JULY 22, 2024

EMPLOYEE ATTRITION ANALYSIS AND PREDICTION

ALISHBA RIZWAN
TECHNOCOLABS SOFTWARES
alishbarizwanakbarmirza@gmail.com

Table of Contents

Introduction 3

Background: 3

Problem Statement: 3

Objectives: 3

Data Overview 3

Data Preprocessing and Cleaning 4

Data Exploration: 4

Data Cleaning: 8

Data Encoding: 14

Data Labeling: 16

Dashboard Building and Reporting 17

Results 19

Visualizations: 19

Key Findings: 20

Recommendations 20

Conclusion 21

Future Work: 21

List of Figures

Figure 1: Acme Corporation Data Overview 3
Figure 2: Data Exploration code with results.................... 4
Figure 3: Summary Statistics in Data Exploration.................... 5
Figure 4: Data Information coding in Data Exploration 6
Figure 5: Printing Heads of Data Set 7
Figure 6: Data Set Information 8
Figure 7: Handling Missing Values.................... 8
Figure 8: handling Missing Values.................... 9
Figure 9: Historic Numerical Features of Data Set.................... 10
Figure 10: Count Plot for Attrition....................
Figure 11: Count Plot for Business Travels ....
Figure 12: Count Plot for Departments....................
Figure 13: Count Plot for Over 18 ....
Figure 14: Count Plot for Educational Field....................
Figure 15: Count Plot for Overtime ....
Figure 16: Count plot with Gender ....
Figure 17: Count Plot for Job Role ....
Figure 18: Count Plot for Marital status ....
Figure 19: Code for Data Encoding ....
Figure 20: Label Encoding and Hot-Encoding 15
Figure 21: Results of Data Encoding ....
Figure 22: Data Labeling 16
Figure 23: Results of Data Labeling 17
Figure 24: Power BI Dashboard representing Attrition trends 19
Figure 25: Power BI Dashboard representing Attrition Trends 20

Introduction

The Employee Attrition Analysis and Prediction project aims to address Acme Corporation's significant challenge with employee turnover. The analysis used historical data to identify key factors influencing attrition and predict future employee departure.

Background:

Acme Corporation faces increasing employee turnover, impacting team dynamics and overall company morale. Understanding and addressing the reasons behind this attrition is crucial for maintaining a stable workforce.

Problem Statement:

The HR department is concerned about the rising attrition rate. This project seeks to analyze employee data to uncover the root causes and predict future turnover.

Objectives:

Analyze historical employee data to identify patterns and trends.
Predict future employee attrition using advanced analytics techniques.

Data Overview

The dataset includes features such as Employee ID, Age, Attrition, Department, Job Satisfaction, Performance Rating, and more. This data spans the last five years and provides insights into employee demographics and job-related factors.

Age	Attrition	BusinessTı	DailyRate	Departmer	DistanceFı	Education	EducationI	EmployeeC	EmployeeN	Environme	Gender	HourlyRate	JobInvolver	JobLevel	JobRole	JobSatisfa	MaritalStat	MonthlyInc	MonthlyRa	NumComp	Over18	OverTime	P
41	Yes	Travel_Rar	1102	Sales	1	2	Life Scienc	1	1	2	Female	94	3	2	Sales Exec	4	Single	5993	19479	8	Y	Yes	
49	No	Travel_Fre	279	Research &	8	1	Life Scienc	1	2	3	Male	61	2	2	Research S	2	Married	5130	24907	1	Y	No	
37	Yes	Travel_Rar	1373	Research &	2	2	Other	1	4	4	Male	92	2	1	Laboratory	3	Single	2090	2396	6	Y	Yes	
33	No	Travel_Fre	1392	Research &	3	4	Life Scienc	1	5	4	Female	56	3	1	Research S	3	Married	2909	23159	1	Y	Yes	
27	No	Travel_Rar	591	Research &	2	1	Medical	1	7	1	Male	40	3	1	Laboratory	2	Married	3468	16632	9	Y	No	
32	No	Travel_Fre	1005	Research &	2	2	Life Scienc	1	8	4	Male	79	3	1	Laboratory	4	Single	3068	11864	0	Y	No	
59	No	Travel_Rar	1324	Research &	3	3	Medical	1	10	3	Female	81	4	1	Laboratory	1	Married	2670	9964	4	Y	Yes	
30	No	Travel_Rar	1358	Research &	24	1	Life Scienc	1	11	4	Male	67	3	1	Laboratory	3	Divorced	2693	13335	1	Y	No	
38	No	Travel_Fre	216	Research &	23	3	Life Scienc	1	12	4	Male	44	2	3	Manufactu	3	Single	9526	8787	0	Y	No	
36	No	Travel_Rar	1299	Research &	27	3	Medical	1	13	3	Male	94	3	2	Healthcare	3	Married	5237	16577	6	Y	No	
35	No	Travel_Rar	809	Research &	16	3	Medical	1	14	1	Male	84	4	1	Laboratory	2	Married	2426	16479	0	Y	No	
29	No	Travel_Rar	153	Research &	15	2	Life Scienc	1	15	4	Female	49	2	2	Laboratory	3	Single	4193	12682	0	Y	Yes	
31	No	Travel_Rar	670	Research &	26	1	Life Scienc	1	16	1	Male	31	3	1	Research S	3	Divorced	2911	15170	1	Y	No	

Figure 1: Acme Corporation Data Overview

Data Preprocessing and Cleaning

Data Exploration:

- Inspect the dataset to understand the structure and identify key features.
- Perform summary statistics to get an overview of data distributions.

The summary of Data is coded and shown in the figures for the purpose of data exploration.

