

Laptop Prices analysis by machine learning

Table of Contents

1. **Abstract**
2. **Introduction**
3. **Dataset Description**
4. **Model Performance**
5. **Analysis**
6. **Methodology**
7. **Future Improvements**
8. **Results**
 - a. Descriptive Statistics:
 - b. Model Performance:
9. **Conclusion**
10. **Visualization**
 - a. Correlation Plot (Heatmap)
 - b. Histogram
 - c. Pair plot
 - d. Boxplot
 - e. Graph Plot
11. **References**

1.Abstract

This report explores the use of machine learning to predict laptop prices based on a dataset of hardware and brand attributes. The predictive model's performance was evaluated using key metrics including R-squared, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

2.Introduction

The rapid advancement of technology and the increasing diversity of laptop features have made it challenging for consumers to assess the value of laptops based on their specifications. Understanding the factors that influence laptop prices is crucial for both manufacturers and consumers. Machine learning provides an effective approach to analyze complex relationships between features and prices, enabling accurate predictions and valuable insights.

3.Dataset Description

The dataset consists of the following columns:

1. Company: Brand of the laptop.
2. TypeName: Type/category of the laptop (e.g., Ultrabook, Gaming, Notebook).
3. Ram: Amount of RAM in GB.
4. Weight: Weight of the laptop in kg.
5. Price: Price of the laptop (target variable).
6. Touchscreen: Binary indicator of whether the laptop has a touchscreen.
7. Ips: Binary indicator of whether the screen has In-Plane Switching (IPS) technology.
8. Ppi: Pixel density of the screen (pixels per inch).
9. CPU brand: Brand of the CPU.
10. HDD: Size of the Hard Disk Drive in GB.
11. SSD: Size of the Solid-State Drive in GB.
12. Gpu brand: Brand of the GPU.
13. Os: Operating System of the laptop.

4. Model Performance

The machine learning model was evaluated using the following metrics:

1. **R-squared (R^2): 0.5496**
This value indicates that approximately 54.96% of the variance in laptop prices can be explained by the model.
2. **Mean Squared Error (MSE): 0.1795**
This metric reflects the average squared difference between the predicted and actual prices.
3. **Root Mean Squared Error (RMSE): 0.4237**
The RMSE provides an interpretation of the prediction error in the same units as the target variable (price).

5. Analysis

The model demonstrates moderate performance with an R^2 score of 0.5496. This suggests that while the model captures some relationships between the features and the target variable, there is room for improvement.

The relatively low MSE and RMSE values indicate that the model's predictions are reasonably close to the actual values, but further refinement could enhance its accuracy.

6. Methodologies

- Data Collection and Processing
- Linear Regression
- Regression Analysis

7.Future Improvements

To improve model performance, consider:

1. **Feature Engineering:** Incorporating interaction terms or polynomial features to capture non-linear relationships.
2. **Hyperparameter Tuning:** Optimizing the model's hyperparameters using grid search or random search.
3. **Data Augmentation:** Expanding the dataset with more samples or introducing additional relevant features.
4. **Algorithm Selection:** Experimenting with more complex models like ensemble methods (e.g., Random Forest, Gradient Boosting).

8.Result

Descriptive Statistics:

- **Average Price:** \$10.828218
- **Price Volatility (Standard Deviation):** \$ 0.619565
- **Maximum Price:** \$12.691441
- **Minimum Price:** \$9.134616

Model Performance:

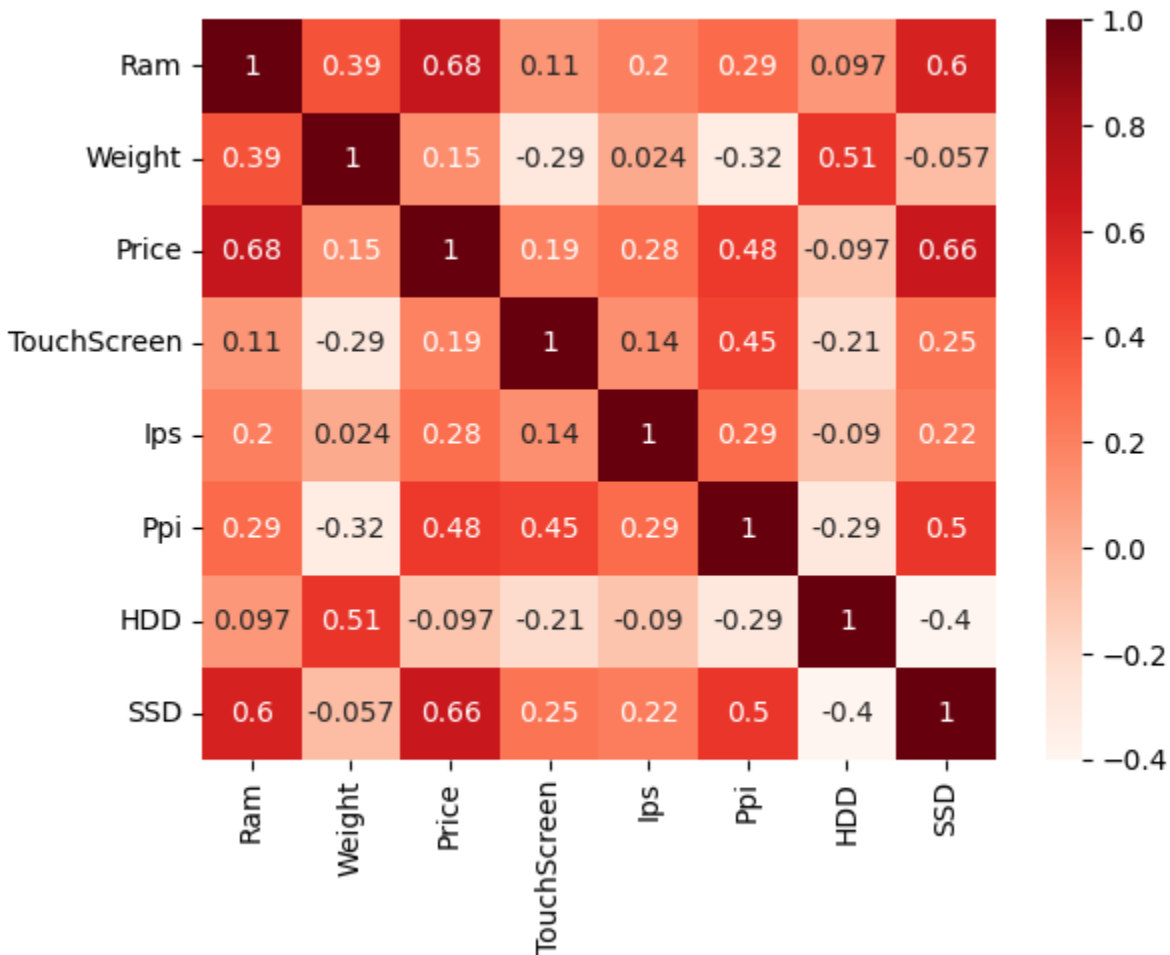
- **R-Squared:** 0.5496
- **Mean Squared Error (MSE):** 0.1795
- **Root Mean Squared Error (RMSE):** 0.4237

9.Conclusion

This preliminary analysis provides a foundation for predicting laptop prices using machine learning. By enhancing feature selection, refining the model, and incorporating additional data, the prediction accuracy can be further improved.

