

**Semester Project**

**Probability and Statistics**

**Unveiling Salary Disparities**

|  |  |  |
| --- | --- | --- |
| **Name** | **Roll No** | **Section** |
| Muhammad Hashim Ali | 22F-3635 | BSE-4A |
| Muhammad Waleed | 22F-3674 | BSE-4A |
| Muhammad Tallal Arif | 22F-3675 | BSE-4A |
| Dameesha Rauf | 22F-3290 | BSCS-4B |
| Laiba Ijaz | 22F-3328 | BSCS-4B |
| Saad Asif | 22F-3338 | BSCS-4B |

**Title:**

Unveiling Salary Disparities: An In-depth Analysis of Department-wise Compensation Structures

**Problem Statement**

Salary disparities across departments within an organization have become a focal point of concern in contemporary workplaces. Understanding the factors contributing to these differences is crucial for fostering equality and ensuring fair compensation practices. However, a comprehensive examination of department-wise salary distributions and the influence of gender remains elusive.

**Objective**

The primary objective of this study is to dissect the salary structures across various departments within the organization, investigating potential variations based on gender. Through rigorous statistical analysis, we aim to unveil any existing pay disparities, identify contributing factors, and provide actionable insights for equitable compensation practices. The following tools have been used to quantify specific relations:

* **Histograms**
* **Bar charts**
* **Pie charts**
* **Box plots**
* **Scatter plots**

And for analysis of this data set we have used:

* Basic descriptive statistics like **mean**,**count**,and **standard deviation**.
* Probability distribution (**Binomial distribution**) because for all data,value of n was **less than 100**
* **Logistic regression** because it is most helpful in determining disease like datasets.

**Data Description**

The dataset under examination comprises seven variables, including:

* Employee ID
* Department
* Division
* Gender
* Salary
* Years of Experience
* Education Level

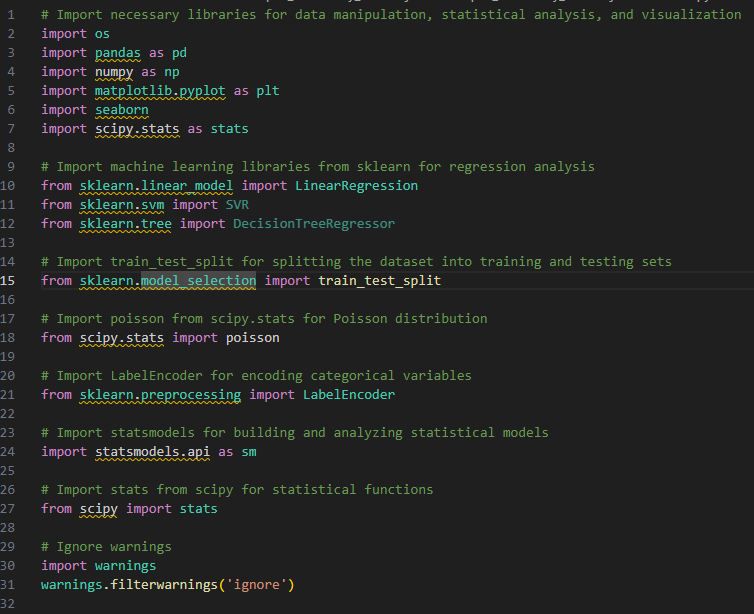
The dataset provides a comprehensive overview of employee demographics, salaries, and other pertinent factors, enabling a granular exploration of compensation patterns across departments and gender groups. This dataset, titled **Unveiling Salary Disparities** is sourced from **Kaggle** and provides a detailed analysis of salaries.The dataset can be accessed at the following link:

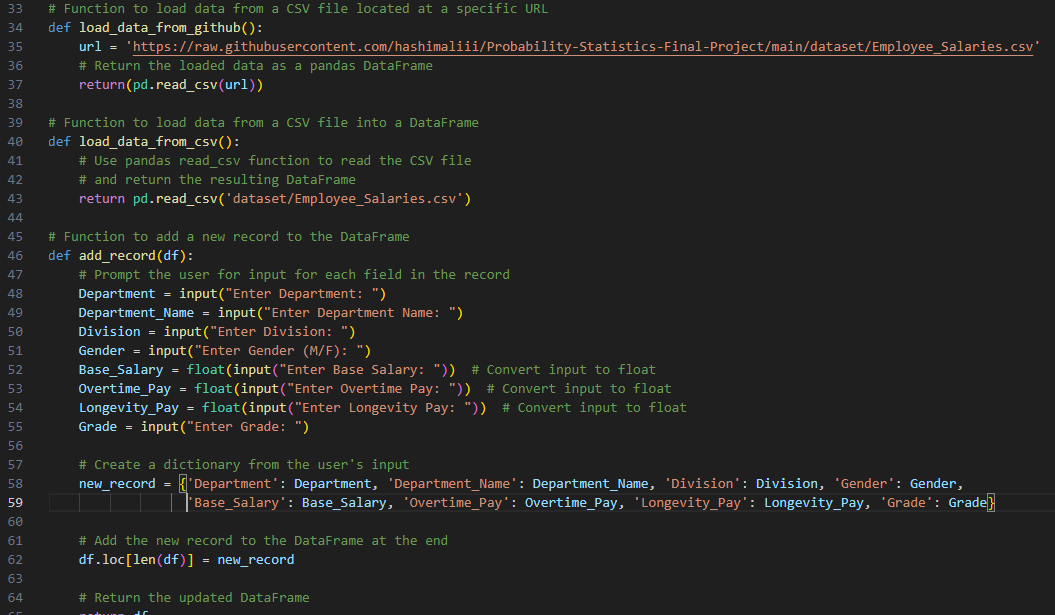
https://www.kaggle.com/datasets/sahirmaharajj/employee-salaries-analysis/data