In [5]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [6]:
```
# Load the dataset
file_path = r"C:\Users\fujisio\Downloads\Project1\EmployeeData.csv"
df = pd.read_csv(file_path)
```

In [7]:
```
# Display the first few rows and column information
print("First few rows of the dataset:")
print(df.head())
```

```
First few rows of the dataset:
   Age Attrition     BusinessTravel  DailyRate              Department  \
0   41       Yes      Travel_Rarely       1102                   Sales
1   49        No  Travel_Frequently        279  Research & Development
2   37       Yes      Travel_Rarely       1373  Research & Development
3   33        No  Travel_Frequently       1392  Research & Development
4   27        No      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  ...                         1            80                 0
1  ...                         4            80                 1
2  ...                         2            80                 0
3  ...                         3            80                 0
4  ...                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
```

Figure 2: Data Exploration code with results

In [8]:
```
# Summary statistics
print("\nSummary statistics:")
print(df.describe())
```

```
Summary statistics:
               Age    DailyRate  DistanceFromHome    Education  EmployeeCount  \
count  1470.000000  1470.000000       1470.000000  1470.000000         1470.0
mean     36.923810   802.485714          9.192517     2.912925            1.0
std       9.135373   403.509100          8.106864     1.024165            0.0
min      18.000000   102.000000          1.000000     1.000000            1.0
25%      30.000000   465.000000          2.000000     2.000000            1.0
50%      36.000000   802.000000          7.000000     3.000000            1.0
75%      43.000000  1157.000000         14.000000     4.000000            1.0
max      60.000000  1499.000000         29.000000     5.000000            1.0

       EmployeeNumber  EnvironmentSatisfaction   HourlyRate  JobInvolvement  \
count     1470.000000              1470.000000  1470.000000     1470.000000
mean      1024.865306                 2.721769    65.891156        2.729932
std        602.024335                 1.093082    20.329428        0.711561
min          1.000000                 1.000000    30.000000        1.000000
25%        491.250000                 2.000000    48.000000        2.000000
50%       1020.500000                 3.000000    66.000000        3.000000
75%       1555.750000                 4.000000    83.750000        3.000000
max       2068.000000                 4.000000   100.000000        4.000000

          JobLevel  ...  RelationshipSatisfaction  StandardHours  \
count  1470.000000  ...               1470.000000         1470.0
mean      2.063946  ...                  2.712245           80.0
std       1.106940  ...                  1.081209            0.0
min       1.000000  ...                  1.000000           80.0
25%       1.000000  ...                  2.000000           80.0
50%       2.000000  ...                  3.000000           80.0
75%       3.000000  ...                  4.000000           80.0
max       5.000000  ...                  4.000000           80.0
```

Figure 3: Summary Statistics in Data Exploration

```
In [9]: # Information about the dataset
        print("\nDataset information:")
        print(df.info())
```

```
Dataset information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64
 24  PerformanceRating         1470 non-null   int64
 25  RelationshipSatisfaction  1470 non-null   int64
 26  StandardHours             1470 non-null   int64
 27  StockOptionLevel          1470 non-null   int64
 28  TotalWorkingYears         1470 non-null   int64
 29  TrainingTimesLastYear     1470 non-null   int64
 30  WorkLifeBalance           1470 non-null   int64
 31  YearsAtCompany            1470 non-null   int64
 32  YearsInCurrentRole        1470 non-null   int64
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
None
```

Figure 4: Data Information coding in Data Exploration

In [22]:

```python
# Display the updated dataframe
print("\nEncoded dataset:")
print(df.head())
```

```
Encoded dataset:
   Age  Attrition  DailyRate  DistanceFromHome  Education  \
0   41          1       1102                 1          2
1   49          0        279                 8          1
2   37          1       1373                 2          2
3   33          0       1392                 3          4
4   27          0        591                 2          1

   EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
0                        2          94               3         2
1                        3          61               2         2
2                        4          92               2         1
3                        4          56               3         1
4                        1          40               3         1

   JobSatisfaction  ...  JobRole_Laboratory Technician  JobRole_Manager  \
0                4  ...                          False            False
1                2  ...                          False            False
2                3  ...                           True            False
3                3  ...                          False            False
4                2  ...                           True            False

   JobRole_Manufacturing Director  JobRole_Research Director  \
0                           False                      False
1                           False                      False
2                           False                      False
```

Figure 5: Printing Heads of Data Set

```
[8 rows x 26 columns]
```

```
]: # Information about the dataset
   print("\nDataset information:")
   print(df.info())
```

```
Dataset information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   Age                       1470 non-null   int64 
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64 
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64 
 6   Education                 1470 non-null   int64 
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64 
 9   EmployeeNumber            1470 non-null   int64 
 10  EnvironmentSatisfaction   1470 non-null   int64 
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64 
 13  JobInvolvement            1470 non-null   int64 
 14  JobLevel                  1470 non-null   int64 
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64 
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64 
 19  MonthlyRate               1470 non-null   int64 
 20  NumCompaniesWorked        1470 non-null   int64 
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64 
 24  PerformanceRating         1470 non-null   int64 
 25  RelationshipSatisfaction  1470 non-null   int64 
 26  StandardHours             1470 non-null   int64 
 27  StockOptionLevel          1470 non-null   int64 
 28  TotalWorkingYears         1470 non-null   int64 
 29  TrainingTimesLastYear     1470 non-null   int64 
 30  WorkLifeBalance           1470 non-null   int64 
 31  YearsAtCompany            1470 non-null   int64 
 32  YearsInCurrentRole        1470 non-null   int64 
 33  YearsSinceLastPromotion   1470 non-null   int64 
 34  YearsWithCurrManager      1470 non-null   int64 
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
None
```

Figure 6: Data Set Information

Data Cleaning:

- Handle missing values by imputing or removing them as appropriate.
- Remove duplicates and irrelevant data points.
- Standardize data formats for consistency.

```
In [10]: # Check for missing values
         print("\nMissing values:")
         print(df.isnull().sum())
```

