

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is light green. They are positioned diagonally, with the blue one partially covering the green one.

Machine Learning

The Prediction of Big Mart Sales

Yeganeh Safari

Mahtab Mohammadi



Abstract



Big Mart Sale uses data on goods, customer information, and assets to predict future sales and improve customer service through machine learning.

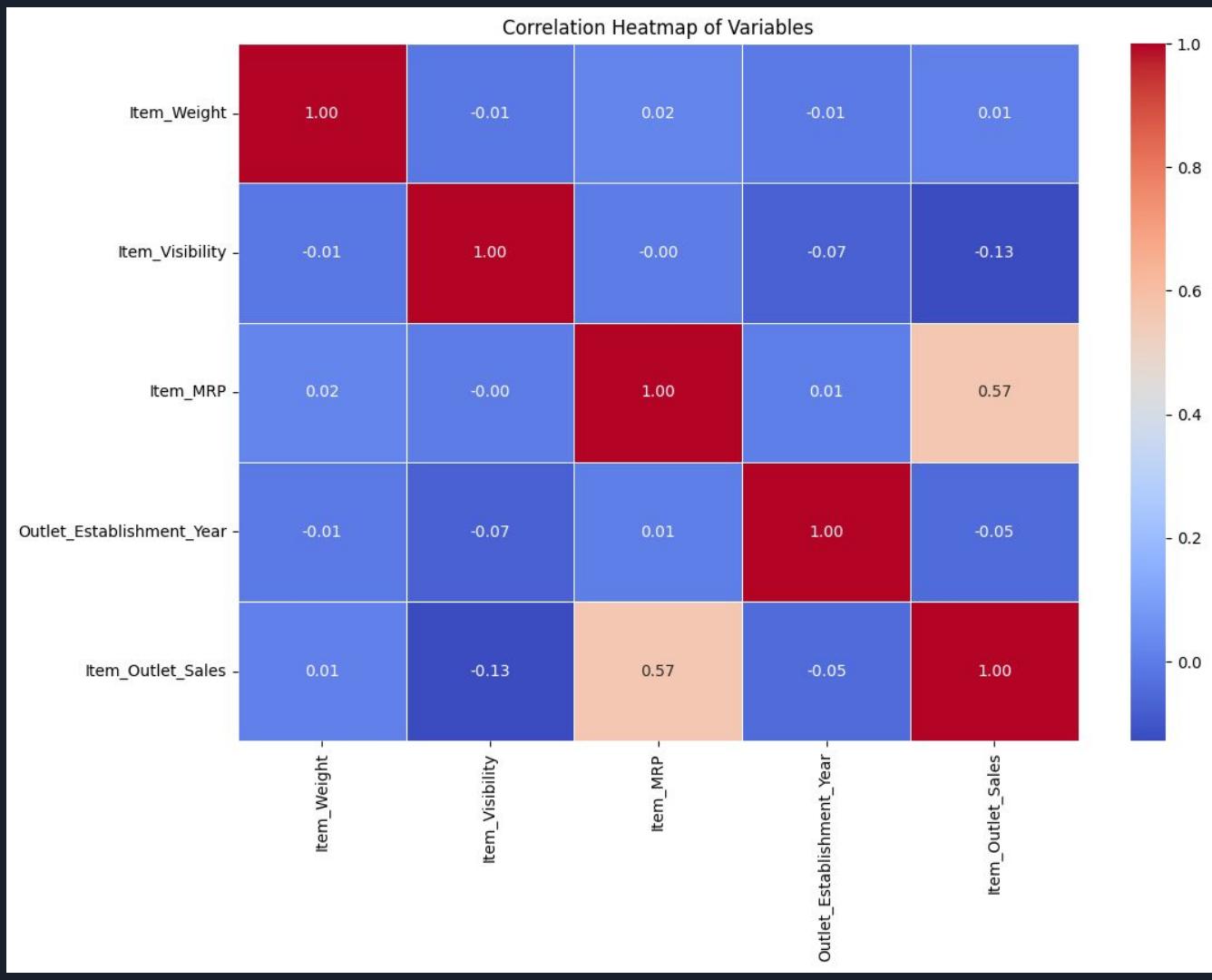
The aim of this project is to build a model that predicts item sales and analyzes what factors impact those sales. We will use the Big Mart dataset, which has many features and a reliable method for prediction.



