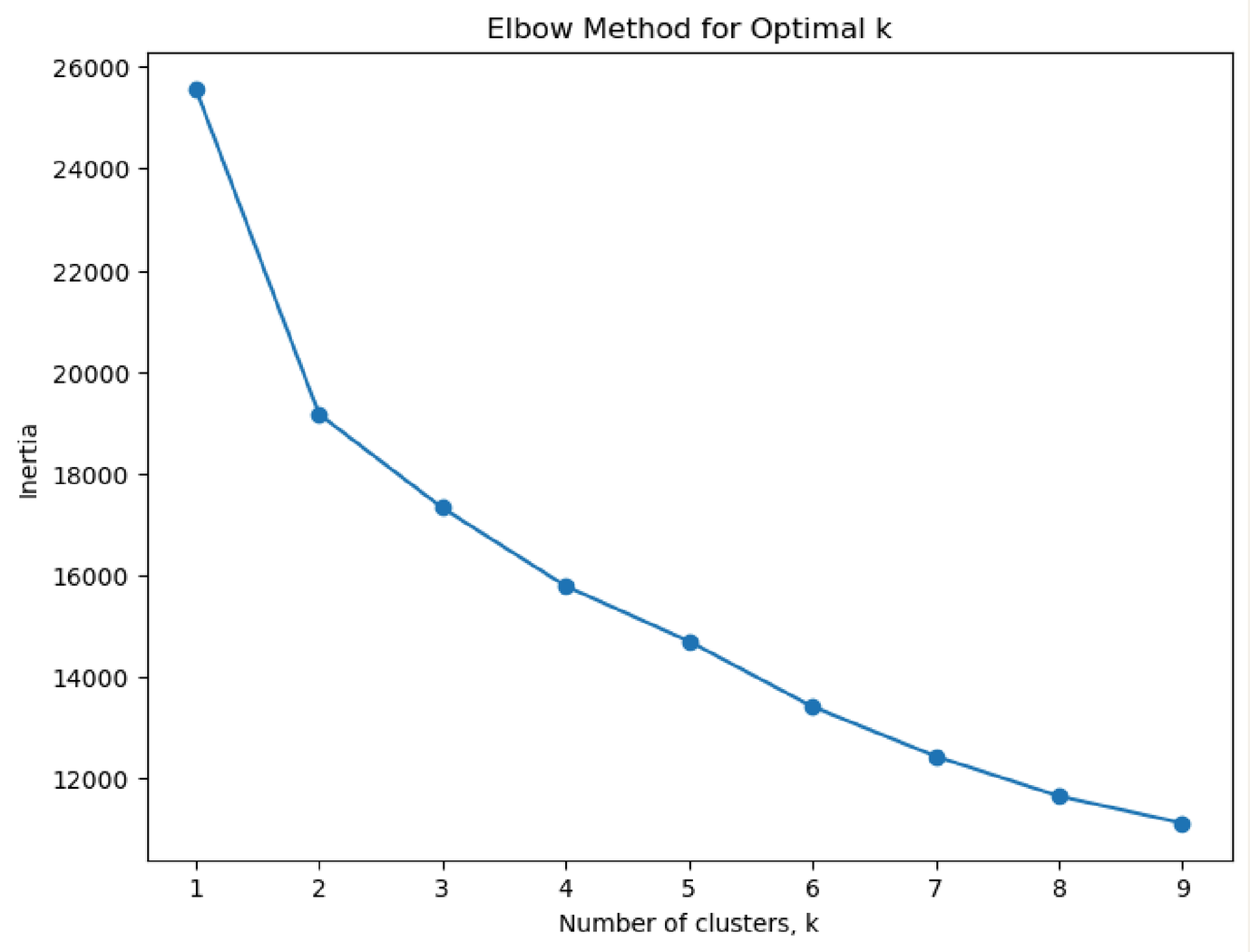
- Decision Trees
- Random Forests
- Gradient-Boosting Trees
- Extreme Gradient Boosting
- Naive Bayes
- KNN

Recommends tree-based approaches for medium and large HR datasets as they tend to have the lowest data variance and a more reliable model can be built using extreme gradient boosting due to its greater predictive power and computational speed

03 - Methodology

EDA & PRE-PROCESSING

- Feature analysis
- Min-Max Scaling
- Label Encoder
- Cluster Analysis
- Elbow Method: k=3
- Hopkins Statistic: 0.9



