ShAI\_BootCamp\_Assignment

# Tasks

The assignment requires do the following tasks on a data set

1. **Basic Data Exploration**: Identify the number of rows and columns in the dataset, determine the data types of each column, and check for missing values in each column.
2. **Descriptive Statistics**: Calculate basic statistics mean, median, mode, minimum, and maximum salary, determine the range of salaries, and find the standard deviation.
3. **Data Cleaning**: Handle missing data by suitable method with explain why you use it.
4. **Basic Data Visualization**: Create histograms or bar charts to visualize the distribution of salaries and use pie charts to represent the proportion of employees in different departments.
5. **Grouped Analysis**: Group the data by one or more columns and calculate summary statistics for each group and compare the average salaries across different groups.
6. **Simple Correlation Analysis**: Identify any correlation between salary and another numerical column and plot a scatter plot to visualize the relationship.
7. **Summary of Insights**: Write a brief report summarizing the findings and insights from the analyses.

# scope

1. Loading the libraries and then loading and reading the salaries dataset
2. Identify the number of rows and columns in the dataset, explore the data types of each column, and check for missing values in each column.
3. **Descriptive Statistics**: Calculate basic statistics mean, median, mode, minimum, and maximum salary, determine the range of salaries, and find the standard deviation.
4. **Data Cleaning**: Handle missing data by suitable method with explain why you use it.
5. **Visualize the data:**
6. **Grouping**
7. **Simple Correlation Analysis**

# Dataset

salaries dataset generally provides information about the employees of an organization in relation to their compensation. It typically includes details such as how much each employee is paid (their salary), their job titles, the departments they work in, and possibly additional information like their level of experience, education, and employment history within the organization.

# Features

* 'Id'
* 'EmployeeName'
* 'JobTitle'
* 'BasePay'
* 'OvertimePay'
* 'OtherPay'
* 'Benefits'
* 'TotalPay' -> salary
* 'TotalPayBenefits'
* 'Year'
* 'Notes'
* 'Agency'
* 'Status'

Graphs after visualizing the data A graph with numbers and a bar

Description automatically generated with medium confidence

A graph showing a salary vs total

Description automatically generated with medium confidence

**Conclusion**

This is a brief report about the task and if you have any questions or feedback email me on   
[mohammedokleh2002@gmail.com](mailto:mohammedokleh2002@gmail.com)

And consider that this is my first report, and I will be happy to join the training.