

# HR Data Analysis: Insights and Implications for Decision-Making

Exploring key insights from HR data to inform strategic decision-making processes.



# Meet Our Team

- Mostafa Bakry AboElhamd
- Esraa Ayman Abd-Elgany Elhofy
- Menna-allah Taha Saadeldin
- Rahma Essam Hassan





# Project's Overview

Our graduation project focuses on HR Data Analysis. The goal of the project was to understand the key factors influencing employee behavior especially attrition and to support HR decision-making using data-driven insights.

We worked with two datasets: employee information and performance ratings. Our objective was to clean, analyze, and visualize this data to discover patterns that can help organizations reduce employee turnover, improve satisfaction, and strengthen overall workforce management.”



## Present Key Findings from HR Data Analysis

The HR data analysis project aims to showcase critical findings that reveal patterns and trends in workforce behavior, performance metrics, and overall organizational health. This analysis serves as the foundation for informed decision-making within the HR department and beyond.

## Discuss Implications for HR Decision-Making

Understanding the implications of the HR data analysis results is crucial for strategic HR decision-making. This includes identifying opportunities for workforce optimization, enhancing talent acquisition strategies, and improving retention efforts, ultimately aligning HR practices with organizational goals.

## Provide Recommendations for Future Analysis

Based on the insights gathered, this presentation will provide actionable recommendations for future HR data analysis initiatives. These recommendations will focus on improving data collection practices, leveraging new analytical tools, and fostering a data-driven culture within the HR team to enhance overall effectiveness.



# Objectives of the Presentation

# Present Key Findings from HR Data Analysis

The HR data analysis project aims to showcase critical findings that reveal patterns and trends in workforce behavior, performance metrics, and overall organizational health. This analysis serves as the foundation for informed decision-making within the HR department and beyond.



# Discuss Implications for HR Decision-Making

Understanding the implications of the HR data analysis results is crucial for strategic HR decision-making. This includes identifying opportunities for workforce optimization, enhancing talent acquisition strategies, and improving retention efforts, ultimately aligning HR practices with organizational goals.

# Provide Recommendations for Future Analysis

Based on the insights gathered, this presentation will provide actionable recommendations for future HR data analysis initiatives. These recommendations will focus on improving data collection practices, leveraging new analytical tools, and fostering a data-driven culture within the HR team to enhance overall effectiveness.