Figure 7: Handling Missing Values

```
Missing values:
Age                         0
Attrition                   0
BusinessTravel              0
DailyRate                   0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
EmployeeNumber              0
EnvironmentSatisfaction     0
Gender                      0
HourlyRate                  0
JobInvolvement              0
JobLevel                    0
JobRole                     0
JobSatisfaction             0
MaritalStatus               0
MonthlyIncome               0
MonthlyRate                 0
NumCompaniesWorked          0
Over18                      0
OverTime                    0
PercentSalaryHike           0
PerformanceRating           0
RelationshipSatisfaction    0
StandardHours               0
StockOptionLevel            0
TotalWorkingYears           0
TrainingTimesLastYear       0
WorkLifeBalance             0
YearsAtCompany              0
YearsInCurrentRole          0
YearsSinceLastPromotion     0
YearsWithCurrManager        0
dtype: int64
```

[11]:

```
# Check unique values for categorical columns
print("\nUnique values for categorical columns:")
for col in df.select_dtypes(include=['object']):
    print(f"{col}: {df[col].unique()}")
```

```
Unique values for categorical columns:
Attrition: ['Yes' 'No']
BusinessTravel: ['Travel_Rarely' 'Travel_Frequently' 'Non-Travel']
Department: ['Sales' 'Research & Development' 'Human Resources']
EducationField: ['Life Sciences' 'Other' 'Medical' 'Marketing' 'Technical Degree'
 'Human Resources']
Gender: ['Female' 'Male']
JobRole: ['Sales Executive' 'Research Scientist' 'Laboratory Technician'
 'Manufacturing Director' 'Healthcare Representative' 'Manager'
 'Sales Representative' 'Research Director' 'Human Resources']
MaritalStatus: ['Single' 'Married' 'Divorced']
Over18: ['Y']
OverTime: ['Yes' 'No']
```

[12]:

```
# Check value counts for target variable 'Attrition'
print("\nValue counts for 'Attrition':")
print(df['Attrition'].value_counts())
```