Where n is the total instances, d is the dimensions in the dataset, k is the number of centroids, m is the cluster size, t is the number of iterations,

Time complexity:

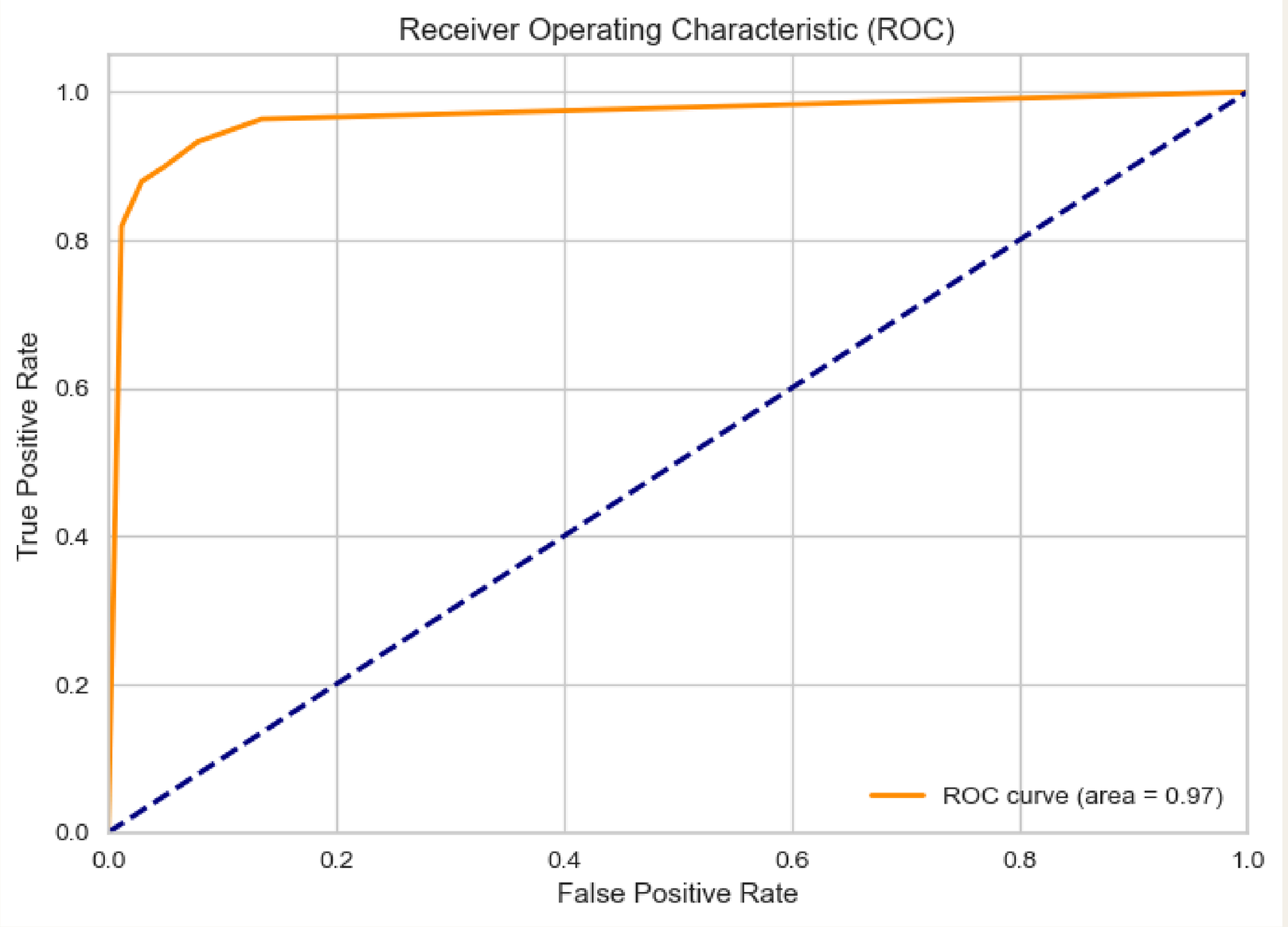
- training is O(tnkd)
- classification, the time complexity is O(kd+Kmd)

KNN & K MEANS

Space Complexity: O(nd+kd)

