**EMPLOYEE ATTRITION**

The Synthetic Employee Attrition Dataset is a simulated dataset designed for the analysis and prediction of employee attrition. It contains detailed information about various aspects of an employee's profile, including demographics, job-related features, and personal circumstances.

The dataset comprises 74,498 samples, split into training and testing sets to facilitate model development and evaluation. Each record includes a unique Employee ID and features that influence employee attrition. The goal is to understand the factors contributing to attrition and develop predictive models to identify at-risk employees.

This dataset is ideal for HR analytics, machine learning model development, and demonstrating advanced data analysis techniques. It provides a comprehensive and realistic view of the factors affecting employee retention, making it a valuable resource for researchers and practitioners in the field of human resources and organizational development.

**FEATURES:**

**Employee ID**: A unique identifier assigned to each employee.  
**Age:** The age of the employee, ranging from 18 to 60 years.  
**Gender:** The gender of the employee  
**Years at Company:** The number of years the employee has been working at the company.  
**Monthly Income:** The monthly salary of the employee, in dollars.  
**Job Role:** The department or role the employee works in, encoded into categories such as Finance, Healthcare, Technology, Education, and Media.  
**Work-Life Balance:** The employee's perceived balance between work and personal life, (Poor, Below Average, Good, Excellent)  
**Job Satisfaction:** The employee's satisfaction with their job: (Very Low, Low, Medium, High)  
**Performance Rating:** The employee's performance rating: (Low, Below Average, Average, High)  
**Number of Promotions:** The total number of promotions the employee has received.

**Distance from Home:** The distance between the employee's home and workplace, in miles.  
**Education Level:** The highest education level attained by the employee: (High School, Associate Degree, Bachelor’s Degree, Master’s Degree, PhD)  
**Marital Status:** The marital status of the employee: (Divorced, Married, Single)  
**Job Level:** The job level of the employee: (Entry, Mid, Senior)  
**Company Size:** The size of the company the employee works for: (Small,Medium,Large)  
**Company Tenure:** The total number of years the employee has been working in the industry.  
**Remote Work:** Whether the employee works remotely: (Yes or No)  
**Leadership Opportunities:** Whether the employee has leadership opportunities: (Yes or No)  
**Innovation Opportunities:** Whether the employee has opportunities for innovation: (Yes or No)  
**Company Reputation:** The employee's perception of the company's reputation: (Very Poor, Poor, Good, Excellent)  
  
**Employee Recognition:** The level of recognition the employee receives:(Very Low, Low, Medium, High)

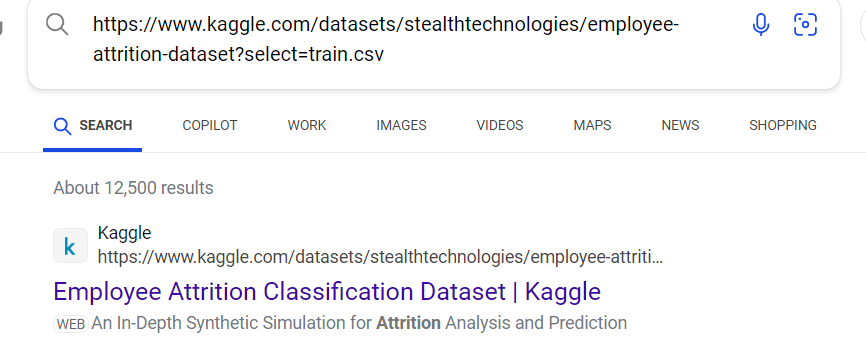
**Attrition:** Whether the employee has left the company, encoded as 0 (stayed) and 1 (Left).

**Task 1**

LINK to DATASET: <https://www.kaggle.com/datasets/stealthtechnologies/employee-attrition-dataset?select=train.csv>

1. Share dataset/s used

* AI Interns to sign up on Kaggle to get datasets



A screenshot of a login page

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Task 2**

1. **Documentation**
   1. **This needs to be reflect the model methodology applied , i.e. Document the entire process, including data preprocessing steps, model selection, evaluation metrics, and any decisions made during development.**
   2. **The key outcome is for the analyst on our end to recreate this model, hence this document needs to be as comprehensive as possible.**

