**DATA ANALYST PROJECT**  
 IBM HR Analytics Employee Attrition & Performance

**Data Set:**

A screenshot of a computer

Description automatically generated

Importing python libraries and setting the pandas options up to 35 columns:  
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

warnings.filterwarnings('ignore')

#setting pandas options(maximum coloumns options)

pd.set\_option('display.max\_columns', 35)

Cleaning the data set (checking for null and duplicate values)  
df.isnull().sum()

df.duplicated().sum()

A screenshot of a computer screen

Description automatically generated

Distribution of Attrition

#Univariate Analysis

df['MonthlyIncome'].hist(bins=50)

plt.show()

# Count plot for categorical data

sns.countplot(x='Gender', data=df)

plt.show()

A graph with red lines

Description automatically generated

Correlation matrix and Heatmap

# Step 1: Encode categorical variables if needed

df['Attrition\_Binary'] = df['Attrition'].map({'Yes': 1, 'No': 0})

# Step 2: Select only numerical columns

numerical\_df = df.select\_dtypes(include=['number'])

# Step 3: Compute correlation matrix

corr = numerical\_df.corr()

# Step 4: Display correlation matrix

print(corr)

# Step 5: (Optional) Plot heatmap for visualization

import seaborn as sns

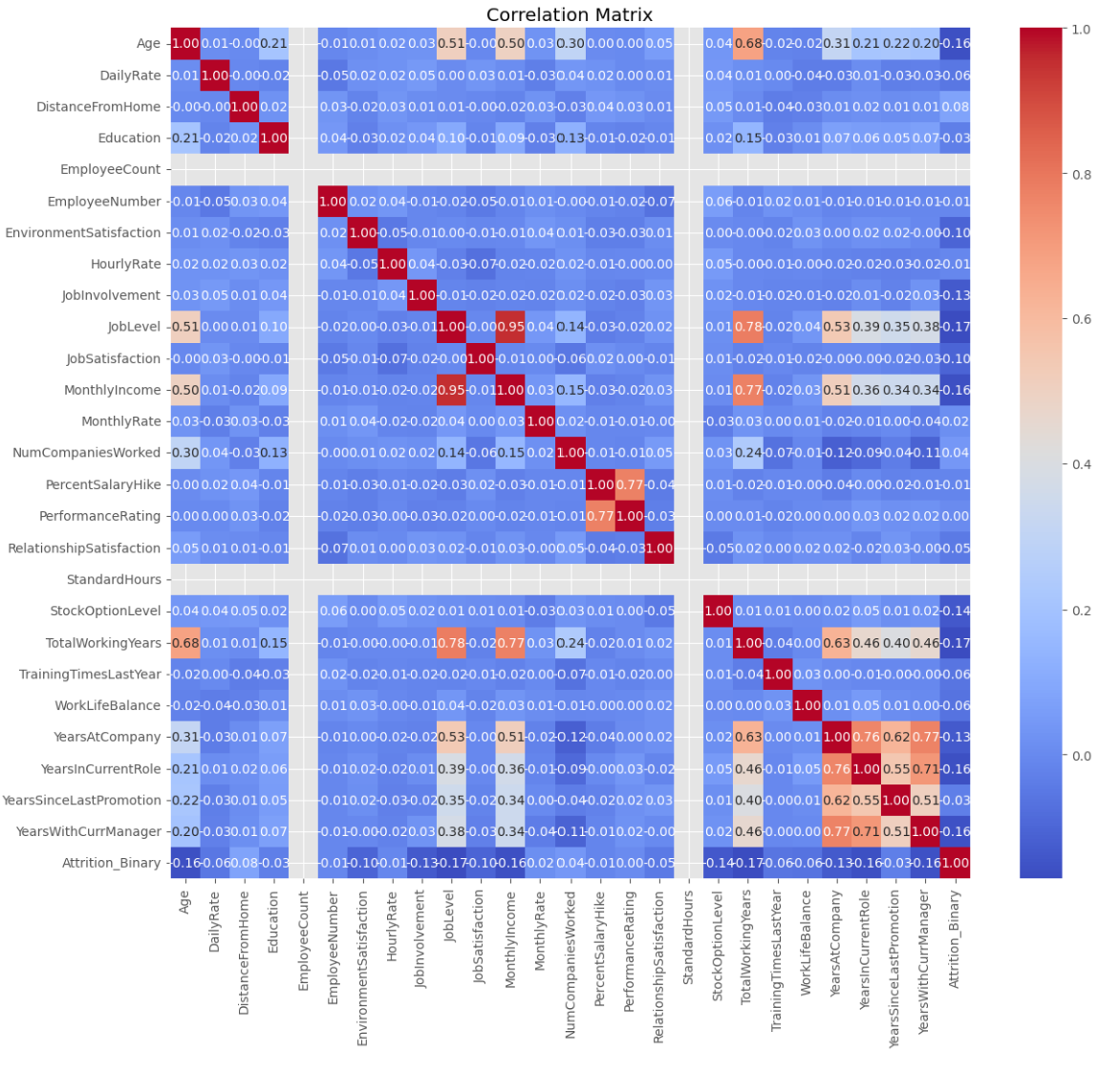