# Project Explanation & Key Features



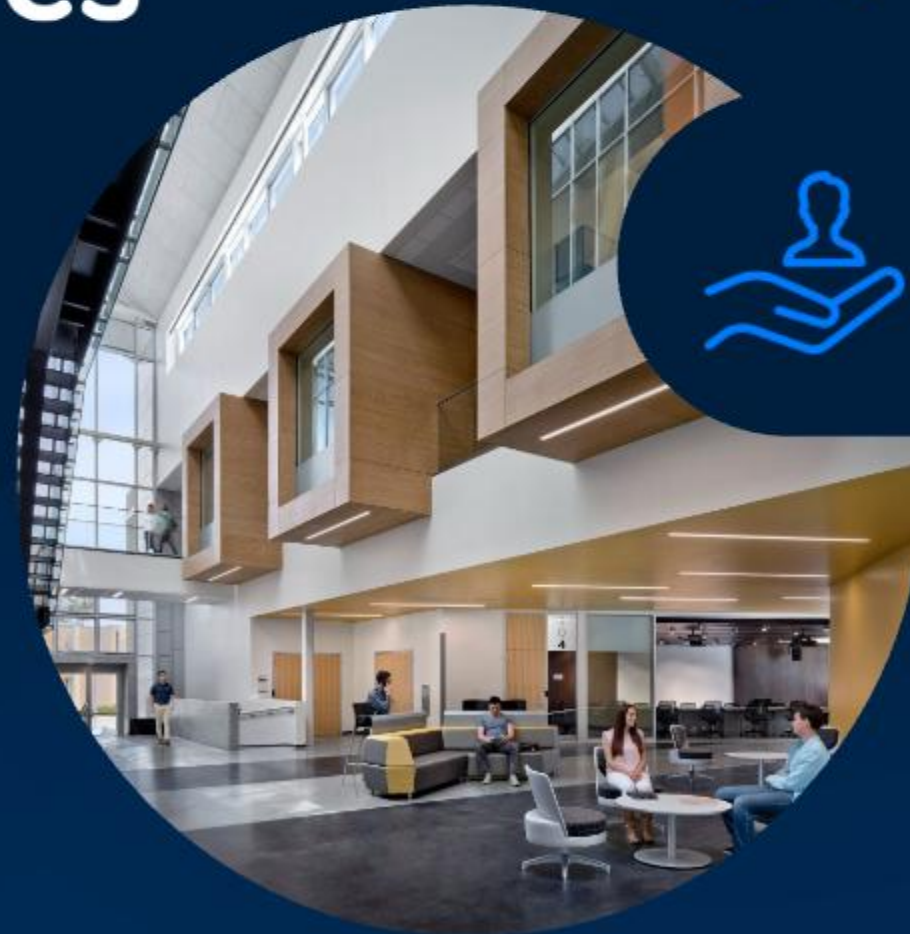
Part 1 — Data Cleaning



Part 2 — Data Analysis



Part 3 — Dashboard







## Data Cleaning

Ensured that both datasets were clean and reliable.

Checked:

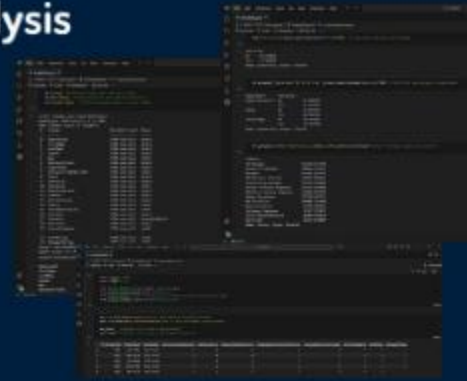
- Data structure and column consistency
- Unique values in categorical columns
- Duplicate records
- Missing values
- Outliers in numerical fields

Also corrected inconsistent values, such as fixing duplicated 'Marketing' variations in the Education Field column. After cleaning, the datasets became fully consistent and ready for merging.



## Data Analysis

- Merged datasets
- Attrition analysis
- Salary trends
- Age groups & attrition
- Promotion impact
- Business travel effect
- Correlation insights



## Rigorous Data Cleaning and Preparation

Prior to analysis, data underwent rigorous cleaning and preparation to ensure accuracy and reliability. This process involved eliminating duplicates, correcting errors, and standardizing formats to create a clean dataset.



## Key Insights

- Attrition rate: 33.7%
- Highest attrition: Sales
- Lower attrition: Technology
- Younger employees leave more
- No promotion  $\Rightarrow$  higher attrition
- Frequent travel  $\Rightarrow$  higher attrition
- Salary has mild negative correlation with attrition

# Project Breakdown & Key Findings





# Data Cleaning

Ensured that both datasets were clean and reliable.

Checked:

- Data structure and column consistency
- Unique values in categorical columns
- Duplicate records
- Missing values
- Outliers in numerical fields

Also corrected inconsistent values, such as fixing duplicated 'Marketing' variations in the Education Field column. After cleaning, the datasets became fully consistent and ready for merging."



## HR Data Analysis

```
In [17]: import pandas as pd

df = pd.read_csv('Employee.csv') # to load csv files
df_second = pd.read_csv('PerformanceRating.csv')
```

## Cleaning process for : Employee.csv

### 1- Initial Review

to confirm the data is loaded successfully and to have an idea about the data we are working with

```
In [18]: print(df.head()) # show first few rows
```

to spot any typos or inconsistencies

```
In [20]: print(df.columns) #Lists all column names

Index(['EmployeeID', 'FirstName', 'LastName', 'Gender', 'Age',
       'BusinessTravel', 'Department', 'DistanceFromHome (KM)', 'State',
       'Ethnicity', 'Education', 'EducationField', 'JobRole', 'MaritalStatus',
       'Salary', 'StockOptionLevel', 'OverTime', 'HireDate', 'Attrition',
       'YearsAtCompany', 'YearsInMostRecentRole', 'YearsSinceLastPromotion',
       'YearsWithCurrManager'],
      dtype='object')
```

define columns with categorical values (non\_numeric values)

```
In [21]: categorical_cols_to_check = [
        'Gender', 'BusinessTravel', 'Department', 'State',
        'Ethnicity', 'EducationField', 'JobRole', 'MaritalStatus',
        'OverTime', 'Attrition'
    ]
```