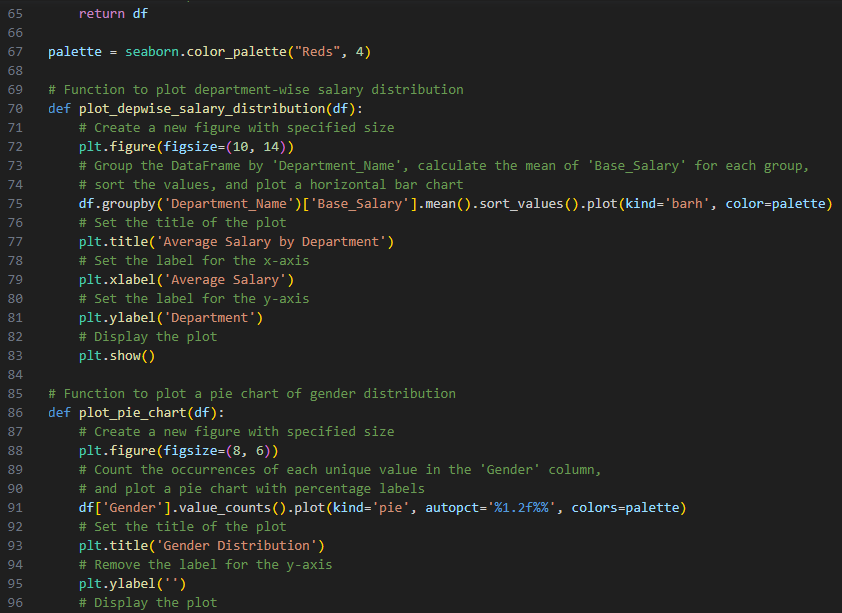
**Results**

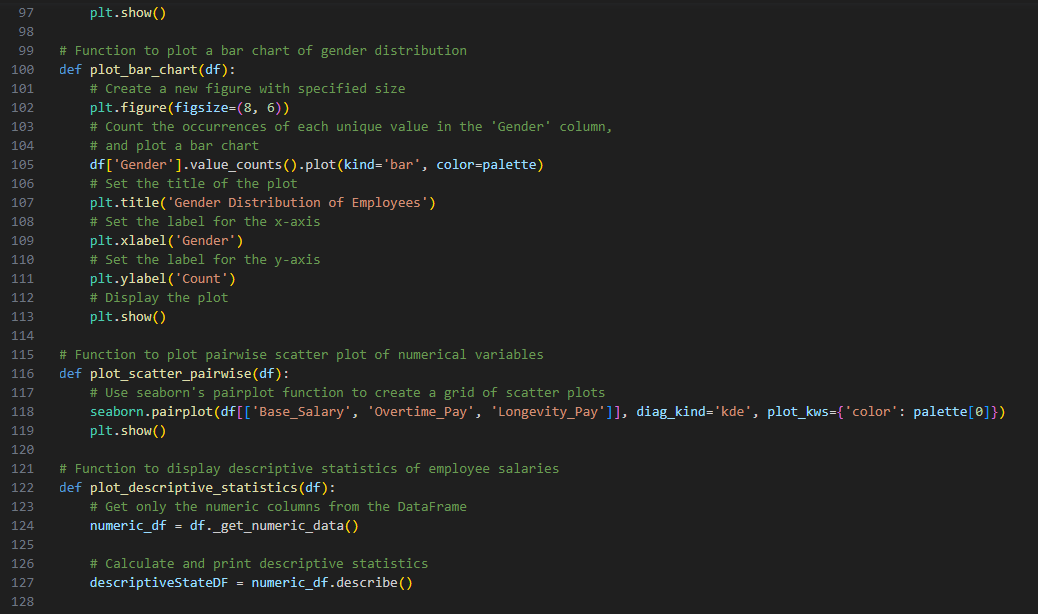
* The **User Interface:**
* For age distribution,we have **Histogram:**
* Then we get age distribution level by **Box Plot:**
* Then we have **Bar Charts,** the prompt first ask for x-axis parameter and then generates the bar chart**:**
* **Pie Charts:**
* **Scatter Plots** for correlation comparison:
* **Poisson distribution:**
* **Logistic Regression:**