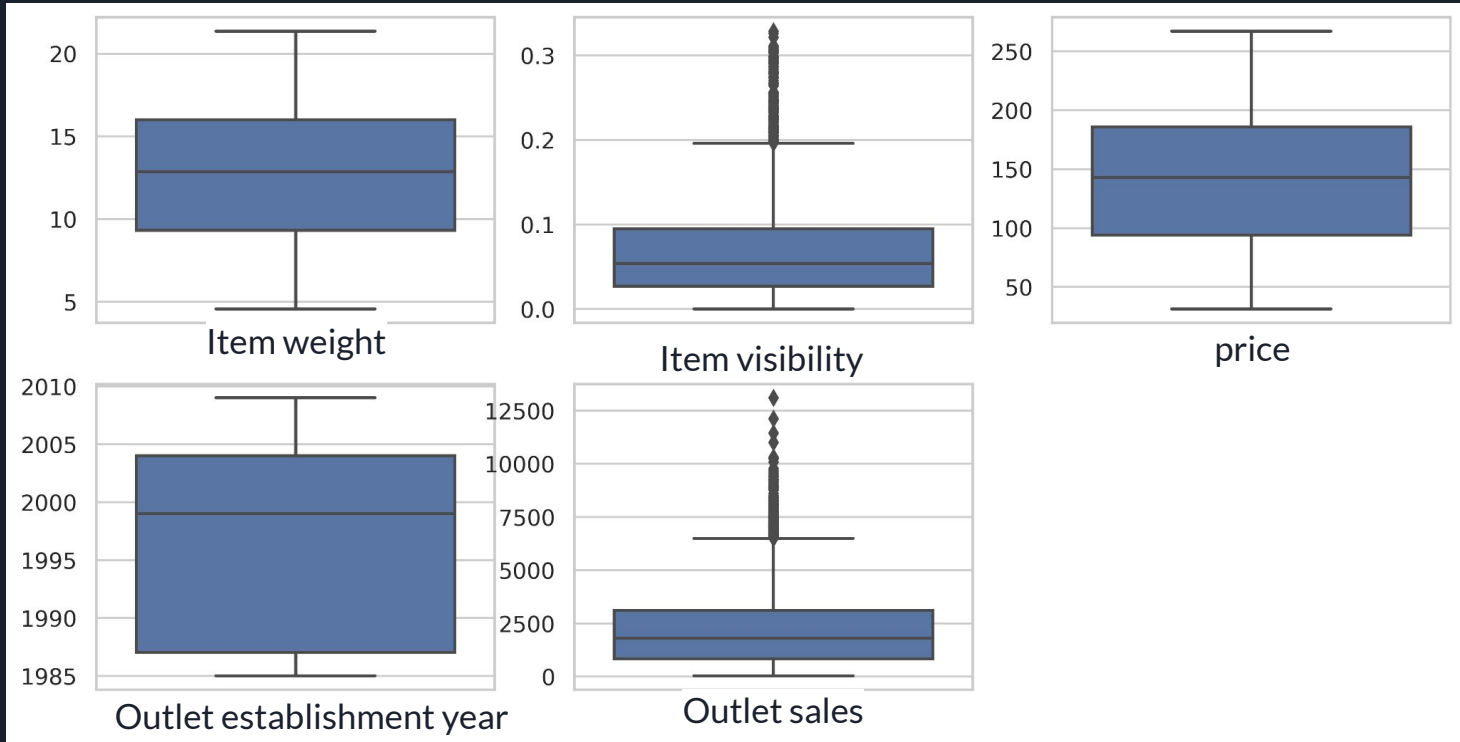
Data Description

The data contains the different items of food sold in other markets. The detailed data dictionary is given below (8 features, including 4 categorical and 4 numerical, and 8523 observations)