03 - Methodology: KNN & K MEANS

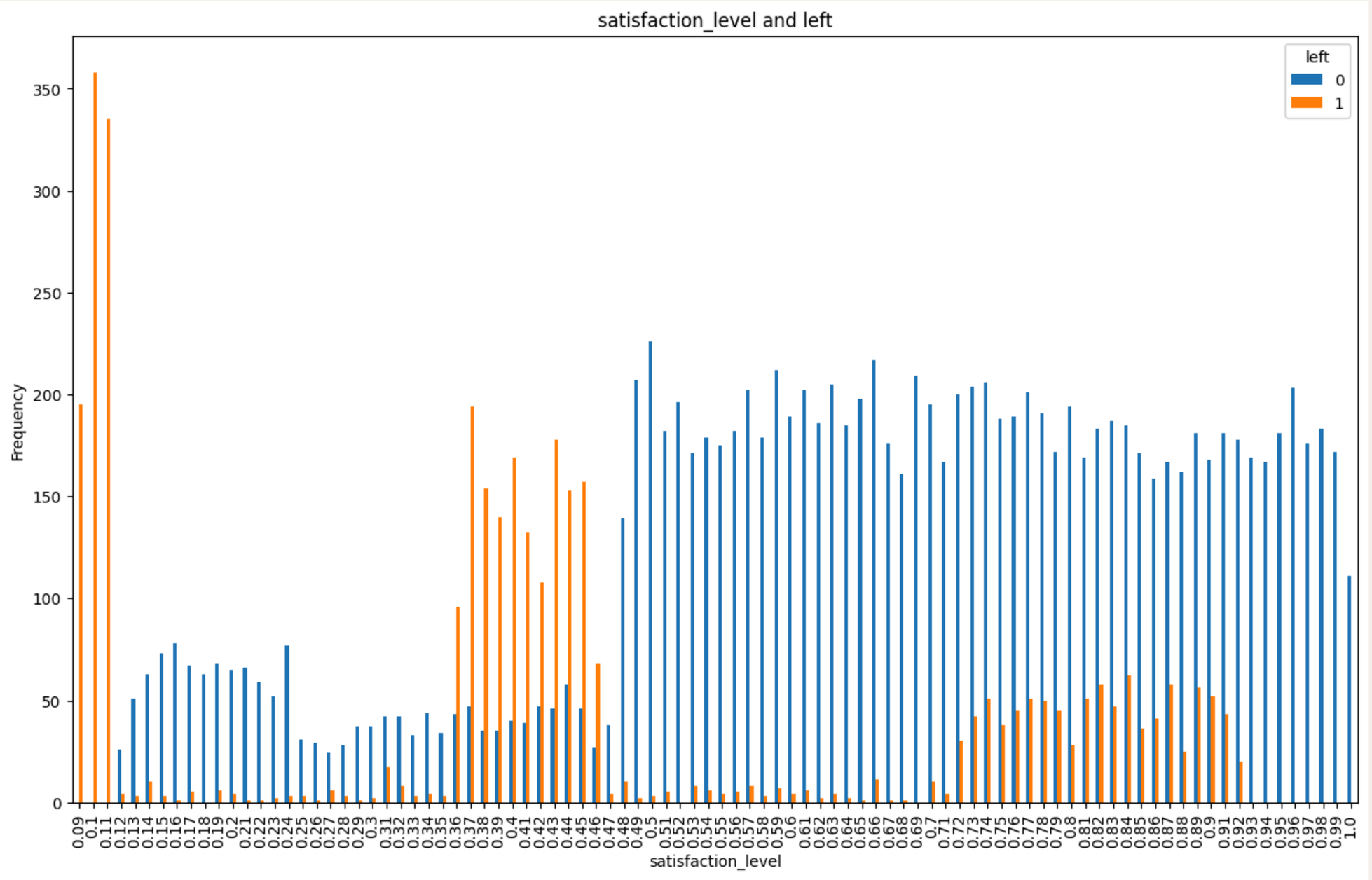
Evaluation


Receiver Operating Characteristic (ROC)
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.97)