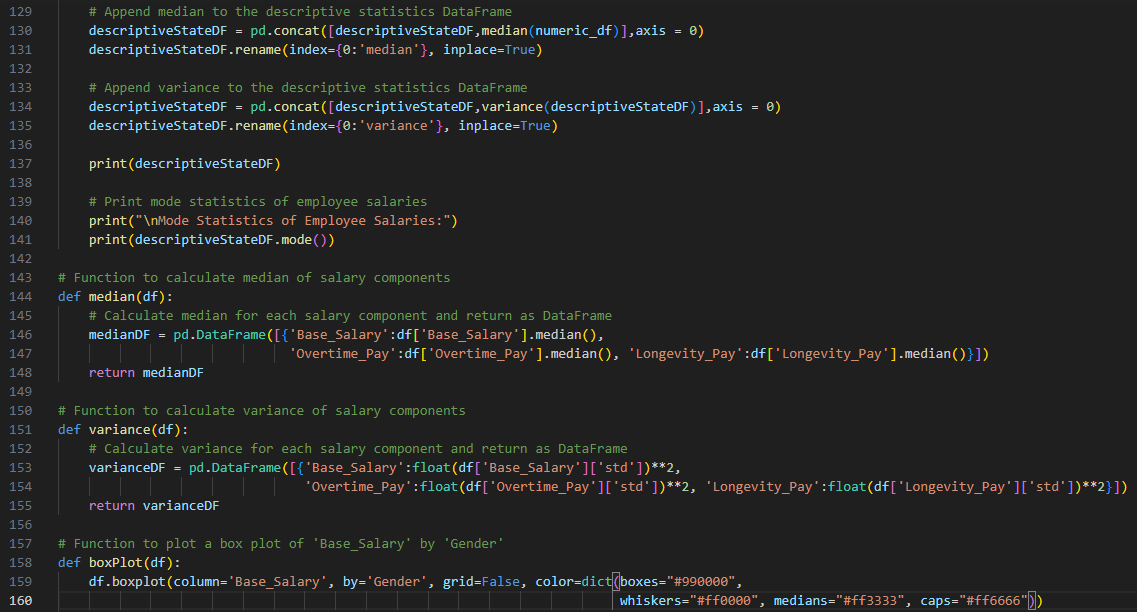
**Code**

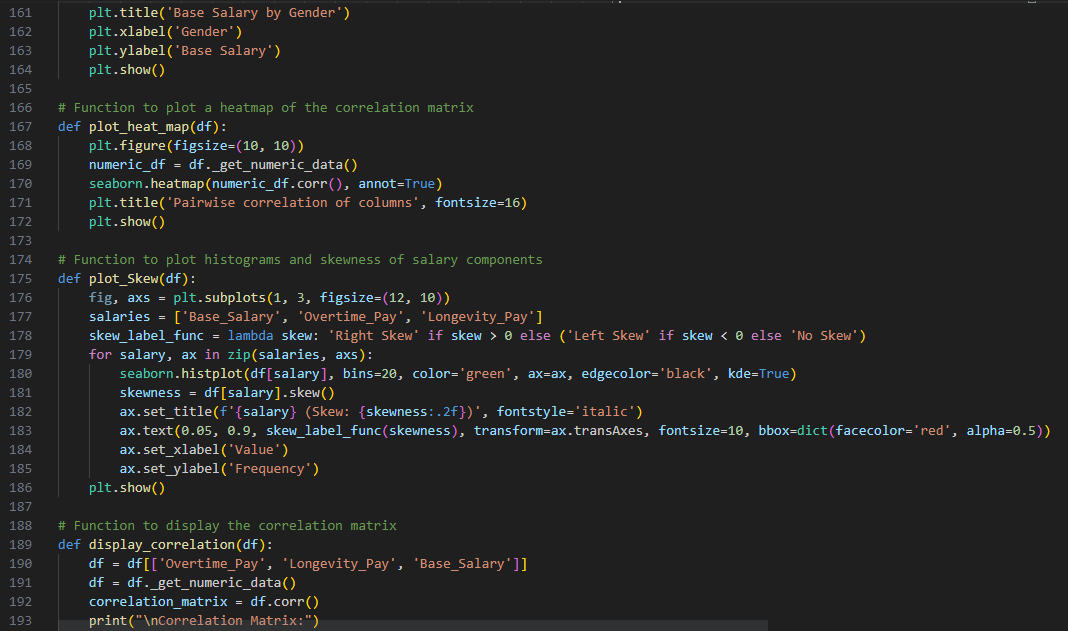
**func.py**

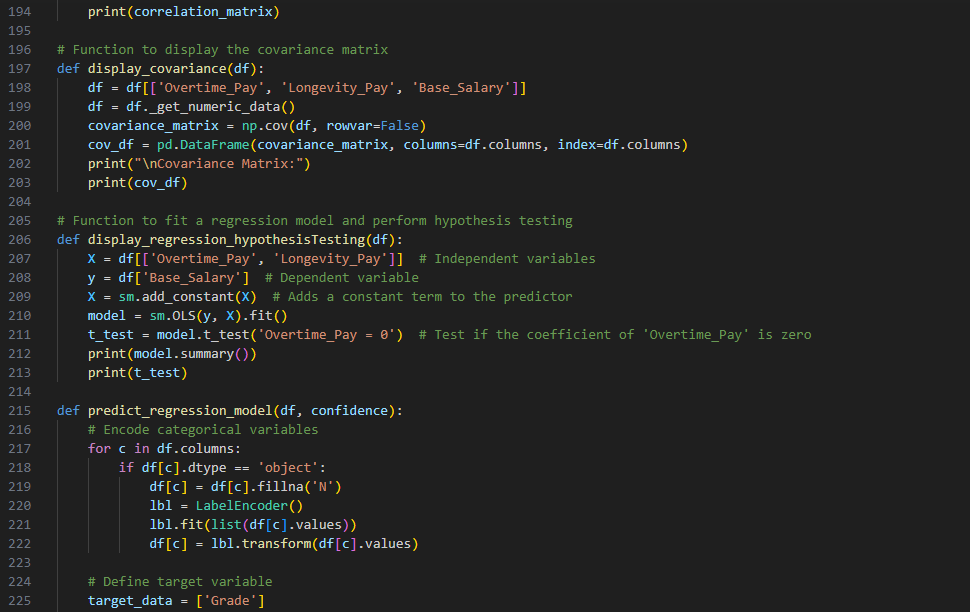
****

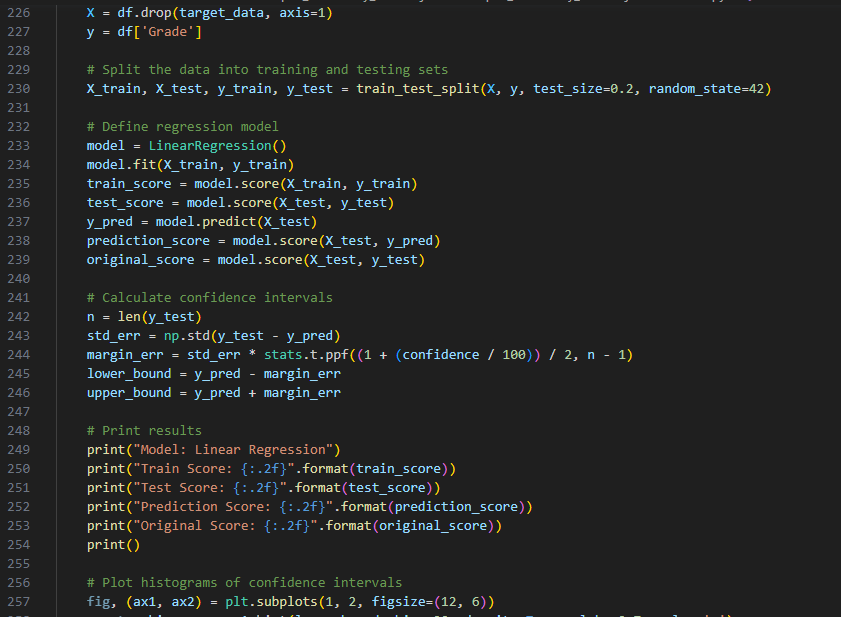
****

****

****

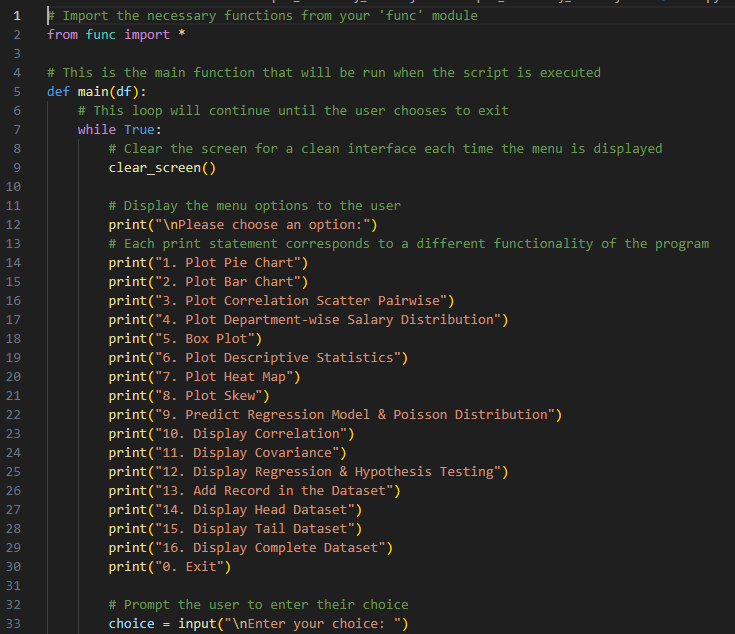
****

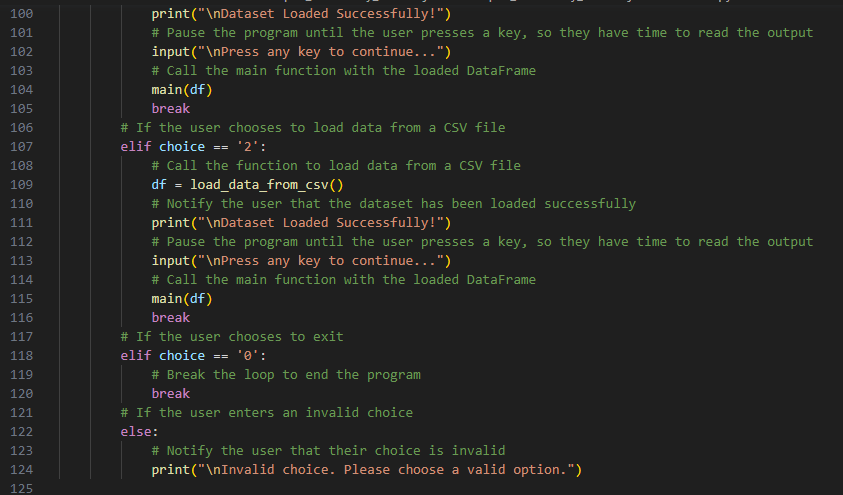
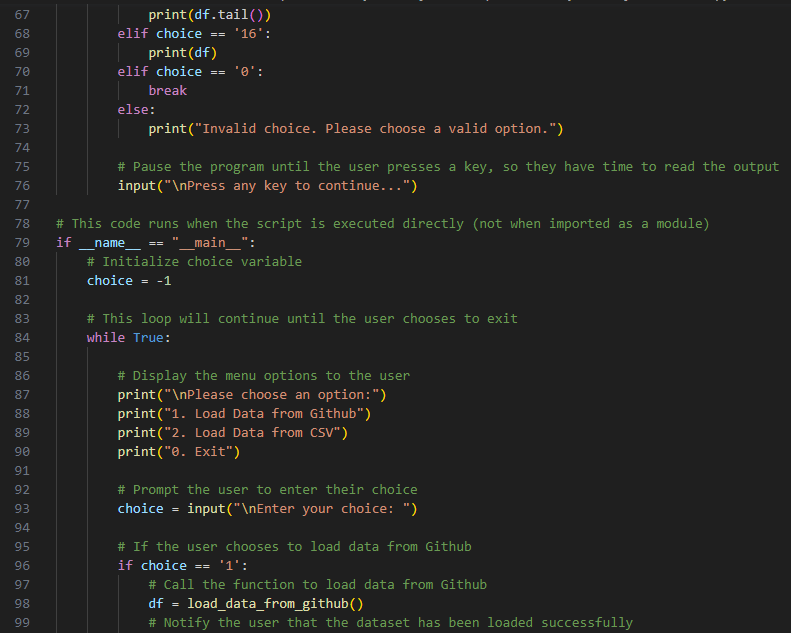
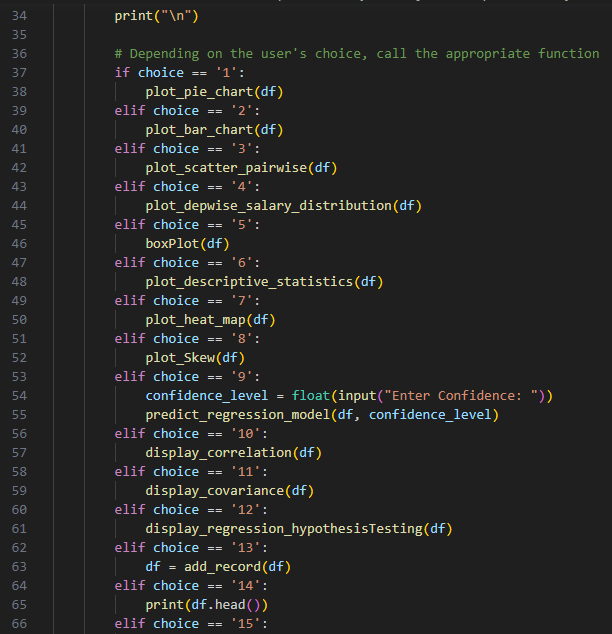
****