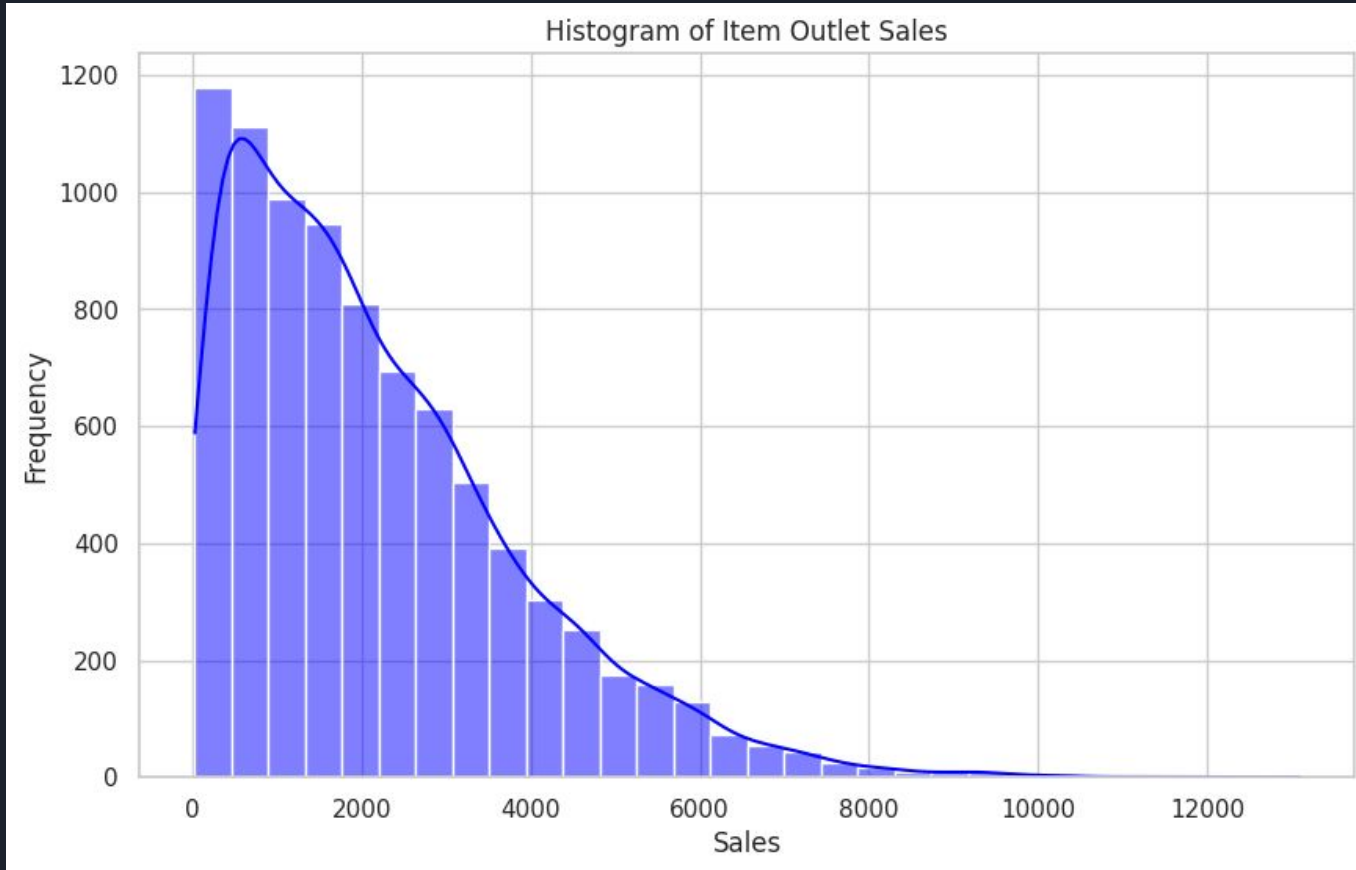
	Item_Weight	Item_Fat_Content	Item_Visibility	Item_MRP	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales
0	9.30	Low Fat	0.016047	249.8092	1999	Medium	Tier 1	Supermarket Type1	3735.1380
1	5.92	Regular	0.019278	48.2692	2009	Medium	Tier 3	Supermarket Type2	443.4228
2	17.50	Low Fat	0.016760	141.6180	1999	Medium	Tier 1	Supermarket Type1	2097.2700
3	19.20	Regular	0.000000	182.0950	1998	NaN	Tier 3	Grocery Store	732.3800
4	8.93	Low Fat	0.000000	53.8614	1987	High	Tier 3	Supermarket Type1	994.7052