Figure 8: handling Missing Values

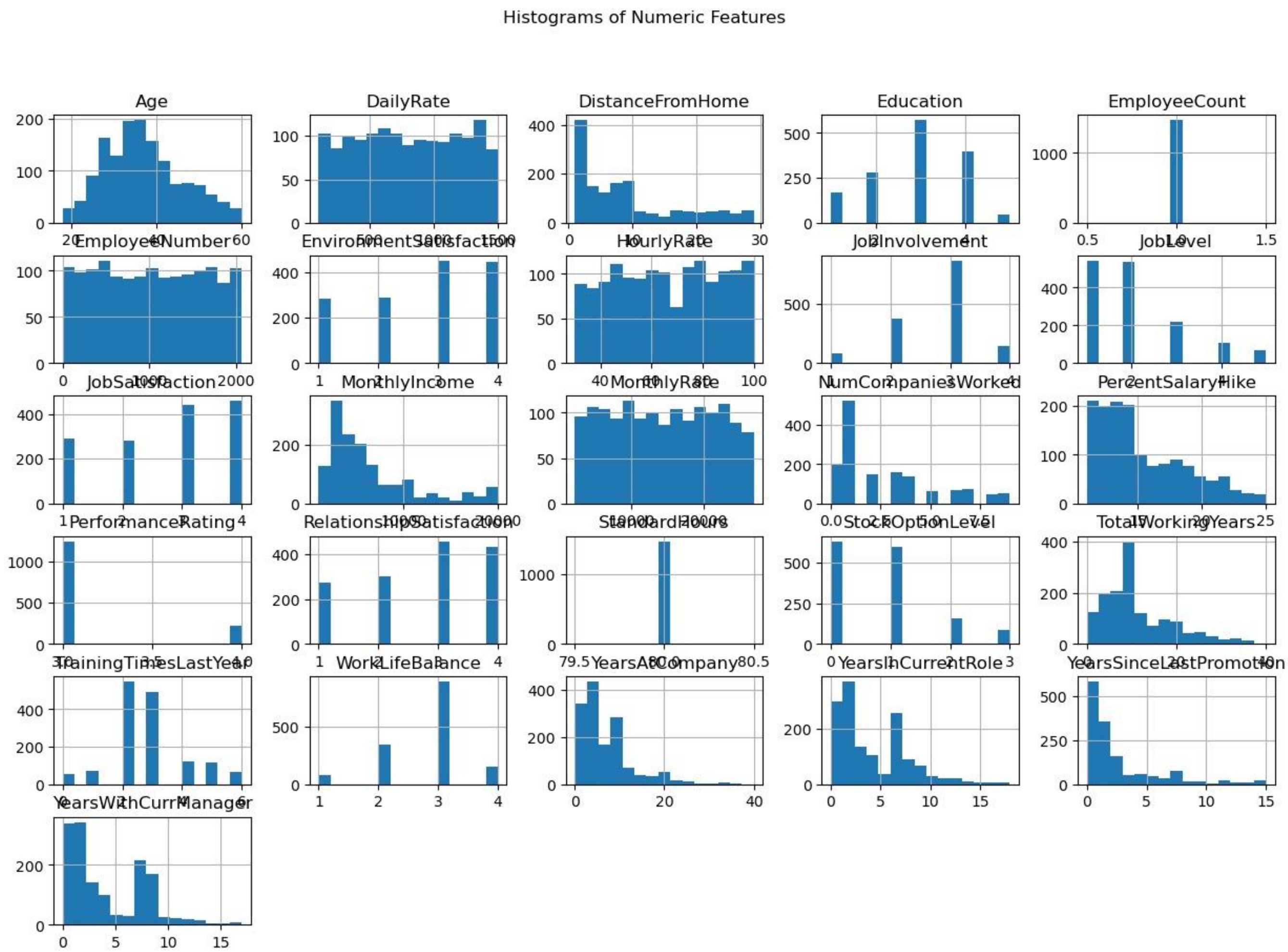


Figure 9: Historic Numerical Features of Data Set

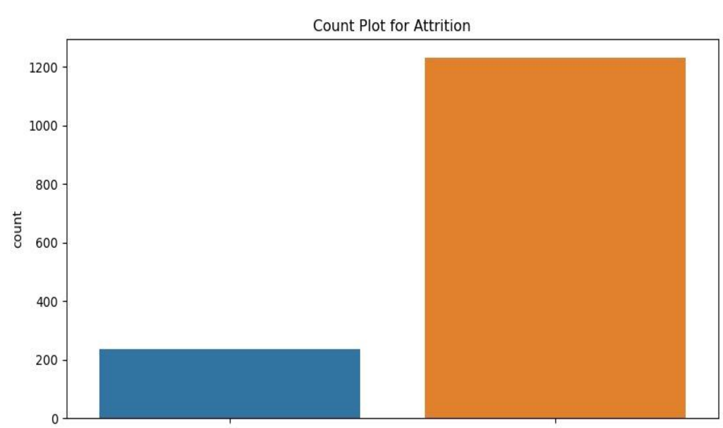


Figure 10: Count Plot for Attrition

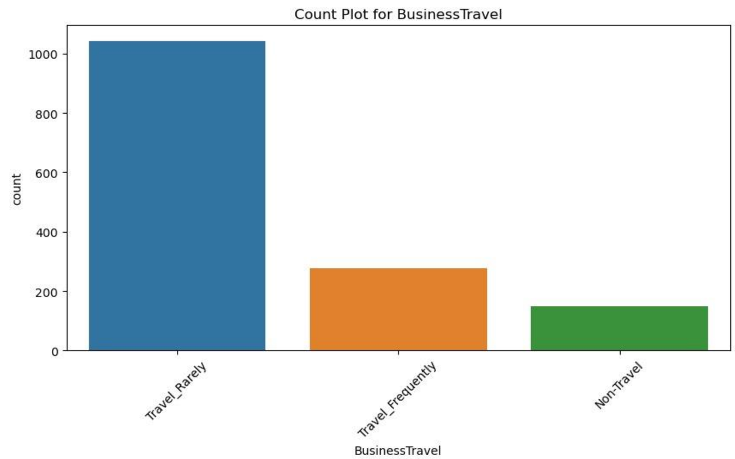


Figure 11: Count Plot for Business Travels

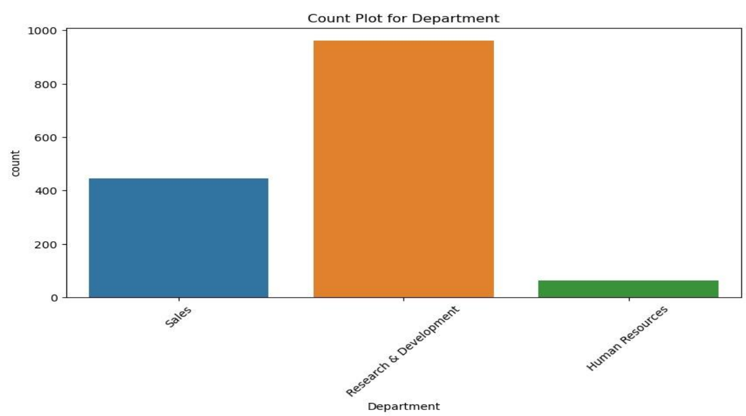


Figure 12: Count Plot for Departments

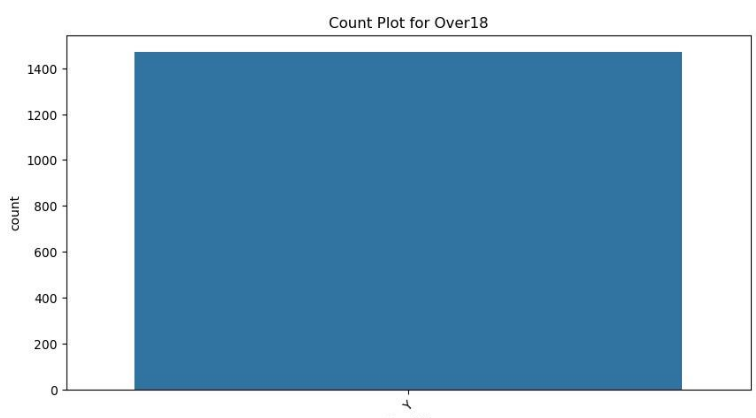


Figure 13: Count Plot for Over 18

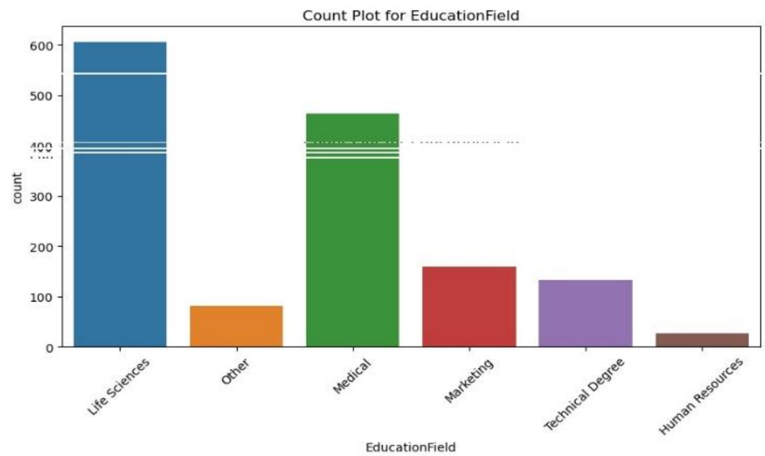