****

****

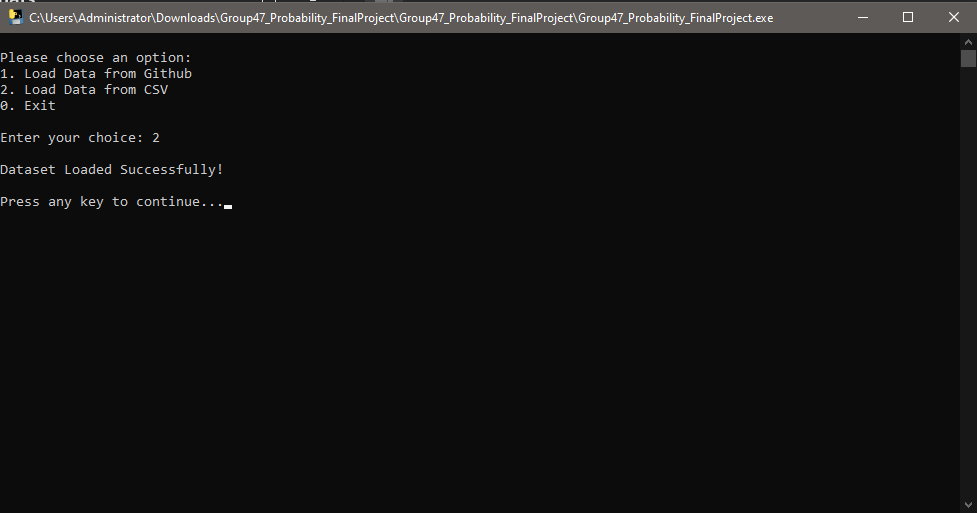
**main.py**

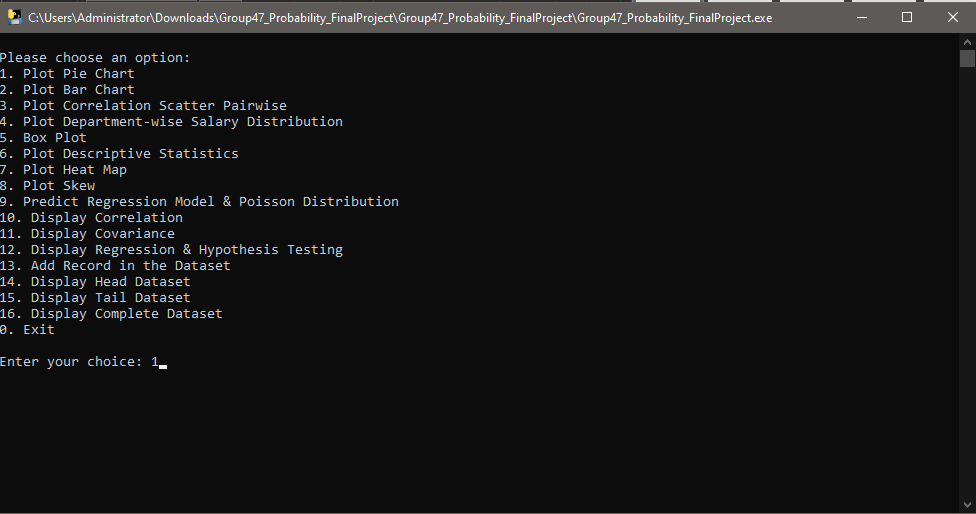
****

****

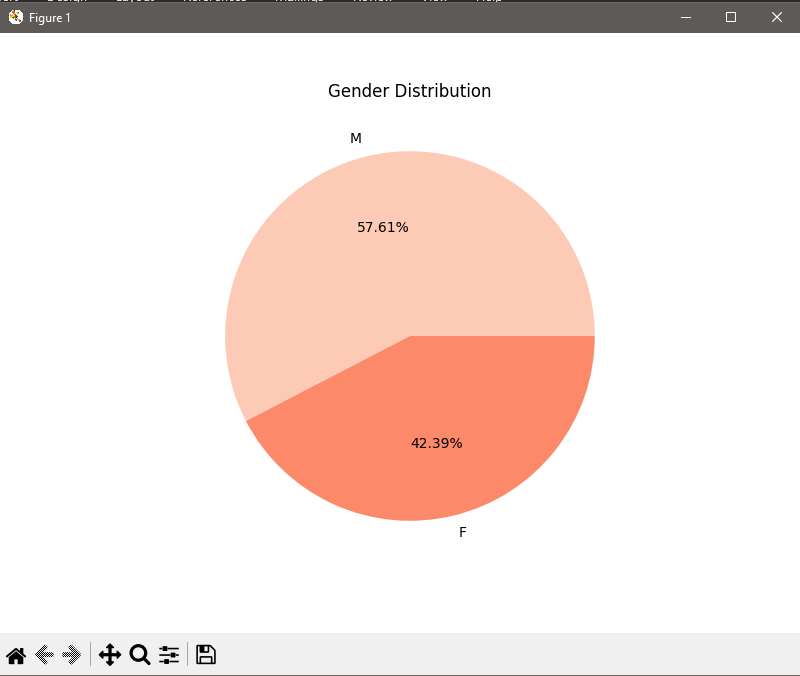
**.exe file**

**Start**

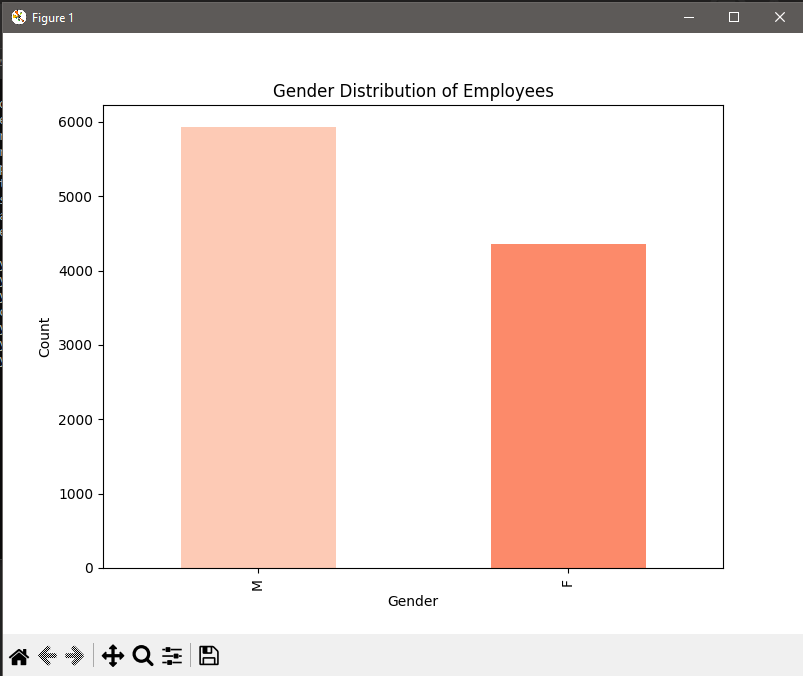
