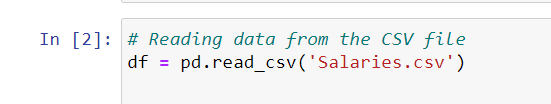
**Summary**

**YAZAN.Y**



Reading data from the CSV file

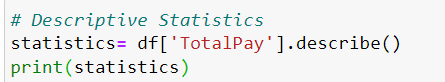


**Data Exploration:**

* Checked the shape of the dataset: The dataset has [number of rows, number of columns] as its shape.
* Checked the data types of each column.

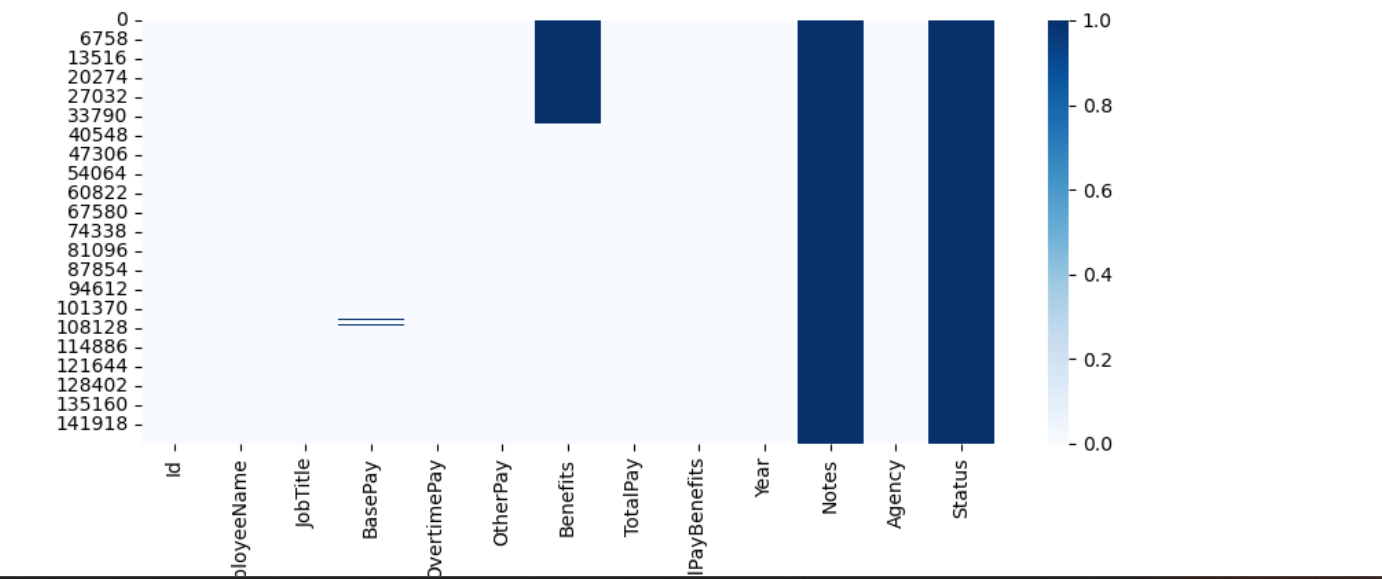
Decision Node

Decision Node



**Statistics:**

Calculated and printed basic statistics for the "TotalPay" column, including mean, median, mode, minimum, maximum, and standard deviation.



**missing values**.

check for missing values in each column



as we see we have missing data

Handle missing data by suitable method with explain why you use it

There are several ways to handle missing data. One of the most common methods is to drop

the column that contains the missing data. However,

this is not a good decision in some cases, as it may lead to the loss of important information or to bias in the results.

Another way to handle missing data is to calculate the mean of the data that does not contain missing values.

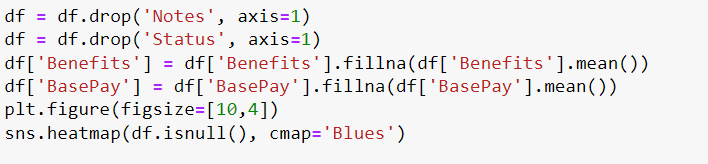
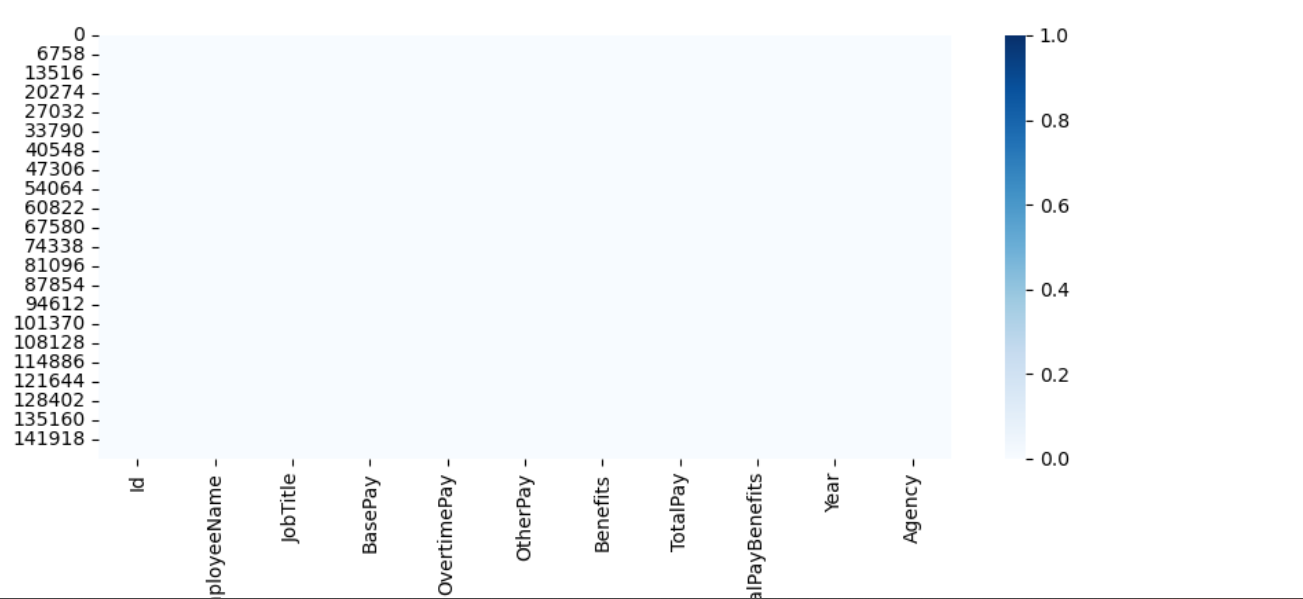
Then, this mean is used to replace the missing values.

