SUBMITION REPORT

By : Mohammed haider

Summary of Insights

1. Basic Data Exploration:

- The dataset contains **148654** rows and **13** columns.

- The data types of each column include [int64 , object , float64].

- There are missing values across the dataset, for each column in the dataset we have saw this :

- BasePay has 609 missing values .
- OvertimePay has 4 missing values .
- OtherPay has 4 missing values .
- Benefits has 36163 missing values .
- Notes has 148654 missing values .
- Status has 148654 missing values .

2. Descriptive Statistics:
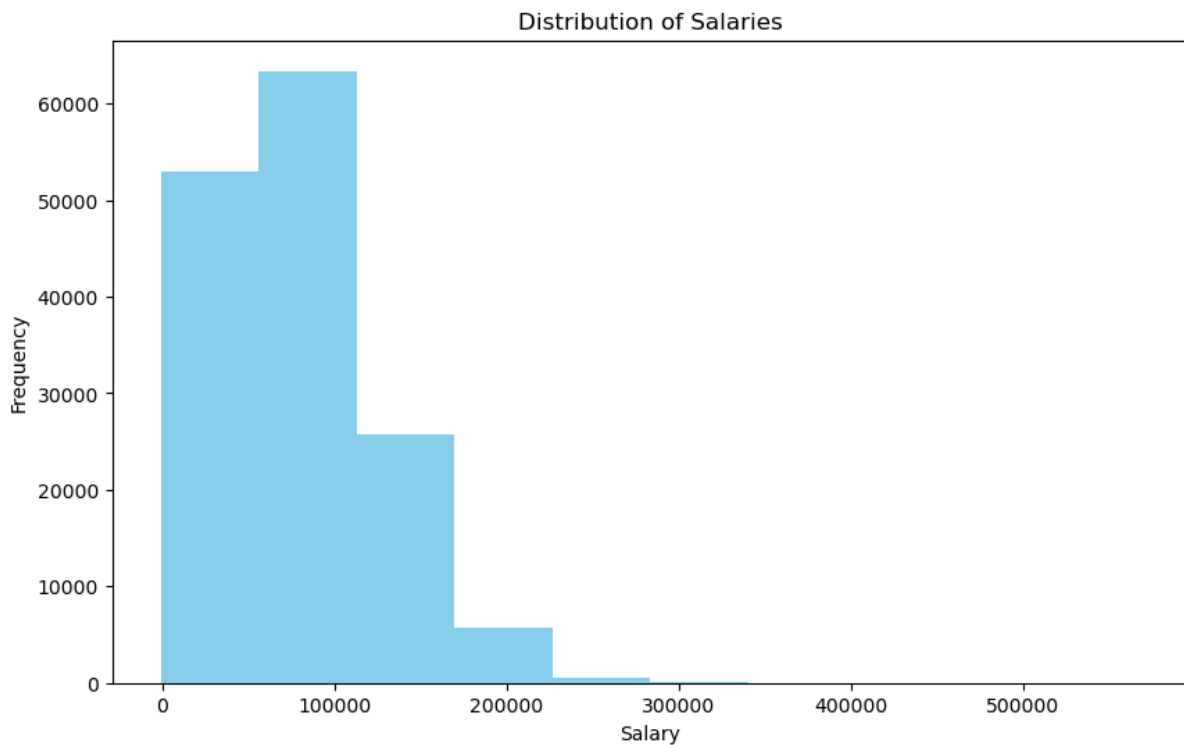
- The average salary in the dataset is 74768.32197169265 ,with a minimum salary of -618.13 and a maximum salary of 567595.43 .

- The range of salaries is 568213.56 , and the standard deviation is [50517.00527394988].

3. Data Cleaning:

- I have dropped the columns that does not have any values .

- I used the mean to fill missing values in the salary column. This is a common approach, especially when the missing values are few and the variable is approximately normally distributed.

4. Basic Data Visualization:

- A histogram was created to visualize the distribution of salaries :

Distribution of Salaries

5. Grouped Analysis:

- some data :

```
Average Salary by Department:

JobTitle
Account Clerk                                    43009.395299
Accountant                                       47429.268000
Accountant I                                     62036.101250
Accountant Ii                                    67355.202699
Accountant Iii                                   79001.071547
                                                      ...
Worker'S Compensation Supervisor I               91020.726000
X-Ray Laboratory Aide                            51212.279048
Youth Comm Advisor                               41414.307500
Youth Commission Advisor, Board Of Supervisors   53632.870000
Zoo Curator                                      66686.560000
```