

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.

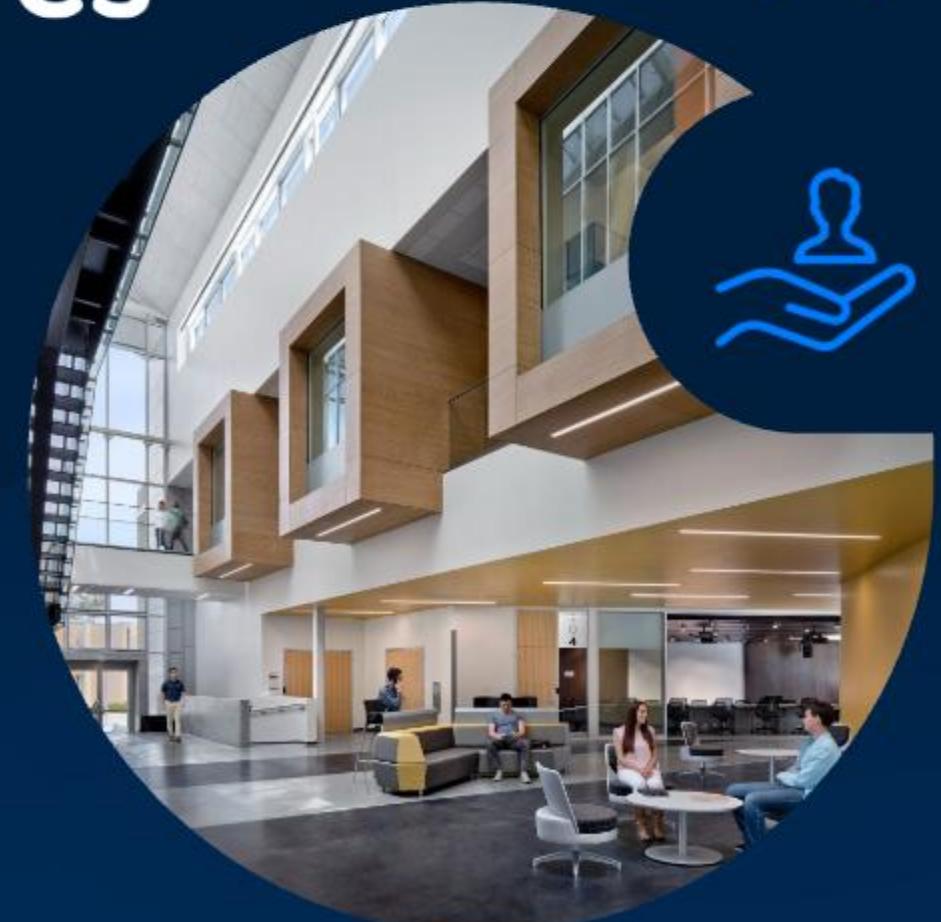
Discuss Implications for HR Decision-Making

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:

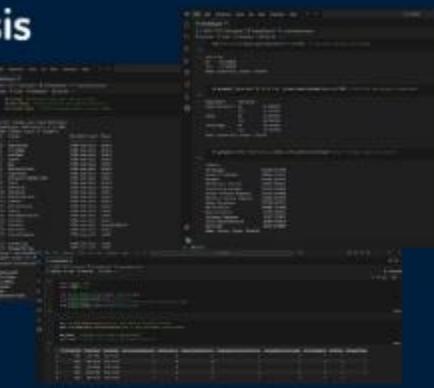
- Data structure and column consistency
- Unique values in categorical columns
- Duplicate records
- Missing values
- Outliers in numerical fields

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 → higher attrition
- Frequent travel → 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

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

The screenshot shows a Jupyter Notebook environment with three code cells and their corresponding outputs.

