SHAI

Share Artificial Intelligence

Summary of Insights

Eng / Mohamed Abdelbaset Bakr

Basic Data Exploration:

- Loaded the dataset using pandas
- Checked the information about the dataset using **info()** to identify data types and missing values.
- Determined the number of rows and columns using the **shape** attribute.
- Checked data types of each column using **dtypes**.
- Checked for missing values in each column using **isnull().sum()**.

Descriptive Statistics:

- Calculated basic statistics for the 'TotalPayBenefits' column, including mean, median, mode, minimum, maximum, range, and standard deviation**.**

Result:

Mean salary: 93692.555
Median salary: 92404.090
Mode salary: 7959.180
Minimum salary: -618.130
Maximum salary: 567595.430
Salary range: 568213.560
Standard deviation of salary: 62793.533

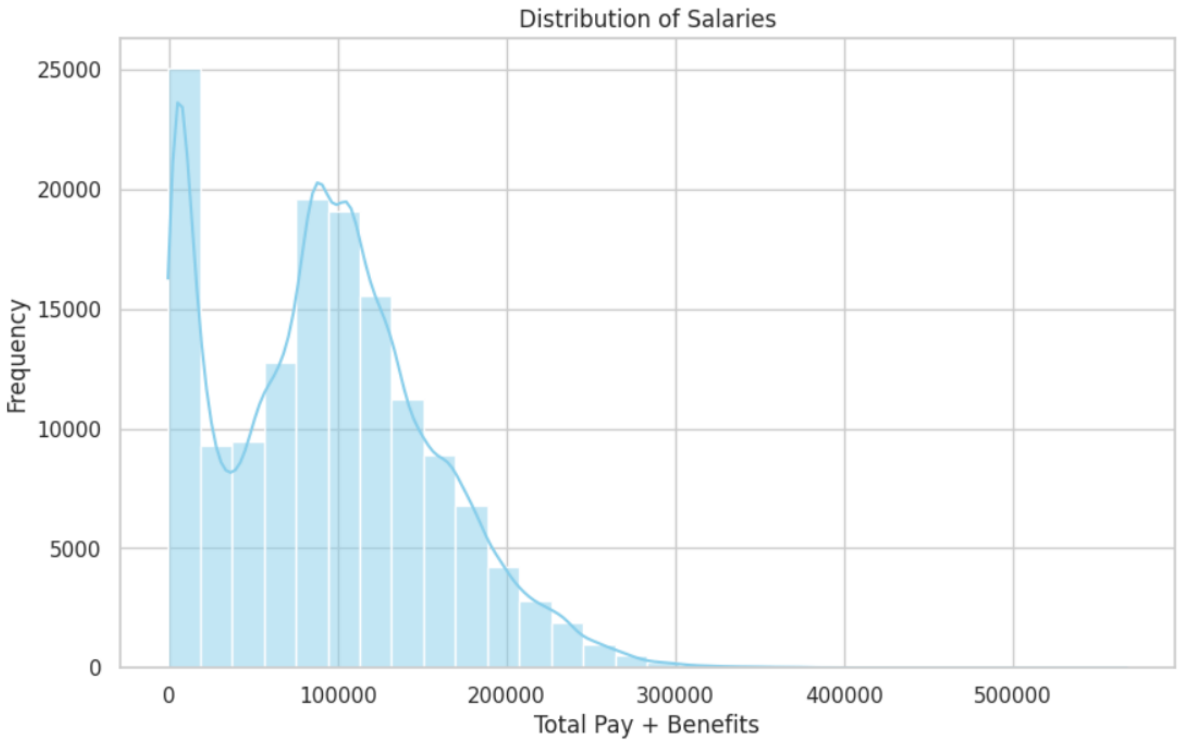
Data Cleaning:

- Identified missing values in the columns and imputed them with the mean.
- Imputed missing values in the 'Notes' and 'Status' columns with appropriate placeholder categories.
- Adapted imputation strategies based on the nature of the columns and the type of missing data.