****

**Main Menu**

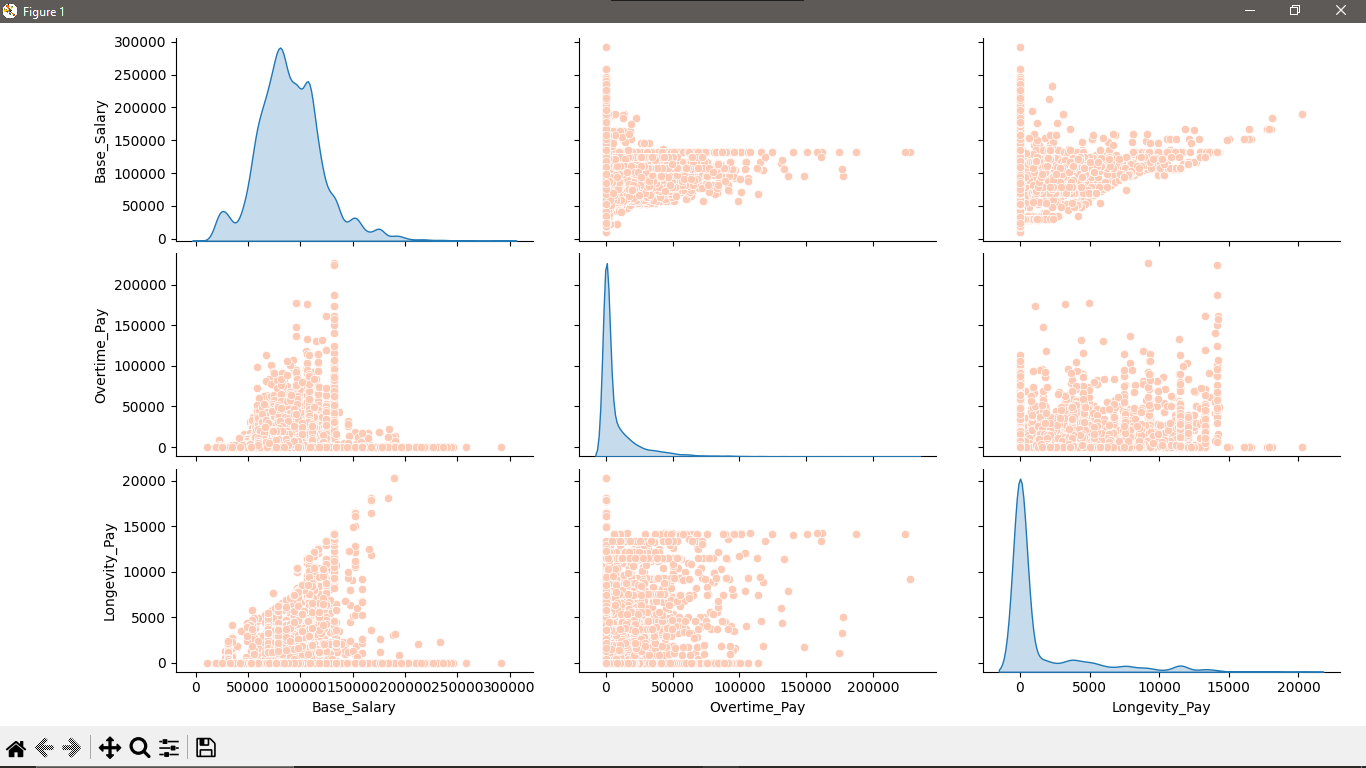
**Pie Chart**

****

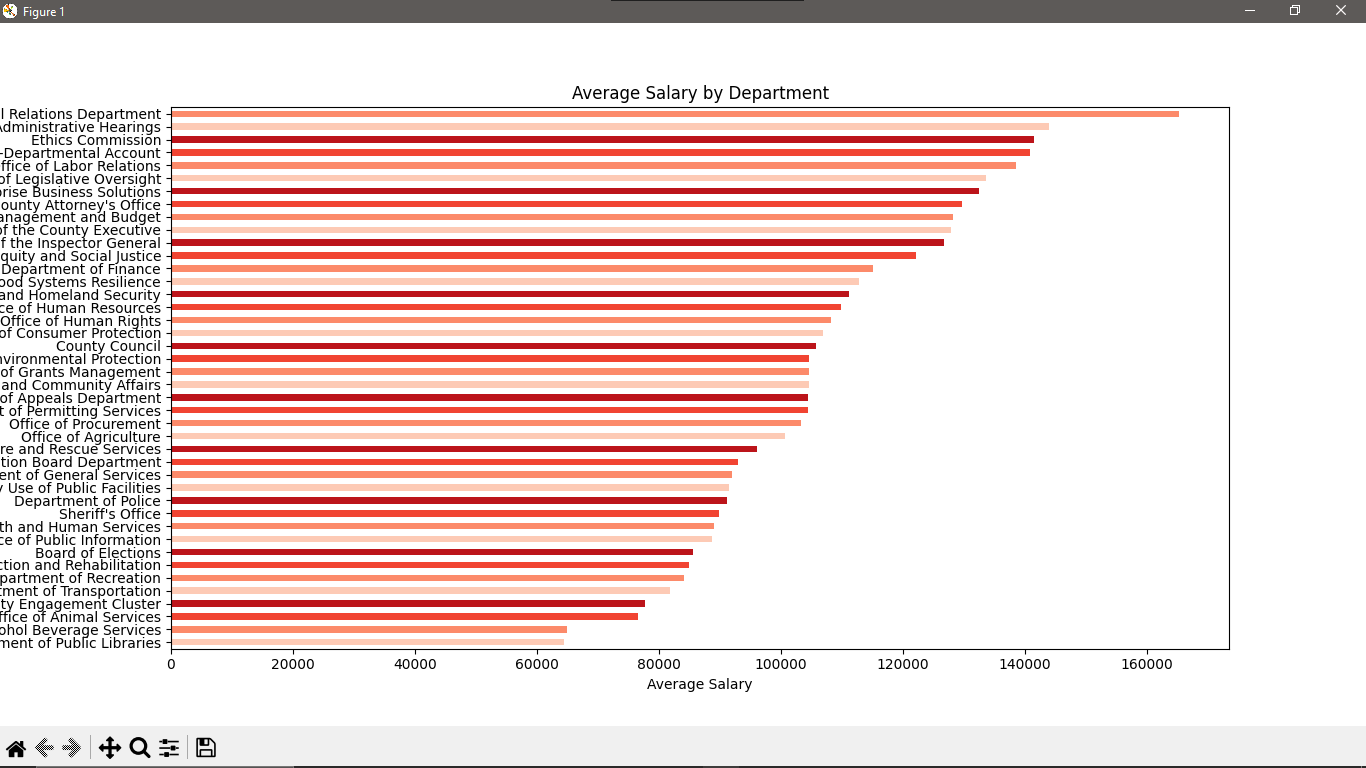
**Bar Chart**

****

**Correlation Scatter**

****

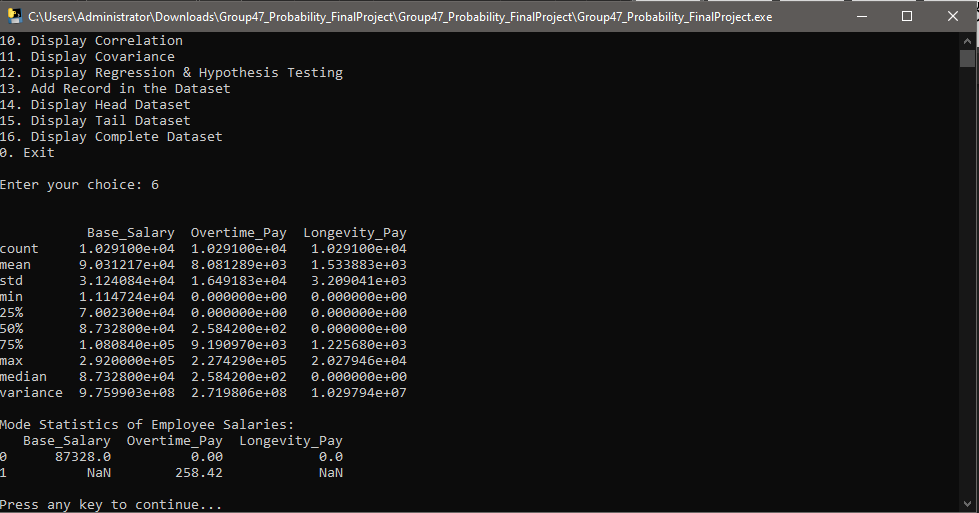
**Department wise Salary Distribution**

****

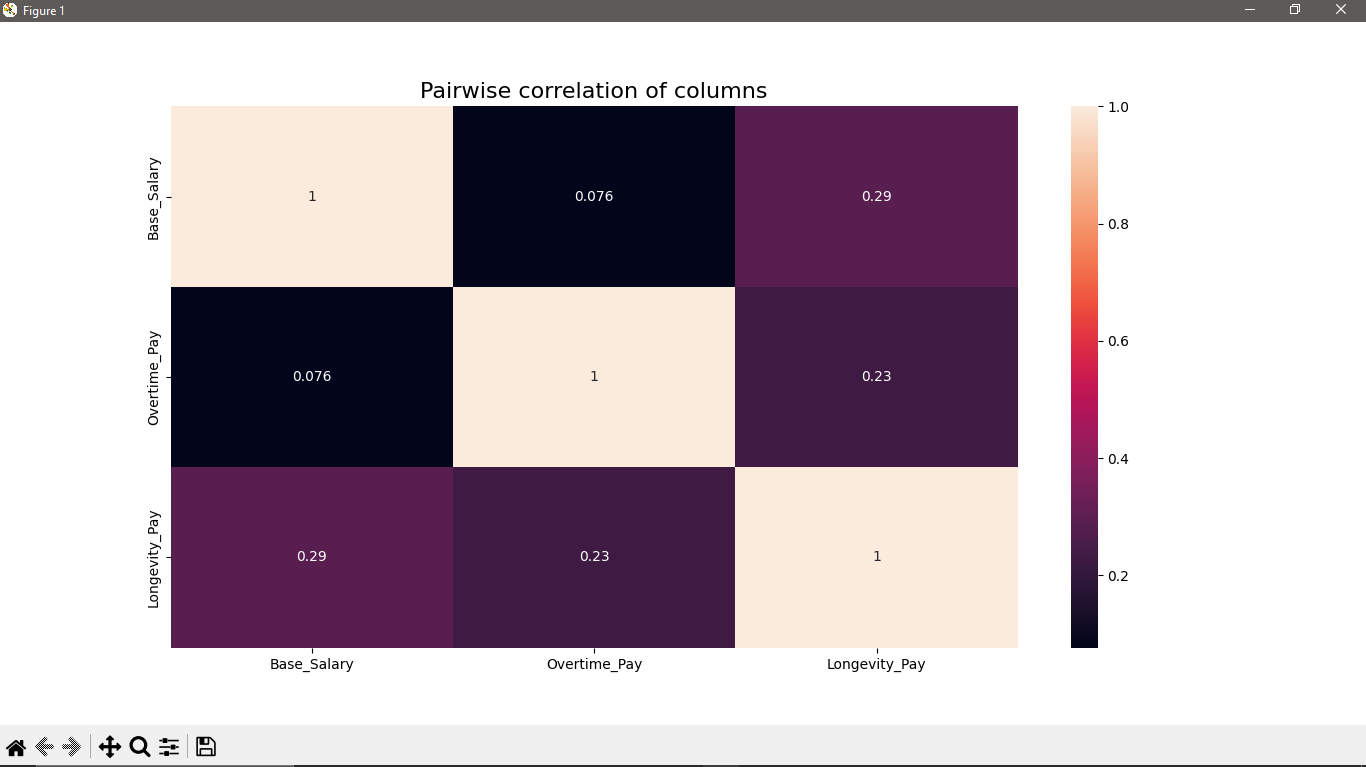