Code Cell 1:

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

Output 1:

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 6799 entries, 0 to 6798  
Data columns (total 33 columns):  
 #   Column           Non-Null Count  Dtype     
---  
 0   EmployeeID      6799 non-null   object    
 1   FirstName       6799 non-null   object    
 2   LastName        6799 non-null   object    
 3   Gender          6799 non-null   object    
 4   Age              6799 non-null   int64     
 5   BusinessTravel  6799 non-null   object    
 6   Department       6799 non-null   object    
 7   DistanceFromHome 6799 non-null   int64     
 8   State            6799 non-null   object    
 9   Ethnicity        6799 non-null   object    
 10  Education         6799 non-null   int64     
 11  EducationField   6799 non-null   object    
 12  JobRole          6799 non-null   object    
 13  MaritalStatus    6799 non-null   object    
 14  Salary            6799 non-null   int64     
 15  StockOptionLevel 6799 non-null   int64     
 16  OverTime         6799 non-null   object    
 17  HireDate         6799 non-null   datetime64[ns]   
 18  Attrition        6799 non-null   object    
 19  YearsAtCompany   6799 non-null   int64     
 ...  
 31  Satisfying       6799 non-null   int64     
 32  ManagerRating    6799 non-null   int64     
dtypes: datetime64[ns](1), float64(1), int64(31), object(2)  
memory usage: 1.7e+09  
Output was truncated. View the full output in the notebook.
```

Code Cell 2:

```
df['Attrition'].value_counts(normalize=True)*100 # Calculate attrition percentage
```

Output 2:

```
No: 66.299991  
Yes: 33.700999  
Name: proportion, dtype: float64
```

Code Cell 3:

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

Output 3:

Department	Attrition	Value
Human Resources	No	62.046295
Human Resources	Yes	37.953795
Sales	No	59.097255
Sales	Yes	40.902745
Technology	No	78.237256
Technology	Yes	29.762744

Code Cell 4:

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

Output 4:

JobRole	Salary
IR Manager	425598.937900
Analytics Manager	356924.745192
Manager	333485.835724
HR Business Partner	313827.869545
Engineering Manager	292449.519848
Senior Software Engineer	134137.556688
Machine Learning Engineer	133849.899749
Sales Executive	127258.367521
HR Executive	103806.991384
Data Scientist	57254.595176
Software Engineer	56247.253073
Sales Representative	41899.194274
Recruiter	41635.651087

Code Cell 5:

```
import pandas as pd  
import numpy as np
```

Code Cell 6:

```
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.ensemble import RandomForestClassifier
```

Code Cell 7:

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

Output 5:

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

Output 6:

PerformanceID	EmployeeID	ReviewDate	EnvironmentSatisfaction	JobSatisfaction	RelationshipSatisfaction	TrainingOpportunitiesWithinYear	TrainingOpportunitiesTaken	WorkLifeBalance	Satisfying	ManagerRating
0	P001	2017-08-01	2013-01-02	5	4	5	1	3	4	4
1	P002	2017-09-01	2013-01-03	5	4	4	1	3	4	3
2	P003	2017-08-08	2013-01-01	3	4	5	3	2	3	5
3	P004	2017-08-04	2013-01-04	5	3	2	2	0	2	3
4	P005	2017-08-10	2013-01-04	5	2	3	1	0	4	3



Key Insights

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

Dashboard



Career

KPIs

Employee Count

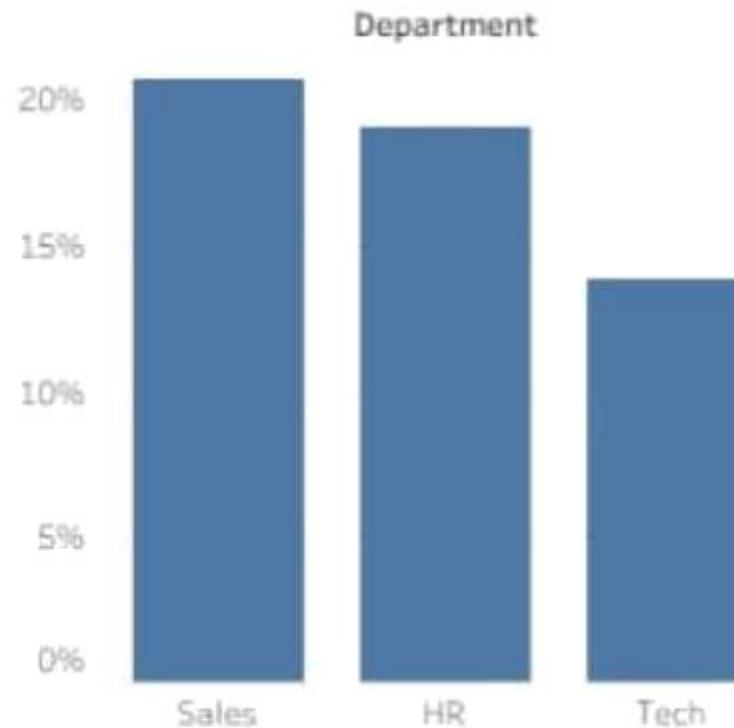
1,470

Attrition Percent

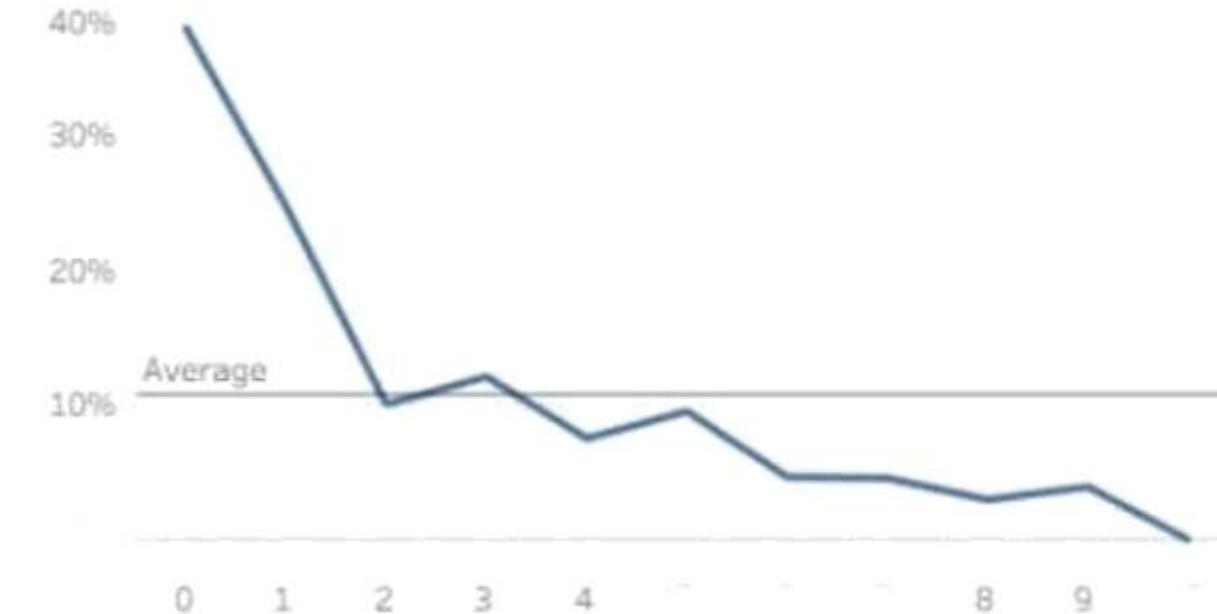
16.12

Average Salary

5113