Basic Data Visualization:

- Used **matplotlib** and **seaborn** libraries to create visualizations.
- Constructed histograms to visualize the distribution of salaries.

- Generated a pie chart to represent the proportion of employees in different departments.
- Adjusted plot settings for clarity and readability.

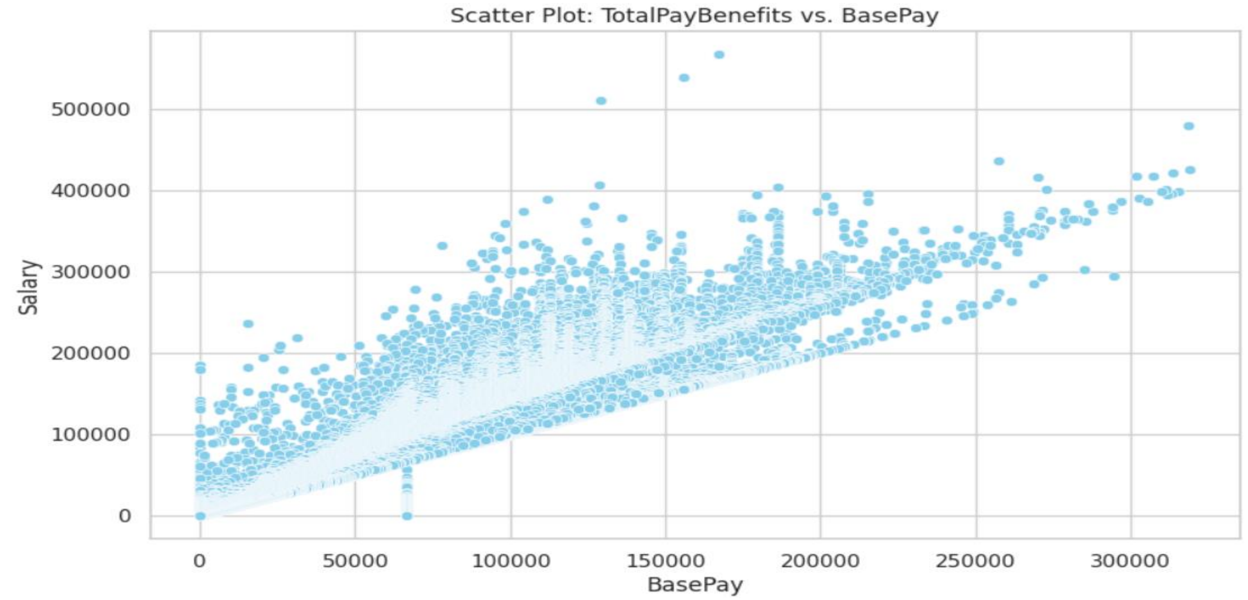


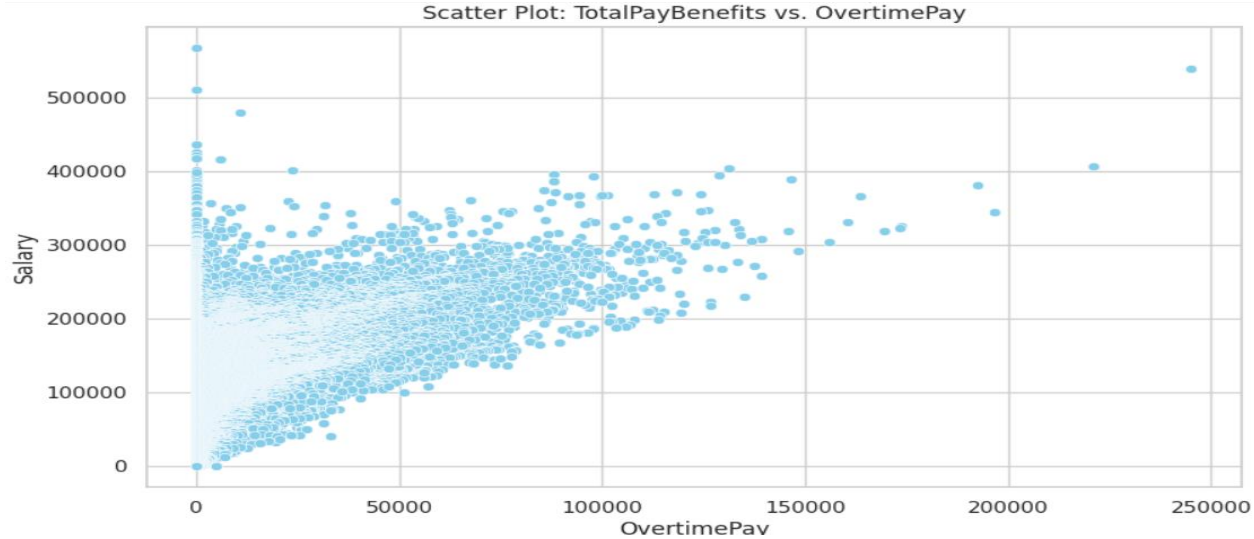
Grouped Analysis:

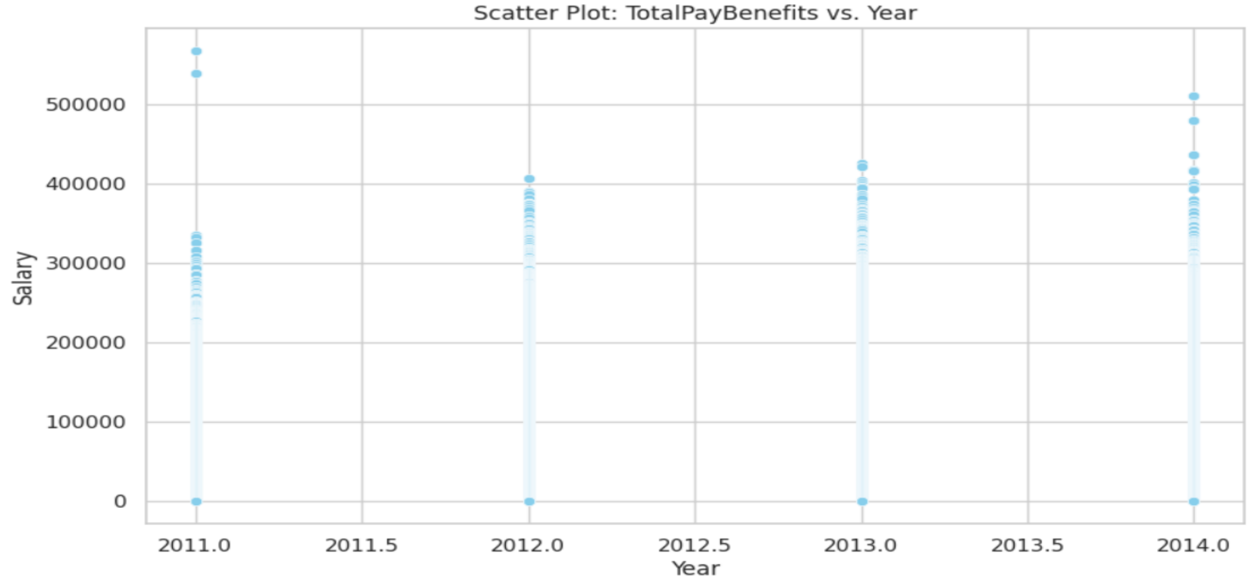
- Utilized the groupby function to group data by the 'JobTitle' column.
- Calculated summary statistics for each group, such as count, mean, median, minimum, maximum, and standard deviation.
- Sorted the grouped data to compare average salaries across different job titles.

Simple Correlation Analysis:

- Calculated the correlation coefficient between 'TotalPayBenefits' and another numerical column (e.g., 'BasePay').
- Created a scatter plot to visualize the relationship between the two variables.


Scatter Plot: TotalPayBenefits vs. BasePay


Scatter Plot: TotalPayBenefits vs. OvertimePay


Scatter Plot: TotalPayBenefits vs. Year