

Advertising Data Analysis Report

Rishabh Verma and Shohom De

31st May

Analysis Report

1. Data Preparation

- **Dataset:** The dataset contains columns for advertising expenditures on TV, Radio, and Newspaper, along with the corresponding Sales.
- **Target Variable:** Sales
- **Predictors:** TV, Radio, and Newspaper advertising expenditures

2. Feature Extraction

Extracted values for TV, Radio, Newspaper, and Sales from the dataset:

```
X = advertise[['TV', 'Radio', 'Newspaper']].values
TV = advertise[['TV']].values
Radio = advertise[['Radio']].values
Newspaper = advertise[['Newspaper']].values
Sales = advertise[['Sales']].values
```

3. Data Splitting

Split the data into training and testing sets for TV, Radio, and Newspaper separately with 80% for training and 20% for testing. Note that after each split, `Sales_train` and `Sales_test` are overwritten:

```
TV_train, TV_test, Sales_train, Sales_test = train_test_split(TV, Sales, test_size=0.2,
                                                                random_state=42)
Radio_train, Radio_test, Sales_train, Sales_test = train_test_split(Radio, Sales,
                                                                      test_size=0.2,
                                                                      random_state=42)
Newspaper_train, Newspaper_test, Sales_train, Sales_test = train_test_split(Newspaper,
                                                                              Sales,
                                                                              test_size=0.2,
                                                                              random_state=42)
```

4. Data Transformation

Applied a square root transformation to the TV training data:

```
TV_train = (TV_train)**0.5
```

5. Standardization

Standardized the Sales and TV data:

```
scaler = StandardScaler()
Sales_train = scaler.fit_transform(Sales_train)
Sales_test = scaler.fit_transform(Sales_test)
TV_train = scaler.fit_transform(TV_train)
TV_test = scaler.fit_transform(TV_test)
```

6. Data Visualization

Created scatter plots to visualize the relationship between Sales and each advertising medium (TV, Radio, Newspaper):

```
plt.scatter(TV, Sales)
plt.xlabel('TV')
plt.ylabel('Sales')
plt.grid(True)
plt.show()
```

```
plt.scatter(Radio, Sales)
plt.xlabel('Radio')
plt.ylabel('Sales')
plt.grid(True)
plt.show()
```

```
plt.scatter(Newspaper, Sales)
plt.xlabel('Newspaper')
plt.ylabel('Sales')
plt.grid(True)
plt.show()
```

7. Linear Regression Model

Built and trained a linear regression model using TV training data and Sales training data:

```
lin_reg = LinearRegression()
lin_reg.fit(TV_train, Sales_train)
```

8. Model Evaluation

Predicted Sales using the test set and calculated the R^2 score:

```
y_pred = lin_reg.predict(TV_test)
R2_score = r2_score(Sales_test, y_pred)
```

9. Results

- **Parameters:**
 - Coefficient (Slope): [0.90779363]
 - Intercept: [-1.51062234e-16]
- **R^2 Score:** 0.9914979860177072

10. Visualization of Predictions

Plotted the predicted values against the actual values for both training and test sets:

```
plt.scatter(TV_test, Sales_test, color='red')
plt.plot(TV_test, y_pred)
plt.show()

y_pred_train = lin_reg.predict(TV_train)
plt.scatter(TV_train, Sales_train, color='red')
plt.plot(TV_train, y_pred_train)
plt.show()
```

Conclusion

The linear regression model trained on the transformed TV advertisement data shows a good fit with an R^2 score of 0.9915. This indicates that approximately 99.15% of the variance in Sales can be explained by the amount spent on TV advertising.

For further analysis, similar steps can be taken for the Radio and Newspaper data to compare their predictive power on Sales.