**Link to coolab:**

https://colab.research.google.com/drive/13DU6PC7ZJme5T0xW8grqiH3QkFEbsM-t?usp=sharing

summary of the findings and insights gained from the analyses conducted on the dataset.

The dataset used for analysis contains information related to employee salaries and job details.

1. Mean or Median Imputation:

To address missing values in the dataset, mean or median imputation was performed. Missing values in numerical columns such as 'BasePay', 'OvertimePay', 'OtherPay', and 'Benefits' were replaced either by their mean or median values. This process allowed us to ensure completeness in the dataset and proceed with further analysis.

2. Basic Data Visualization:

Histograms and bar charts were utilized to visualize the distribution of salaries and the proportion of employees in different Jobs.

- Salary Distribution:

A histogram was created to visualize the distribution of salaries. The 'TotalPayBenefits' column was used to represent the salary information. The histogram showcased the frequency of salaries within different ranges, giving an overview of the salary distribution across the dataset.

- Jobs Proportions:

Bar charts were used to represent the proportion of employees in different jobs. The 'JobTitle' column was used to identify the jobs, and the bar chart showed the number of employees in the top 10 Jobs by count. Additionally, a pie chart was created to represent the proportion of employees in the top 5 jobs, providing a visual representation of the job composition.

3. Grouped Analysis:

The dataset was grouped by one or more columns to calculate summary statistics for each group. The average salaries were compared across different groups using the 'TotalPayBenefits' column. This analysis provided insights into the variations in average salaries based on different groupings, such as departments and years.

4. Simple Correlation Analysis:

A correlation analysis was conducted to identify any correlation between salary (basepay) and another numerical column (e.g., benefits). The correlation coefficient was calculated, indicating the strength and direction of the relationship between the variables. A scatter plot was generated to visualize the relationship, allowing for a better understanding of the correlation between them.

**Conclusion:**

The analyses performed on the dataset provided several key insights:

- The distribution of salaries was visualized, giving a clear understanding of how salaries are spread across different ranges.

- The proportion of employees in different jobs was identified, highlighting the jobs with the highest number of employees.

- Grouped analysis allowed for a comparison of average salaries across different groups, providing insights into salary variations.

- The correlation analysis between basepay and benefits revealed the relationship between the two variables, with the scatter plot providing a visual representation.

These findings and insights can be used to gain a deeper understanding of the dataset and inform decision-making processes related to employee compensation, departmental analysis, and the impact of experience on salary.