10.Visualization:

Correlation Matrix(Heat Map):

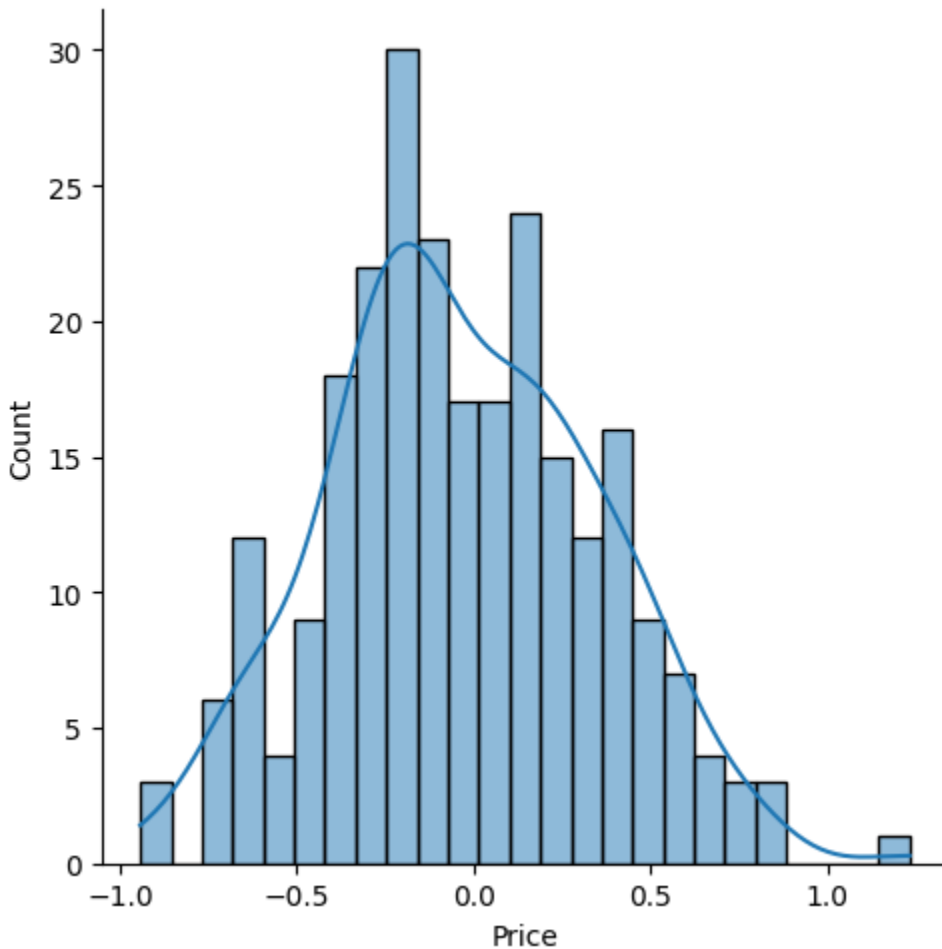


Dark Red shows positive Correlation between variables.

Light Red shows negative Correlation. between variables.

