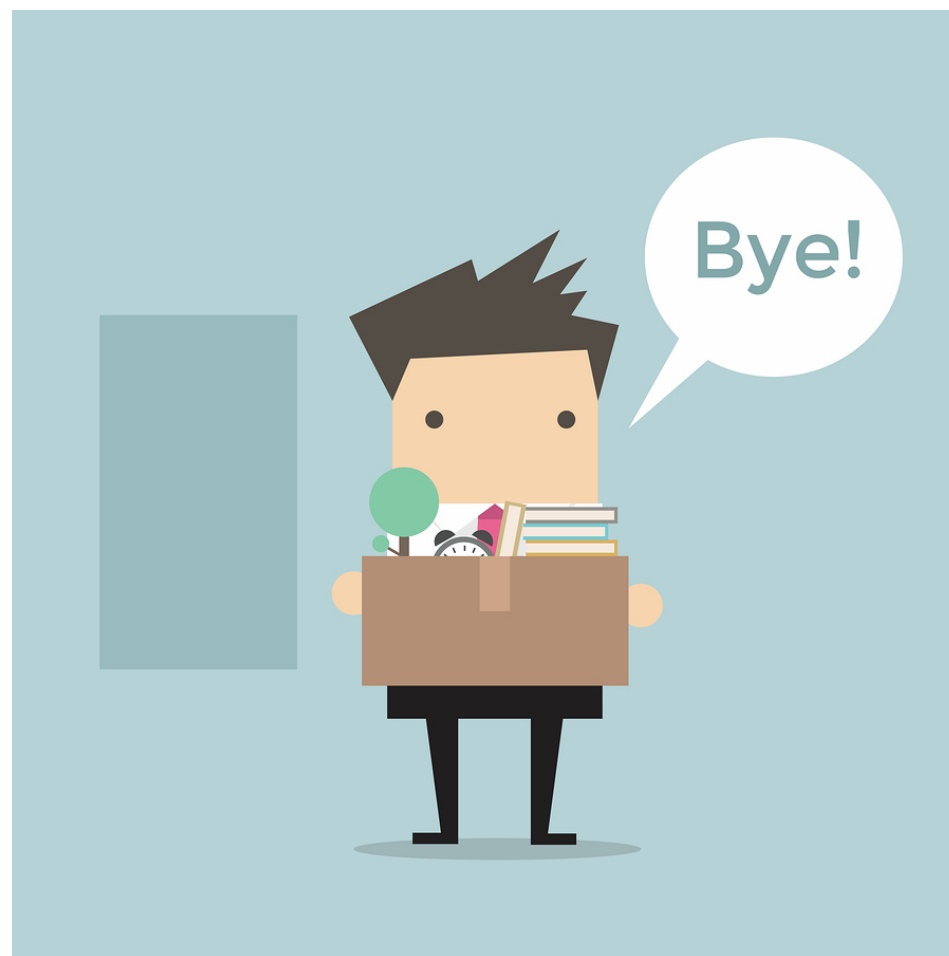
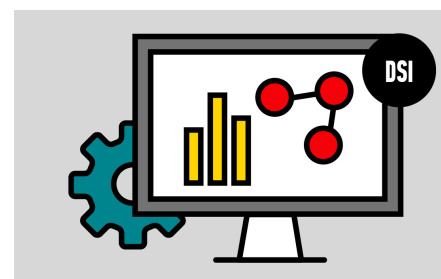


# Attrition Prediction

(resignation prediction)



Atigon Hongchumpol / Mar 1, 2023

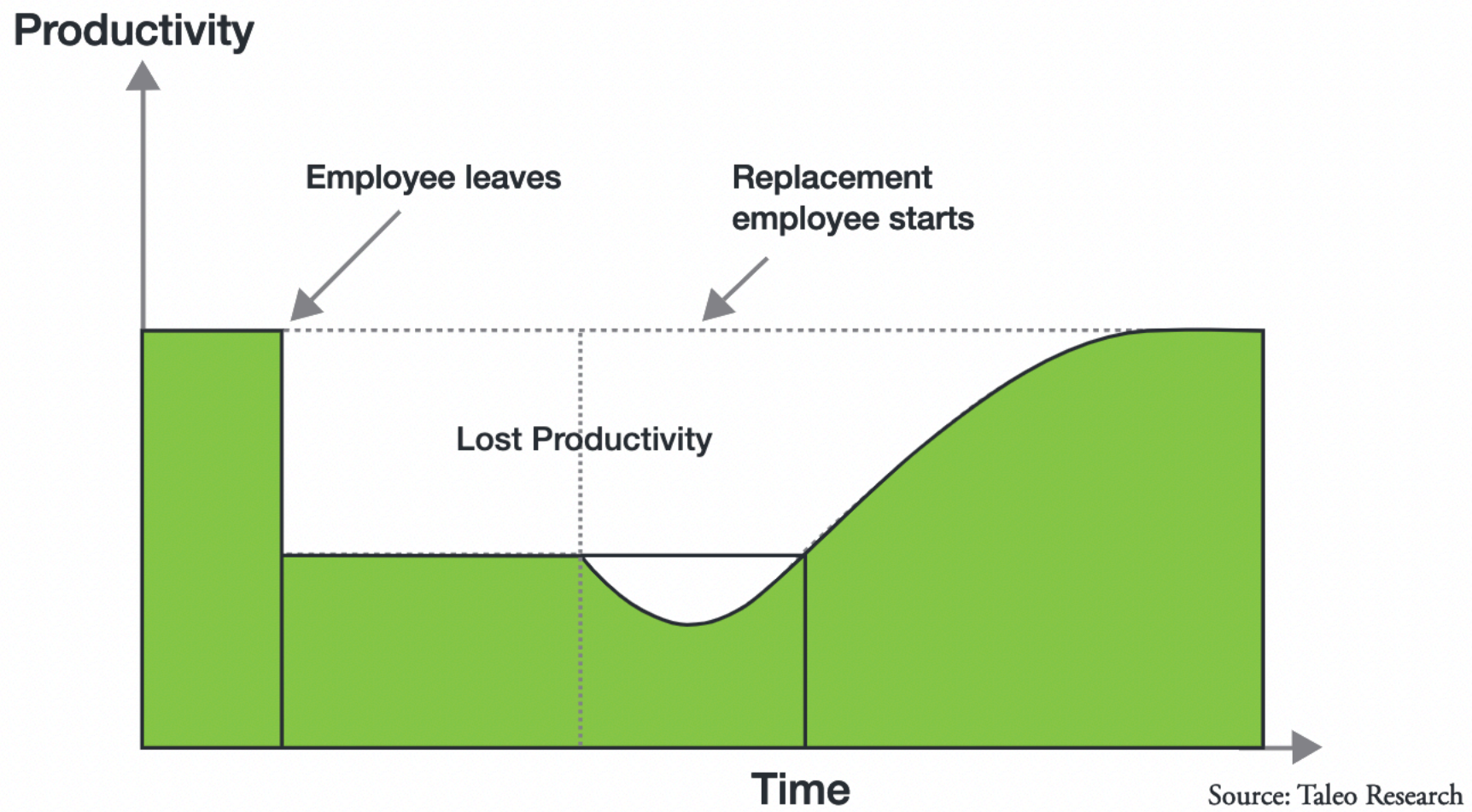


# Problem Statement

# Problem statement

- Employee attrition is a significant challenge for many organizations, leading to increased costs and decreased productivity.
- The sample company has experienced a high employee turnover rate, and we want to develop an attrition prediction model that can identify employees at risk of leaving the company.
- This project aims to develop a predictive model that can accurately identify employees who are likely to leave, allowing the company to take proactive measures to retain valuable employees and reduce the negative impact of attrition on the organization.

## Productivity Losses Tied to Turnover



# Data collection & Cleaning

# Data collection & Cleaning

## Acknowledgements

- The dataset was obtained from Kaggle, a popular platform for data science competitions and data analysis projects. The dataset was provided by Pavansubhash and is publicly available for download on the Kaggle website.
- The dataset contains 1470 rows and 35 columns, each row representing each employee.
- **Acknowledgments:** We would like to thank IBM data scientists for providing the dataset used in this analysis. Their contribution was invaluable in enabling us to carry out this research.