Figure 14: Count Plot for Educational Field

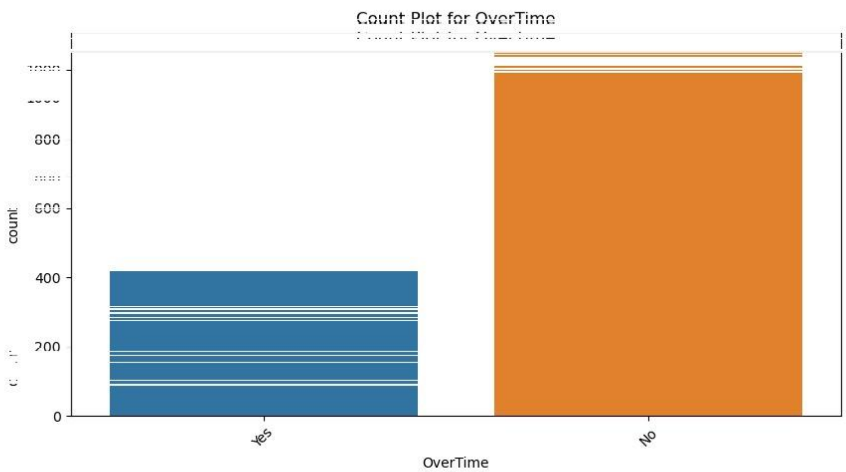


Figure 15: Count Plot for Overtime

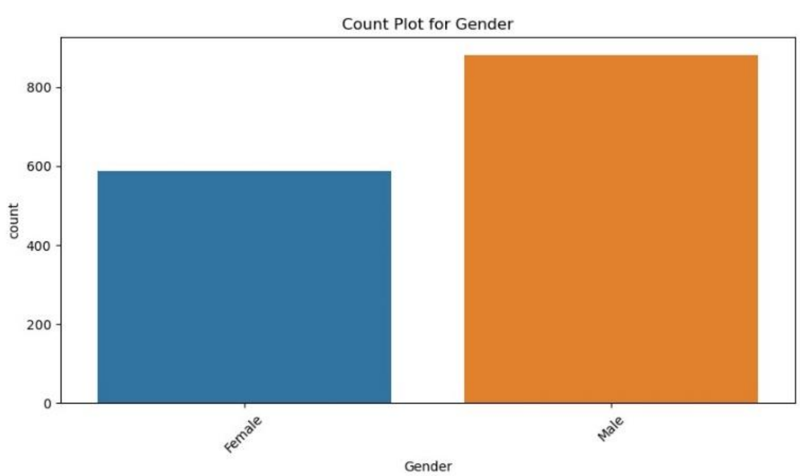


Figure 16: Count plot with Gender

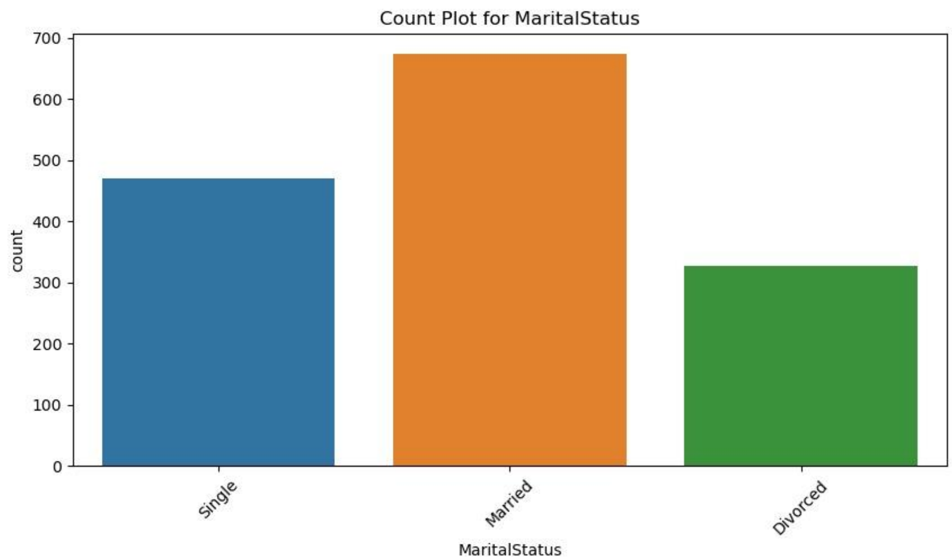


Figure 18: Count Plot for Marital status

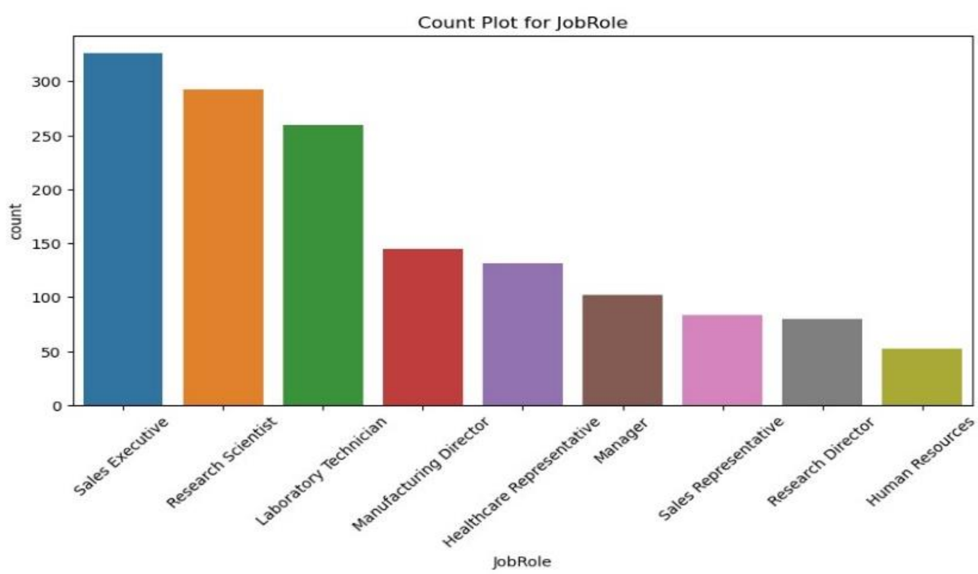


Figure 17: Count Plot for Job Role

Data Encoding:

— Convert categorical variables into numerical representations using one-hot encoding or label encoding.