Loop through the list to Check for Unique Values to find (Inconsistencies & Typos)

### 2- Statistical Overview

```
In [24]: print(df.shape) #number of rows, columns

(1470, 23)
```

to spot outliers

```
In [25]: print(df.describe()) #Provides summary statistics
```

### 3- checking for duplicates and missing values

```
In [26]: # Handling duplicates
df.duplicated().sum()
```

```
Out[26]: np.int64(0)
```

there are no duplicates

```
In [27]: # checking missing value
df.isnull().sum()
```





# Data Analysis

- Merged datasets
- Attrition analysis
- Salary trends
- Age groups & attrition
- Promotion impact
- Business travel effect
- Correlation insights

```
df.info() # Show data types and non-null counts
df.describe() # Generate basic descriptive statistics
df.isna().sum() # Count missing values in each column
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6709 entries, 0 to 6708
Data columns (total 33 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   EmployeeID          6709 non-null  object
 1   FirstName           6709 non-null  object
 2   LastName            6709 non-null  object
 3   Gender              6709 non-null  object
 4   Age                 6709 non-null  int64
 5   BusinessTravel      6709 non-null  object
 6   Department          6709 non-null  object
 7   DistanceFromHome (KM) 6709 non-null  int64
 8   State               6709 non-null  object
 9   Ethnicity           6709 non-null  object
10   Education            6709 non-null  int64
11   EducationField      6709 non-null  object
12   JobRole             6709 non-null  object
13   MaritalStatus       6709 non-null  object
14   Salary              6709 non-null  int64
15   StockOptionLevel    6709 non-null  int64
16   OverTime            6709 non-null  object
17   HireDate            6709 non-null  datetime64[ns]
18   Attrition           6709 non-null  object
19   YearsAtCompany      6709 non-null  int64
...
31  SelfRating          6709 non-null  int64
32  ManagerRating       6709 non-null  int64
dtypes: datetime64[ns]
memory usage: 1.7+ MB
Output is truncated. View
```

```
df.groupby("Department")["Attrition"].value_counts(normalize=True)*100 # Attrition percentage by department
```

Department	Attrition	Proportion
Human Resources	No	62.046205
	Yes	37.953795
Sales	No	59.097255
	Yes	40.902745
Technology	No	70.237256
	Yes	29.762744

```
df.groupby("JobRole")["Salary"].mean().sort_values(ascending=False) # Average salary by job role
```

JobRole	Average Salary
HR Manager	425598.537500
Analytics Manager	350824.745192
Manager	331405.825714
HR Business Partner	313027.869565
Engineering Manager	292449.519808
Senior Software Engineer	134137.556688
Machine Learning Engineer	133849.879749
Sales Executive	127258.367521
HR Executive	103086.591384
Data Scientist	57254.591176
Software Engineer	54347.353073
Sales Representative	41899.194274
Recruiter	41635.651007

```
import pandas as pd
import numpy as np

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.metrics import roc_auc_score
```

```
emp = pd.read_excel("employee.xlsx") # Read employee information dataset
perf = pd.read_excel("performance.xlsx") # Read performance rating dataset

emp.head() # Display first 5 rows of employee data
perf.head() # Display first 5 rows of performance rating data
```

PerformanceID	EmployeeID	ReviewDate	EnvironmentSatisfaction	JobSatisfaction	RelationshipSatisfaction	TrainingOpportunitiesWithinYear	TrainingOpportunitiesTaken	WorkLifeBalance	SelfRating	ManagerRating
0	PR01	2017-01-02	5	4	5	1	0	4	4	4
1	PR02	2017-01-03	5	4	4	1	3	4	4	3
2	PR03	2017-01-01	3	4	5	3	2	3	5	4
3	PR04	2017-01-04	5	3	2	2	0	2	3	2
4	PR05	2017-01-04	5	2	3	1	0	4	4	3



# Key Insights

- Attrition rate: 33.7%
- Highest attrition: Sales
- Lower attrition: Technology
- Younger employees leave more
- No promotion  $\square$  higher attrition
- Frequent travel  $\square$  higher attrition
- Salary has mild negative correlation with attrition



