Summary Report: Exploratory Data Analysis of Salaries Dataset

**Basic Data Exploration:**

The dataset contains information about employees' salaries, job titles, and other details.

It consists of a total of 148654 rows and 13 columns.

Data types range from integers, floats and strings, indicating a mix of numerical and categorical features.

There are missing values in the dataset; in OvertimePay 4 cells, OtherPay 4cells, for “Not provided data” for Name and Job, while Benefits 36163 cells, Notes 148654 cells , and in Status 148654 cells are empty.

**Descriptive Statistics:**

The mean salary ('TotalPay') is $74768.3, and the standard deviation is $50517.0 is very high.

The salary range is from 567595.43 to -618.13

**Clean Data**

* Rows with “Not provided data” for all columns were deleted ,
* 0 and negative values for Total\_Pay is not logic, therefore Rows deleted when 'BasePay' <=0 and 'OvertimePay' <=0 and 'OtherPay'<=0 and 'Benefits'<=0 and 'TotalPay'<=0 and 'TotalPayBenefits'<=0 as the
* A duplicate noticed in EmployeeName.

**Data Visualization:**

1. The distribution of 'TotalPay' is slightly skewed to the lift, with a majority of employees earning less than 150,000 $.
2. The pie chart shows the proportion of employees in the top 10 job titles, with the highest proportion in the 'Transit Operator' category.
3. The top 10 job titles with the highest average salary are 'JobTitle1', 'JobTitle2', ..., 'JobTitle10'.
4. There is a positive correlation between 'TotalPay' and 'OvertimePay,' suggesting that employees who receive higher overtime pay also tend to have higher total pay.
5. Average Salary Trend Over the Years:

* The line plot indicates how the average salary has increased over the years from 2011 to 2013 and suddenly decreased in in 2014.
* Understanding these trends can be crucial for workforce planning, budgeting, and identifying potential areas for salary adjustments.