Data quality management was ensured

Data profiler packages was created

Data quality packages were created

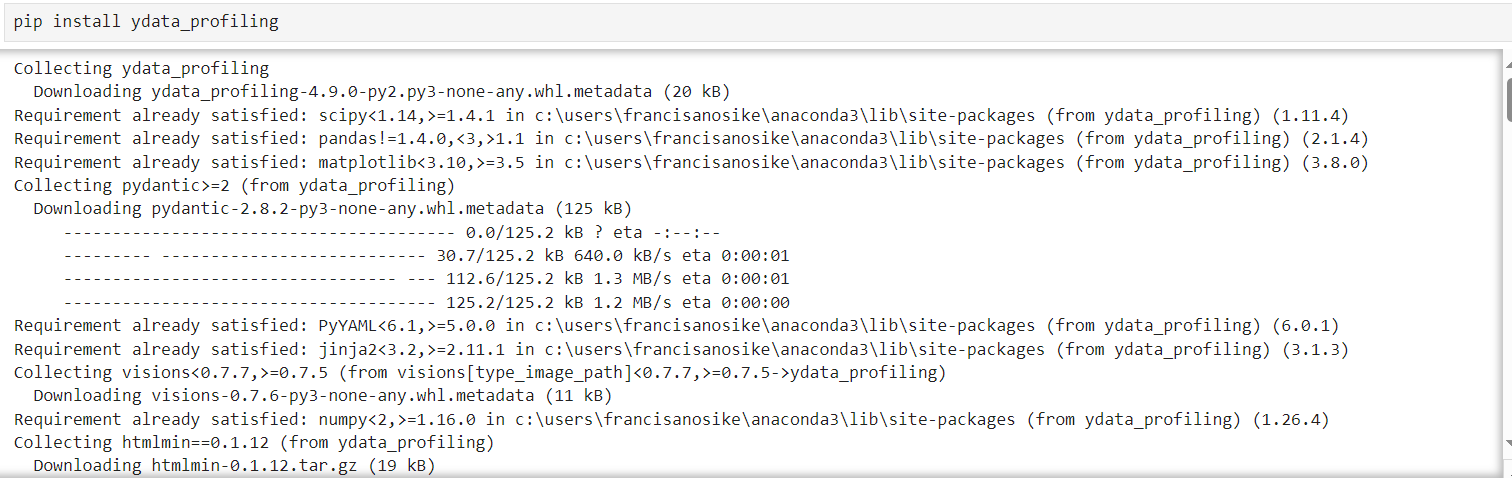
Employee Attrition model created

A screenshot of a computer

Description automatically generated

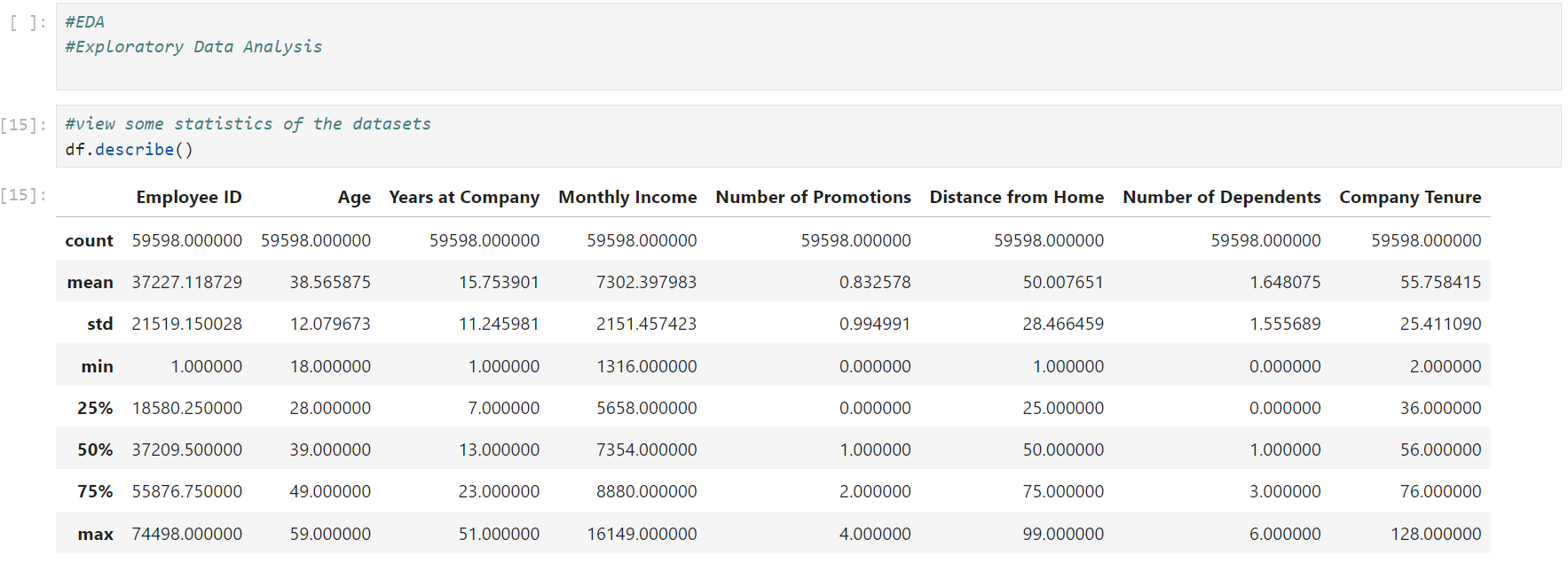
A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated



A screenshot of a phone

Description automatically generated

A white screen with black text

Description automatically generated

A screenshot of a computer

Description automatically generated

A white screen with text

Description automatically generated

A screenshot of a computer

Description automatically generated



A white rectangular object with a black border

Description automatically generated with medium confidence

A graph with numbers and lines

Description automatically generated

A graph with blue bars

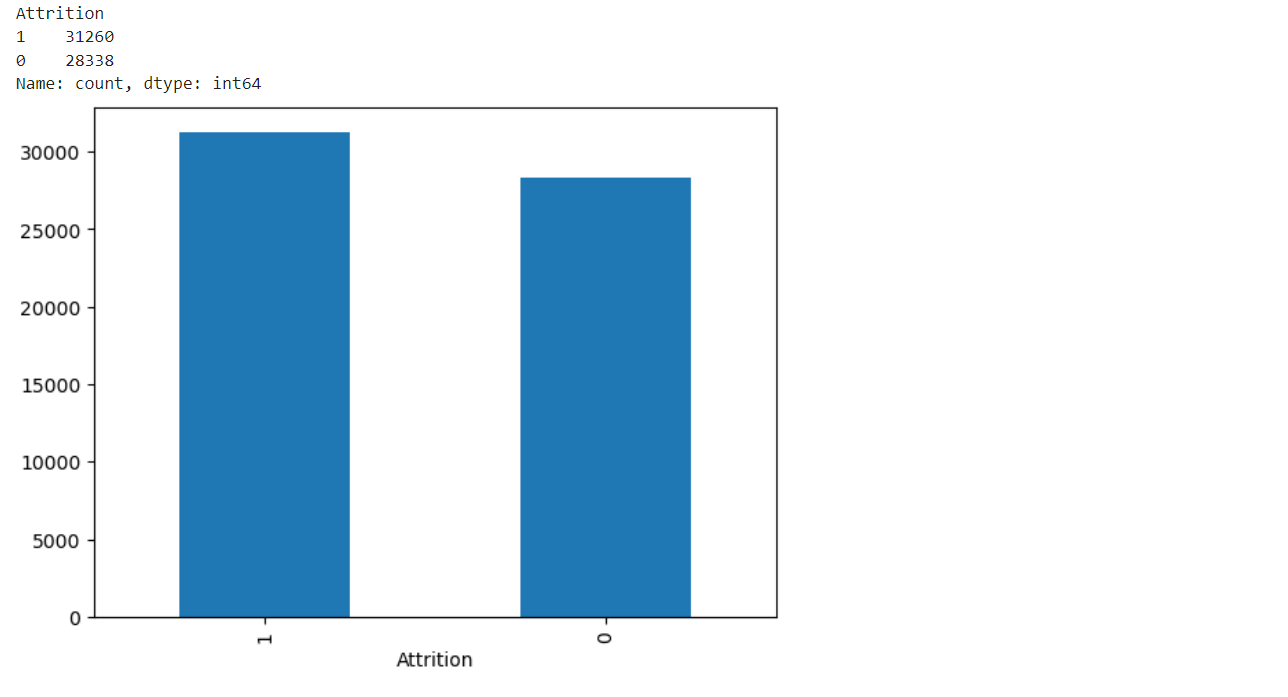
Description automatically generated

A graph with numbers and lines

Description automatically generated

A screenshot of a computer

Description automatically generated

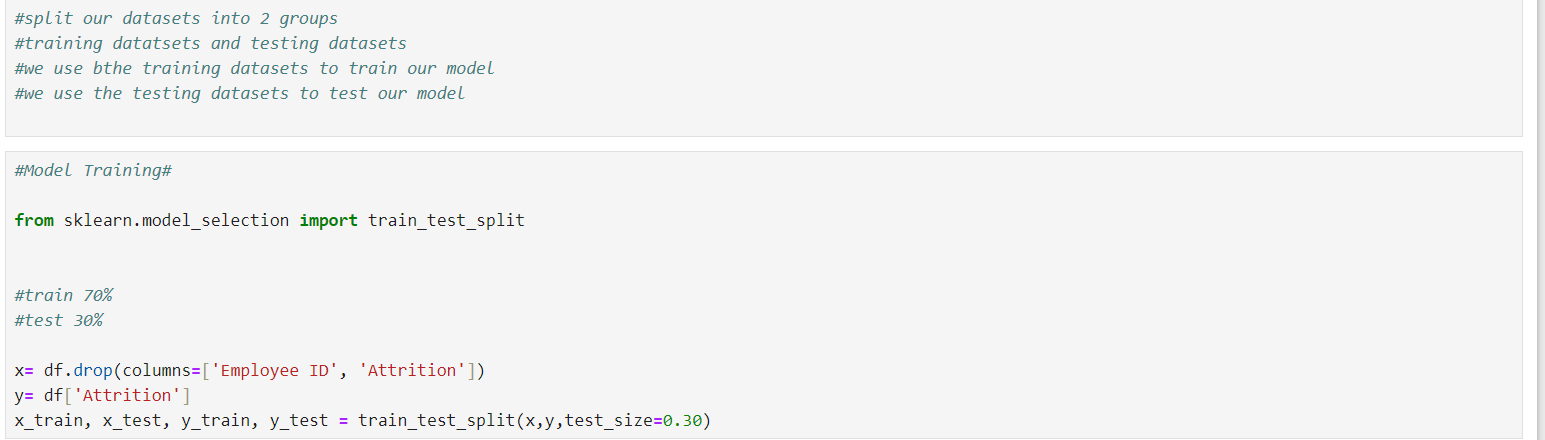


A close-up of a computer screen