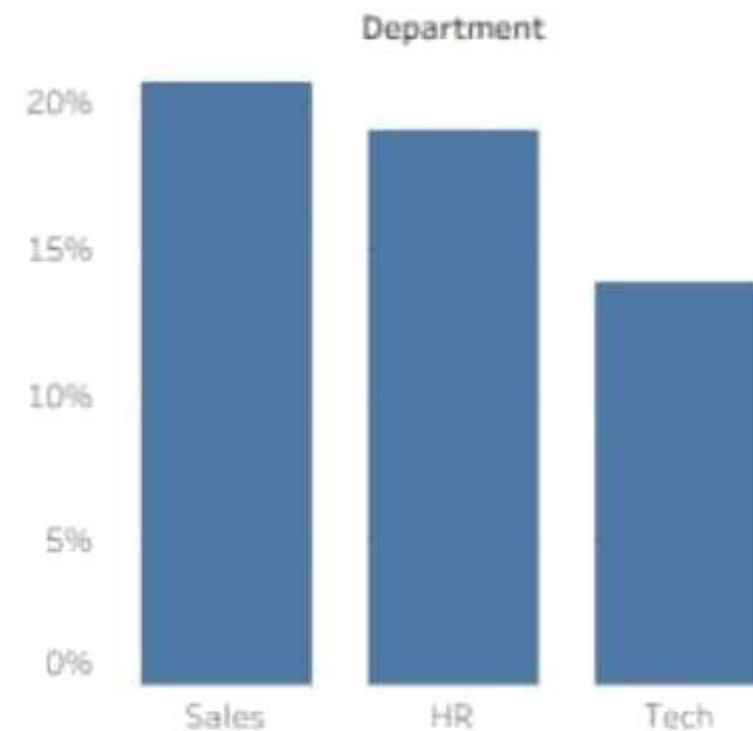
# Dashboard



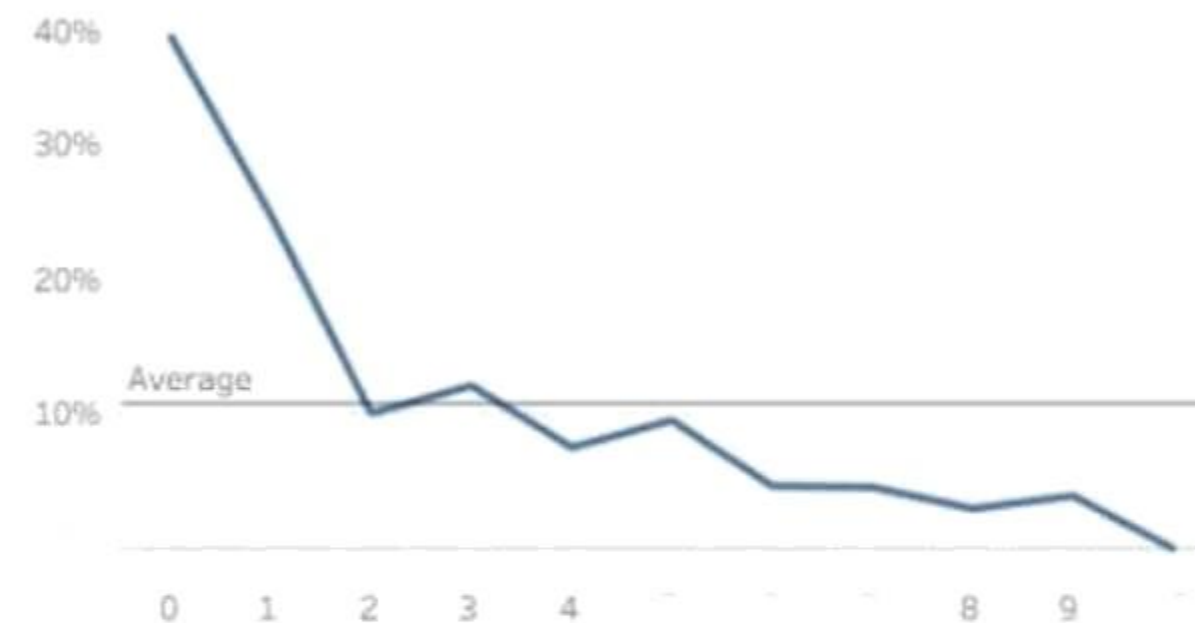
## Career

## KPIs

Employee Count	1.470
Attrition Percent	16.12
Average Salary	5113



## Attrition by Years since Last Promotion



Department

(All)