import matplotlib.pyplot as plt

plt.figure(figsize=(14, 12))

sns.heatmap(corr, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0)

plt.title('Correlation Matrix')

plt.show()



Distribution of Attrition rate in the company

A graph with a red rectangular bar

Description automatically generated A pie chart with numbers and a few words

Description automatically generated with medium confidence

Employee Demographics

A graph of a person and person

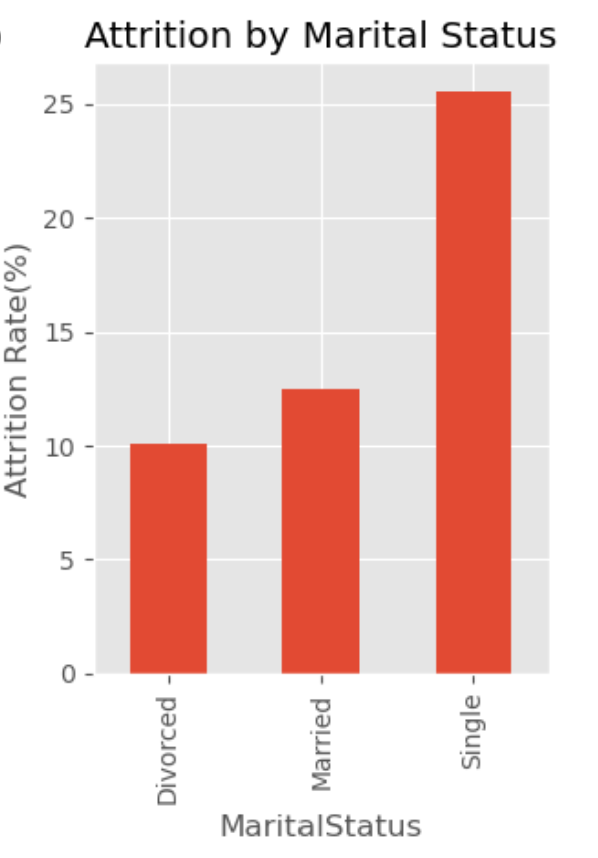
Description automatically generated with medium confidence

Distribution of attrition by gender and education

A graph of a person and person

Description automatically generated

1 Younger workers seem to be more prone than other age groups to quit a company, particularly those between the ages of 30 and 35. A more alluring work offer elsewhere, discontent with the pay or career path, or a desire for fresh experiences could all be contributing causes.   
2. Job stability is generally higher among older workers. The presence of mandated retirement benefits, the difficulty of finding a new employment at an older age, or a higher level of devotion to the organization could all be contributing factors.

Attrition by Maritial Status:  


As we can see the attrition rate is higher in singles than compared to divorced or married people.