Attrition by Years since Last Promotion



Department

(All)

Years Since Last Promotion (bin)

(All)

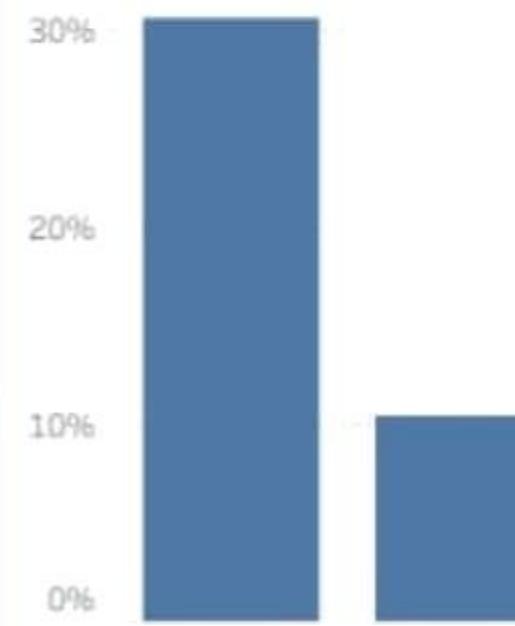
Avg. Salary

\$449K

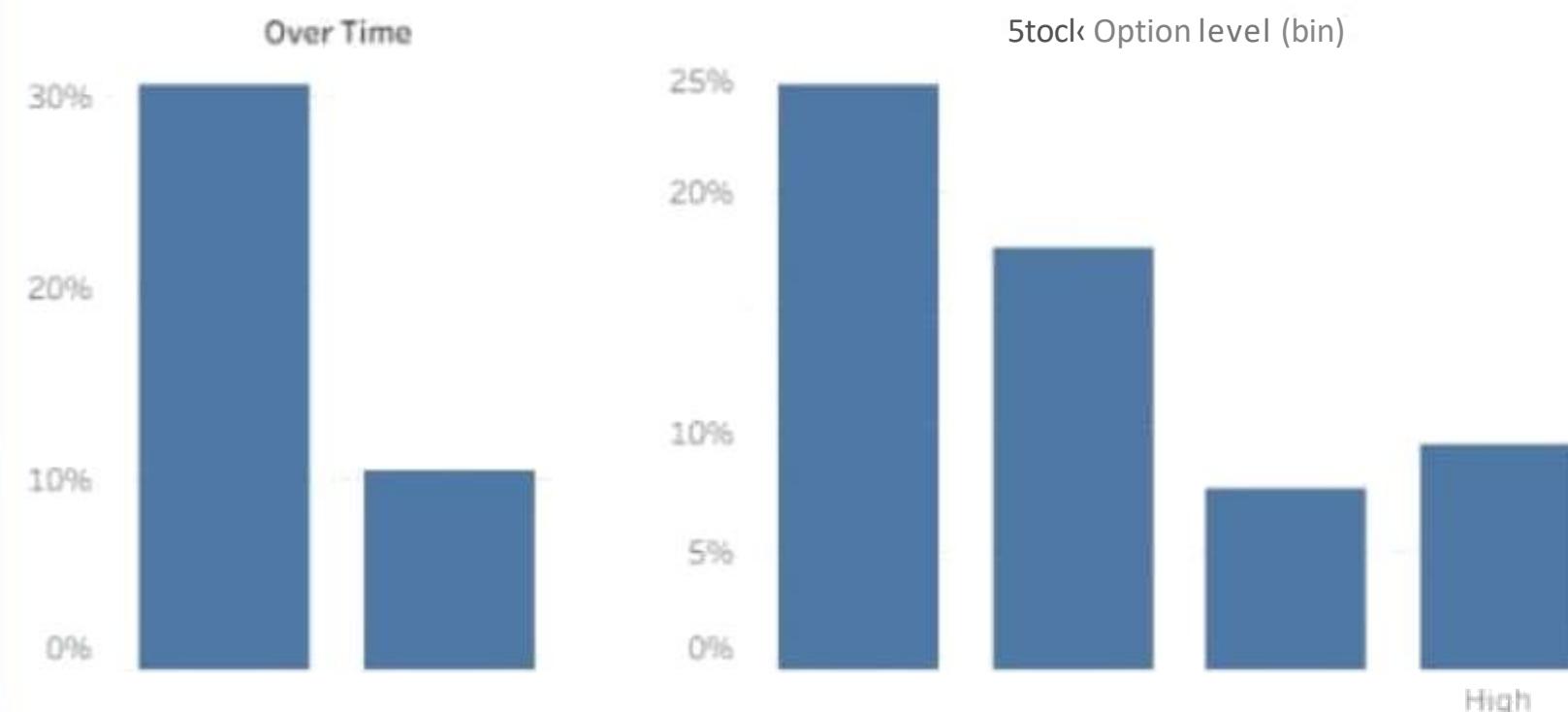
Over Time

(All)

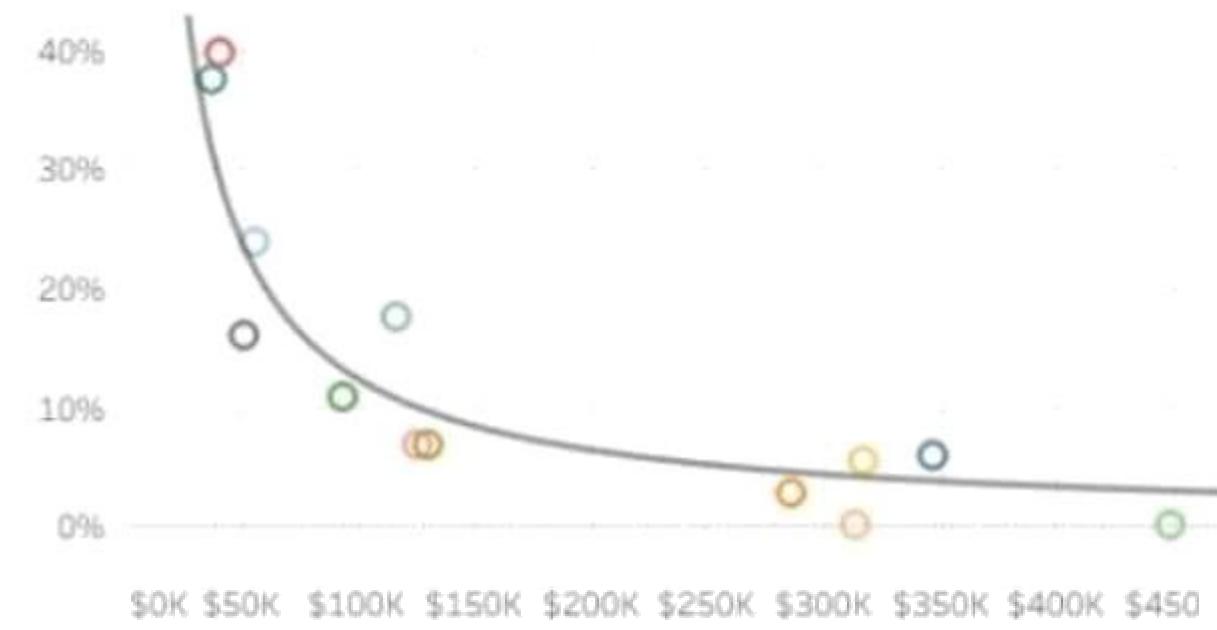
Over Time



Stock Option level (bin)



Attrition by Salary for Job Roles