— Ensure all features are in a format suitable for analysis.

```
File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:10054, in DataFrame.corr(sel
f, method, min_periods, numeric_only)
  10052 cols = data.columns
  10053 idx = cols.copy()
> 10054 mat = data.to_numpy(dtype=float, na_value=np.nan, copy=False)
  10056 if method == "pearson":
  10057     correl = libalgos.nancorr(mat, minp=min_periods)

File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:1838, in DataFrame.to_numpy(s
elf, dtype, copy, na_value)
   1836 if dtype is not None:
   1837     dtype = np.dtype(dtype)
-> 1838 result = self._mgr.as_array(dtype=dtype, copy=copy, na_value=na_value)
   1839 if result.dtype is not dtype:
   1840     result = np.array(result, dtype=dtype, copy=False)

File ~\anaconda3\Lib\site-packages\pandas\core\internals\managers.py:1732, in BlockMa
nager.as_array(self, dtype, copy, na_value)
   1730         arr.flags.writeable = False
   1731 else:
-> 1732     arr = self._interleave(dtype=dtype, na_value=na_value)
   1733     # The underlying data was copied within _interleave, so no need
   1734     # to further copy if copy=True or setting na_value
   1736 if na_value is not lib.no_default:

File ~\anaconda3\Lib\site-packages\pandas\core\internals\managers.py:1794, in BlockMa
nager._interleave(self, dtype, na_value)
   1792     else:
   1793         arr = blk.get_values(dtype)
-> 1794     result[rl.indexer] = arr
   1795     itemmask[rl.indexer] = 1
   1797 if not itemmask.all():
```

Figure 19: Code for Data Encoding

```
Missing values after cleaning:
Age                          0
Attrition                    0
BusinessTravel               0
DailyRate                    0
Department                   0
DistanceFromHome             0
Education                    0
EducationField               0
EnvironmentSatisfaction      0
Gender                       0
HourlyRate                   0
JobInvolvement               0
JobLevel                     0
JobRole                      0
JobSatisfaction              0
MaritalStatus                0
MonthlyIncome                0
MonthlyRate                  0
NumCompaniesWorked           0
Over18                       0
OverTime                     0
PercentSalaryHike            0
PerformanceRating            0
RelationshipSatisfaction     0
StockOptionLevel             0
TotalWorkingYears            0
TrainingTimesLastYear        0
WorkLifeBalance              0
YearsAtCompany               0
YearsInCurrentRole           0
```

```
# Encode categorical variables using label encoding or one-hot encoding
from sklearn.preprocessing import LabelEncoder

# Example: Encode 'Attrition' column (target variable)
label_encoder = LabelEncoder()
df['Attrition'] = label_encoder.fit_transform(df['Attrition'])

# Example: One-hot encode other categorical columns
df = pd.get_dummies(df, columns=['BusinessTravel', 'Department', 'EducationField', 'Ge

# Display the updated dataframe
print("\nEncoded dataset:")
print(df.head())
```

Figure 20: Label Encoding and Hot-Encoding

```
Encoded dataset:
   Age  Attrition  DailyRate  DistanceFromHome  Education  \
0   41          1       1102                 1          2
1   49          0        279                 8          1
2   37          1       1373                 2          2
3   33          0       1392                 3          4
4   27          0        591                 2          1

   EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
0                        2          94               3         2
1                        3          61               2         2
2                        4          92               2         1
3                        4          56               3         1
4                        1          40               3         1

   JobSatisfaction  ...  JobRole_Laboratory Technician  JobRole_Manager  \
0                4  ...                          False            False
1                2  ...                          False            False
2                3  ...                           True            False
3                3  ...                          False            False
4                2  ...                           True            False

   JobRole_Manufacturing Director  JobRole_Research Director  \
0                           False                      False
1                           False                      False
2                           False                      False
3                           False                      False
4                           False                      False

   JobRole_Research Scientist  JobRole_Sales Executive  \
0                       False                     True
1                        True                    False
2                       False                    False
3                        True                    False
4                       False                    False

   JobRole_Sales Representative  MaritalStatus_Married  MaritalStatus_Single  \
0                         False                  False                  True
1                         False                   True                 False
2                         False                  False                  True
3                         False                   True                 False
4                         False                   True                 False

   OverTime_Yes
0          True
1         False
2          True
3          True
4         False

[5 rows x 46 columns]
```

Figure 21: Results of Data Encoding

Data Labeling:

- Create labels for the target variable (e.g., Attrition Yes/No).
- Define any derived features that could enhance model performance.

```python
# Example: Bin 'Age' into categories
bins = [18, 25, 35, 45, 55, 65]
labels = ['18-25', '26-35', '36-45', '46-55', '56-65']
df['AgeCategory'] = pd.cut(df['Age'], bins=bins, labels=labels, right=False)
```

```python
# Display the updated dataframe with labels
print("\nUpdated dataset with labels:")
```

