ICP-3 S.Ramyasri

7007727100

GitHub Link: <https://github.com/Ramyasri0123/ICP3>

Video Link[: https://drive.google.com/file/d/1gGPksx0nt7FWzd4RJHGe8SXd4ssL\_2v6/view?usp=sharing](https://d.docs.live.net/602b75a2e2a56016/Documents/:%20%20%20%20https:/drive.google.com/file/d/1gGPksx0nt7FWzd4RJHGe8SXd4ssL_2v6/view?usp=sharing%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20)

Q1

CODEA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

**EXPLATION**

**Loading Data:**

* **Action**: Loads a CSV file (data.csv) into a **Pandas DataFrame** (df).
* **Purpose**: To read and process the data stored in the CSV file.

**Basic Statistical Summary:**

* **Action**: Displays a summary of the dataset, including basic statistics like count, mean, standard deviation, min, max, and quartiles for numerical columns.
* **Purpose**: To get a quick overview of the data's distribution and range.

**Filling Missing Values:**

* **Action**: Fills the missing values in the DataFrame with the **mean** of their respective columns.
* **Purpose**: To handle missing data by replacing it with an estimate. This ensures the analysis doesn't get disrupted by missing data.

**Creating a Scatter Plot (Duration vs. Calories):**

* Generates a scatter plot comparing the Duration and Calories Burned from the data.
* To visually inspect the relationship between exercise duration and calories burned. The plot helps to see if there's any correlation between the two variables.

**OUTPUT**

A screen shot of a graph

Description automatically generated

Q-2

CODE

A screenshot of a computer program

Description automatically generated

A computer code on a black background

Description automatically generated

**Explanation**

1. **Data Loading**: The dataset is loaded from a CSV file containing information about years of experience and salary.
2. **Data Preparation**: The dataset is split into features (years of experience) and the target variable (salary), and further divided into training and testing sets.
3. **Model Training**: A Linear Regression model is created and trained using the training data to learn the relationship between experience and salary.
4. **Prediction**: The trained model is used to predict salary values for the testing set.
5. **Model Evaluation**: The performance of the model is assessed by calculating the Mean Squared Error (MSE), which quantifies the accuracy of the predictions.
6. **Data Visualization**: Scatter plots are created to visually compare actual vs. predicted salary values for both the training and testing datasets, helping to assess model fit.

**OUTPUT**

A screenshot of a computer

Description automatically generated