**Summary of Insights**

**1. Data Overview:**

- The dataset contains information on employee salaries with columns like 'Id', 'EmployeeName', 'JobTitle', 'BasePay', 'OvertimePay', 'OtherPay', 'Benefits', 'TotalPay', 'TotalPayBenefits', 'Year', 'Agency', 'Notes', and 'Status'.

- Various data types are present, including integers, floats, and objects.

- Missing values are observed in columns such as 'BasePay', 'OvertimePay', 'OtherPay', 'Benefits', 'Notes', and 'Status'.

**2. Descriptive Statistics:**

- Descriptive statistics provide insights into the distribution of numerical columns:

- Mean TotalPayBenefits: $93,692.55

- Minimum TotalPayBenefits: $0

- Maximum TotalPayBenefits: $567,595.43

- The standard deviation indicates variability in columns such as 'BasePay', 'OvertimePay', 'OtherPay', 'Benefits', 'TotalPay', and 'TotalPayBenefits'.

**3. Data Cleaning:**

- Columns 'Notes' and 'Status' are dropped due to being completely empty, providing no meaningful information.

- Rows with negative values in salary-related columns ('BasePay', 'OvertimePay', 'OtherPay', 'Benefits') are removed to ensure the integrity of salary data.

- Missing values in 'BasePay' are handled by dropping corresponding rows due to their limited number, preserving the dataset's accuracy.

- Missing values in 'Benefits', 'BasePay', 'OvertimePay', and 'OtherPay' are replaced with the mean values of their respective columns, maintaining a representative distribution of the data.

- After the cleaning process, no missing values or negative values are present in the dataset, enhancing its completeness and suitability for further analysis.

**4. Basic Data Visualization:**

- A histogram of 'TotalPayBenefits' shows the salary distribution.

A graph showing a distribution of benefits

Description automatically generated

- A pie chart illustrates the proportion of employees in different job titles ('JobTitle').

A pie chart with numbers and text

Description automatically generated

**5. Grouped Analysis:**

- The dataset is grouped by 'JobTitle,' and average 'TotalPayBenefits' is calculated for each group.

- The top 10 job titles with the highest average salaries are presented, revealing positions like 'Chief Investment Officer' and 'Chief of Police' as top earners.

**6. Simple Correlation Analysis:**

Correlation Matrix:

A graph of a graph

Description automatically generated with medium confidenceUtilizing the DataFrame.corr() function in Python, a correlation matrix was generated for various variables. The correlation values with respect to 'TotalPayBenefits' are as follows:

TotalPayBenefits: 1.000000

TotalPay: 0.977308

BasePay: 0.943008

Benefits: 0.842997

OvertimePay: 0.467961

OtherPay: 0.422329

Year: 0.151959

Id: -0.092144

**Conclusion:**

- The dataset provides valuable insights into salary distributions, job titles with the highest average salaries, and correlations between salary components.

- Cleaning steps were effective in handling missing values.

- Visualizations enhance the understanding of salary distribution and departmental composition.

- Further analysis and exploration could yield deeper insights into specific job roles and their impact on total pay and benefits.

**Name:** Ahmed Talat Abd El-Mohsen