Feature Distribution



Distribution of Target Variable





MODEL BUILDING

Training Model

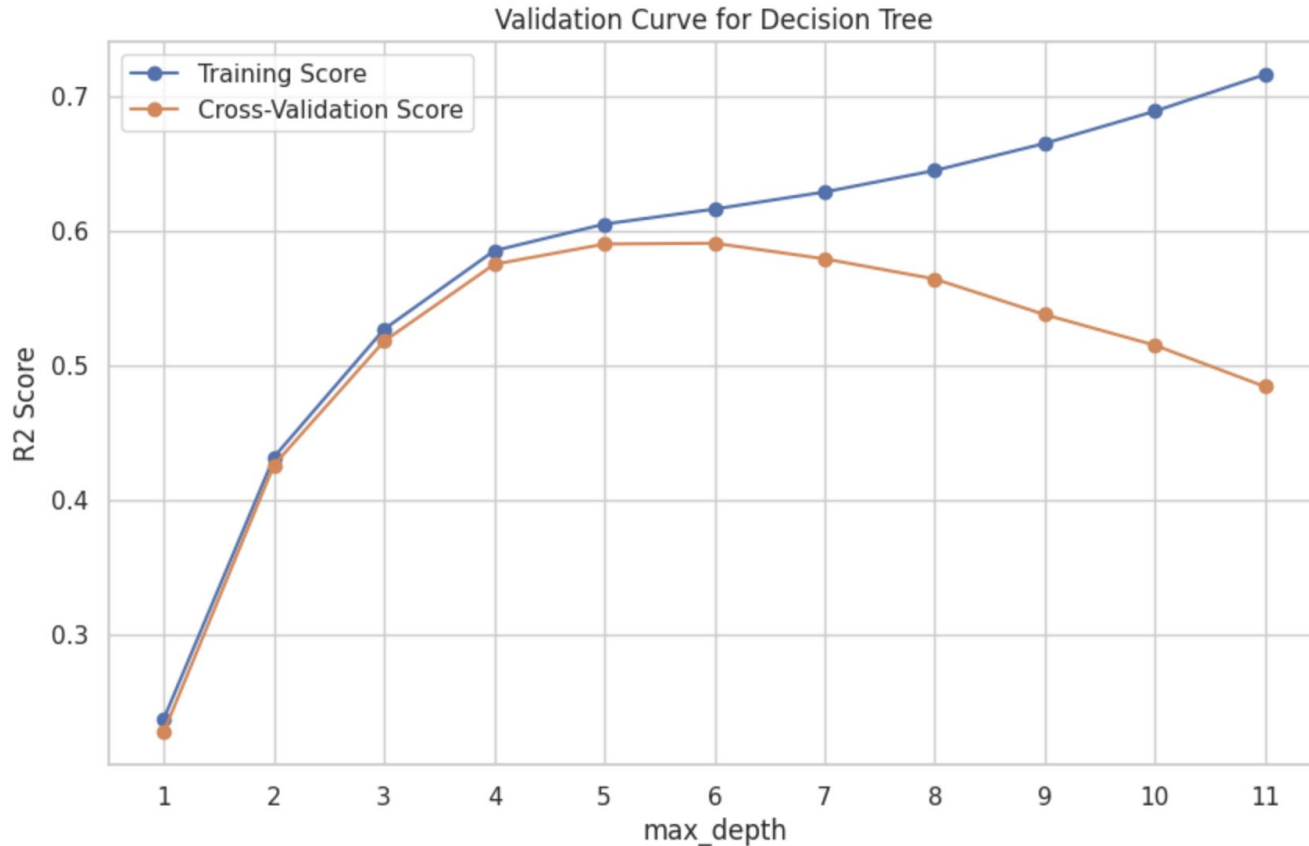
- Linear Regression
- Decision Tree
- Random Forest