Description automatically generated

A table of numbers and letters

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

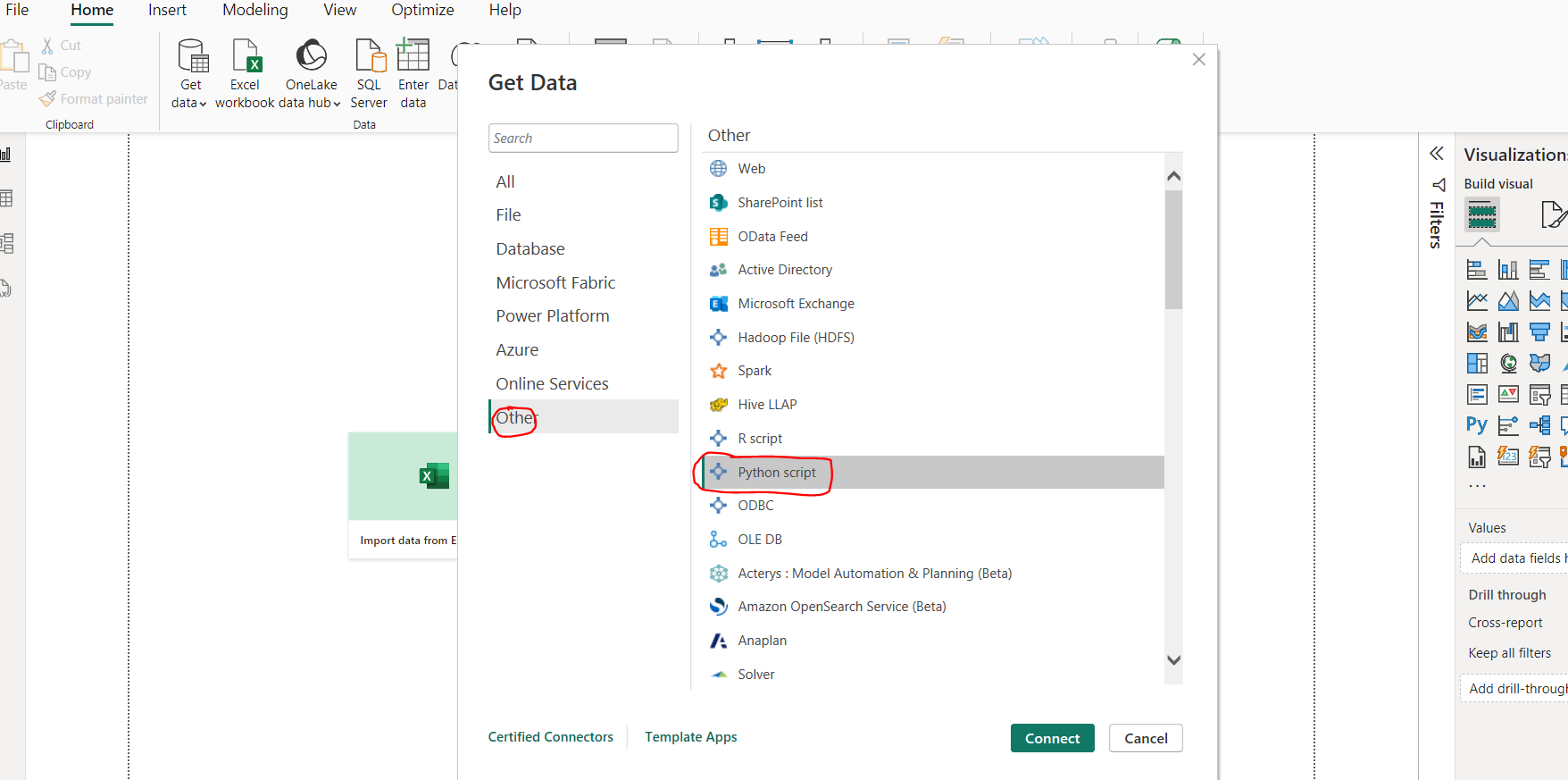
A screenshot of a chat

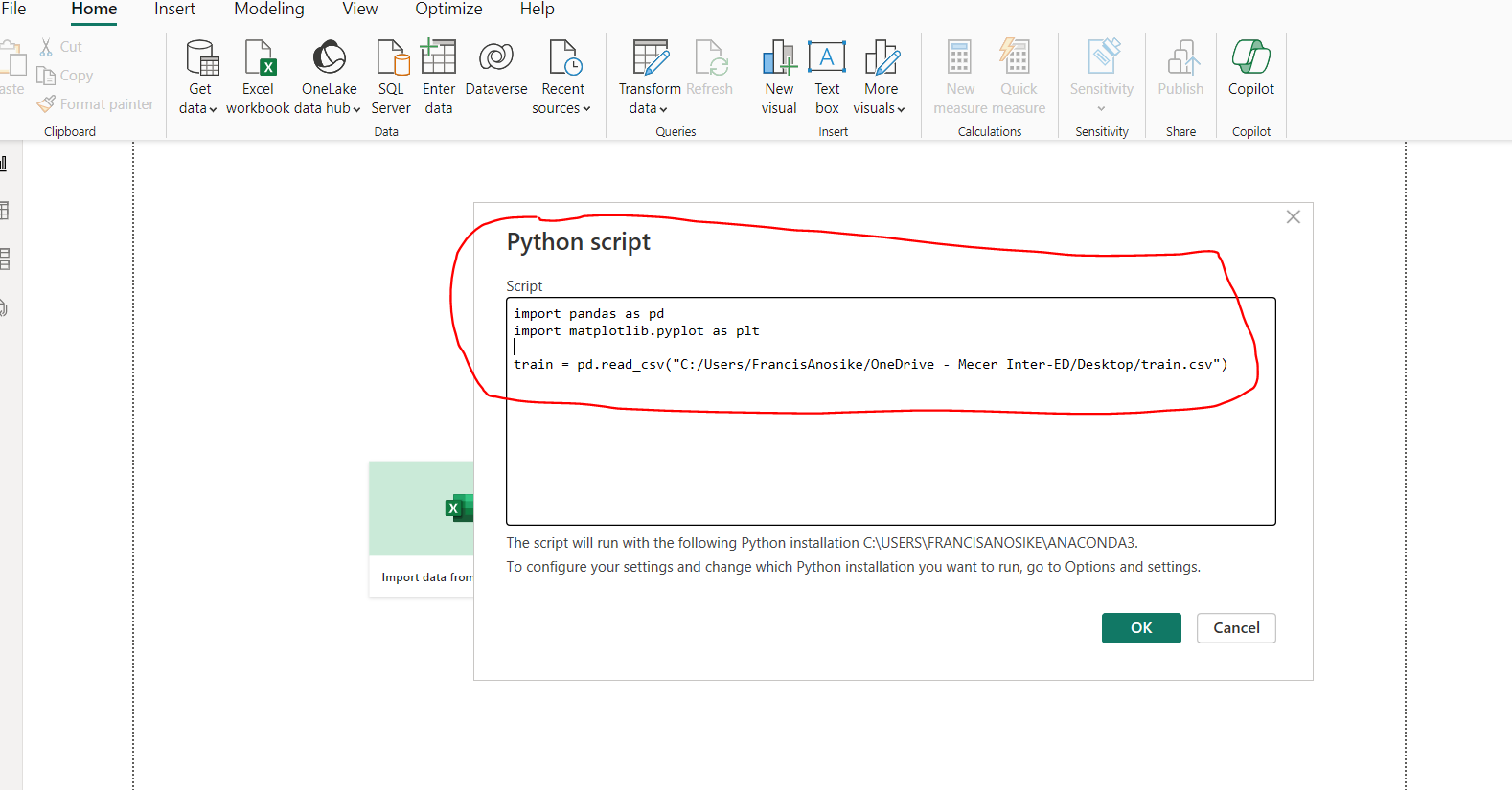
Description automatically generated

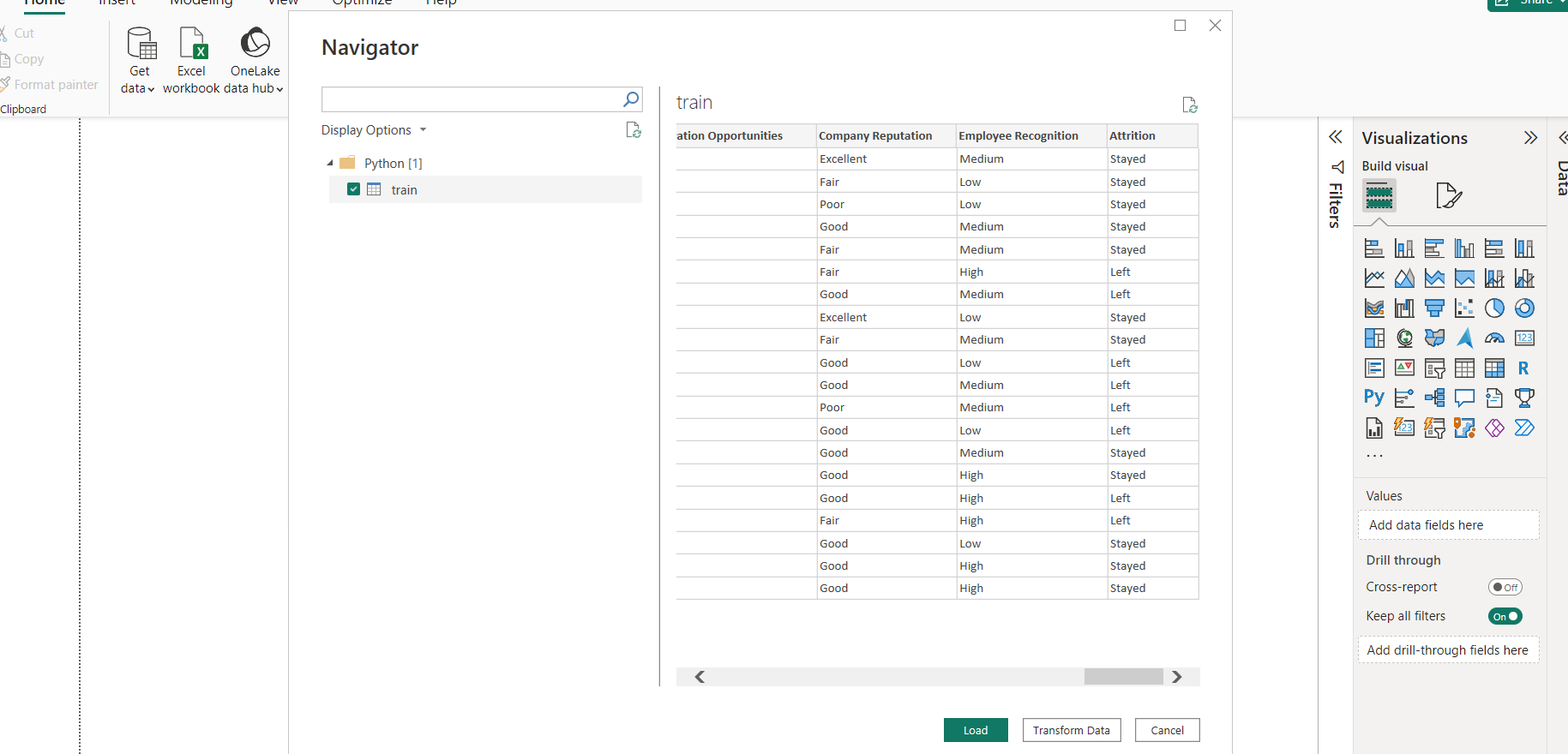
**Task 3**

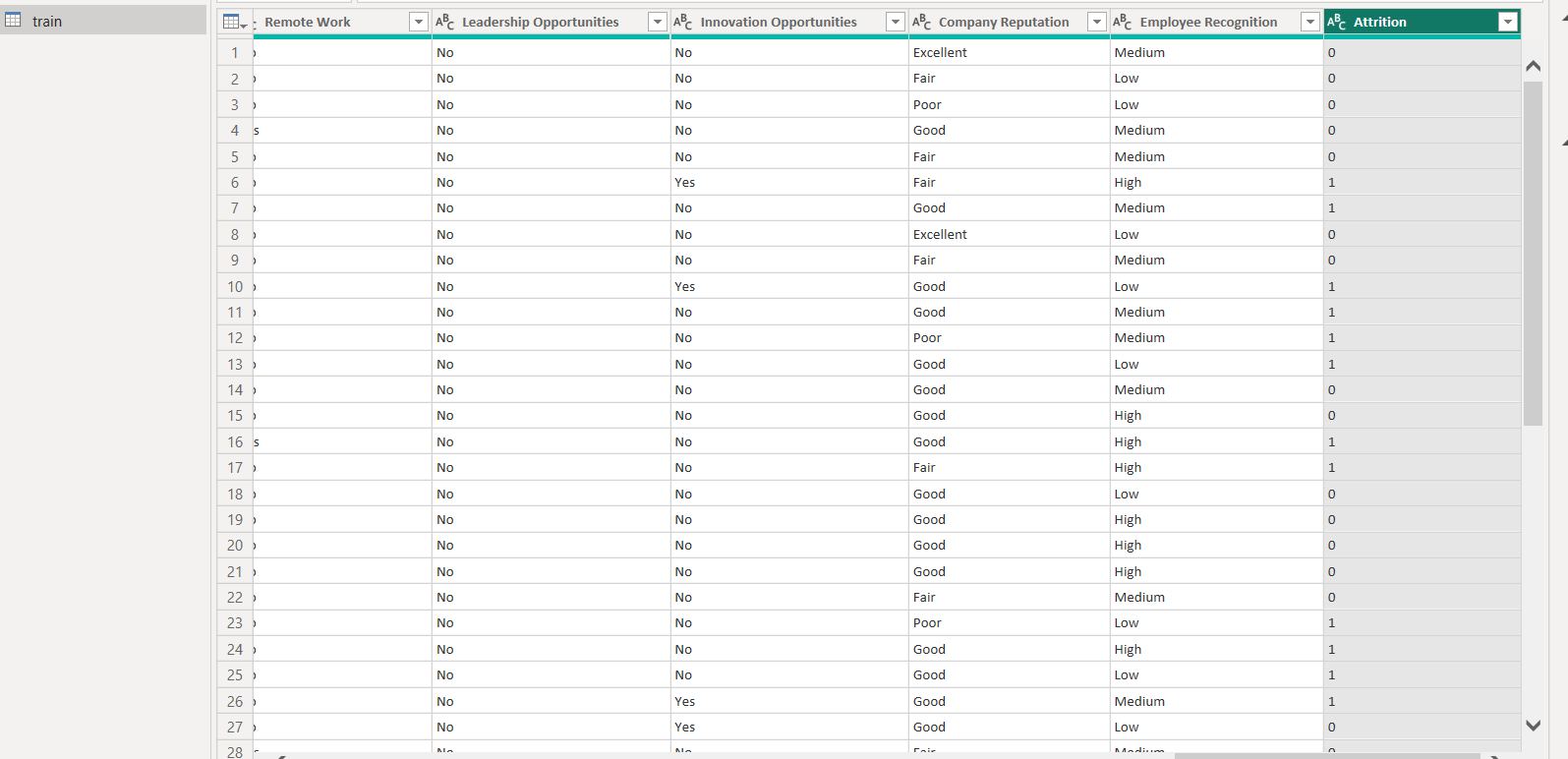
Create visuals : Using Python Scripts embedded in PowerBI

