Predict Total Spends On Advertising

Parth Hareshbhai kacha
Department of Information Technology,
Government Polytechnic Daman

National Institute Of Electronics and Information Technology

Problem Statement

How do investments in television, radio, and newspaper advertising influence sales outcomes? Our objective is to create a model that predicts future sales figures based on the spending patterns across these different advertising channels

Data Validation

Column Name	Dtypes	Missing value	Outliers	Typing errors
Sr. No	Float64	0	No	No
Tv	Float64	0	No	No
Radio	Float64	0	No	No
Newspaper	Float64	0	No	No
Sales	Float64	0	No	No

No Missing values ,outliers and Typing errors are found.

Data Analysis

Hyperlink report

Report Result

```
import pandas as pd
from ydata_profiling import ProfileReport
import warnings
warnings.filterwarnings("ignore")
```

```
report=ProfileReport(data)
report.to_file("C:\int 24\my report")
```

Summarize dataset: 100% 39/39 [00:03<00:00, 12.49it/s, Completed]

Generate report structure: 100% 1/1 [00:01<00:00, 1.93s/it]

Render HTML: 100% 1/1 [00:01<00:00, 1.05s/it]

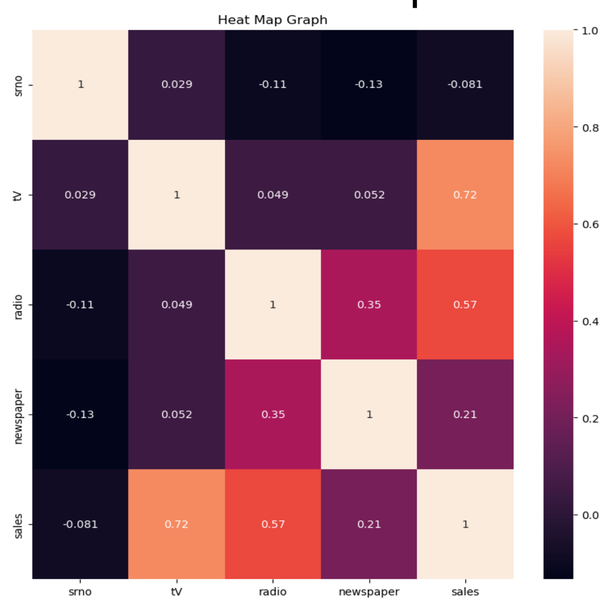
Export report to file: 100% 1/1 [00:00<00:00, 18.55it/s]

Libraries Used

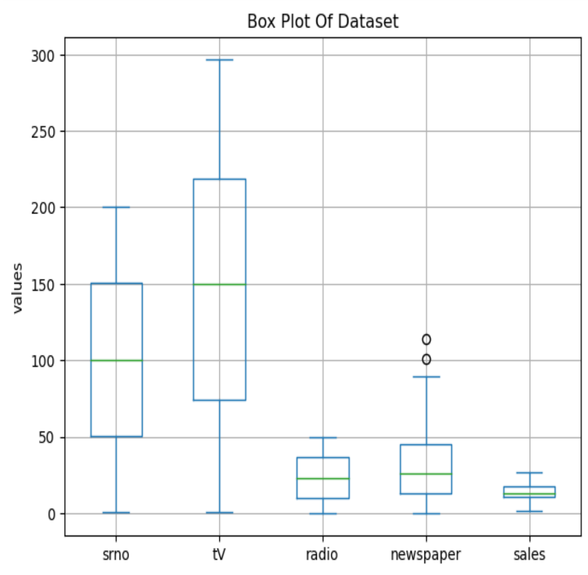
- Pandas
- Numpy
- Sklearn
- Matplotlib
- Seaborn

Data Visualization

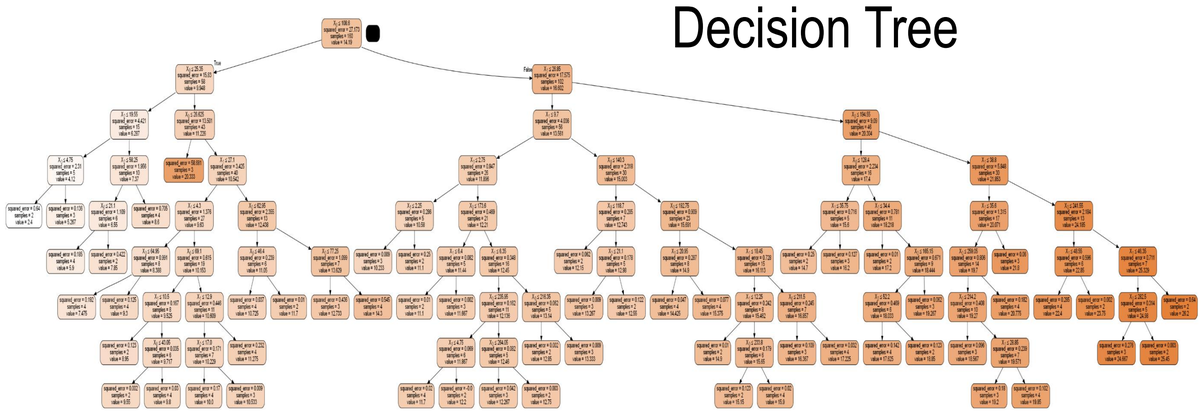
Heatmap



Box plot



Decision Tree



Model Development Equation

- Linear regression: $y = mx + c$ where, c=constant
- Decision tree regressor& Random forest regressor:

→ Entropy: (D)=∑ pi $\log_2$(pi)

→Gini index: $Gini = 1 - \sum_{i=1}^{n} (p_i)^2$

→ Information Gain: Entropy(Parent)−Weighted Average Entropy(Children)

Comparison And Conclusion

```
result=[my_rmse,my_r_square,adj_r2_square_value]
result
```

```
[[1.4983414997969908, 1.1177280403668064, 1.1696855086352065],
 [0.9091048238828131, 0.959460390751816, 0.9495586167430304],
 [0.9015302258730475, 0.956082089981134, 0.945355168138283]]
```

```
a=pd.DataFrame(data=result,columns=["LinearRegression","DecisionTree","RandomForest"],index=["my_rmse","my_r_square","adj_r2
a
```

	LinearRegression	DecisionTree	RandomForest
my_rmse	1.498341	1.117728	1.169686
my_r_square	0.909105	0.959460	0.949559
adj_r2_square_value	0.901530	0.956082	0.945355

The Decision Tree model has the lowest RMSE(1.117728), highest R^2 value (0.959460) and highest Adjusted R^2 value (0.956082)

The **Decision Tree** model appears to be the best-performing model among the three, based on all three evaluation metrics (RMSE, R^2, and Adj. R^2).

Deployment url

Github url: https://github.com/Djdjsjsjndnznsmsmm-Regression

Streamlit url: https:// /Djdjsjsjndnznsmsmm.streamlit.app/