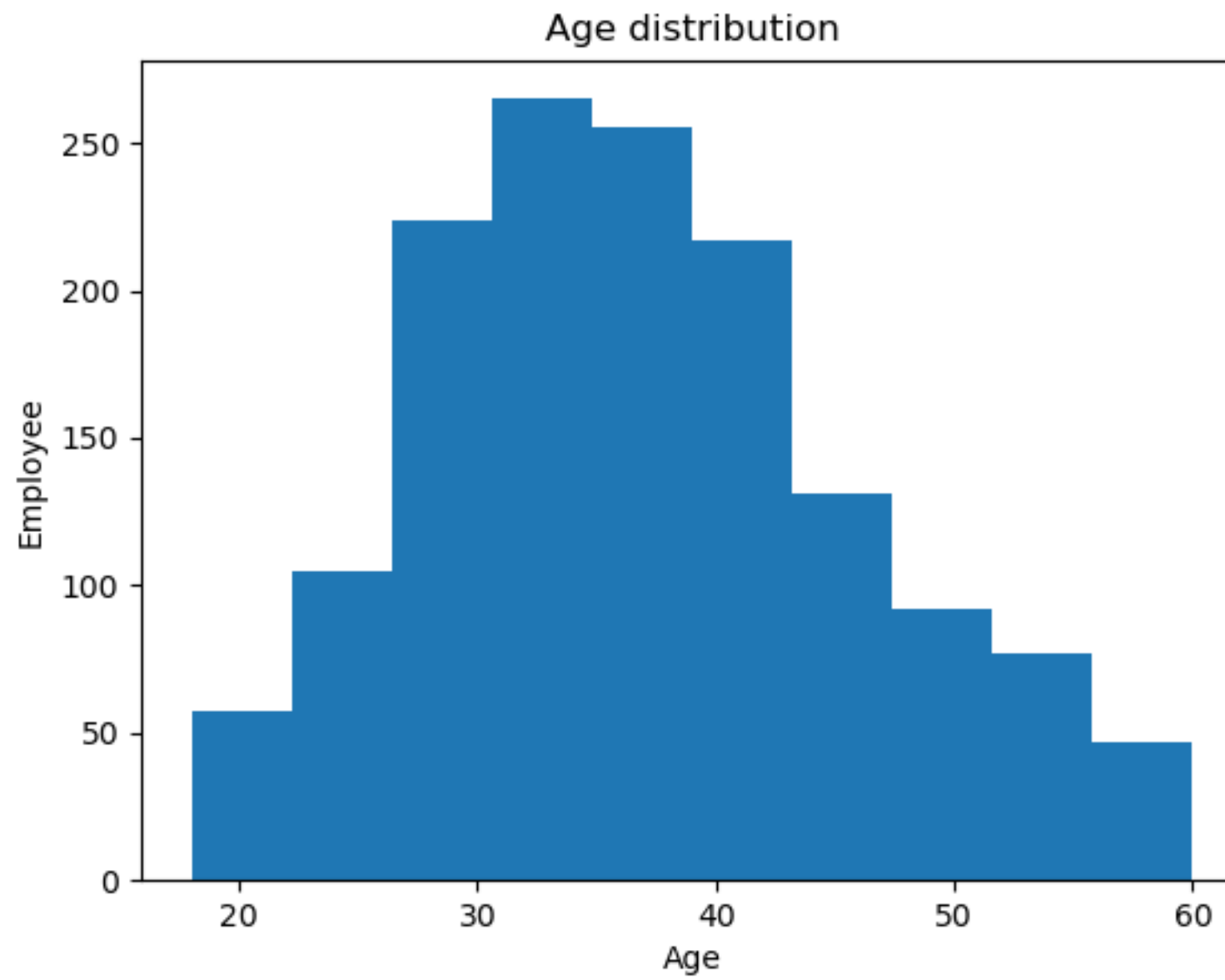
# Data cleaning

- This dataset is clean and ready for analysis, as it does not contain any null values or missing data. Therefore, no data cleaning is required prior to analysis.