Years Since Last Promotion (bin)

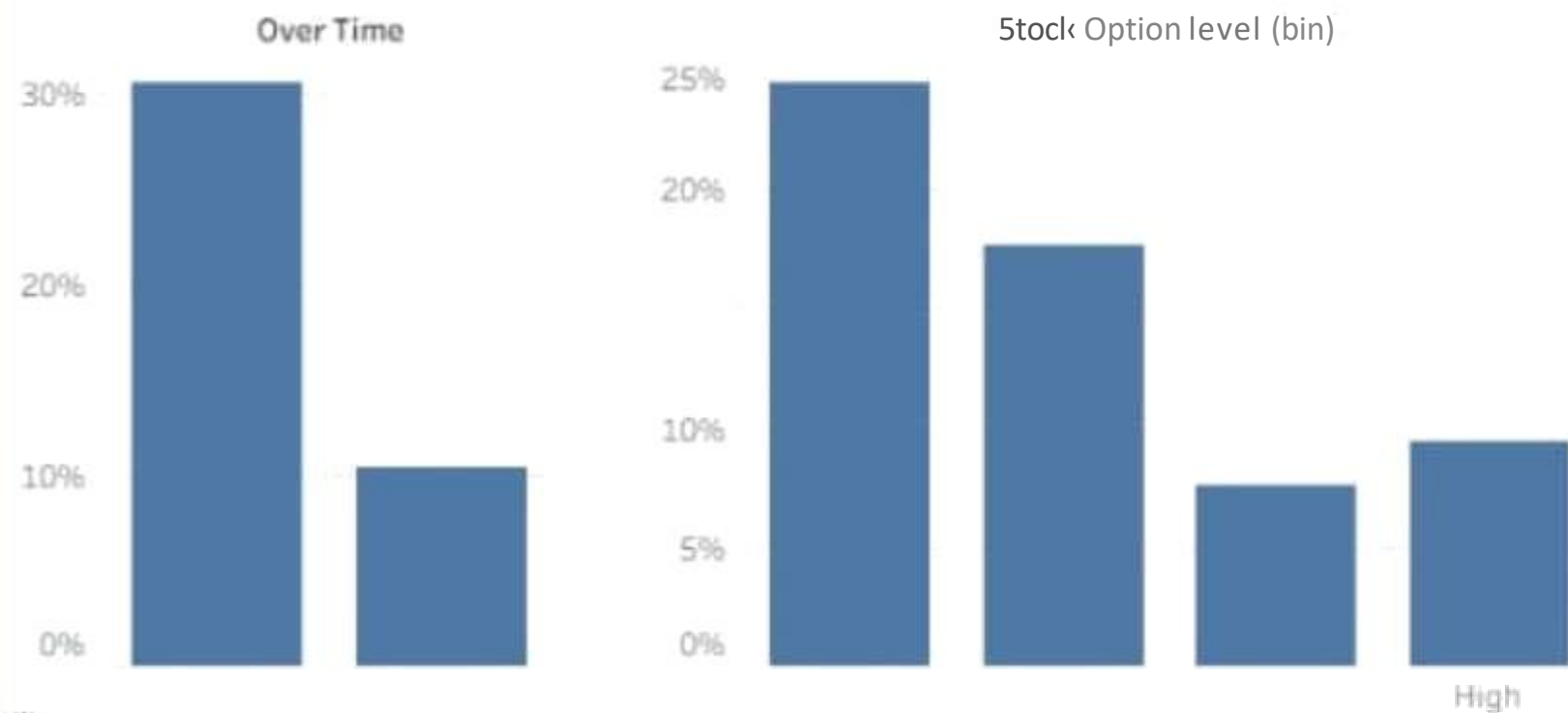
(All)

