Basic Salaries Analysis

By Eng. Mary Alrayes

January 31, 2024

General information and handling missing values

The data consists of 148654 employees and 13 features about their jobs and salaries.

The graph below is a visual representation for the missing values in each column. It shows that two columns (Notes and Status) were entirely empty, therefore; I chose to drop these columns from the data. There were other columns that contain some missing values, mostly in Benefits column as the graph shows below. These missing values were handled either by adding the mean or the mode in place of the null values.

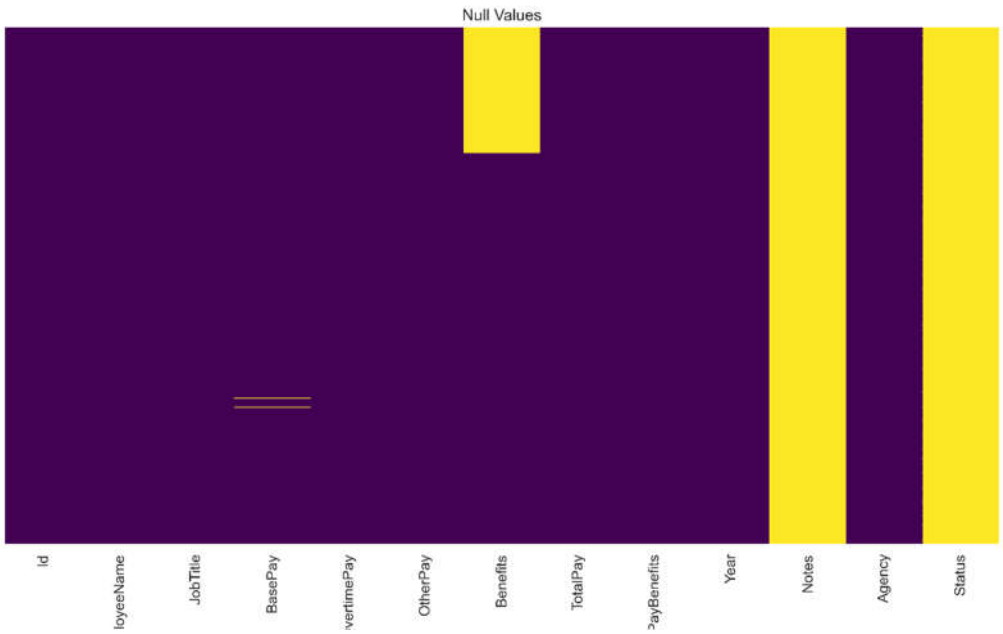

Figure 1: Missing Values in each Column

Salaries Distribution

The two graph below show the salary values distribution. It is clearly that salaries range from 0 to 300000. Most of the salaries were under 100000.