Figure 22: Data Labeling

```
print(df.head())
```

```
Updated dataset with labels:
   Age  Attrition  DailyRate  DistanceFromHome  Education  \
0   41          1       1102                 1          2
1   49          0        279                 8          1
2   37          1       1373                 2          2
3   33          0       1392                 3          4
4   27          0        591                 2          1

   EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
0                        2          94               3         2
1                        3          61               2         2
2                        4          92               2         1
3                        4          56               3         1
4                        1          40               3         1

   JobSatisfaction  ...  JobRole_Manager  JobRole_Manufacturing Director  \
0                4  ...            False                           False
1                2  ...            False                           False
2                3  ...            False                           False
3                3  ...            False                           False
4                2  ...            False                           False

   JobRole_Research Director JobRole_Research Scientist  \
0                      False                      False
1                      False                       True
2                      False                      False
3                      False                       True
4                      False                      False

   JobRole_Sales Executive  JobRole_Sales Representative  \
0                     True                         False
1                    False                         False
2                    False                         False
3                    False                         False
4                    False                         False

   MaritalStatus_Married  MaritalStatus_Single  OverTime_Yes  AgeCategory
0                  False                  True          True        36-45
1                   True                 False         False        46-55
2                  False                  True          True        36-45
3                   True                 False          True        26-35
4                   True                 False         False        26-35

[5 rows x 47 columns]
```

Figure 23: Results of Data Labeling

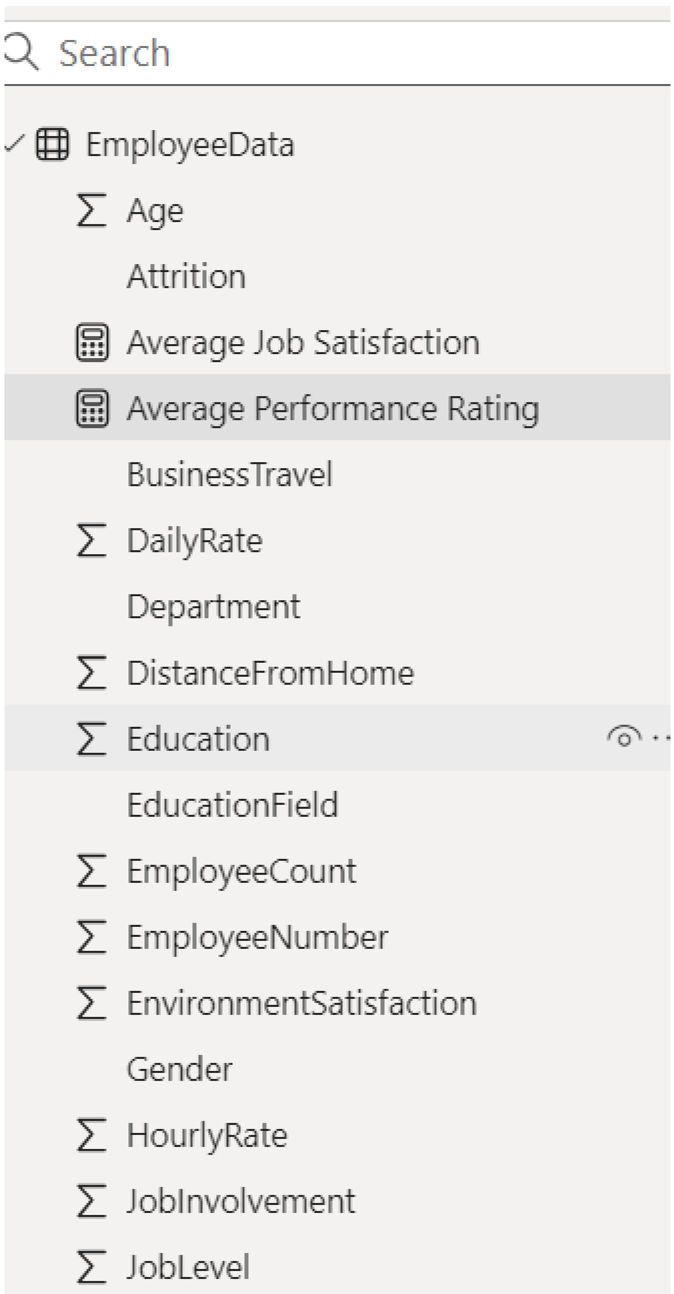
Dashboard Building and Reporting

Generate reports highlighting:

- Reasons for attrition.
- Employee demographics.

Steps for creating a dashboard in Power BI:

1. Import the cleaned dataset into Power BI.
2. Create the required visualizations as outlined in the Gitbook.
3. Arrange the visuals in a logical layout to tell a cohesive story.



Results

Visualizations:

Trend Analysis: Line charts reveal attrition trends over time.

Departmental Insights: Clustered bar charts highlight departments with the highest turnover.

Attrition Reasons: Pie charts and tree maps illustrate common reasons for leaving.

Job Satisfaction and Performance: Scatter plots demonstrate the correlation between job satisfaction and performance ratings.