Attrition with respect to the job role in the company:

A graph of a job role distribution with attrition

Description automatically generated

*Data Processing*

*1*)Feature Selection

A screen shot of a computer code

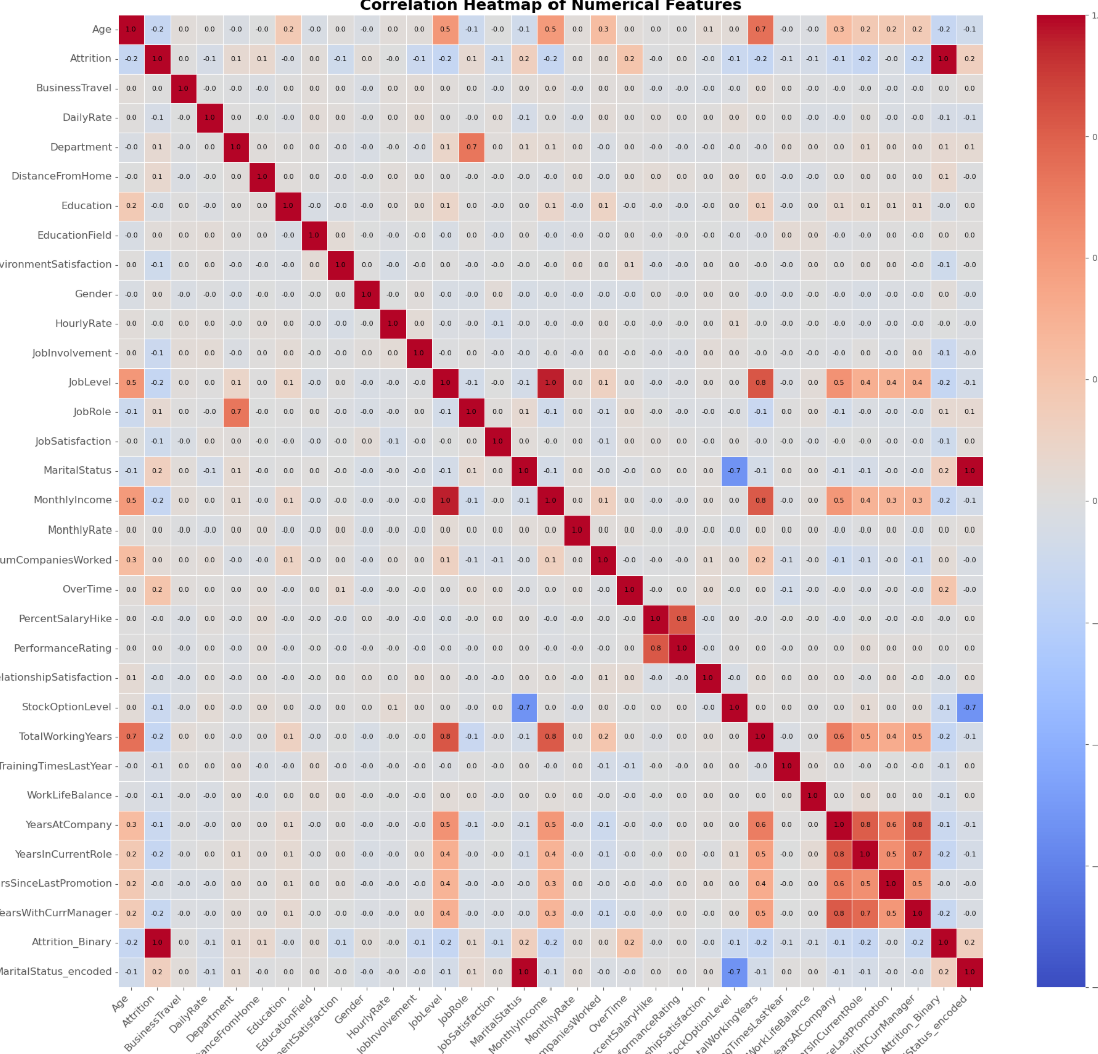
Description automatically generated

*2)*Categorical Conversion (to make categorical data interpretable for machine learning algorithms)

*A screenshot of a computer program

Description automatically generated*

*3) Heatmap*

**