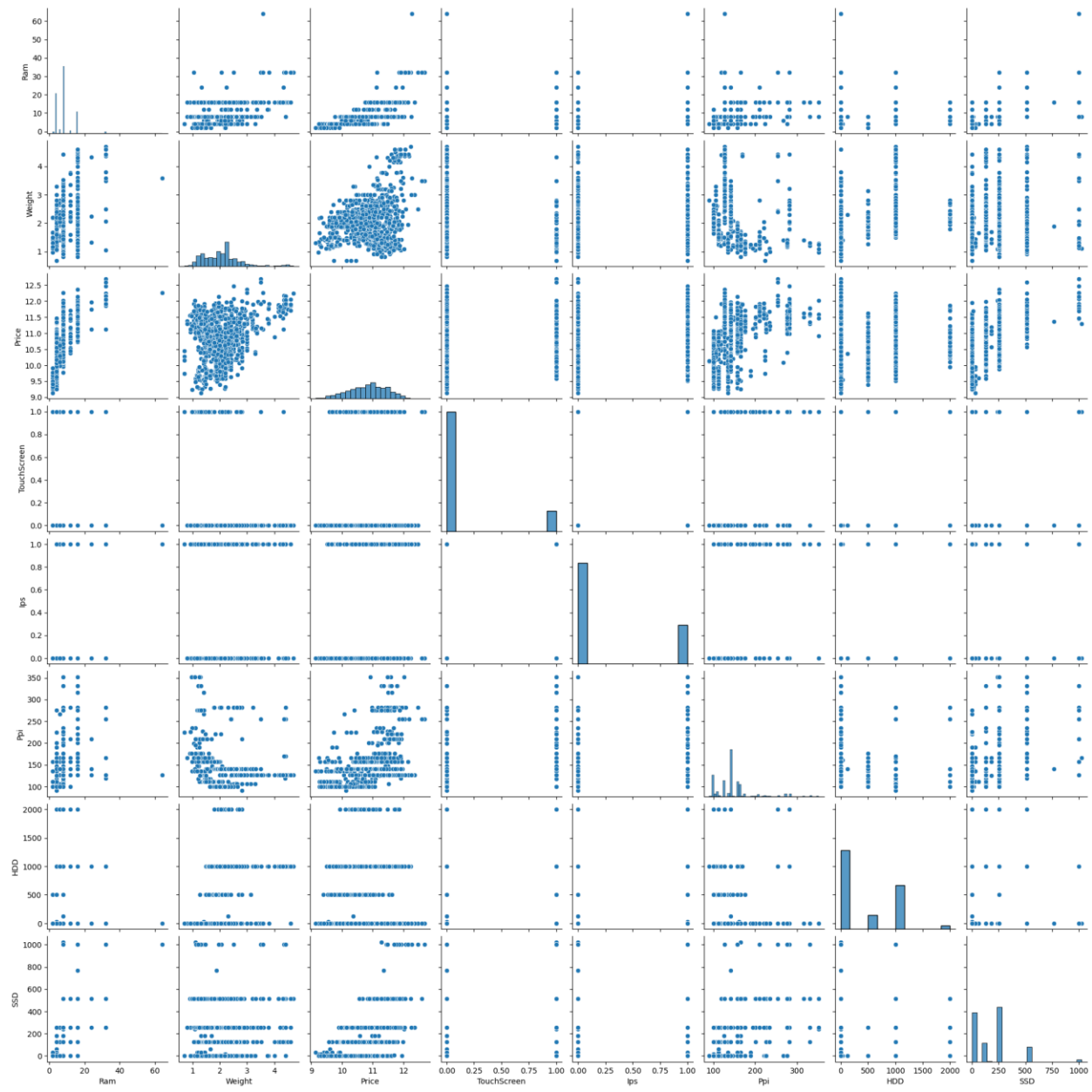
The heatmap reveals correlations between laptop features. Price and SSD capacity have a strong positive correlation, while Weight and HDD capacity show a moderate negative correlation.

Histogram:



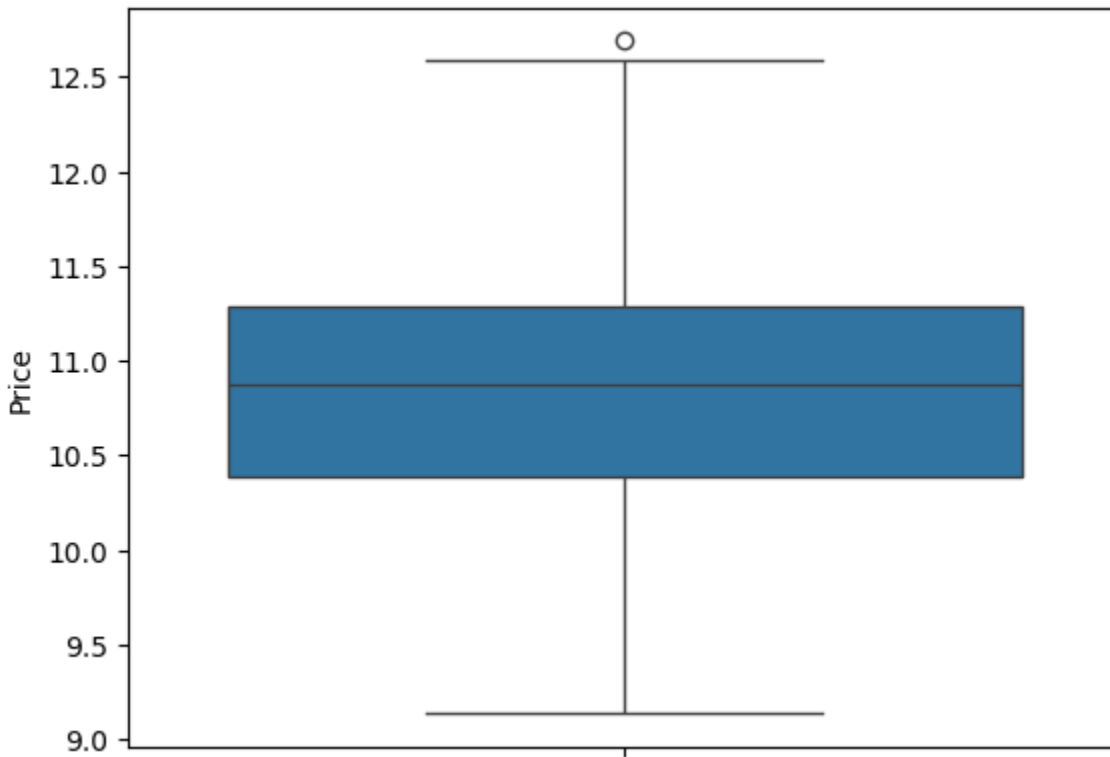
The graph displays the distribution of a variable called "Price." The data is plotted as a histogram, showing the frequency of observations within different price ranges. A smooth curve is overlaid on the histogram, suggesting a normal distribution. The distribution appears to be centered around 0, with most observations falling within the range of -0.5 to 0.5. There are a few outliers on both ends of the distribution, indicating some extreme price values.