➡ 80% Training (6818)

➡ 20% Testing (1705)



Best Parameters from Model Tuning





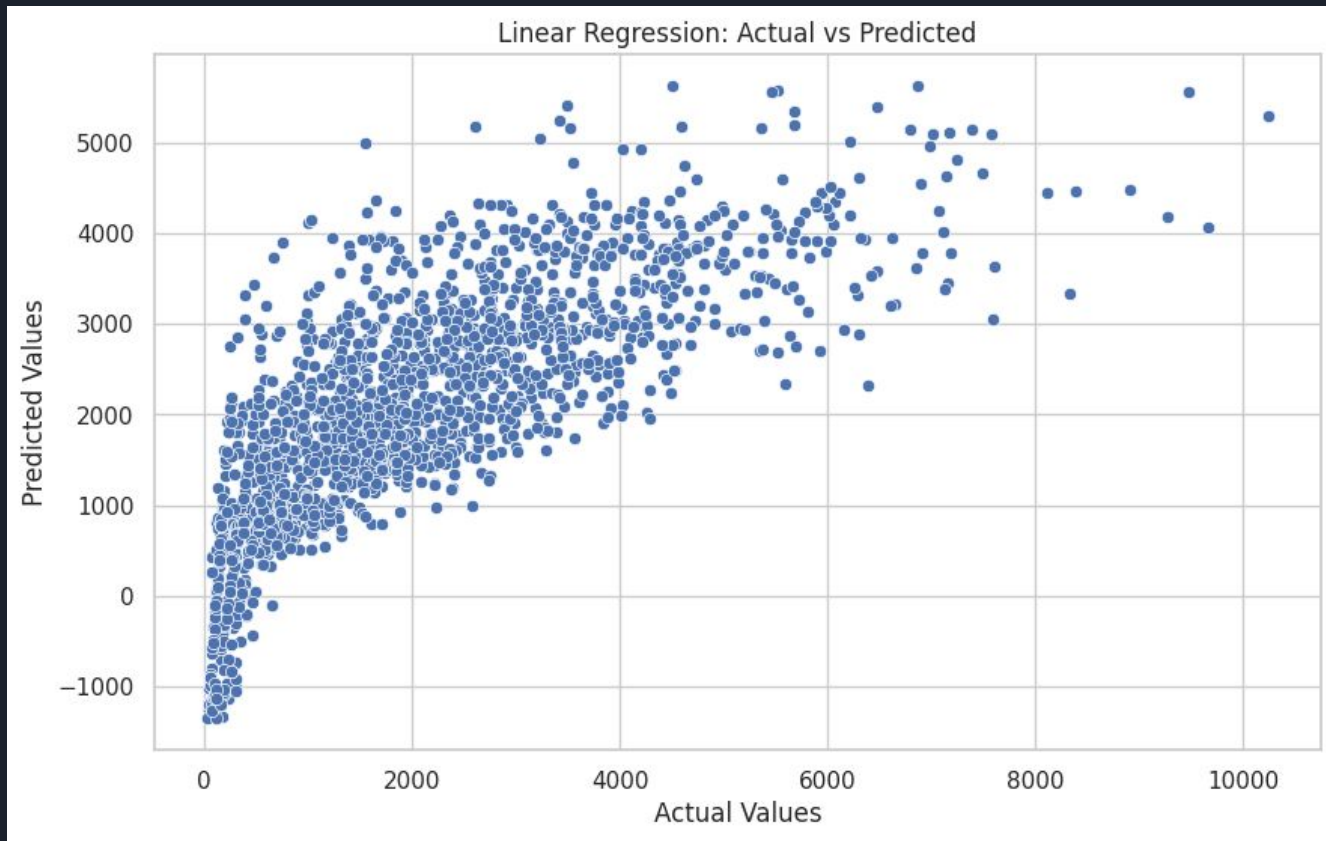
RESULTS

Predicting Regression Model

r2_score: 0.5793792269143474

RMSE: 1069.2215255240556

MAE: 791.8347251203975

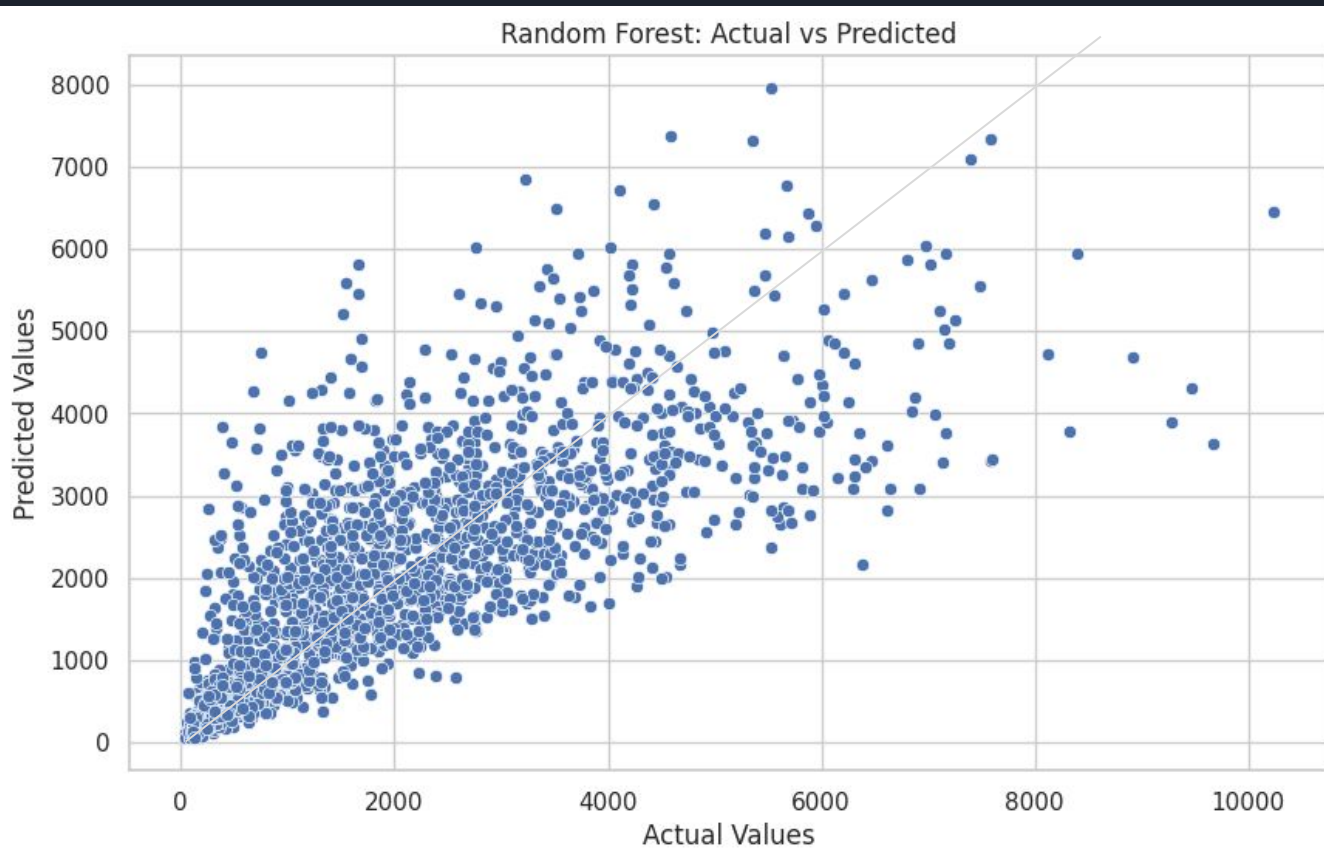