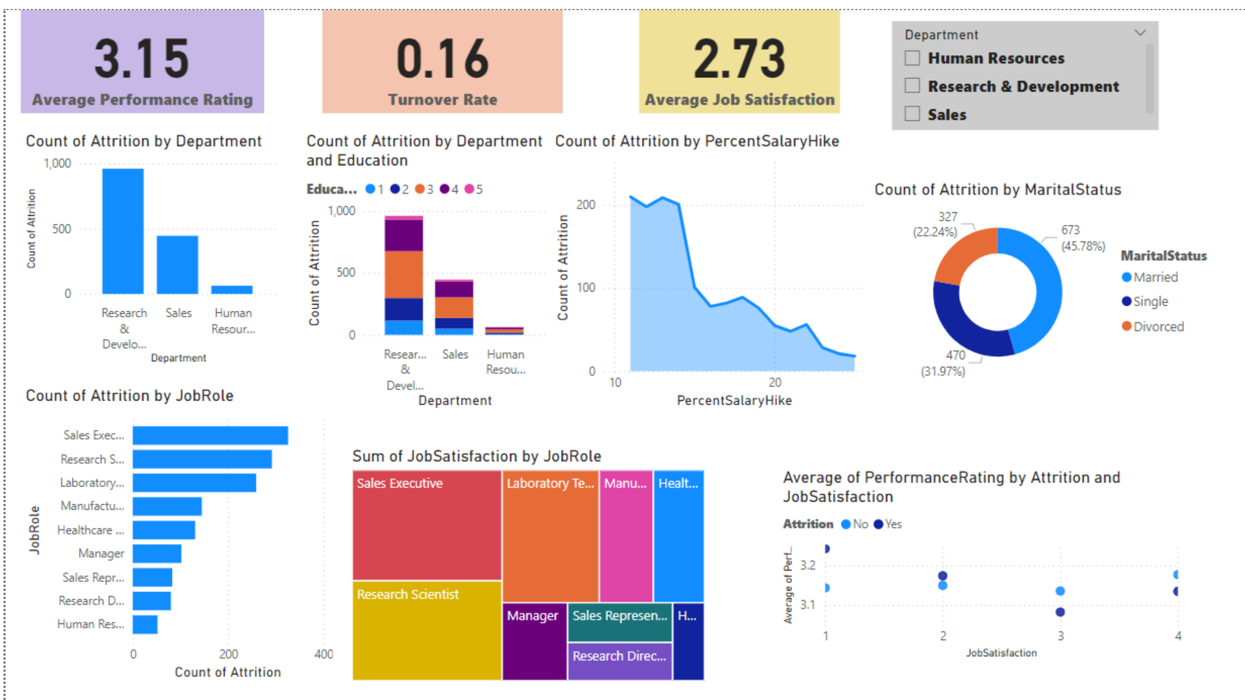


Figure 24: Power BI Dashboard representing Attrition trends

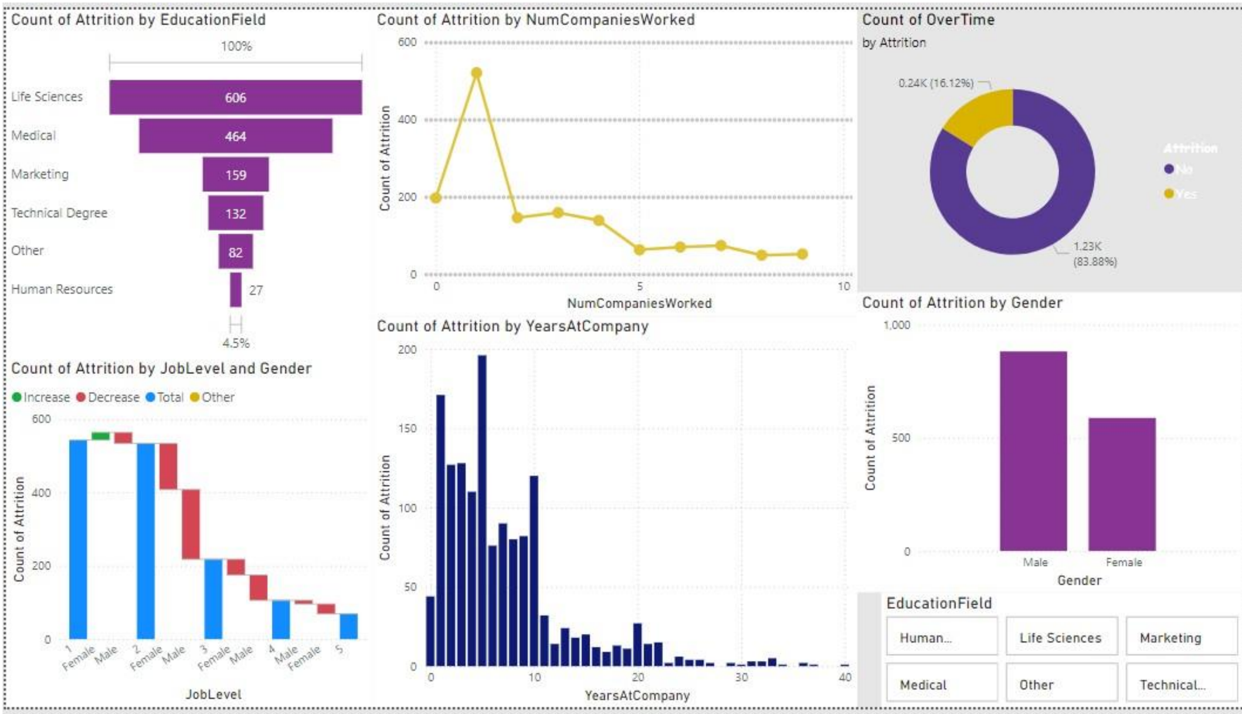


Figure 25: Power BI Dashboard representing Attrition Trends

Key Findings:

- Higher attrition rates in departments with lower job satisfaction.
- Employees with longer tenure and lower performance ratings are more likely to leave.

Recommendations

Actionable Insights:

— Improve job satisfaction and work-life balance, especially in high-turnover departments.

— Develop targeted retention programs and career development initiatives.

Strategic Initiatives:

— Introduce regular employee feedback mechanisms.

— Enhance training and development opportunities.

Conclusion

The analysis provides a comprehensive understanding of employee attrition patterns and factors influencing turnover. Predictive modeling offers valuable insights for future attrition management.

Future Work:

Further research could explore additional factors influencing attrition or test alternative predictive models for improved accuracy.