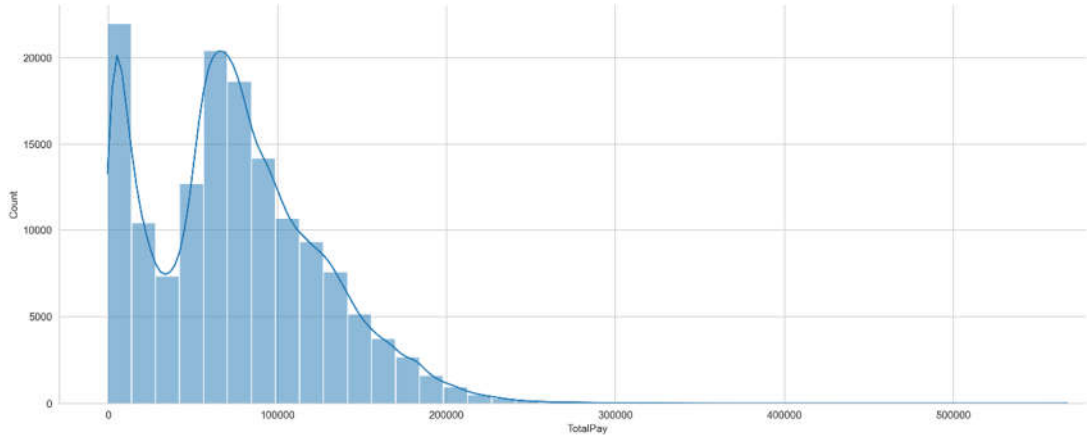

Figure 2: Salaries histogram along with KDB

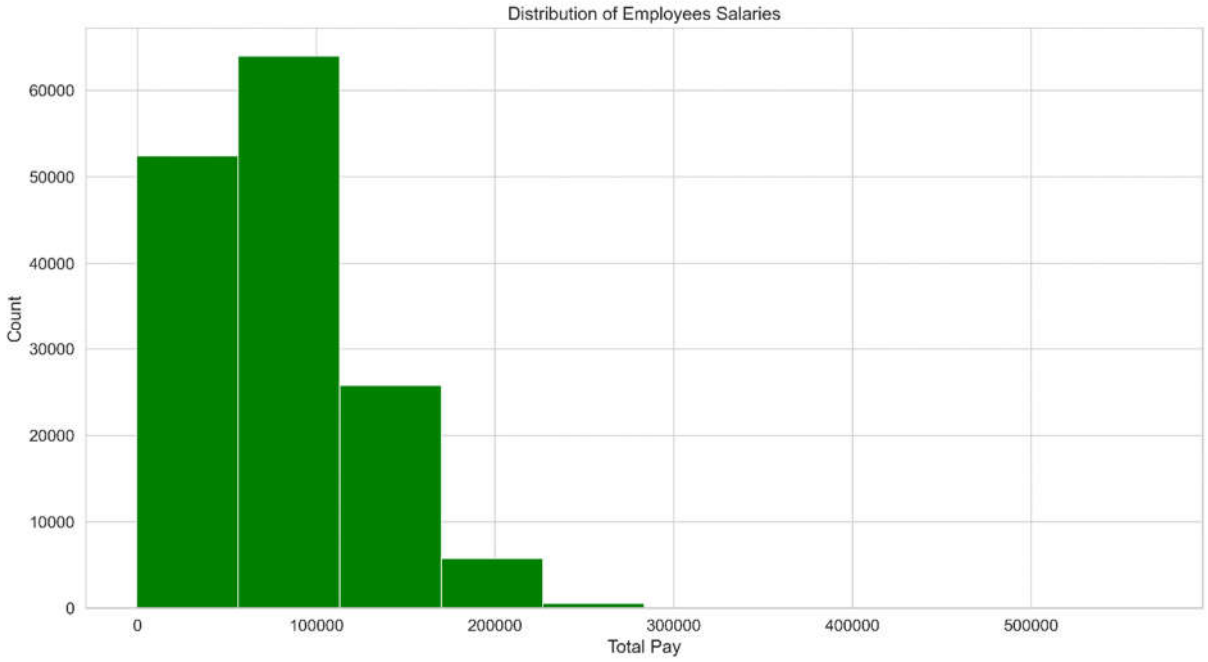

Figure 3: Salaries Distribution

Employees Jobs

There are 2159 different jobs. The pie chart depicts the employees' proportion in the 10 jobs with highest number of employees. The job with the highest number of employees is "Transit Operator", 4.73% of the total employees.

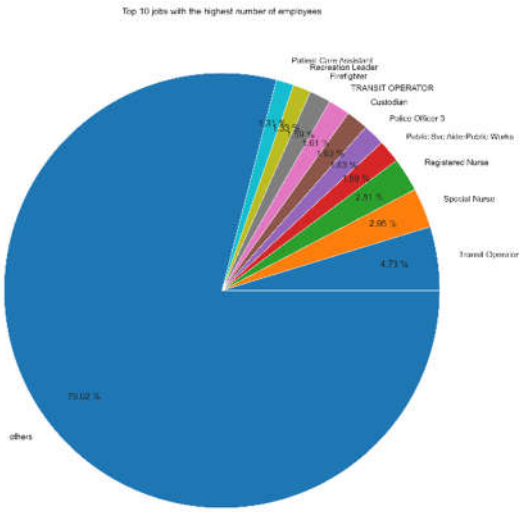

Figure 4: Top 10 jobs with the highest number of employees

Employees distribution in each year

The data was collected to represent employees' salaries in 4 years. The pie chart illustrates that there was almost an equal percentage of employees in all the 4 years.

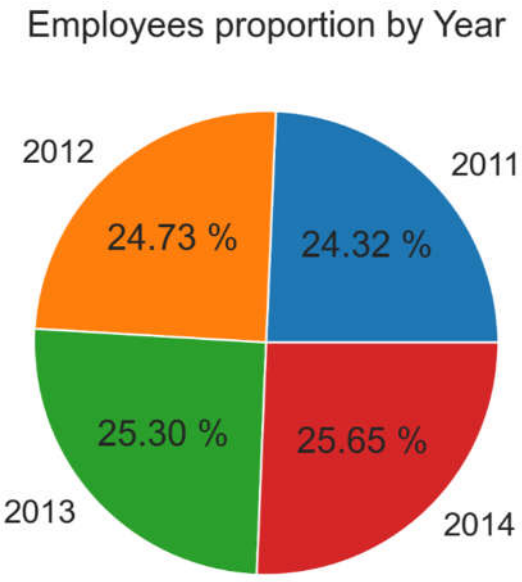

Figure 5: Employees proportion by Year

Average Salaries

The bar chart below depicts that the average salaries in all 4 years were very close to each other. The highest average salary was in 2013, accounting for over than 78000.