Avg. Salary

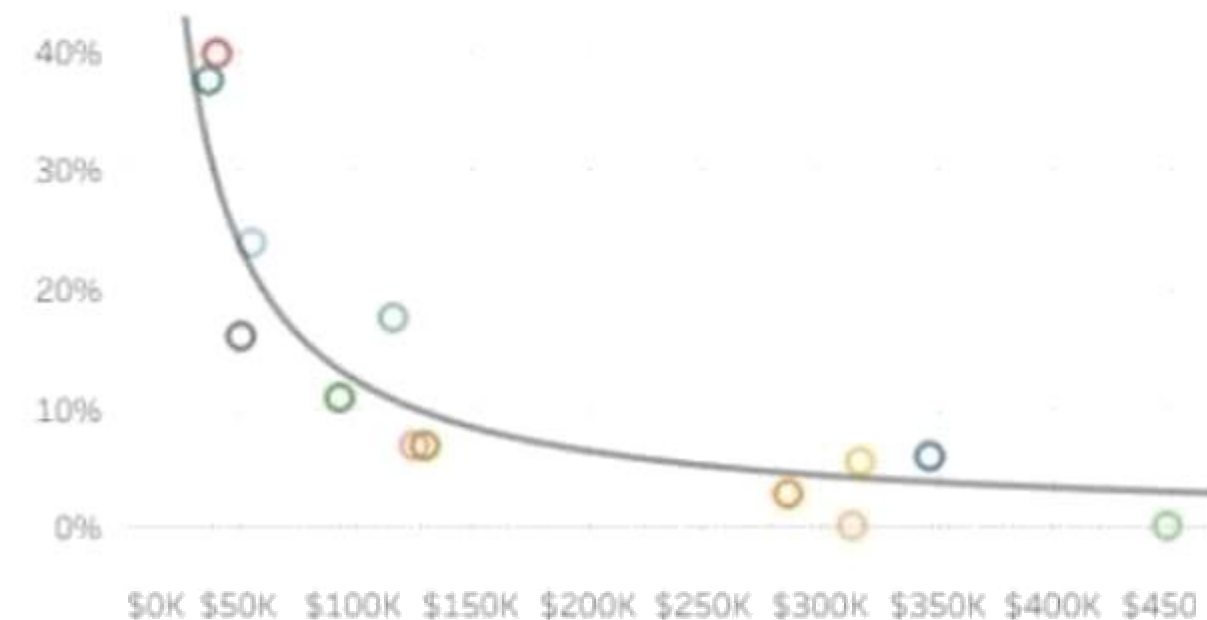
\$449K

Over Time

(All)



## Attrition by Salary for Jop Roles



Stock Option Level (bin)

(All)

Job Role

- Analytics Manager
- Data Scientist
- Engineering Manager
- HR Business Partner
- HR Executive
- HR Manager
- Machine Learning Engin..



Prezi

Opportunities...