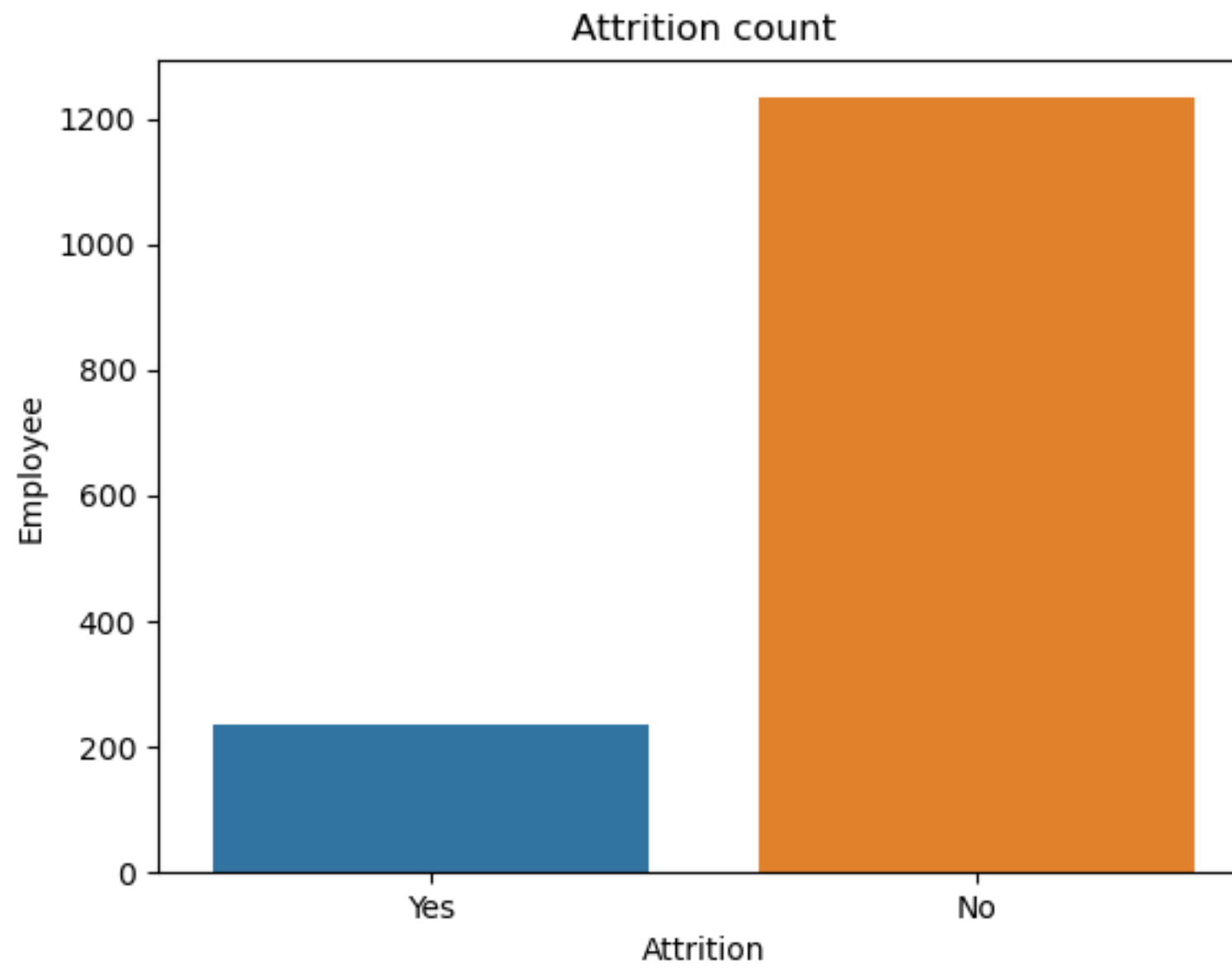
# Exploratory Data Analysis



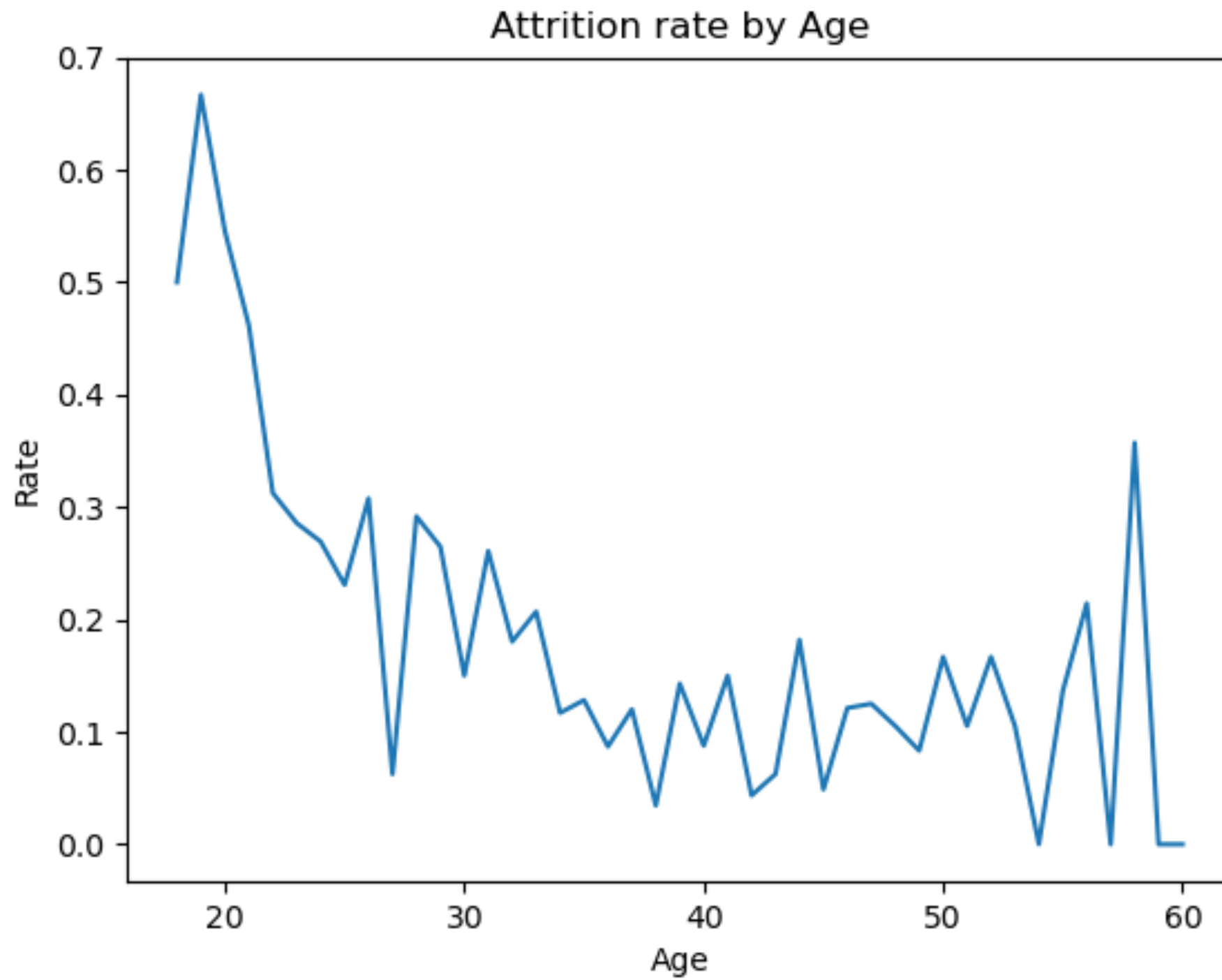
# EDA



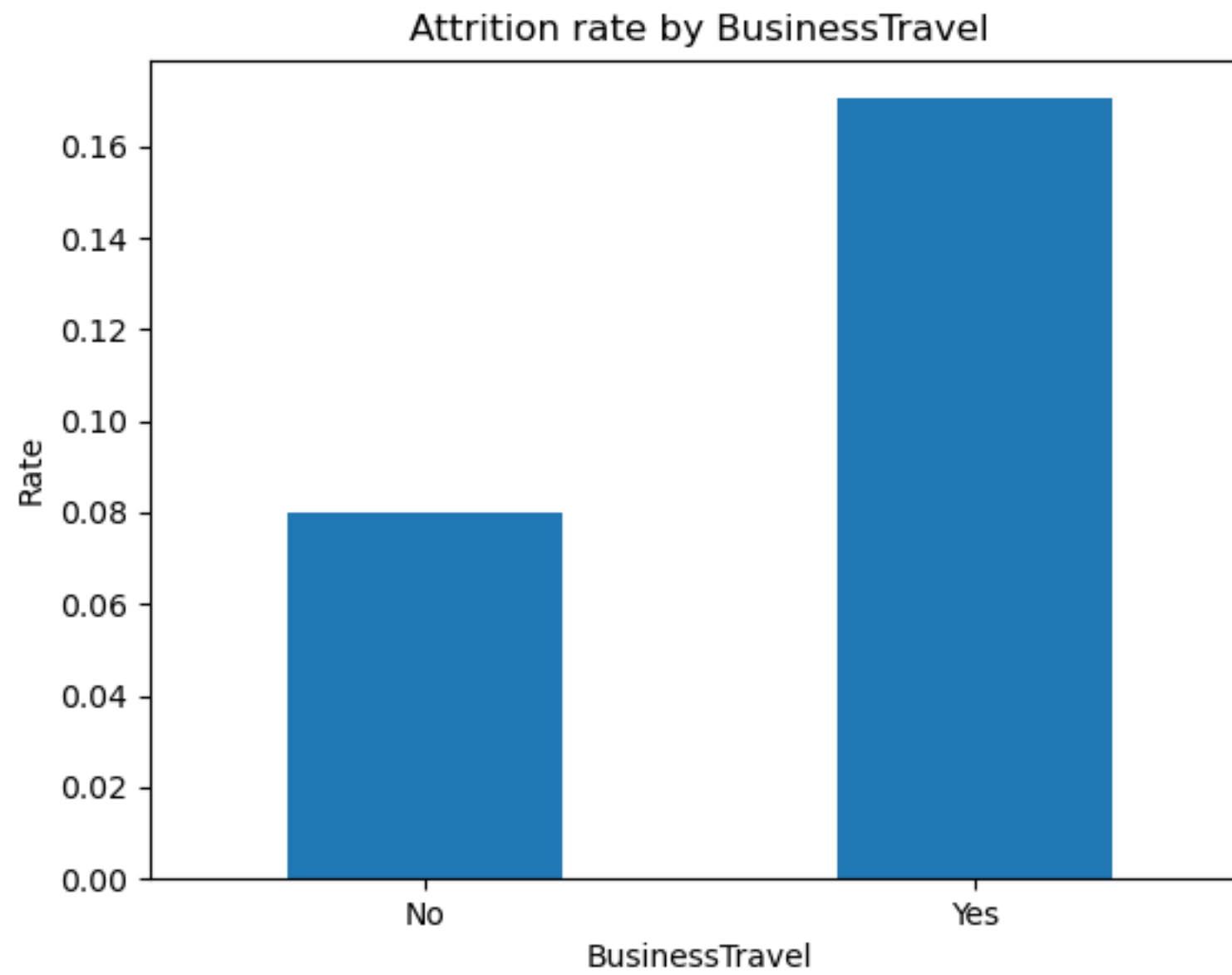
# EDA



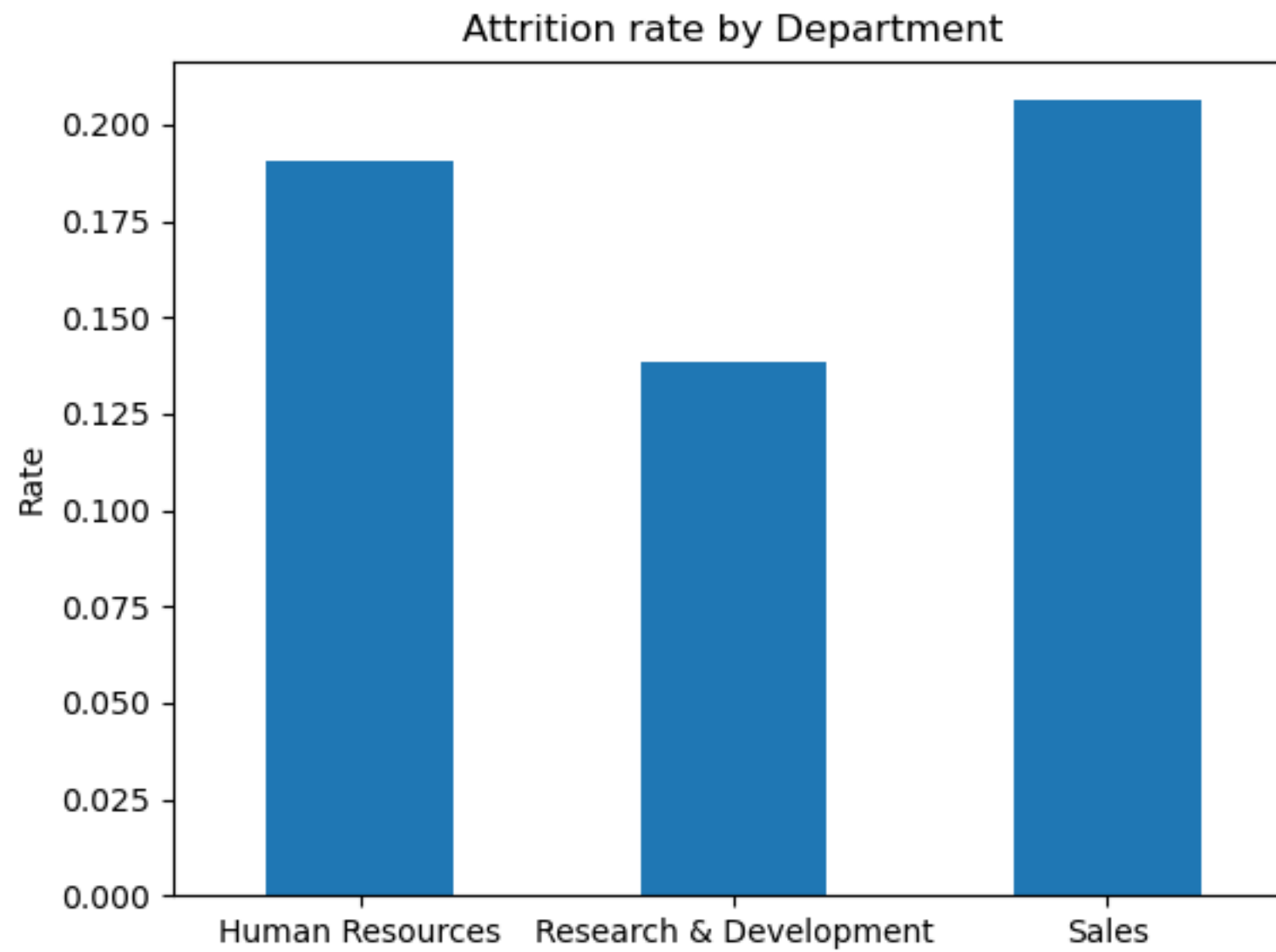
# EDA



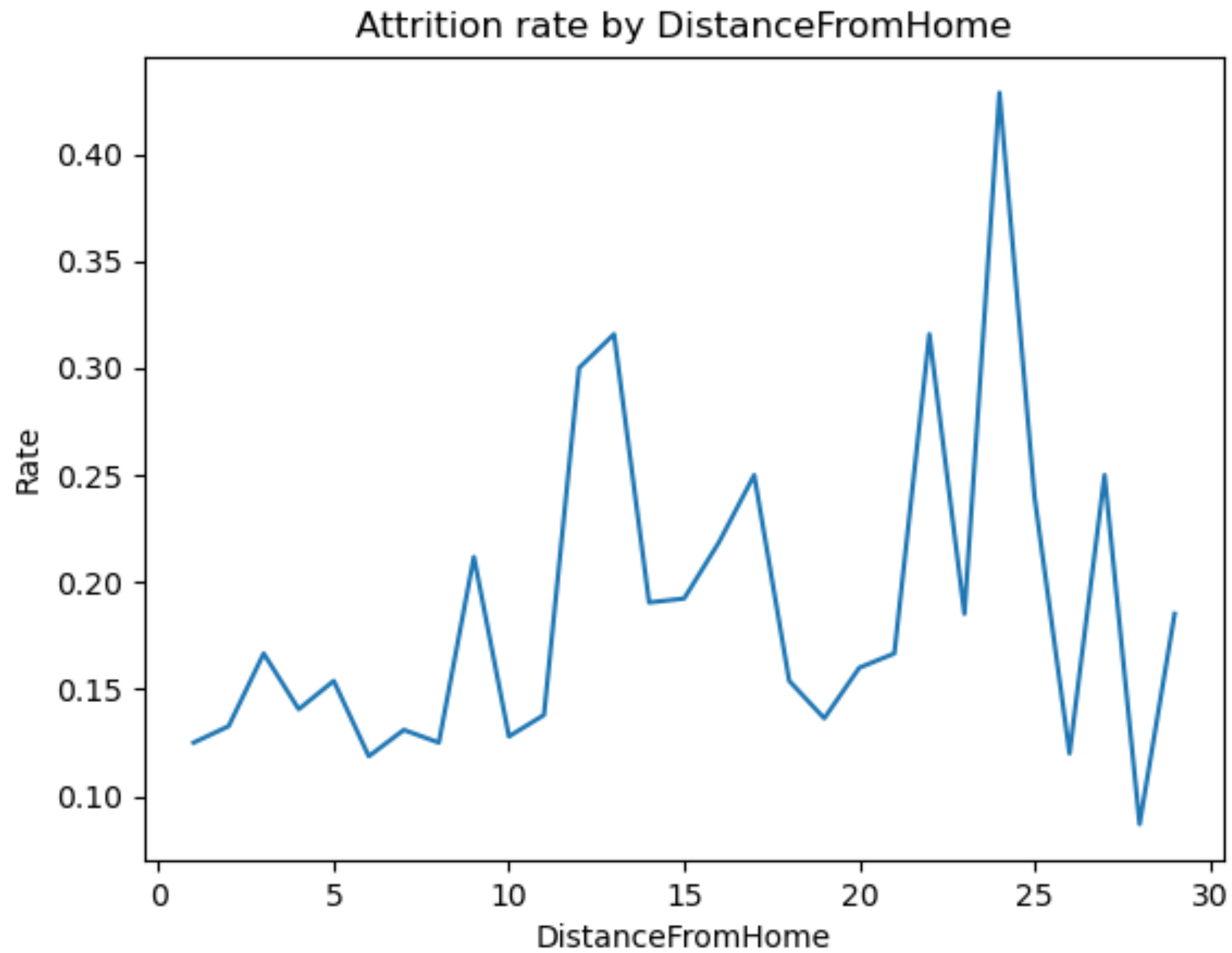
# EDA



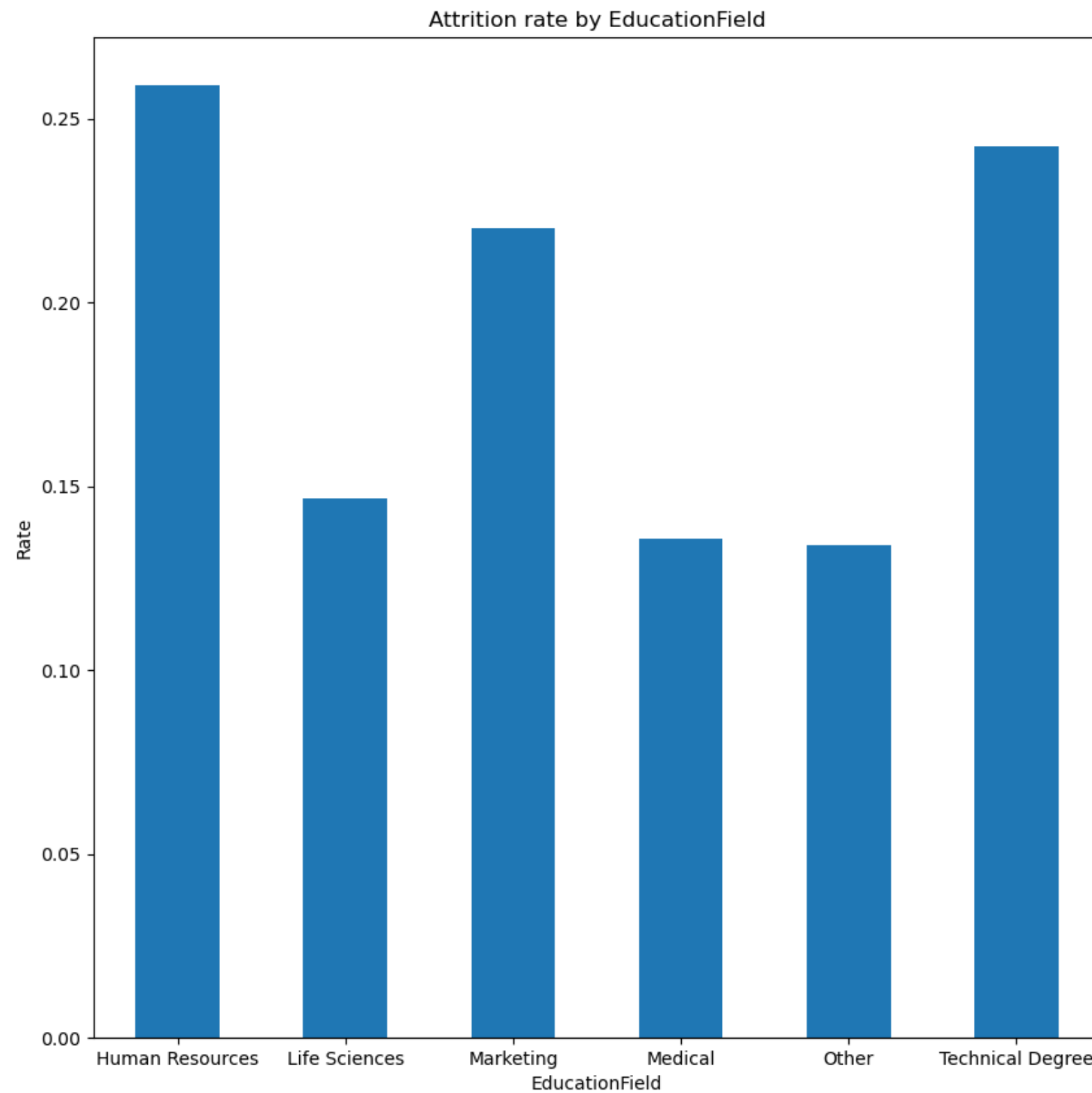
# EDA



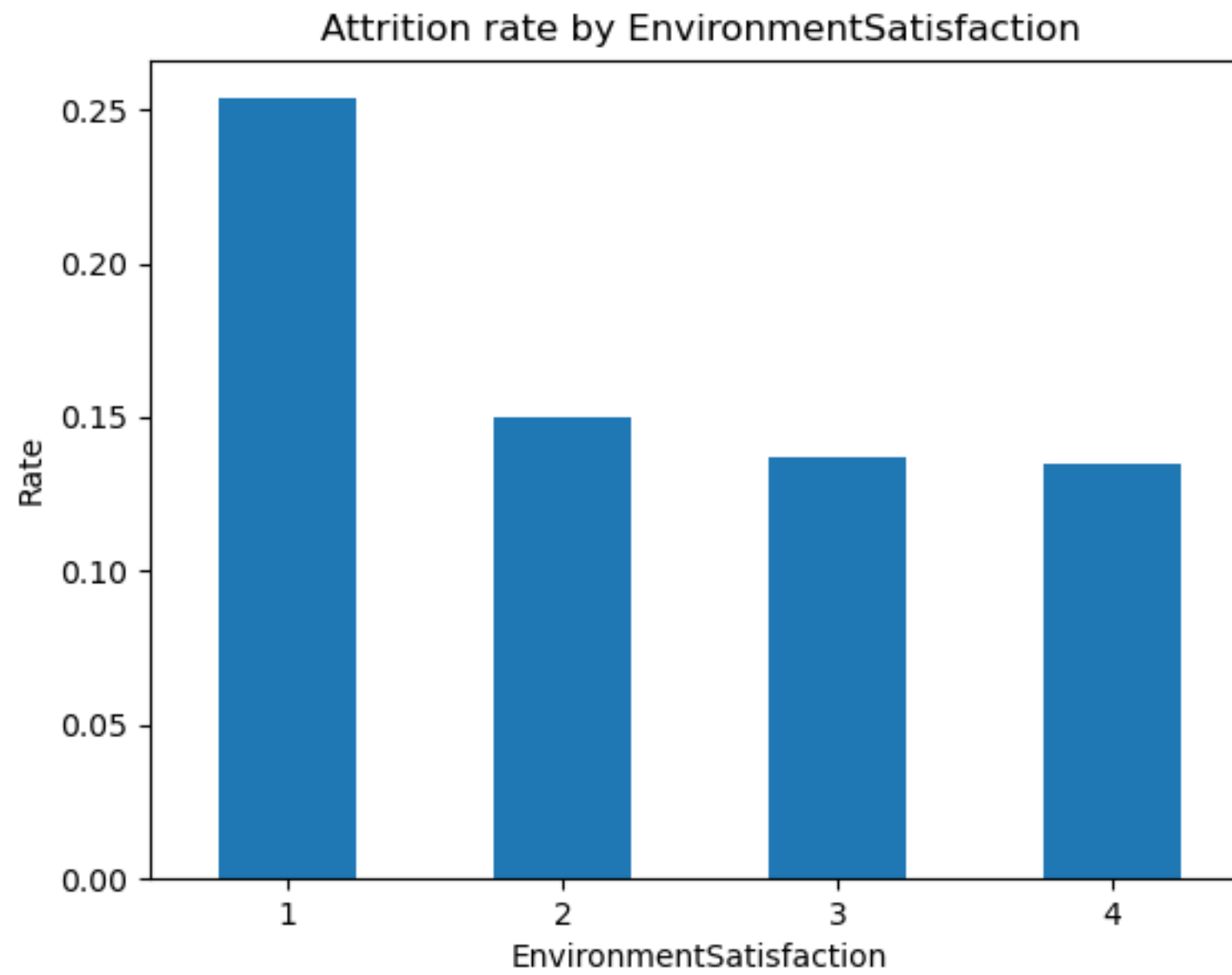
# EDA



# EDA

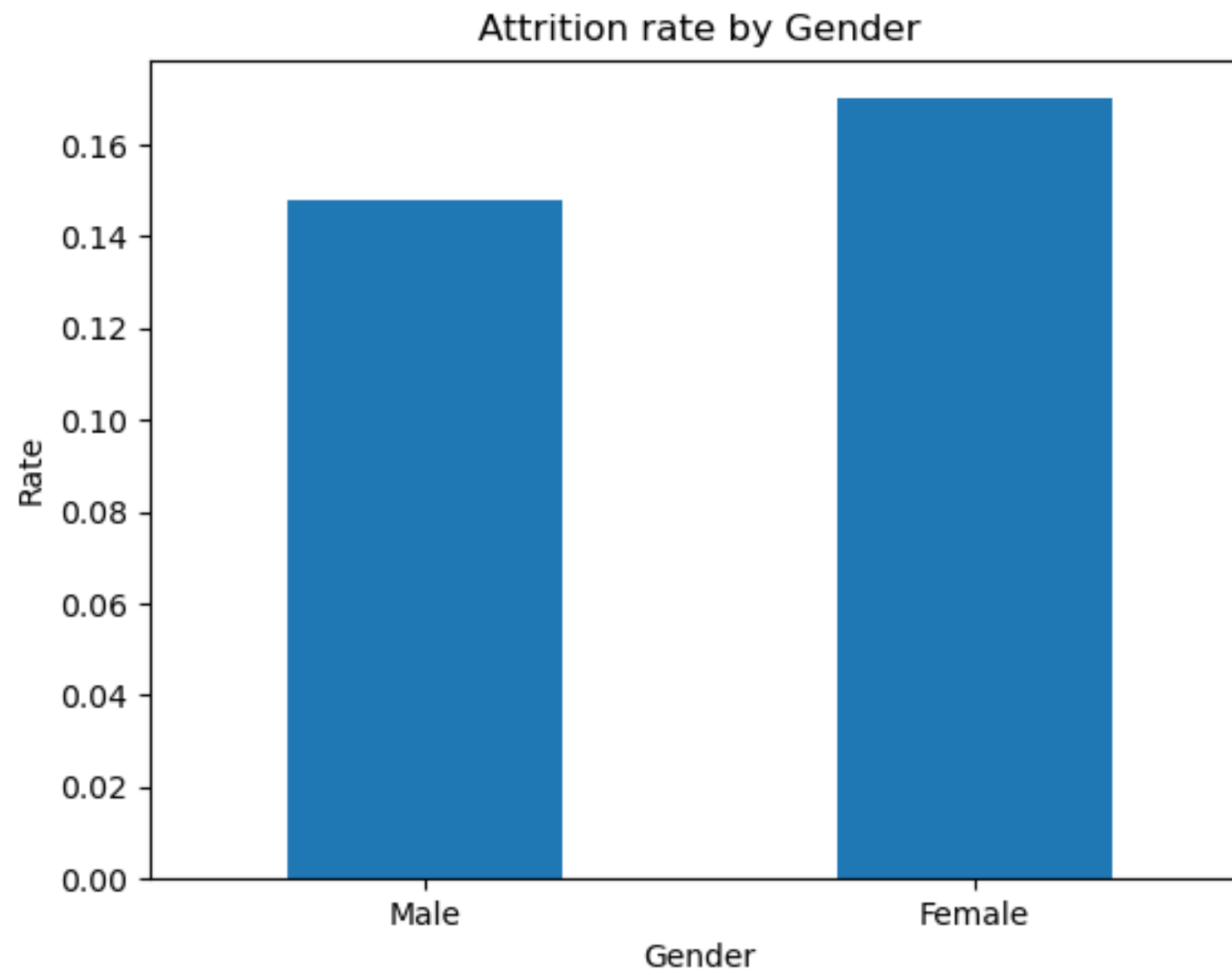


# EDA

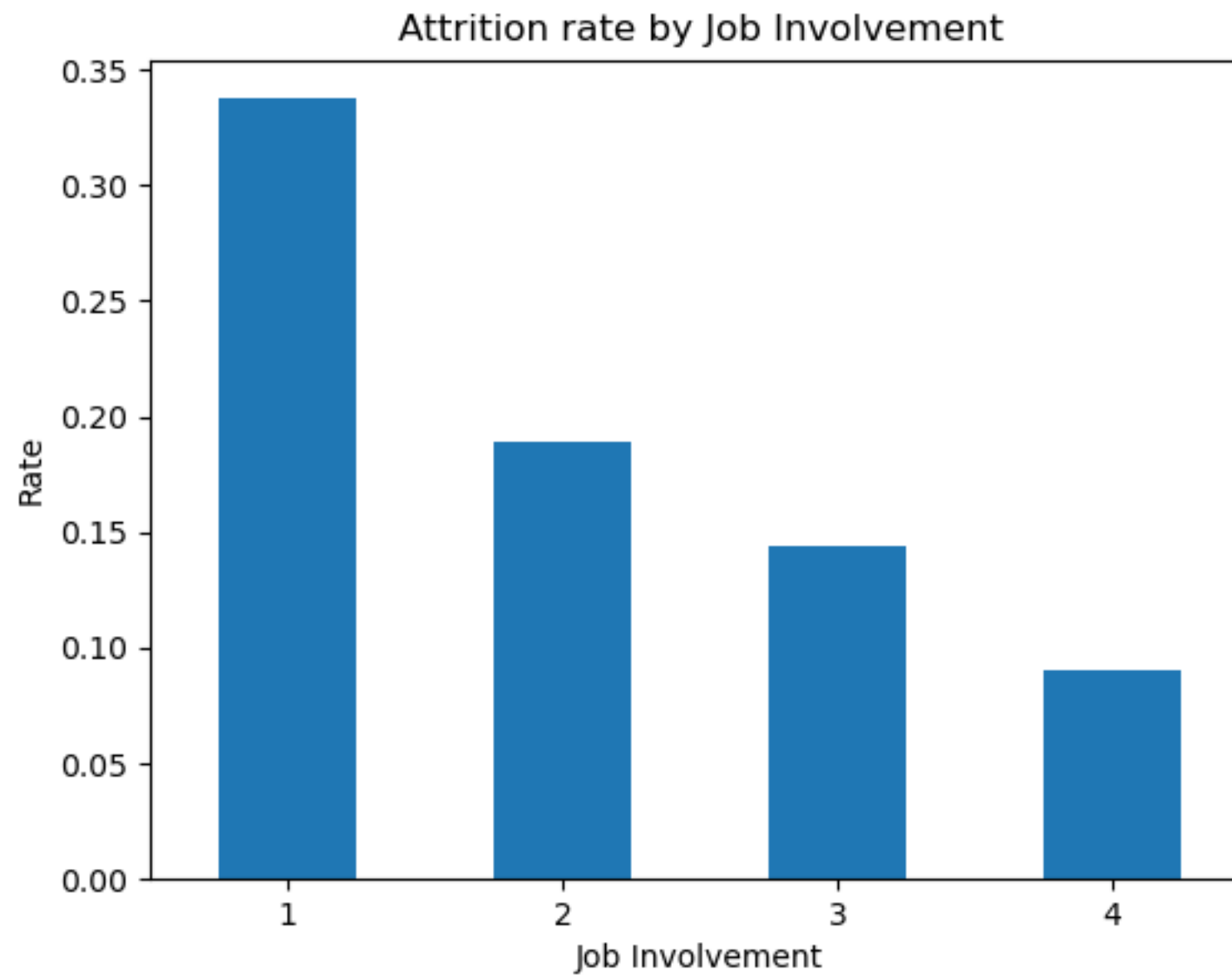




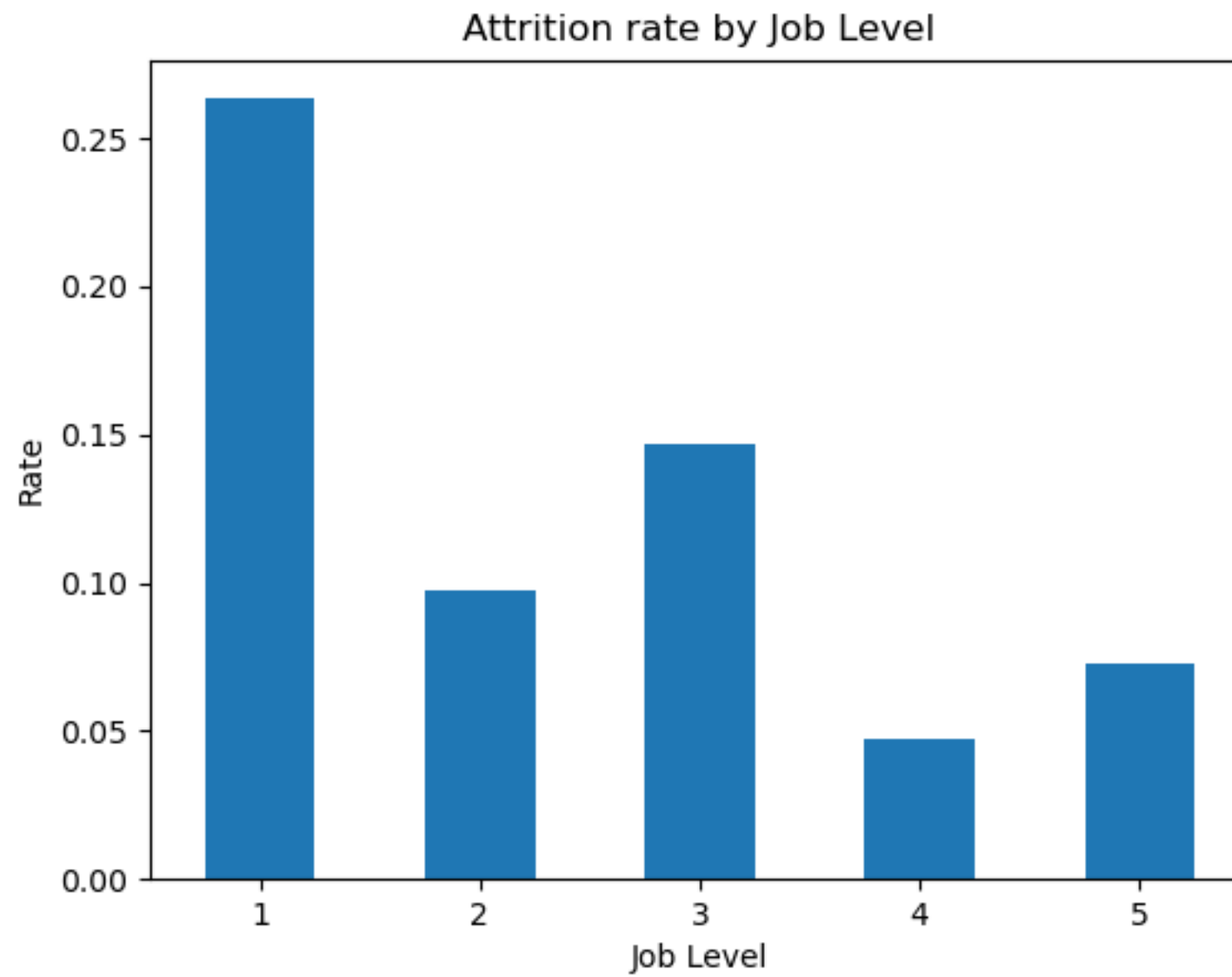
# EDA



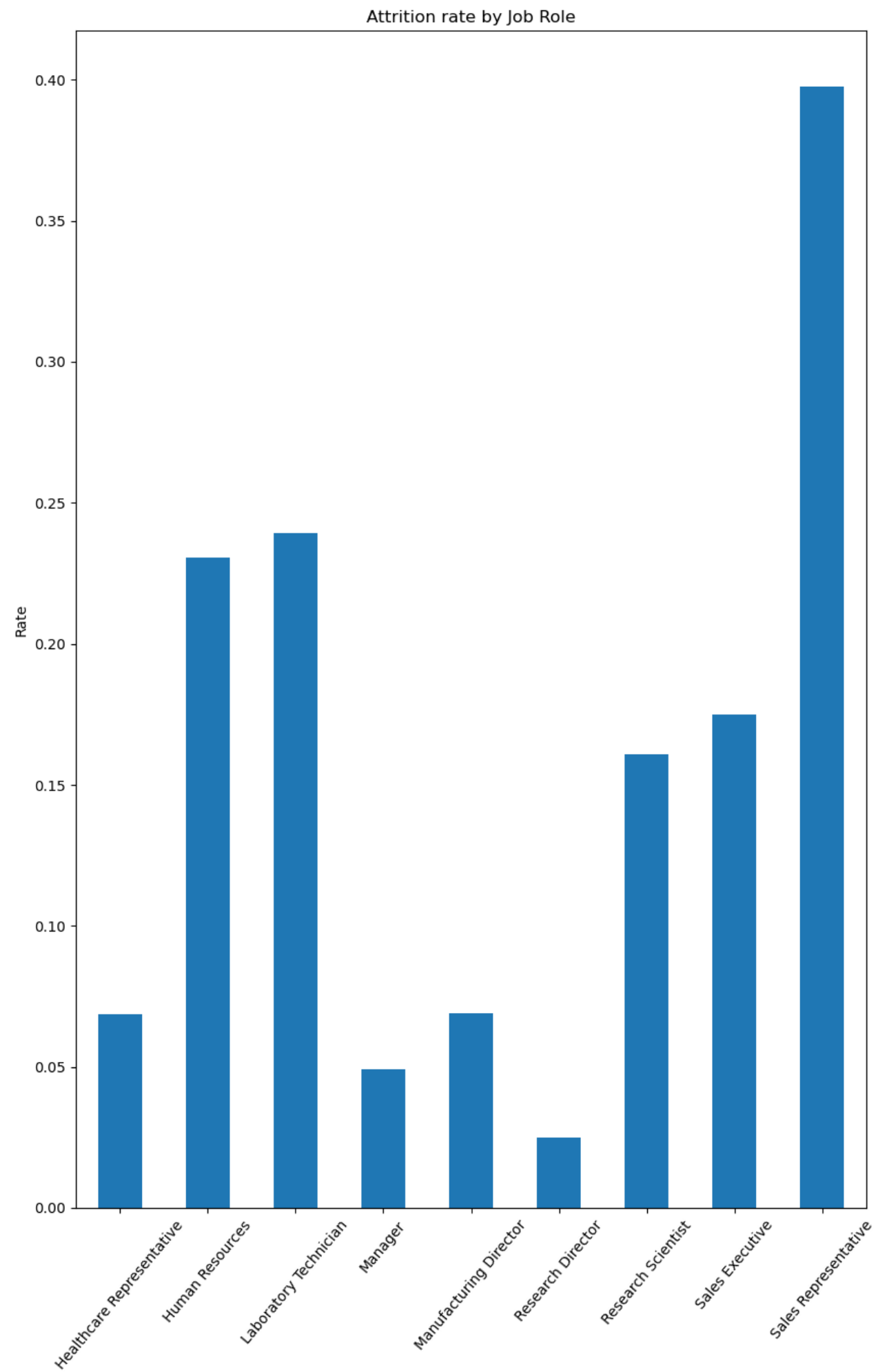
# EDA



# EDA



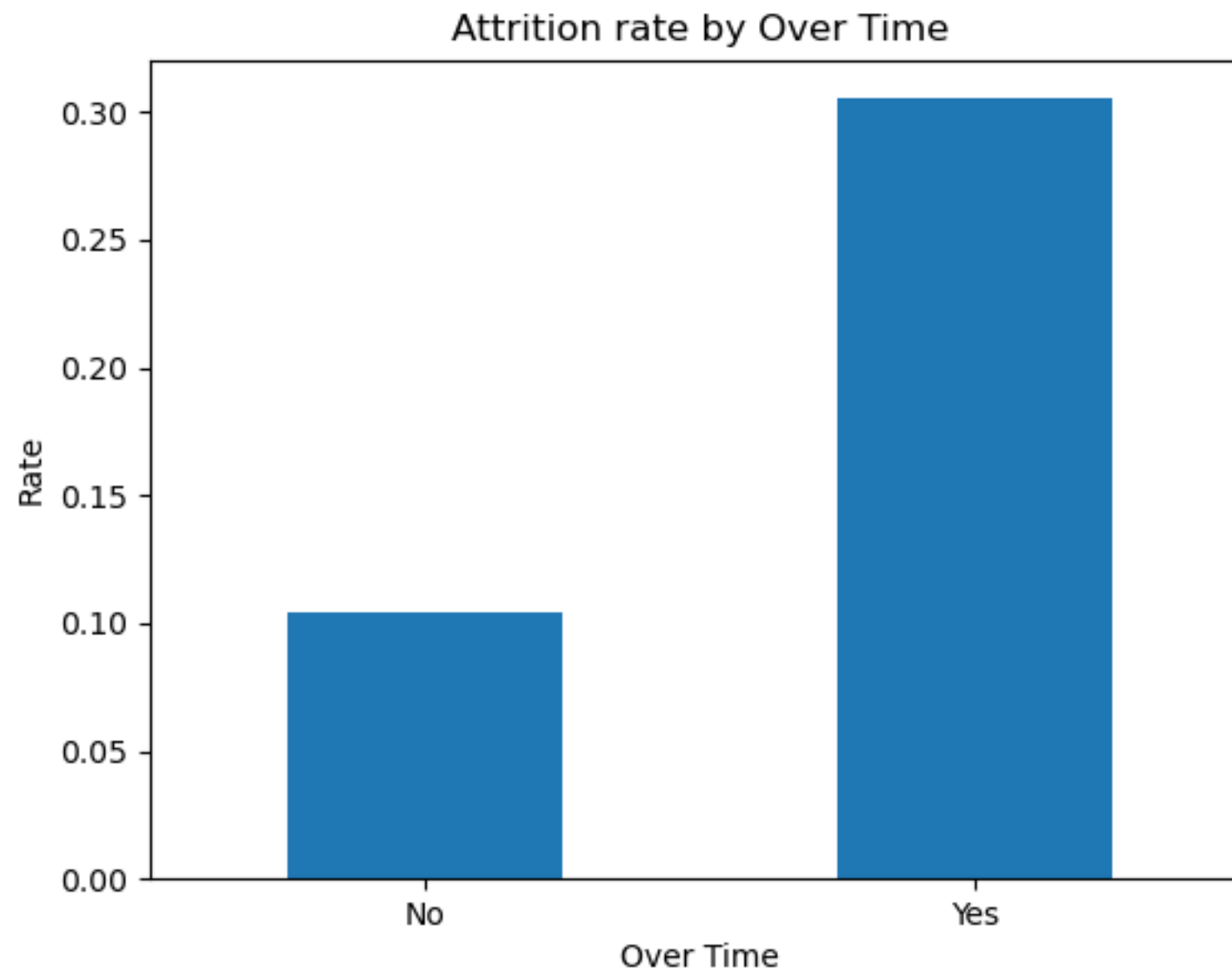
# EDA



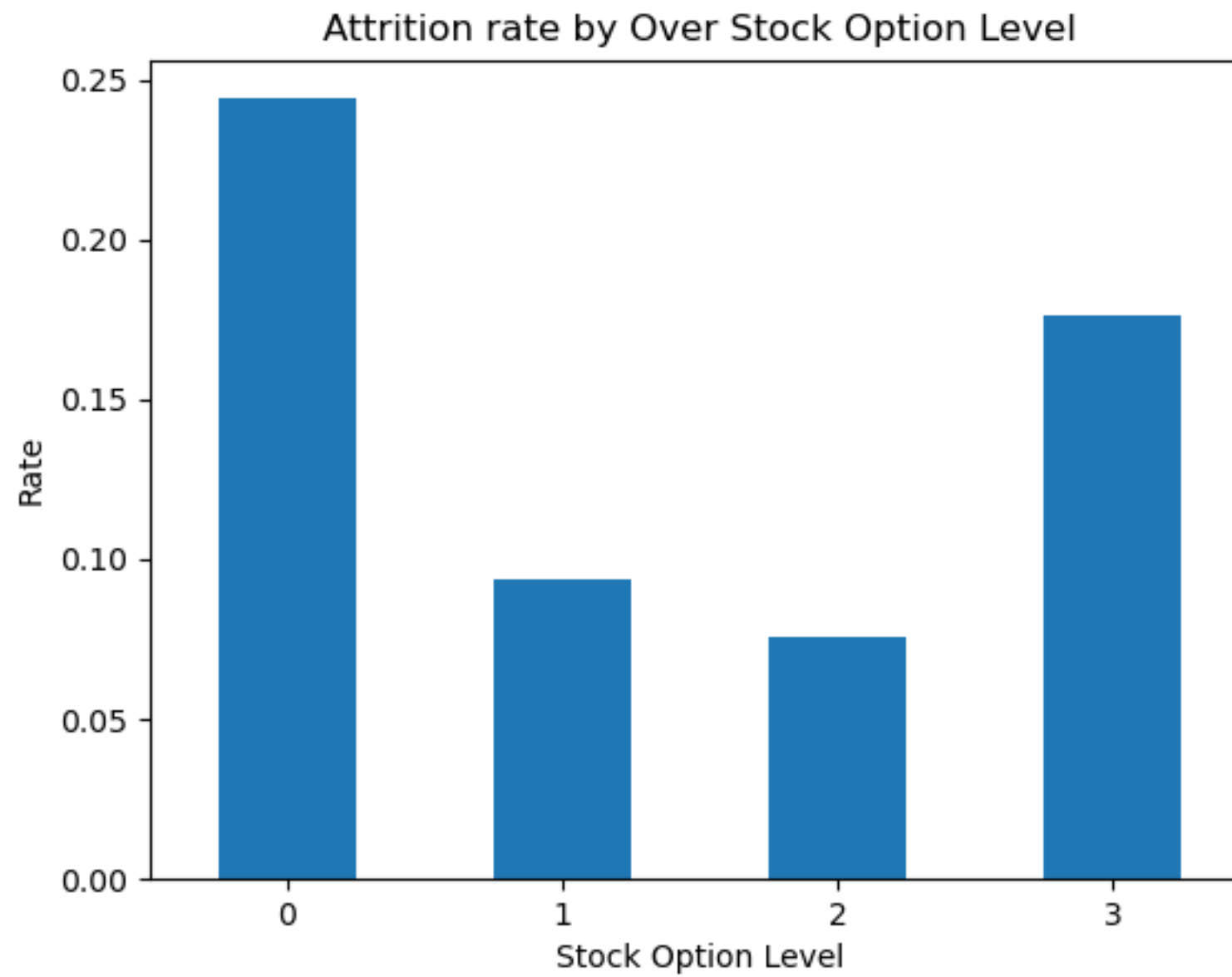
# EDA



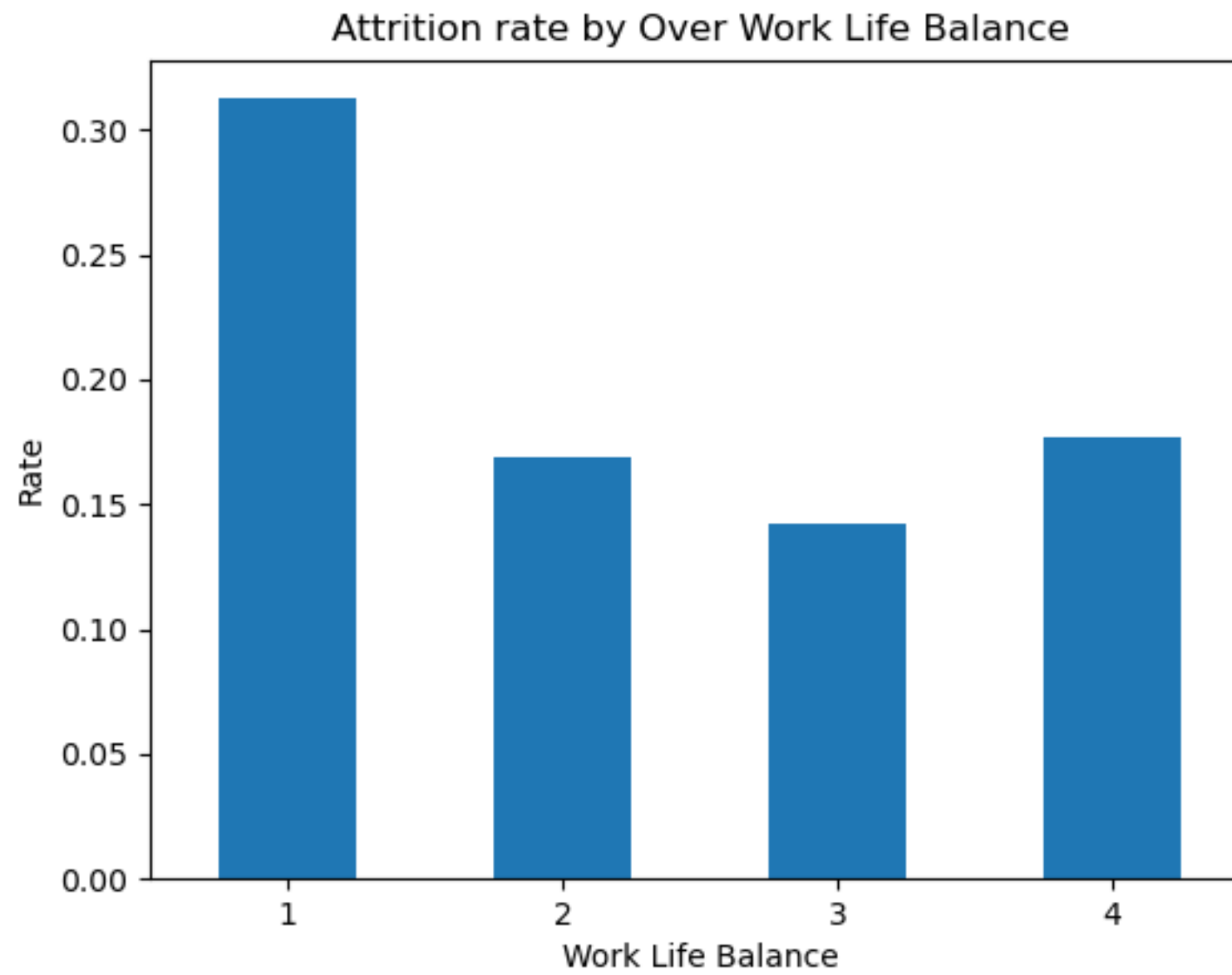
