Documentation

1. Project Title

- Sales Data Regression Analysis

2. Objective

- To analyze the impact of various advertising mediums (TV, Radio, Newspaper) on sales using linear regression. The goal is to predict sales based on advertising expenditures.

3. Dataset Description

- **Source**: sales price prediction_TV.xlsx.
- **Columns**:
    - Sales: Target variable representing sales figures.
    - TV: Expenditure on TV advertisements.
    - Radio: Expenditure on radio advertisements.
    - Newspaper: Expenditure on newspaper advertisements.

4. Methodology

1. **Data Loading**:
    - Imported the dataset using pandas.read_excel().
2. **Data Preprocessing**:
    - Checked for null values using isnull().sum().
    - Reviewed dataset structure with info() and describe().
3. **Exploratory Data Analysis (EDA)**:
    - Visualized relationships between Sales and each feature (TV, Radio, Newspaper) using scatter plots with trendlines.
4. **Model Development**:
    - Split the data into training and testing sets using train_test_split().
    - Built a linear regression model using LinearRegression() from sklearn.

5. Visualizations

- Scatter plots were generated to highlight trends:
    - Sales vs. TV advertisements.
    - Sales vs. Newspaper advertisements.
    - Sales vs. Radio advertisements.
- Trendlines (OLS) were added to show the regression fit.

6. Results

- Linear regression results (coefficients, R^2 score, etc.) will be documented here after model evaluation.

7. Conclusion

- Summarize insights, such as which advertising medium has the strongest impact on sales.
- Discuss potential applications of the model in marketing budget allocation.

8. Future Work

- Explore more complex models (e.g., polynomial regression, decision trees).
- Include more features or external factors (e.g., seasonality, regional data).

9. References

- Any resources used, such as data sources or libraries.