Attr\_by\_StockOption

Attr\_by\_OverTime

Corr

JobSatis

WorkLifeBalance

EnvironmentSatis

RelationshipSatis

SatisfactionLevel

Personal Information

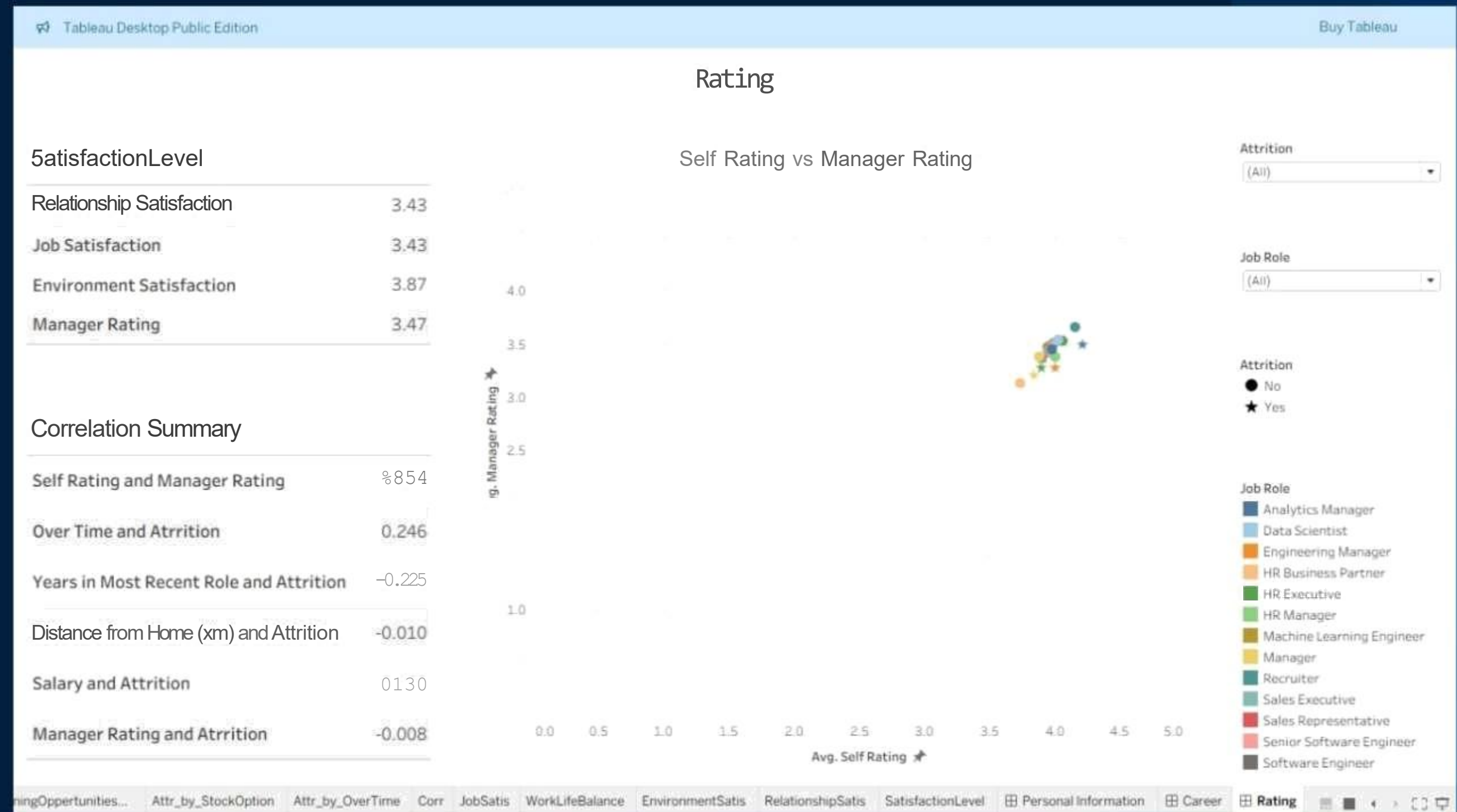
Career

Rating



# Dashboard

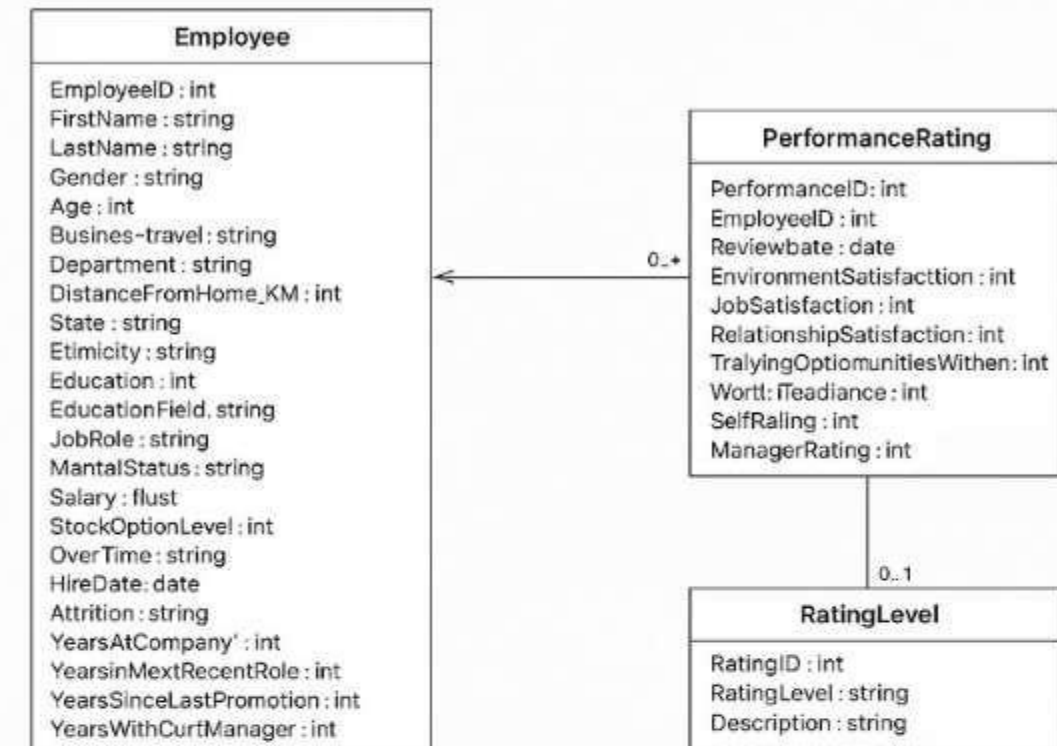
- Interactive KPIs
- Attrition by department & role
- Salary visualization
- Satisfaction metrics
- Promotions & travel insights