**Box Plot**

****

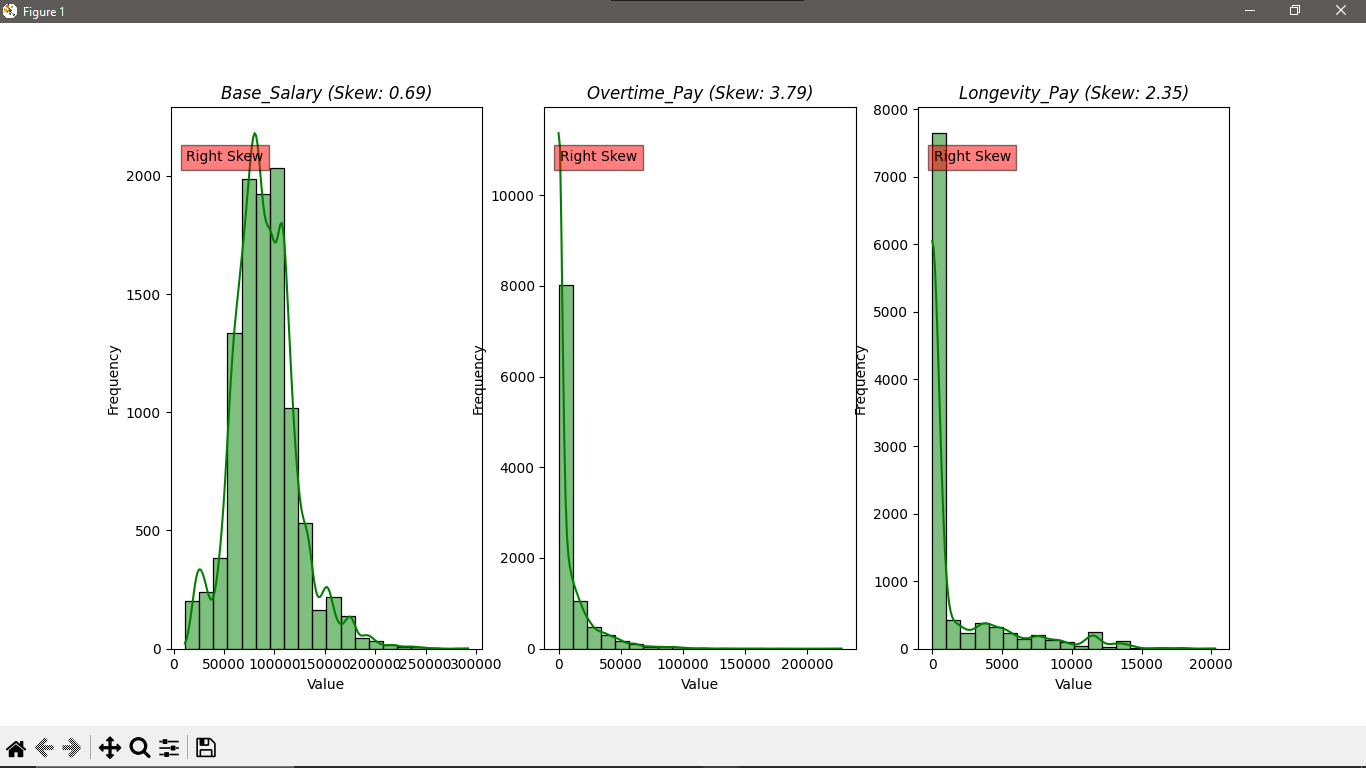
**Descriptive statistics**

****

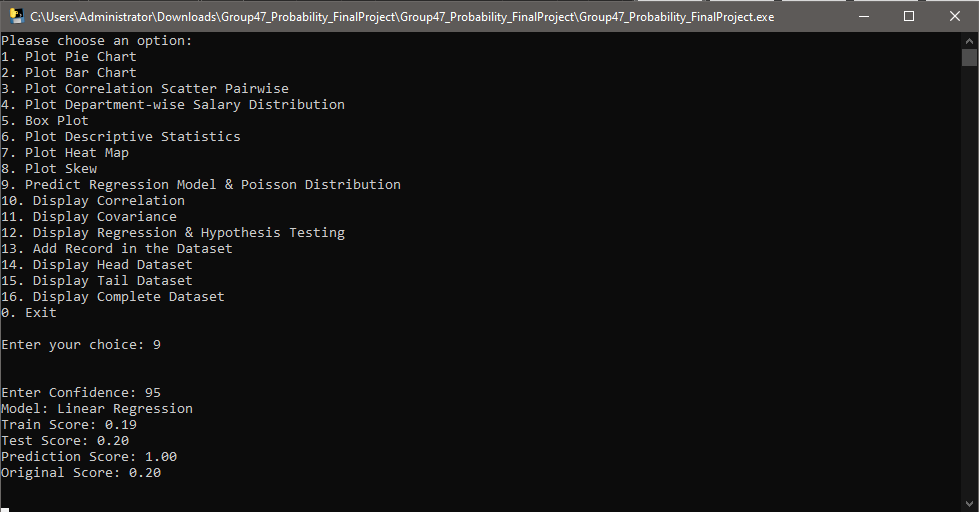
**Heat Map**

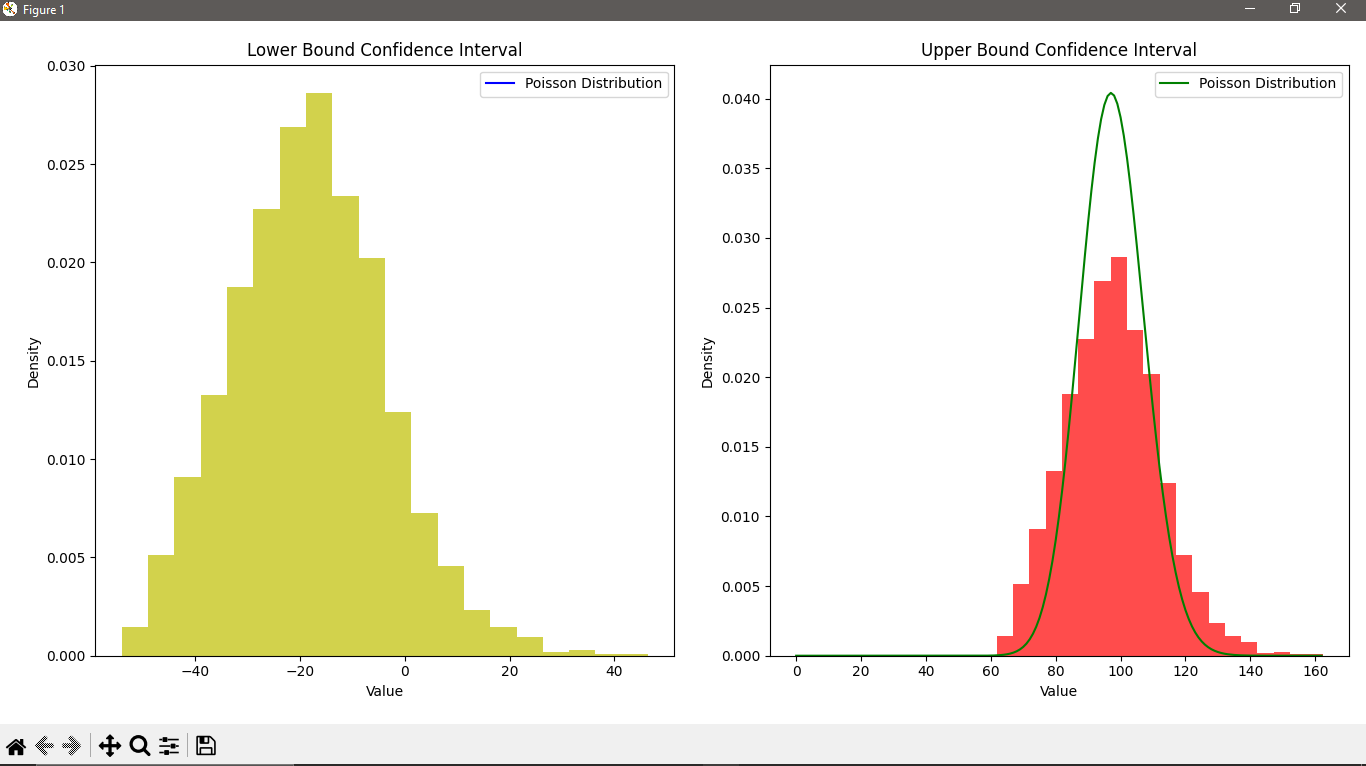
****

**Skew**

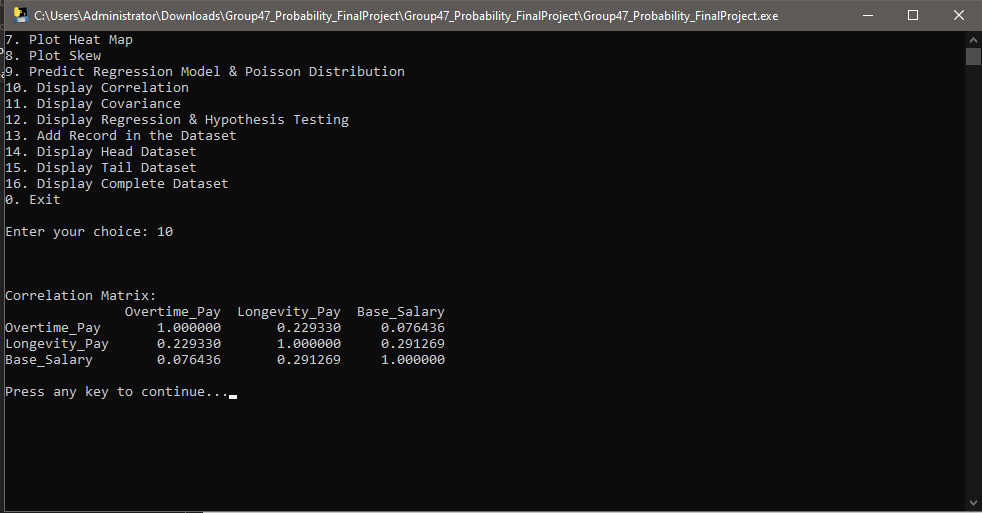