Stock Option Level (bin)

(All)

Job Role

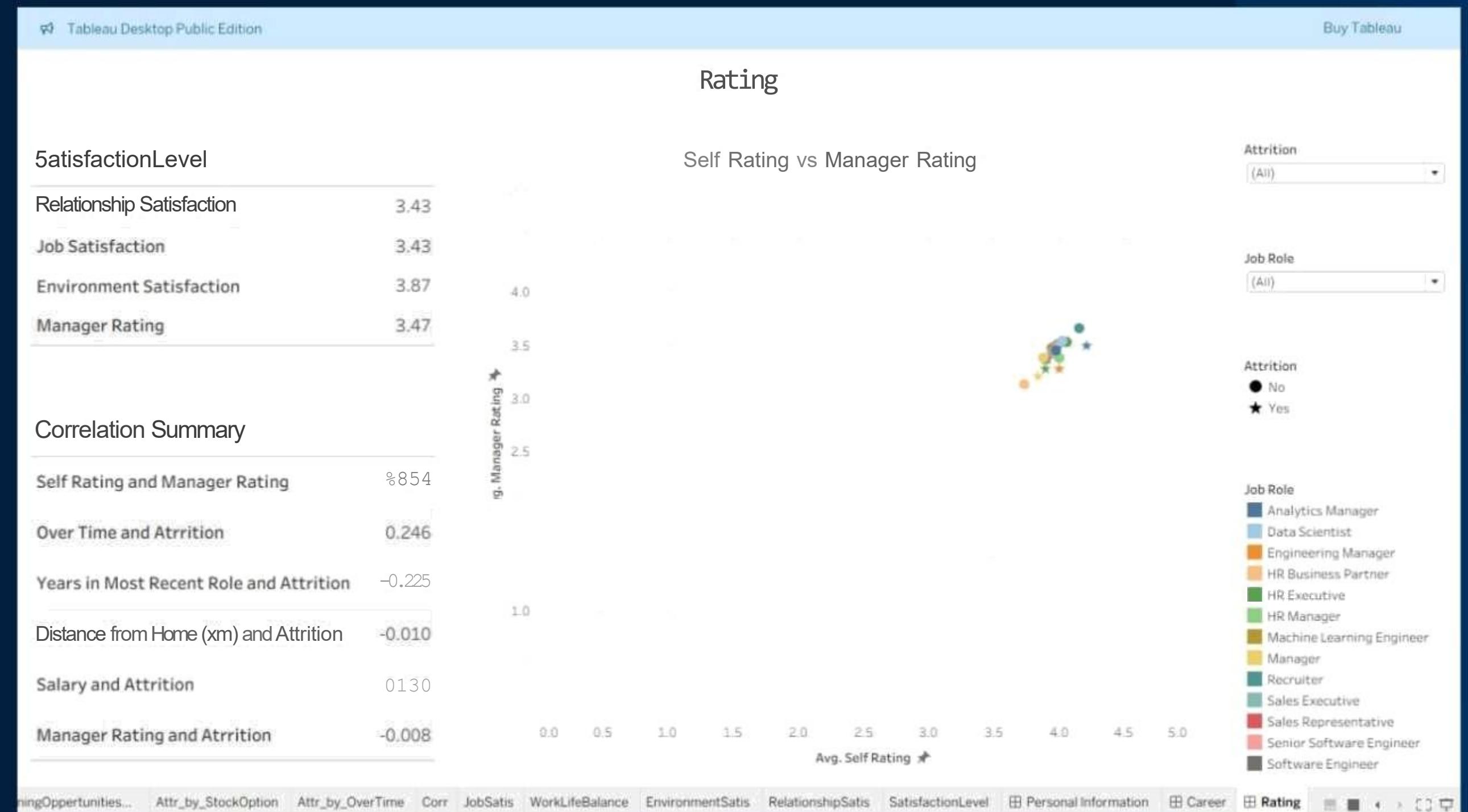
- Analytics Manager
- Data Scientist
- Engineering Manager
- HR Business Partner
- HR Executive
- HR Manager
- Machine Learning Engin..



Prezi

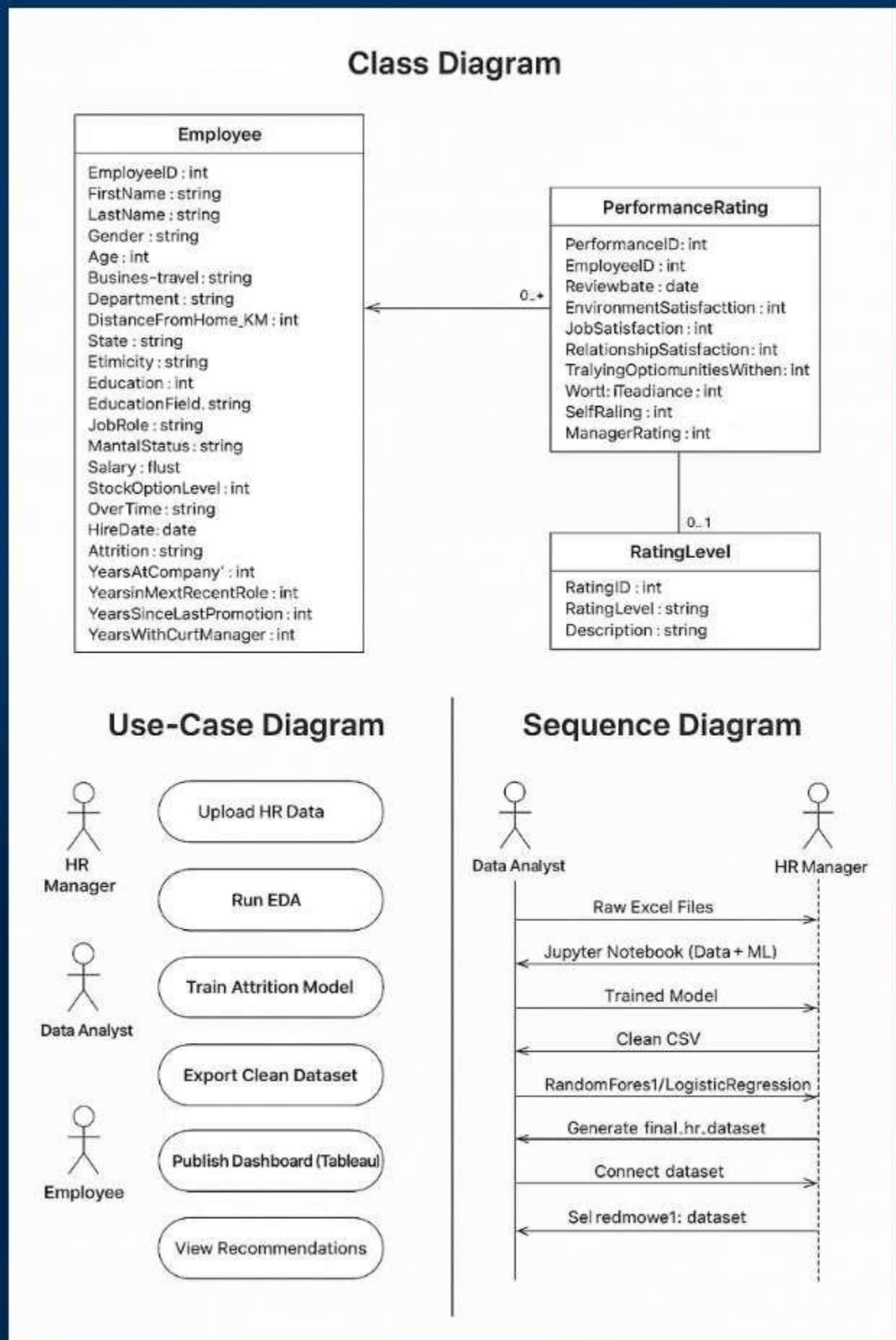
Dashboard