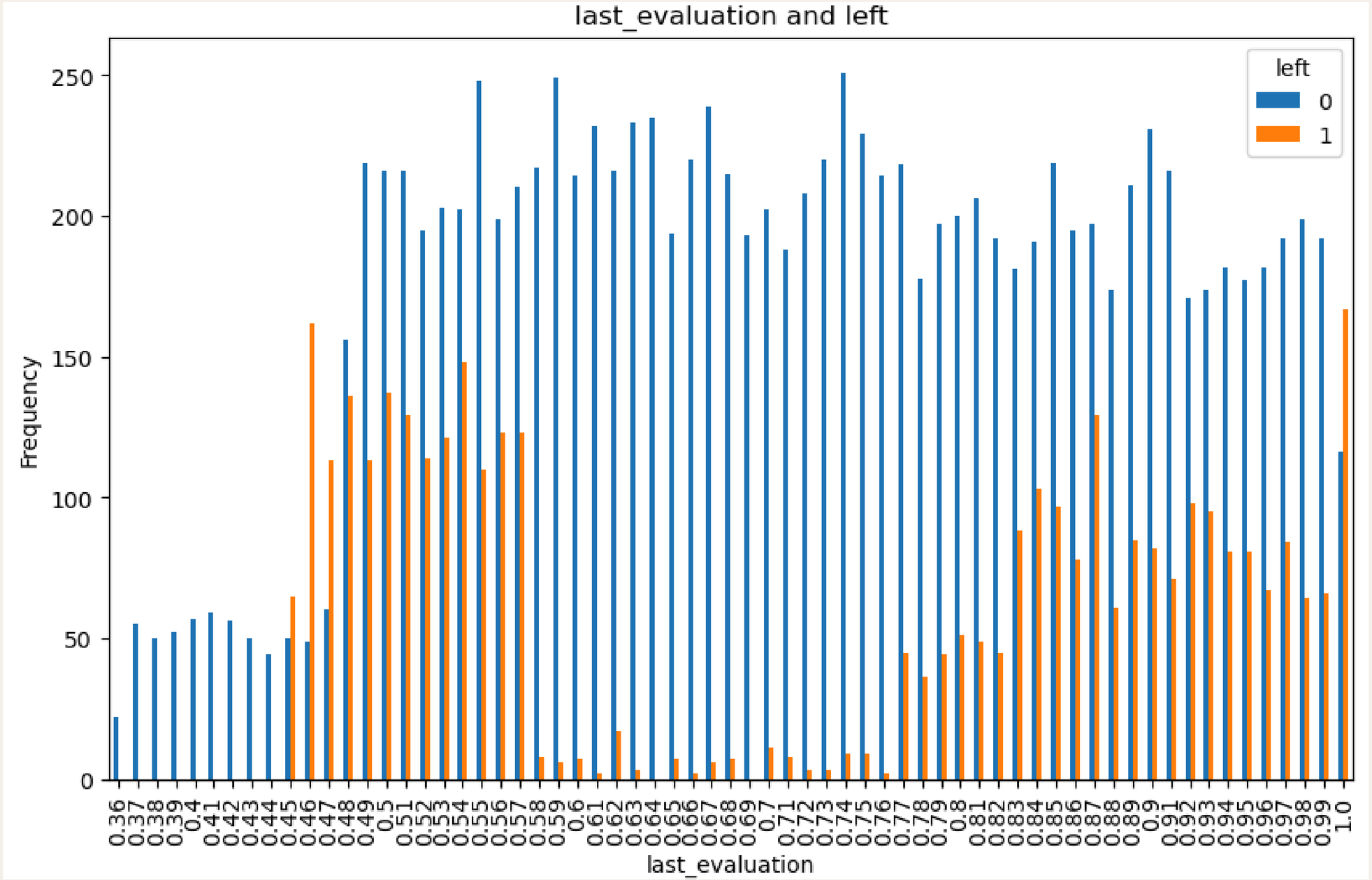
04 - Conclusions

- **Hypothesis 1:** The relationship between turnover and satisfaction level is negative. [DISPROVED]
- **Hypothesis 2:** The relationship between employee turnover and the last evaluation is negative. [DISPROVED]
- **Hypothesis 3:** There is a negative relationship between time spent at the company and turnover. [PROVED]
- **Hypothesis 4:** There is a positive relationship between working hours and turnover. [PROVED]
- **Hypothesis 5:** There is a negative relationship between the level of salary and turnover. [PROVED]

Hypothesis 1



Hypothesis 2



Thanks

Alba Castro