



ADVERTISEMENT AND SALES DATA FOR ANALYSIS

HIBA U.H



Introduction

The goal of this project is to develop predictive models for advertising sales based on different marketing channels such as TV, Radio, and Newspaper.

The project involves data exploration, visualization, and building two regression models: Linear Regression and XGBoost. The predictive performance of these models is evaluated, and the better-performing model is saved for future use.

Objective: "The goal of this project is to analyze advertisement spending across different channels and predict sales performance using machine learning techniques."

Data Overview

The advertising dataset captures the sales revenue generated with respect to advertisement costs across multiple channels like radio, tv, and newspapers. It is required to understand the impact of ad budgets on the overall sales.

The dataset is taken from Kaggle

Methodology

- Linear Regression
- Regression Plots: Simple linear regression plots are created for each marketing channel (TV, Radio, Newspaper) against Sales using Plotly Express.
- Model Building: Linear regression analysis is performed using the statsmodels library. The relationship between TV ad spend and Sales is emphasized, and key statistics are analyzed.
- Model Evaluation: Residual analysis is conducted to validate assumptions, and predictions on the test set are evaluated using metrics such as Mean Squared Error (MSE) and R-squared.

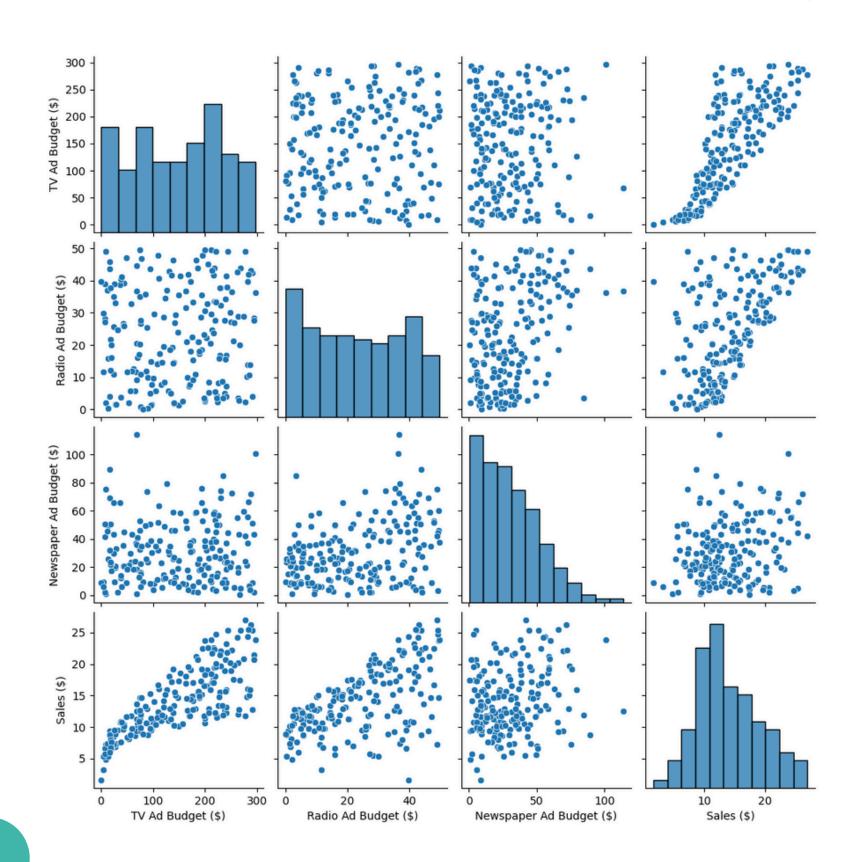


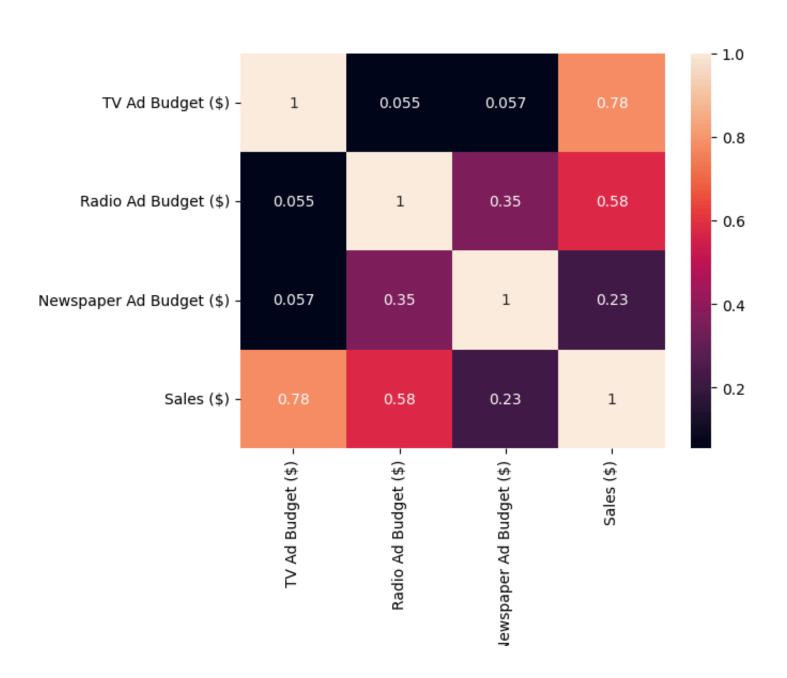
Implementation TOOLS

- Python 3
- Jupyter Notebook
- Libraries: pandas, scikit-learn, seaborn, matplotlib

Results

Data Visualization





Conclusion

The Polynomial Regression model showed improved performance compared to Multiple Linear Regression, with a higher R-squared value indicating better fit to the data.

The project concludes by summarizing the key findings and insights from both regression models.

References

- kaggle
- github
- sklearn,numpy,matplotlib,pandas,seaborn documentation