****

**Regression Model And Poisson Distribution**

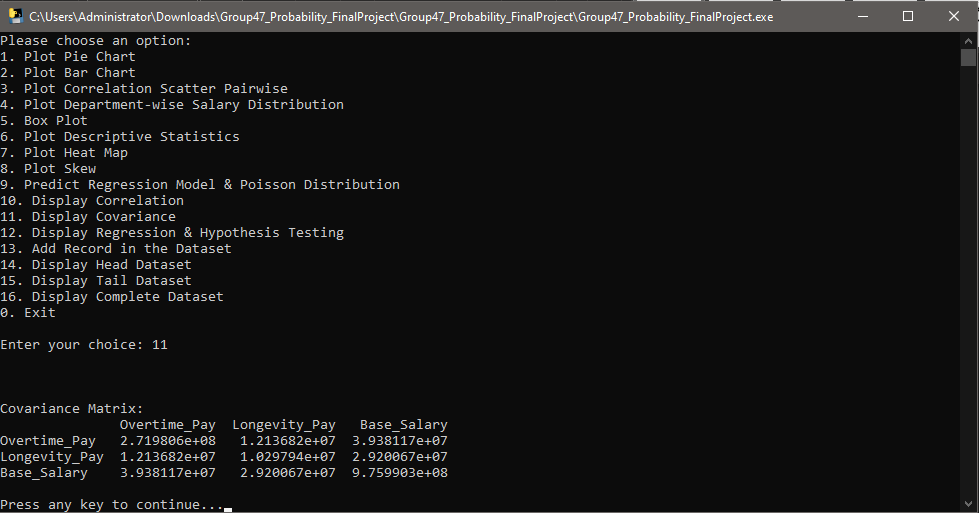
****

****

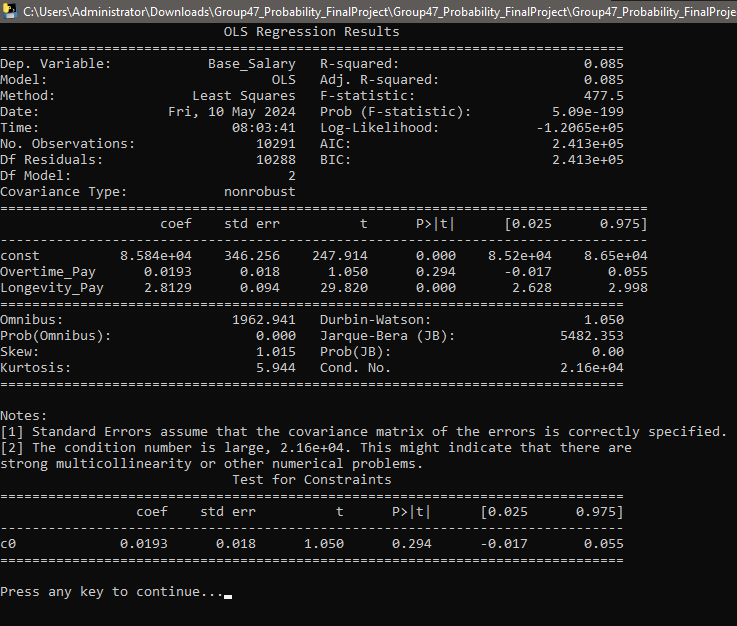
**Correlation**

****

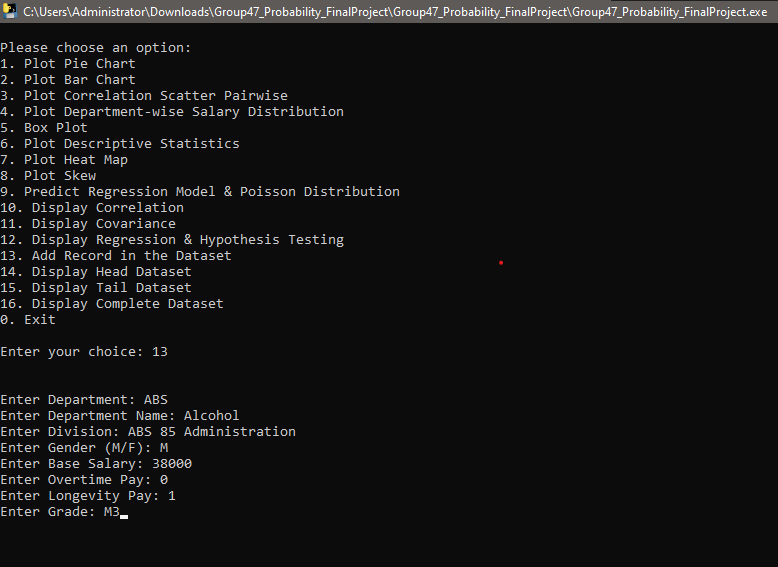