# EDA



# EDA



# EDA





**Baseline model**

# Baseline model

	model	test_accuracy	test_recall	tn	fp	fn	tp	precision	F1-score
0	LogisticRegression	0.846154	0.055556	185	0	34	2	1.000000	0.105263
1	RandomForestClassifier	0.859729	0.194444	183	2	29	7	0.777778	0.311111
2	AdaBoostClassifier	0.855204	0.361111	176	9	23	13	0.590909	0.448276
3	KNeighborsClassifier	0.782805	0.027778	172	13	35	1	0.071429	0.040000

- The best baseline score came from the AdaBoost model
  - Recall 36%
  - Precision 59%
  - F1-score 45%
- Our goal is to improve F1-score by 20%

# Feature selection model

# Selected features

```
selected_features = [  
    'Age',  
    'BusinessTravel',  
    'Department',  
    'DistanceFromHome',  
    'EducationField',  
    'EnvironmentSatisfaction',  
    'JobInvolvement',  
    'JobLevel',  
    'JobRole',  
    'JobSatisfaction',  
    'MaritalStatus',  
    'salary_range',  
    'NumCompaniesWorked',  
    'num_comp_work',  
    'OverTime',  
    'RelationshipSatisfaction',  
    'StockOptionLevel',  
    'TotalWorkingYears',  
    'TrainingTimesLastYear',  
    'WorkLifeBalance',  
    'YearsAtCompany',  
    'YearsInCurrentRole',  
    'YearsWithCurrManager'  
]
```

# Feature selection model

	model	test_accuracy	test_recall	tn	fp	fn	tp	precision	F1-score
0	LogisticRegression	0.855204	0.085714	186	0	32	3	1.000000	0.157895
1	RandomForestClassifier	0.850679	0.142857	183	3	30	5	0.625000	0.232558
2	AdaBoostClassifier	0.873303	0.428571	178	8	20	15	0.652174	0.517241
3	KNeighborsClassifier	0.832579	0.142857	179	7	30	5	0.416667	0.212766