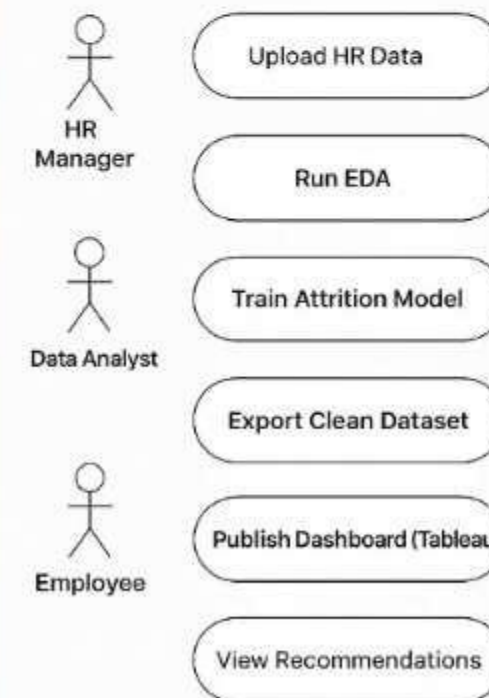
# Future Development

- Predict attrition with ML
- HR system automation
- Alerts & recommendations
- Enterprise-level analytics product
- Dashboard integration with HR databases

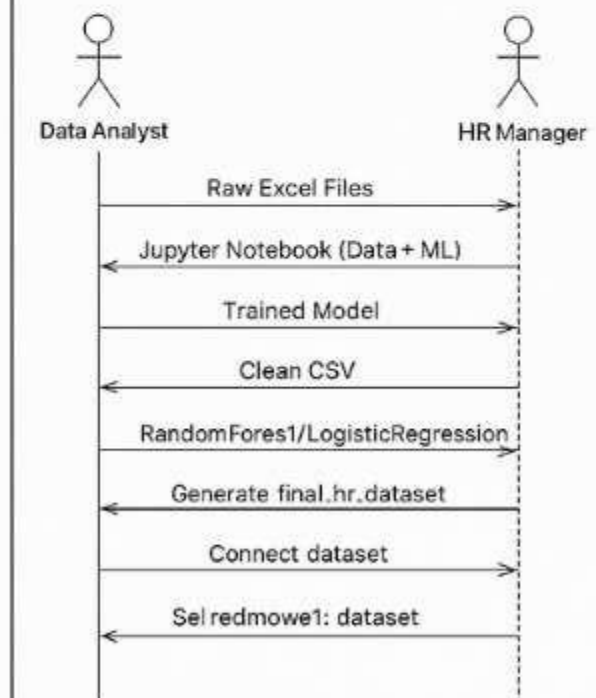
Class Diagram



Use-Case Diagram



Sequence Diagram







# Conclusion

- Full HR analytics pipeline
- Clear insights
- Ready for real-world application
- Strong foundation for future expansion

A person in a dark suit is holding a large, detailed architectural model of a modern building. The model is made of clear material and shows the internal structure and layout of the building. The person is standing in front of a blurred background that appears to be an office or a construction site.

# Thank You