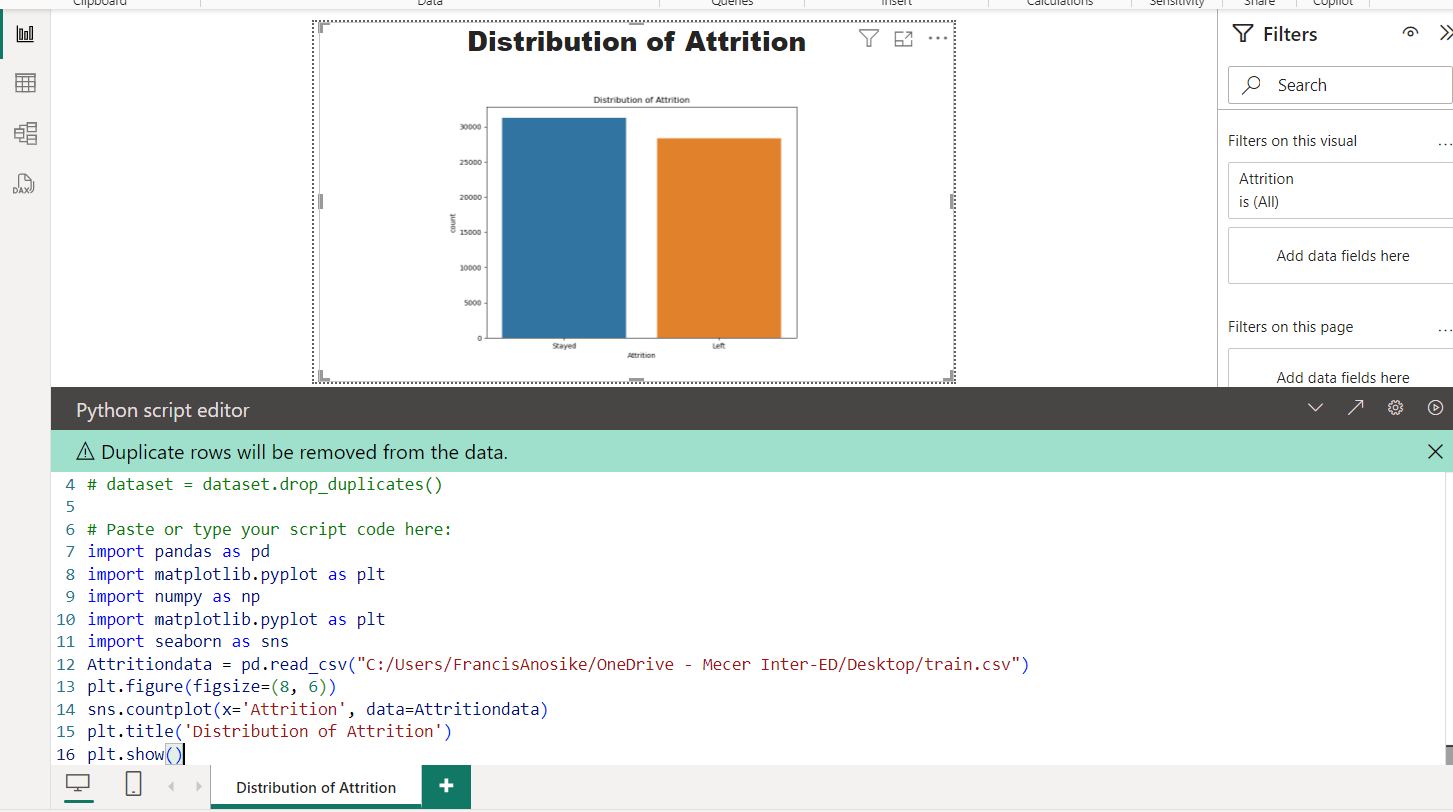
Visualizing the dataset using Power Bi





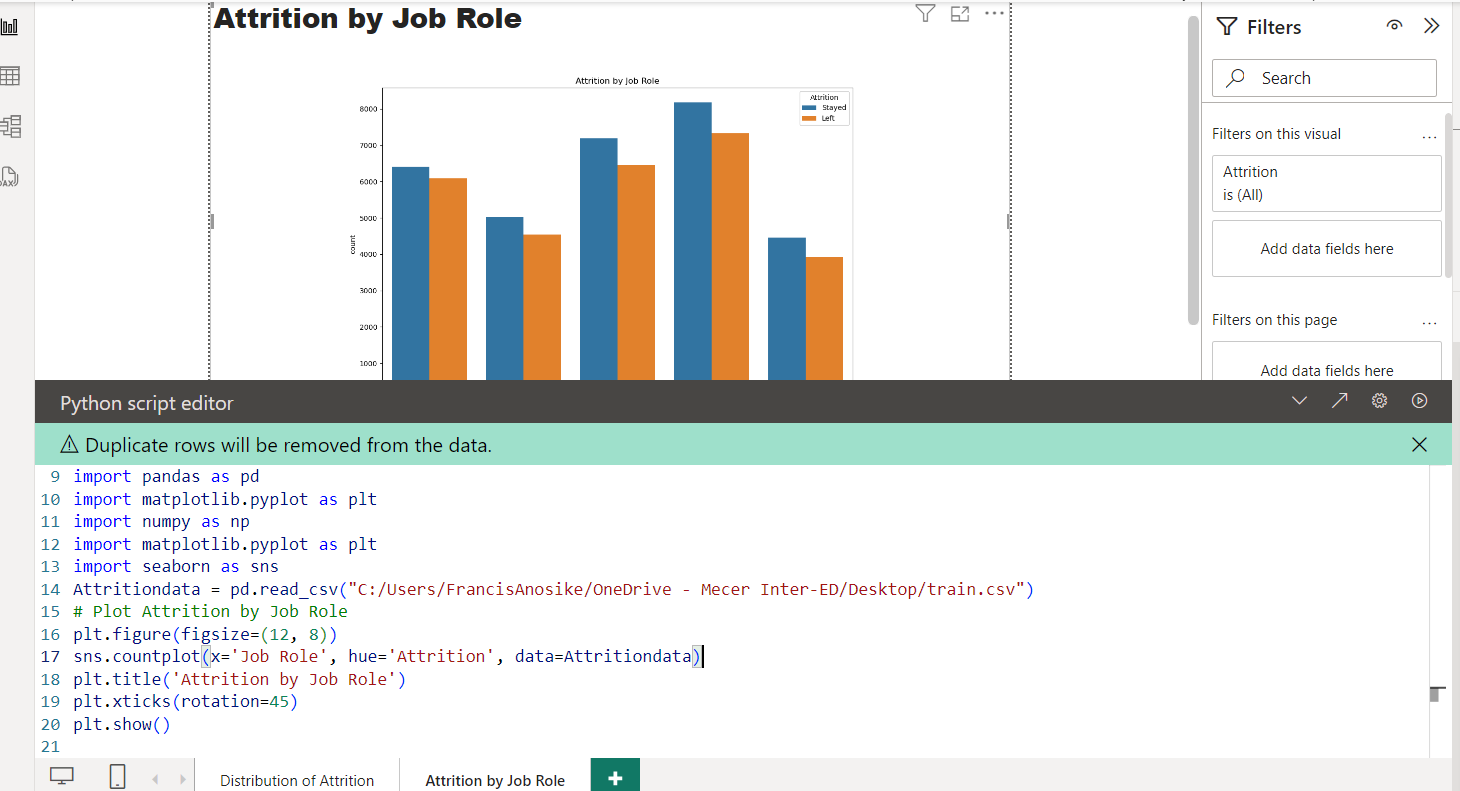






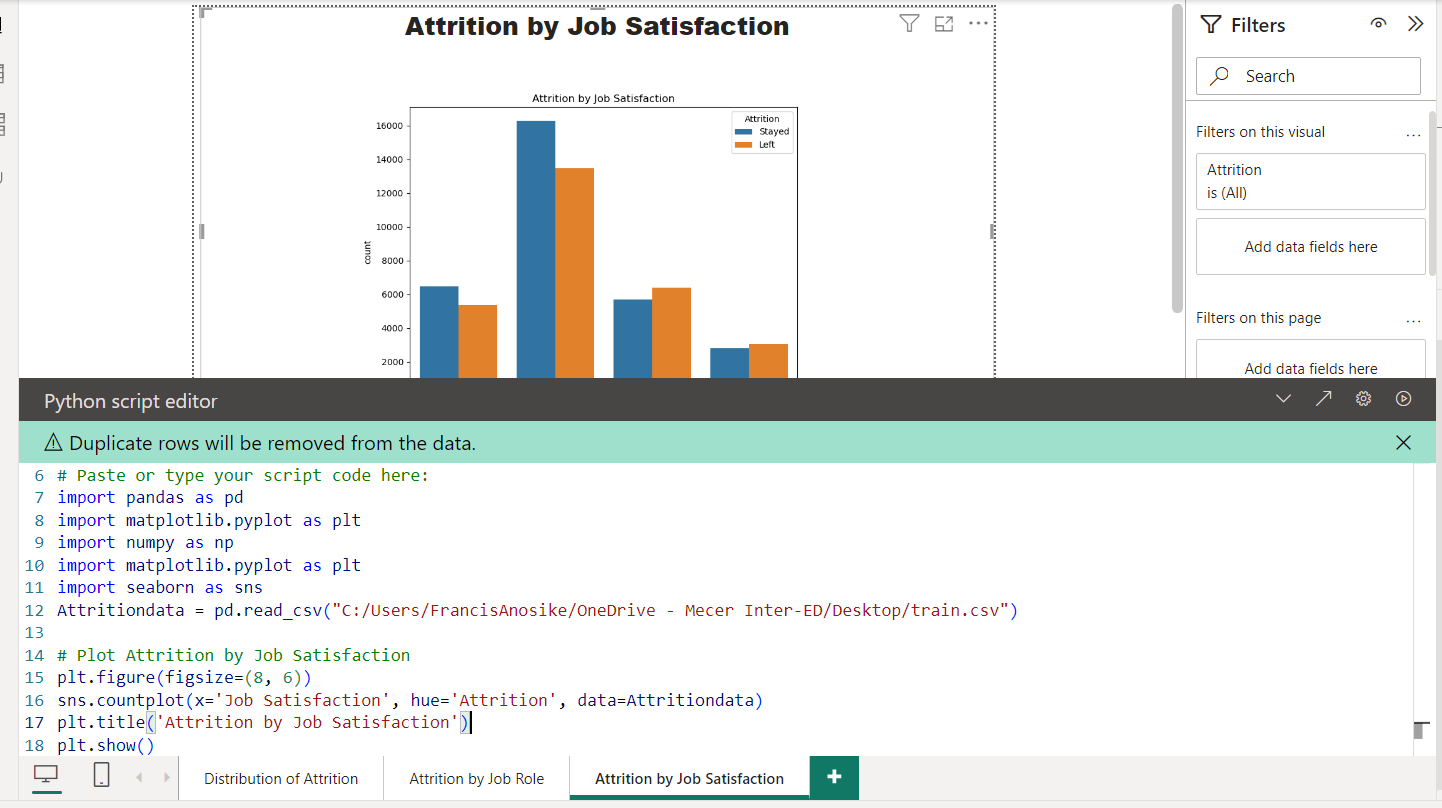
A blue and orange bars

Description automatically generated



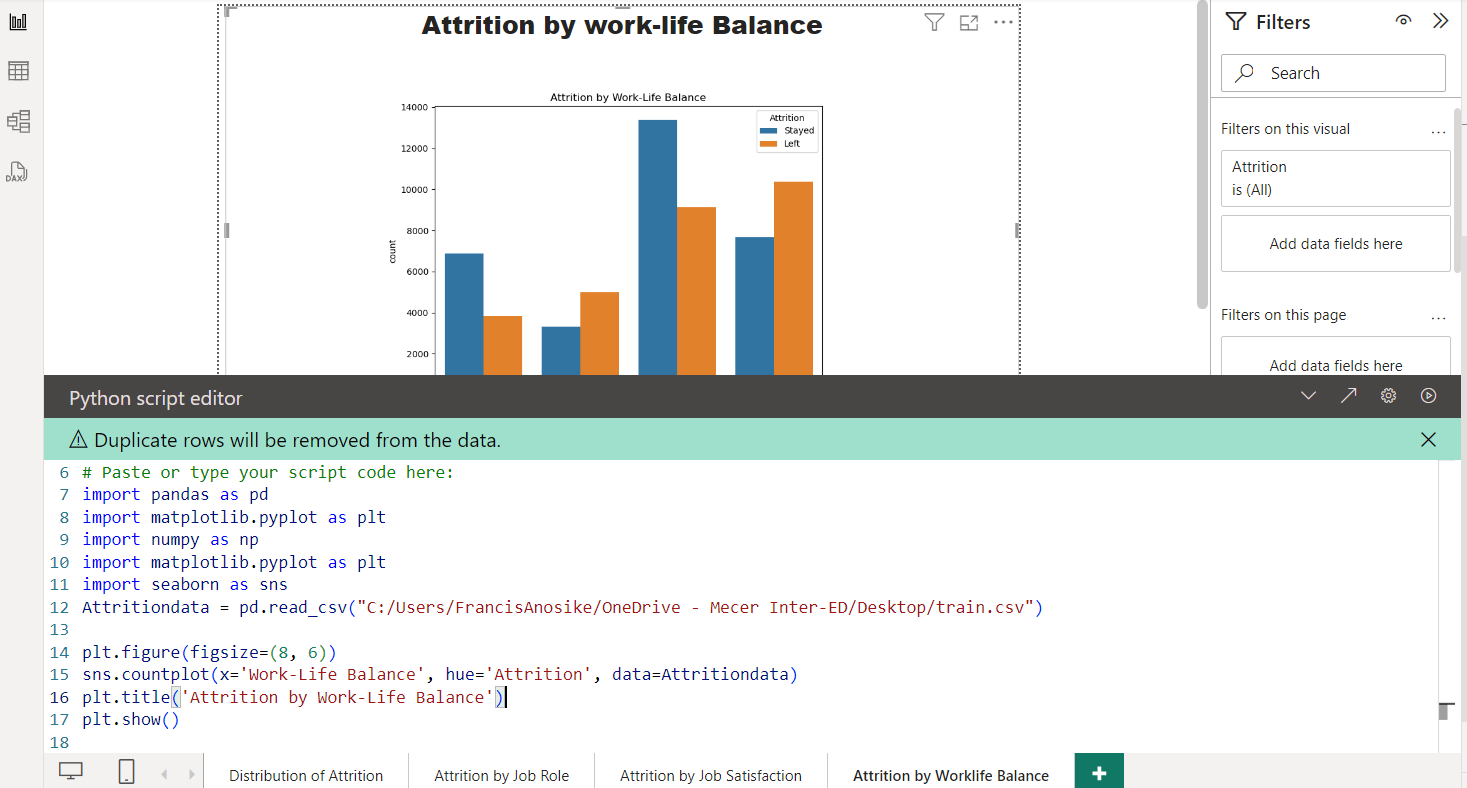
A graph of blue and orange bars

Description automatically generated



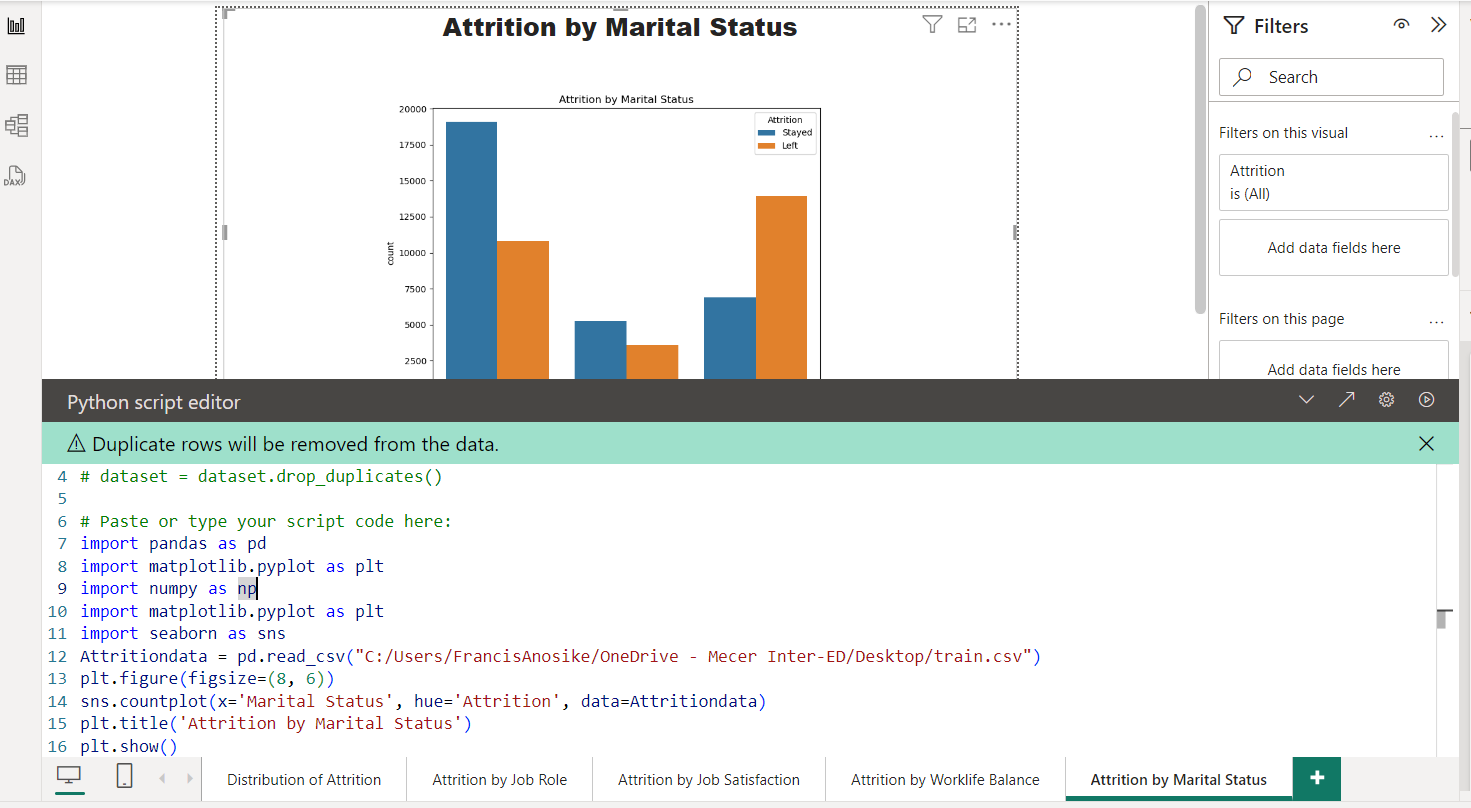
A graph with blue and orange bars

Description automatically generated