- The best production score came from the AdaBoost model
  - Recall 43% (baseline 36%)
  - Precision 65% (baseline 59%)
  - F1-score 52% (baseline 45%) **+15%**

# Hyper Tuning & Final Model

# Best hyper tuning model

	model	test_accuracy	test_recall	tn	fp	fn	tp	precision	f1-score
0	AdaBoostClassifier	0.895928	0.4	184	2	21	14	0.875	0.54902

- The best production score came from the AdaBoost model
  - Recall 40% (baseline 36%)
  - Precision 88% (baseline 59%)
  - F1-Score 55% (baseline 45%) **+22%**

# Streamlit



# Streamlit

## Online

- <https://atigonh-resignation-streamlitattrition-app-lgwauk.streamlit.app/>

**Input Features**

Age

18 37 59

Business Travel

☒ Yes  
☐ No

Department

Research & Development

Distance from Home (miles)

1 5 29

Education Field

### Resignation Prediction

(Edit features on left panel)

**Prediction: Stay**

(Probability to Leave = 0.476)

server: online

	0
Age	37
BusinessTravel	1
Department	Research & Development
DistanceFromHome	5
EducationField	Life Sciences
EnvironmentSatisfaction	3
JobInvolvement	3
JobLevel	2
JobRole	Sales Executive
JobSatisfaction	3
MaritalStatus	0
salary_range	6250
NumCompaniesWorked	2

# Evaluation

# Evaluation

Actual:	Actual:0	181 +2% TN	5 -44% FP
	Actual:1	21 -8% FN	14 -7% TP
		Predicted:0	Predicted:1

# Evaluation

- Recall
  - Baseline 0.36
  - Production model 0.40
- Precision
  - Baseline 0.59
  - Production model 0.88
- F1
  - Baseline 0.45
  - Production model 0.55 (+22%)

# Summary & Recommendations

# Summary

- The model met the expectations by improving the F1-Score by 22%. (Expected 20%)
- Precision is 88% (guess 9 leave, correct 8)

# Recommendation

- Any future improvements should prioritize high precision and acceptable recall.
- To approach the employee, using this model carefully and with caution is important.

good luck  
in your new job



# Thank you



- **Source**
- **Dataset:** <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>
- Image: <https://silo.tips/download/hidden-roi-of-talent-acquisition-and-mobility>
- **Image:** <https://www.google.co.th/url?sa=i&url=https%3A%2F%2Fwww.pinterest.com%2Fpin%2F678917712545864019%2F&psig=AOvVaw32QiSD437M9wg29P3NnISQ&ust=1677346125350000&source=images&cd=vfe&ved=0CA8QjRxqFwoTCOjkvbHXrv0CFQAAAAAdAAAAABAE>
- ChatGPT