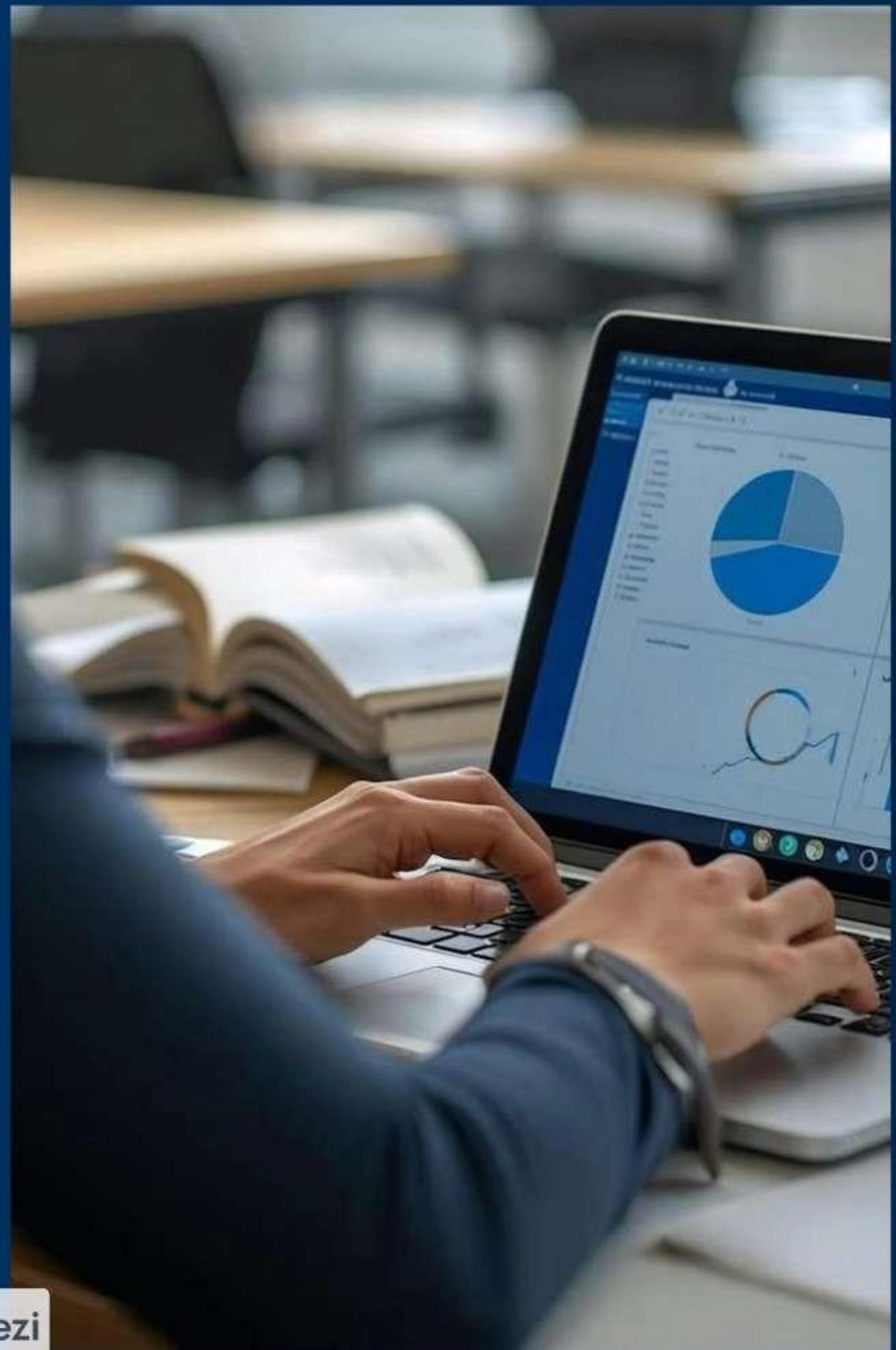
- Interactive KPIs
- Attrition by department & role
- Salary visualization
- Satisfaction metrics
- Promotions & travel insights



Future Development

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





Conclusion

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



Thank You