A graph of a bar chart

Description automatically generated with medium confidence



A graph of a bar chart

Description automatically generated with medium confidence

Task 4

1. Deployment
   1. Deploy the model in a controlled environment, continuously monitor its performance, and update it as necessary based on new data and changing business needs.
   2. Evidence of the above needs to form part of the Documentation in point 2.a above.

Deployment of Trained Model



A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

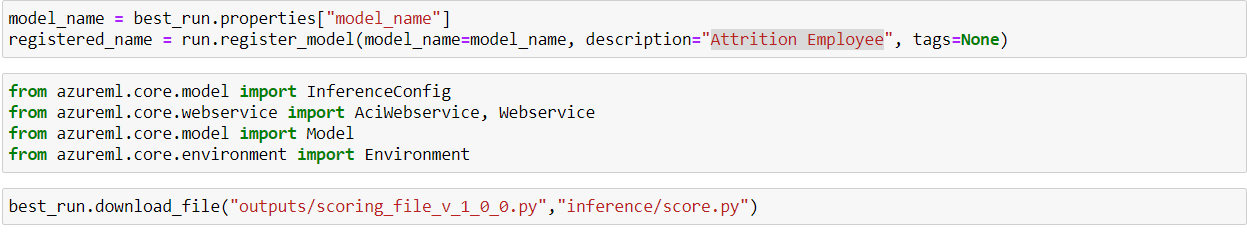
Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

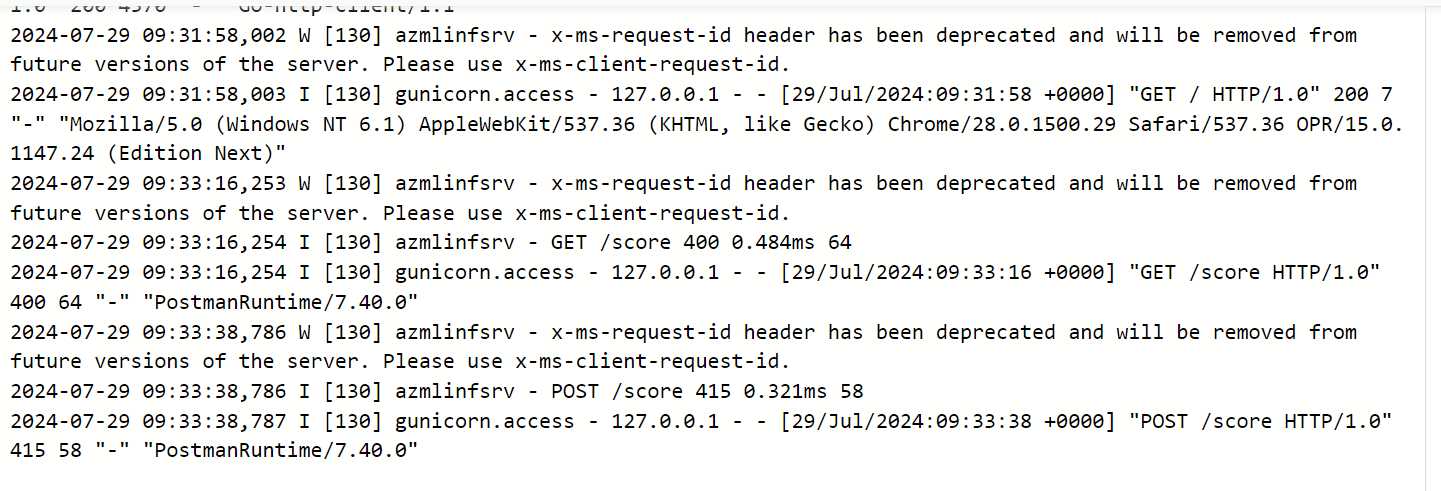
Description automatically generated



A screenshot of a computer

Description automatically generated





A screenshot of a computer

Description automatically generated

<http://bba6b5ab-7969-42c3-848f-72f69800a406.eastus.azurecontainer.io/score>