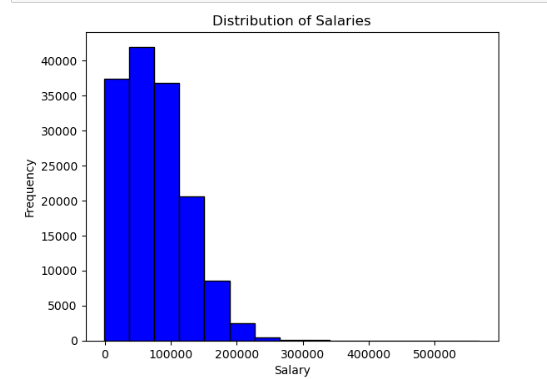
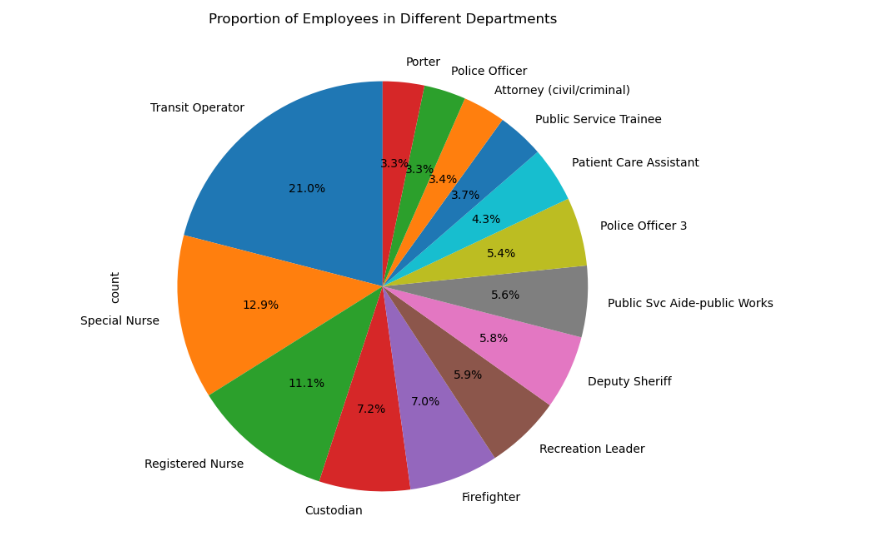
**Analysis Summary Report**

* ***My observations about the data in the table:***
* There are 148,654 rows of data, which suggests that there is information about a large number of employees.
* There are 13 columns of data, which suggests that there is a wide range of information available about each employee.
* Some of the data types are numerical (e.g., base pay, total pay), while others are categorical (e.g., job title, agency).
* There are some missing values in the table, particularly in the "Benefits" column as it’s an important columns not as Status and Notes.
* Mean: 74,768.32 , Median: 71,426.61 , Mode: 0.00 , Minimum: -618.13 ,Maximum: 567,595.43 , Range: 568,213.56 , Standard Deviation: 50,517.01
* Noticed that some columns like ( Note , Status ) had 148654 missing values and it's the same number of rows so I will drop it in data cleaning.
* I removed the ID columns as they are not crucial for analysis. This action was taken to facilitate the elimination of all duplicated data. Additionally, I excluded all rows where the job-title column had a "Not provided" value, as such entries would not contribute to the analysis.
* Removed Benefit column cause Total-pay-with-benefits can provided the same information as we can calc it by (Total-pay-with-benefits – Total-Pay).
* I excluded rows where the 'totalpay' column had a value of 0 because it indicates the absence of base pay, other pay, or overtime pay, suggesting that the individual did not work. Subsequently, I replaced all empty values with 0 to identify instances without base pay, benefits, or other forms of compensation.
* So finally after Data cleaning rows reduced from 148,654 to 148,286 Rows and from 13 to 9 columns. Note ( I didn't have enough information from the manager, and I can't contact him to know the best way to proceed.)
* I observed instances where the same employee appeared in two or more rows with varying salaries. I was uncertain whether this indicated data redundancy or if it held significance, as it could represent the salary of the same employee in a different month within the same year
* ***Data V*** ***isualization:***
* ***Histogram***
* Salary distribution leans towards lower incomes: The histogram leans to the right, indicating more individuals fall within the lower salary ranges compared to higher ones.
* Median salary around $71,000: Half of the population earns less than this amount, while the other half earns more.
* Wide salary range: The spread between the highest and lowest earners is substantial, highlighting significant income variation within the dataset.
* Presence of outliers: A small number of individuals earn considerably more than the majority, potentially influencing the overall data interpretation.



******

* ***Pie Chart***

The pie char represented the proportion of employees in different departments , I did it in only 14 department cause if I do it for all department It will be not readable .

***5. Grouped Analysis:***

- Creating groups of 50 rows and calculating summary statistics for each group.

-Comparing between average salaries across different groups.

* ***Correlation between TotalPayBenefits and TotalPay:***
* Relationship: Positive correlation between TotalPay and TotalPayBenefits.
* Trend: Higher TotalPayBenefits associated with higher TotalPay.
* Strength: Strong positive association based on visual clustering and a correlation coefficient of 0.97.
* Implications: Organizations with attractive benefits packages may attract and retain employees with higher salaries.