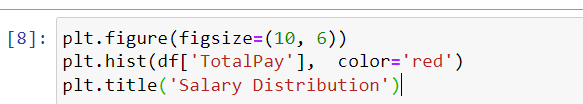
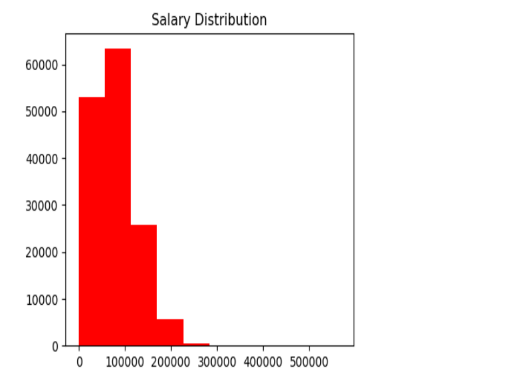
The choice of the right method to handle missing data depends on several factors,

including the percentage of missing data, the type of data, and the nature of the task being performed.'''

**Data Cleaning:**

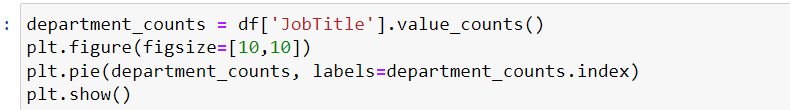
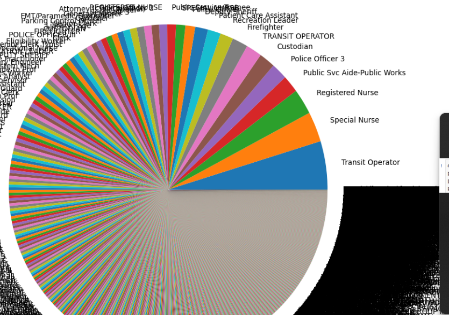
* Dropped the "Notes" and "Status" columns from the dataset.
* Handled missing data in the "Benefits" and "BasePay" columns:

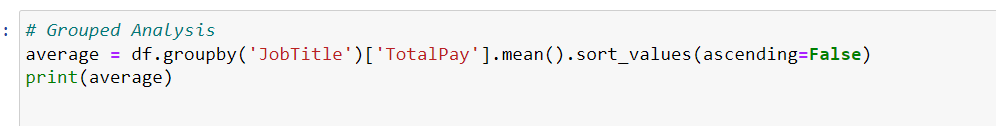




**Data Visualization:**

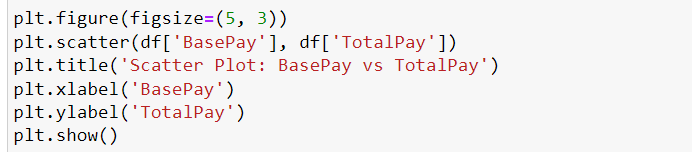
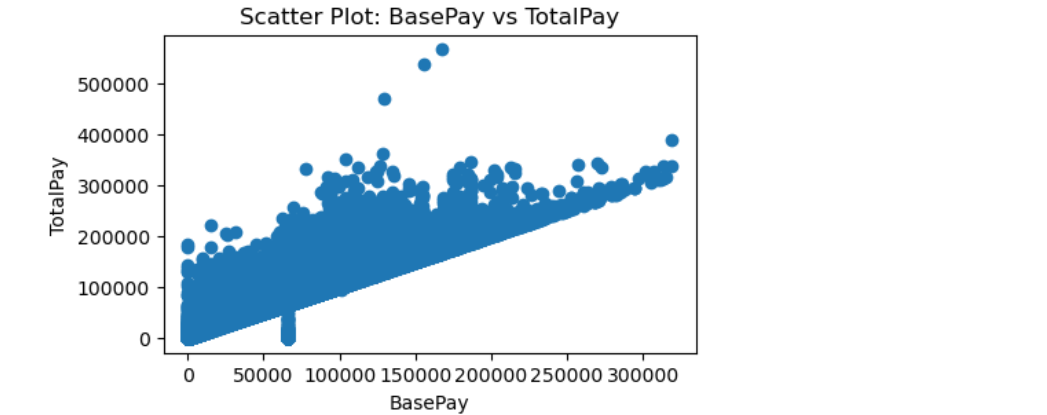
* Created a histogram to visualize the distribution of salaries in the "TotalPay" column.
* Created a pie chart to represent the proportion of employees in different departments based on the "JobTitle" column.





**Grouped Analysis:**

Grouped the data by "JobTitle" and calculated the average salary for each group.



**Correlation Analysis**

* Calculated the correlation between the "BasePay" and "TotalPay" columns.
* Plotted a scatter plot to visualize the relationship between "BasePay" and "TotalPay".