Predicting Random Forest

r2_score: 0.556680754983667

RMSE: 1097.6923238820236

MAE: 764.4891753888563

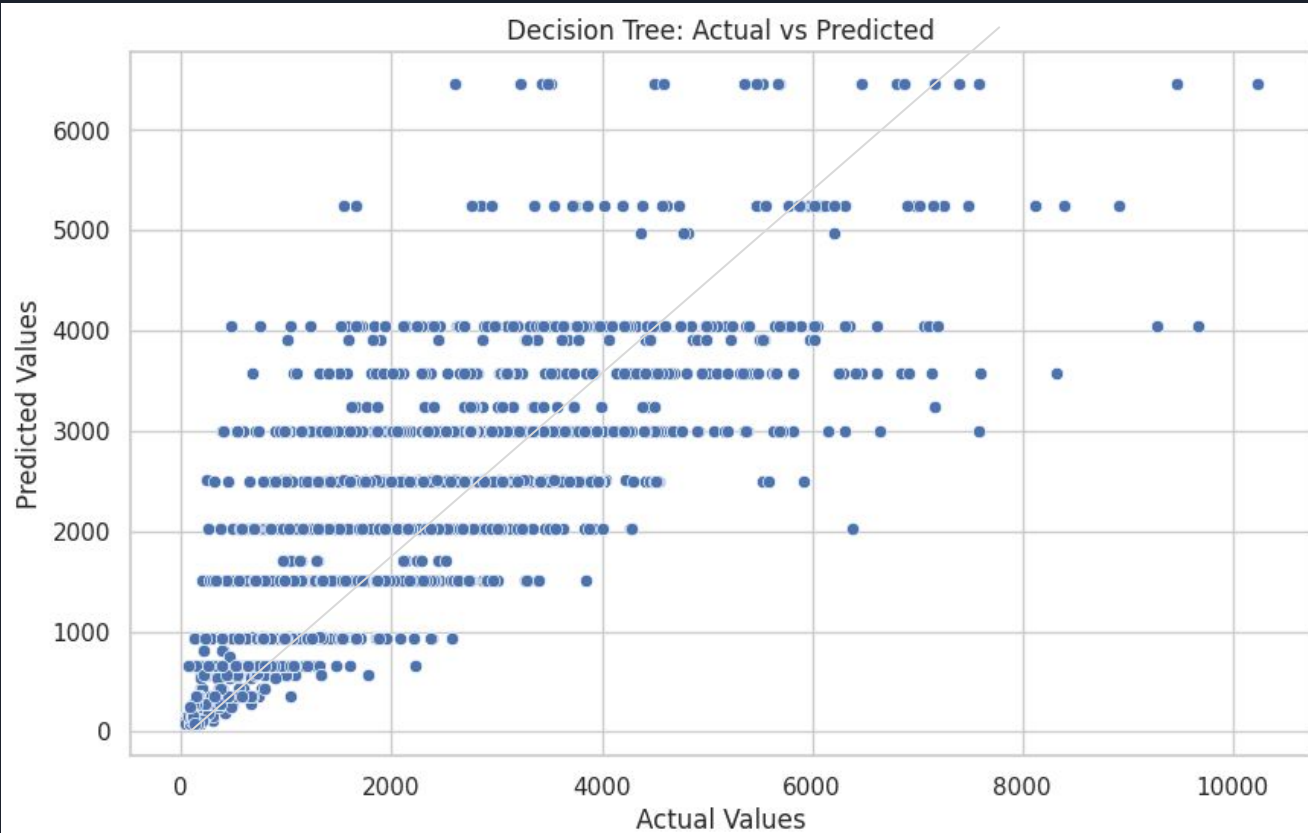


Predicting Decision Tree

r2_score: 0.6113896069457742

RMSE: 1027.7311982700267

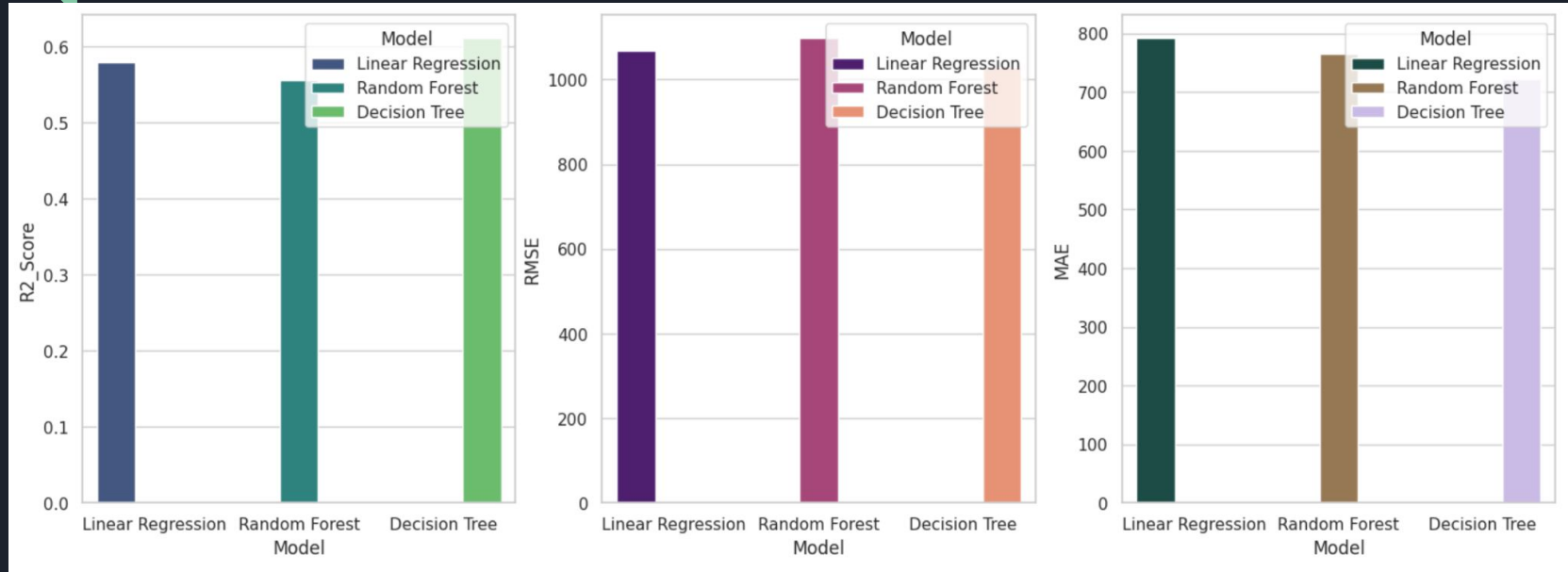
MAE: 721.6447426931057



Comparing the Models

	Model	R2_Score	RMSE	MAE
2	Decision Tree	0.611390	1027.731198	721.644743
0	Linear Regression	0.579379	1069.221526	791.834725
1	Random Forest	0.556681	1097.692324	764.489175

Comparing the Models





Conclusion

Decision Tree model Performed Best after parameter tuning and will be monitored and used for future work.

Future Work

- Explore More Advanced Models such as XGBoost or Neural Networks
- Since sales data is likely time-dependent, perform time series analysis:
- Explore trends, seasonality, and cycles in the data.
- More data such as Previous 5 years or more





Thank You!

