

PYTHON FINAL PROJECT

**Data Analysis and Visualization report**



Project by:

kenzy younis

kamal rabie

gana ayman

mariam mansour

***Report***

This project processes the data, performs analysis, and visualizes the insights of the dataset using Python. The dataset used in this project is 'Womens\_Reviews.csv'.

- To start the project, we Import the necessary libraries such as Pandas, Seaborn, NumPy, and Matplotlib.

-The dataset is loaded using the **DataProcessor** class.

* The processing steps involve handling missing values and remove rows with missing values from the dataset.

-The **DataAnalyzer** class provides 4 methods to assess statistics from the dataset.

* Average: The average value of a numeric column was calculated using the **get\_average\_of\_column** method.
* Distribution: The standard deviation of a numeric column was obtained using the **get\_distribution\_of\_column** method.
* Median: The median value of a numeric column was computed using the **get\_median\_of\_column** method.
* Mode: The mode value of a numeric column was determined using the **get\_mode\_of\_column** method.

-The **DataVisualizer** class has 4 visualization methods to represent the data.

* **Line Chart**: A line chart was plotted to visualize the relationship between 'Rating' and 'Age'. This plot helps identify any trends or patterns.
* **Bar Plot**: A bar plot was used to display the distribution of 'Rating' across different age groups. This plot provides an overview of how ratings are distributed among different age categories.
* **Pie Chart**: A pie chart was utilized to represent the distribution of 'Rating' categories. This visualization provides a clear understanding of the proportion of each rating category in the dataset.
* **Scatter Plot**: A scatter plot was generated to illustrate the relationship between 'Rating' and 'Age'. This plot helps identify any correlations between these two variables.

- And finally, in cell 6 the data processing, analysis, and visualization steps are executed in the following order:

1. An instance of the **DataProcessor** class is created as **data\_processor**.
2. The data is loaded using the **load\_data** method of **data\_processor** and assigned to the variable **data**.
3. The loaded data is cleaned using the **clean\_data** method of **data\_processor**, and the cleaned data is assigned to the variable **cleaned\_data**.
4. An instance of the **DataAnalyzer** class is created as **data\_analyzer**, taking the **cleaned\_data** as input.
5. An instance of the **DataVisualizer** class is created as **data\_visualizer**, also taking the **cleaned\_data** as input.
6. The **DataAnalyzer** methods are called to compute the average, distribution, median, and mode of the 'Age' column. The results are stored in the variables **average\_ratings**, **rating\_distribution**, **recommendation\_distribution**, and **feedback\_count\_distribution**, respectively.
7. The **DataVisualizer** methods are called to generate line, bar, pie, and scatter plots, showcasing different aspects of the data.
8. The average, distribution, median, and mode values are printed to the console.