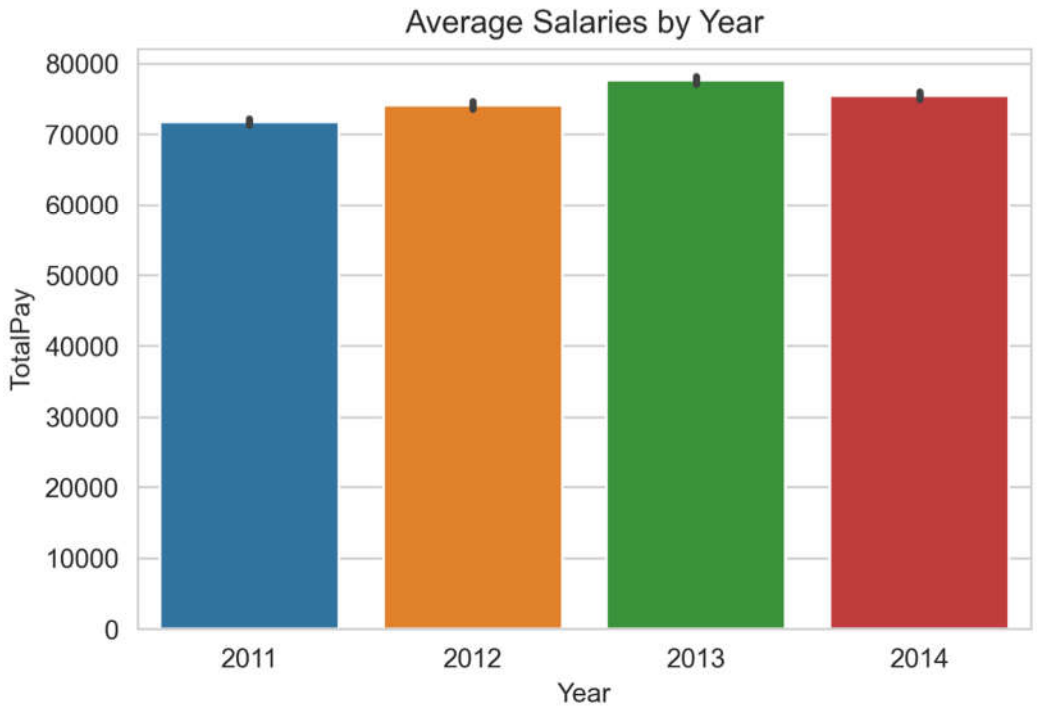

Figure 6: Average Salaries by Year

Salaries Correlation

The bar chart below compares the correlation between the salary and each of the other features. It is clearly that the Base Pay and Total Pay Benefits features have the highest correlation, approximately 0.97 and 0.95 respectively.

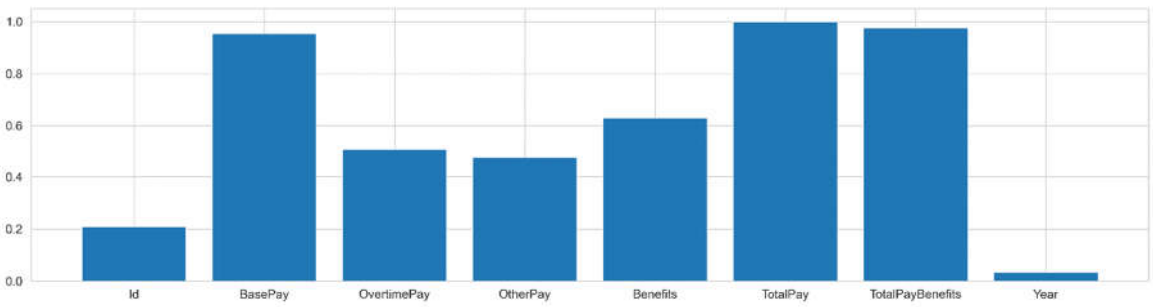

Figure 7: Salaries correlation with other features

Salary relation with other Features

The scatter graphs below show the relation between the salary's values and the other features' values. There is a linear relation between the values in all the graphs except for the year graph. There are some outliers in the values, and it is noticeable that many features' values are zero.

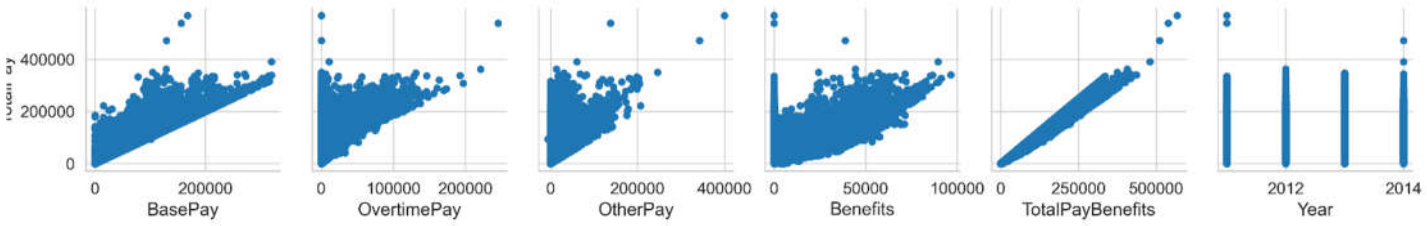

Figure 8: Salary relation with each feature