**Covariance**

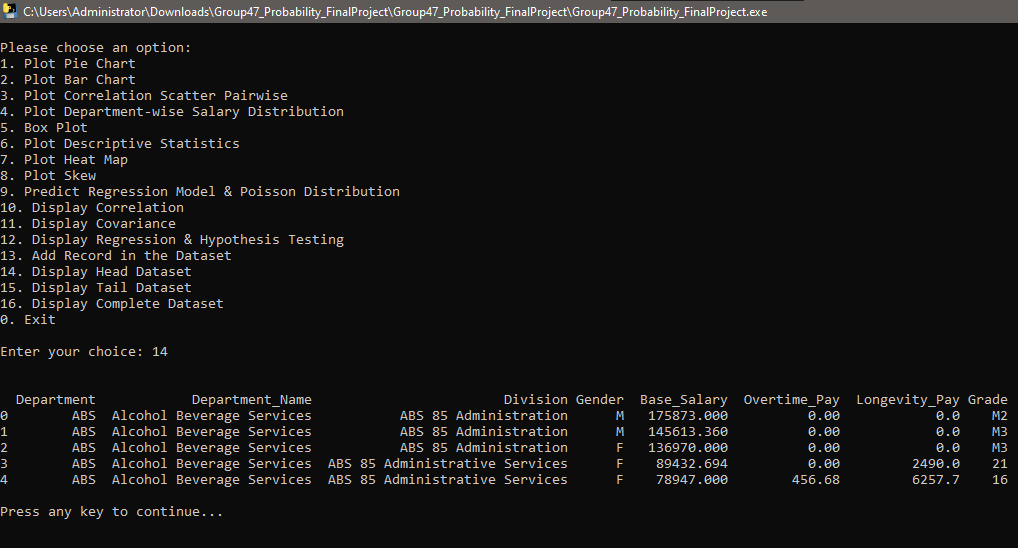
****

**Regression and Hypothesis testing**

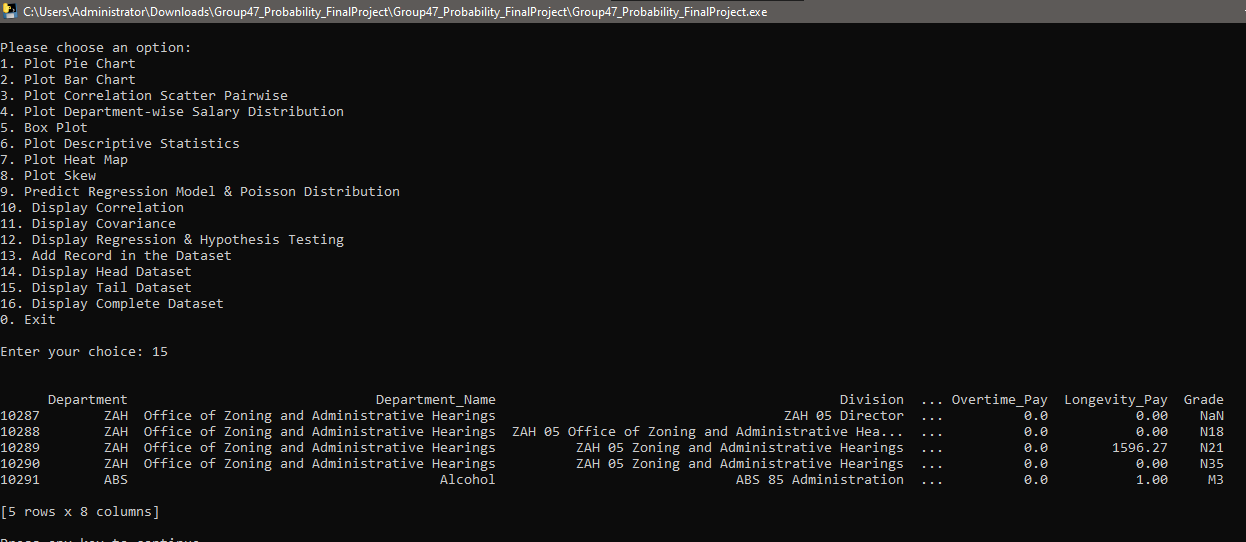
****

**Add Record**

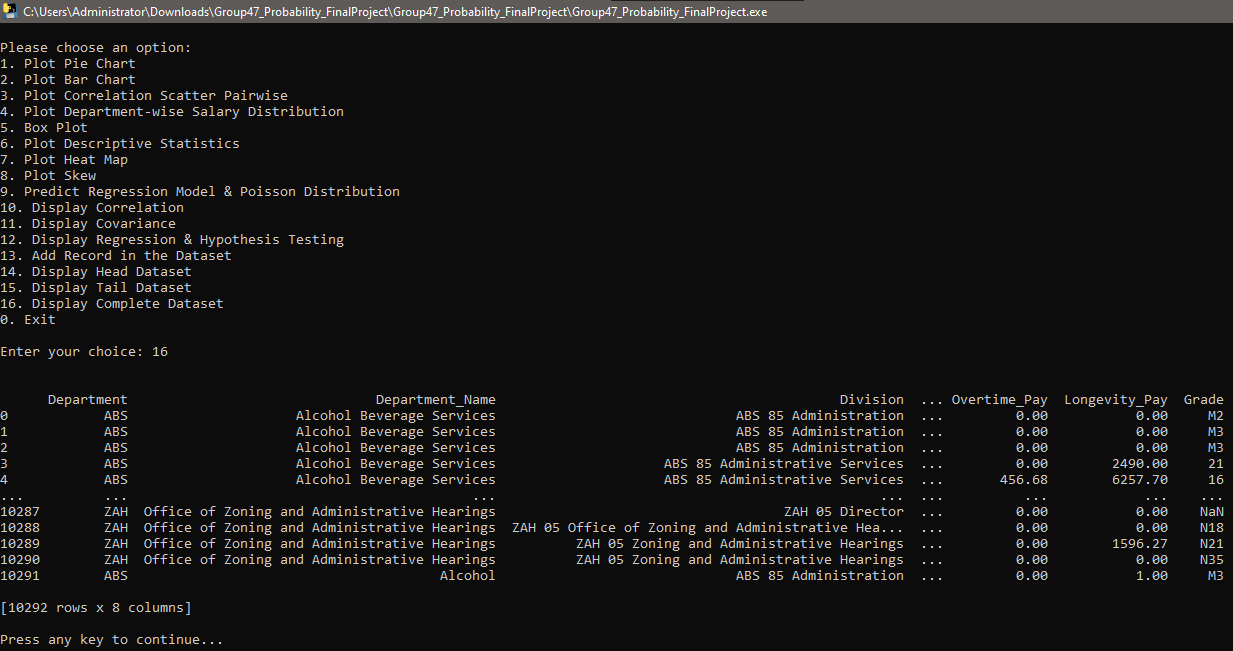
****

**Head Dataset**

**Tail Dataset**

****

**Complete Dataset**

****