Pair Plot:



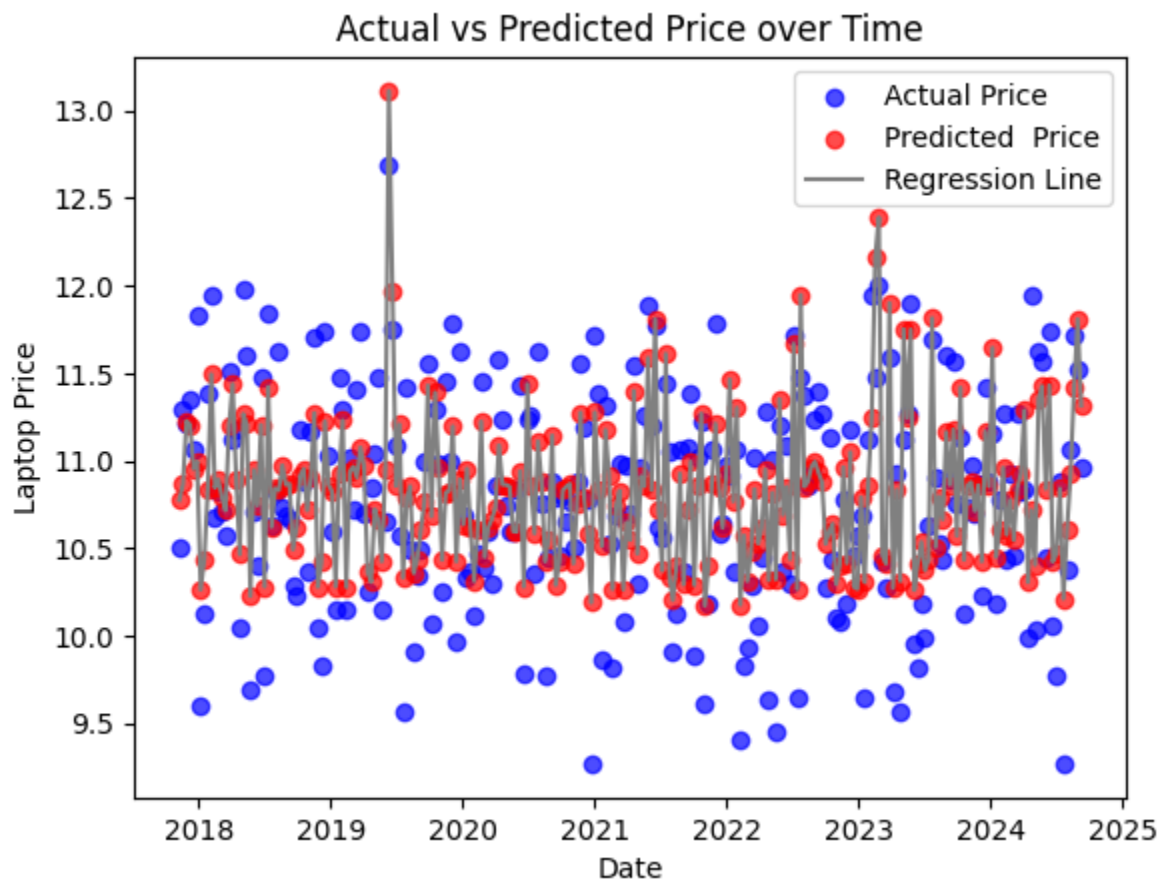
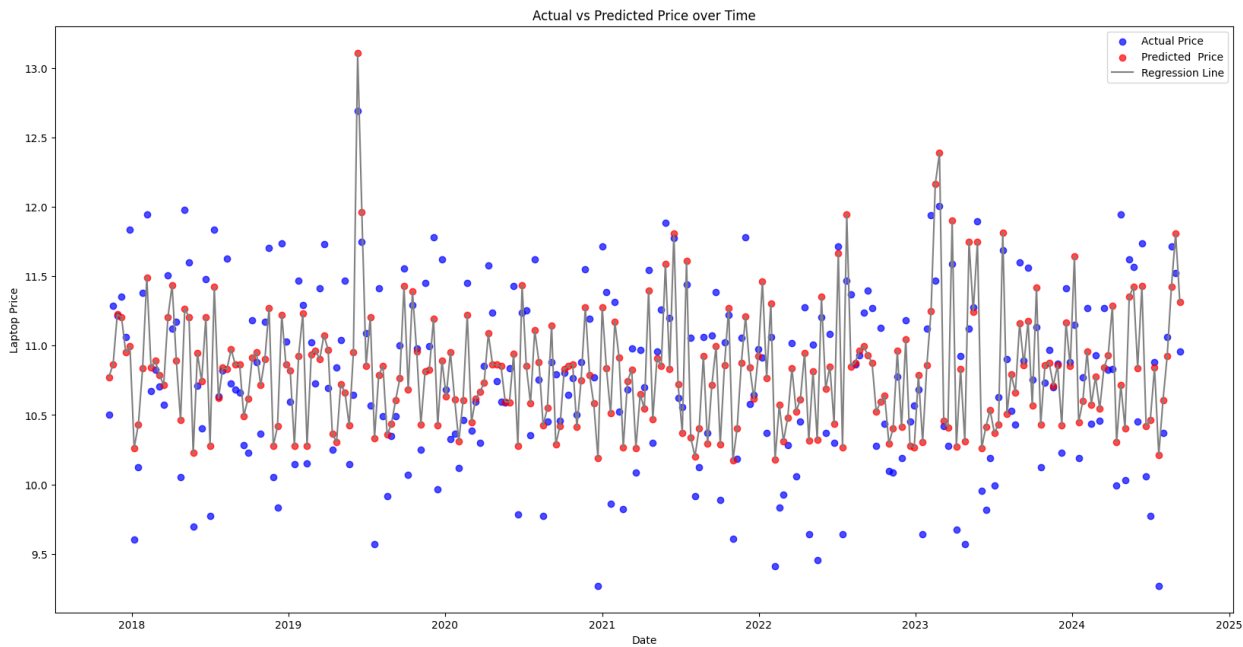
All columns relation with each other (Scatter plot form). This is a pair plot, which shows scatter plots and histograms of multiple variables in a dataset to explore their relationships. Each diagonal element represents the distribution (histogram) of a single variable, while the scatter plots below the diagonal compare two variables. The data contains features RAM, weight, price, and other attributes, with relationships and trends between some variables being more evident.

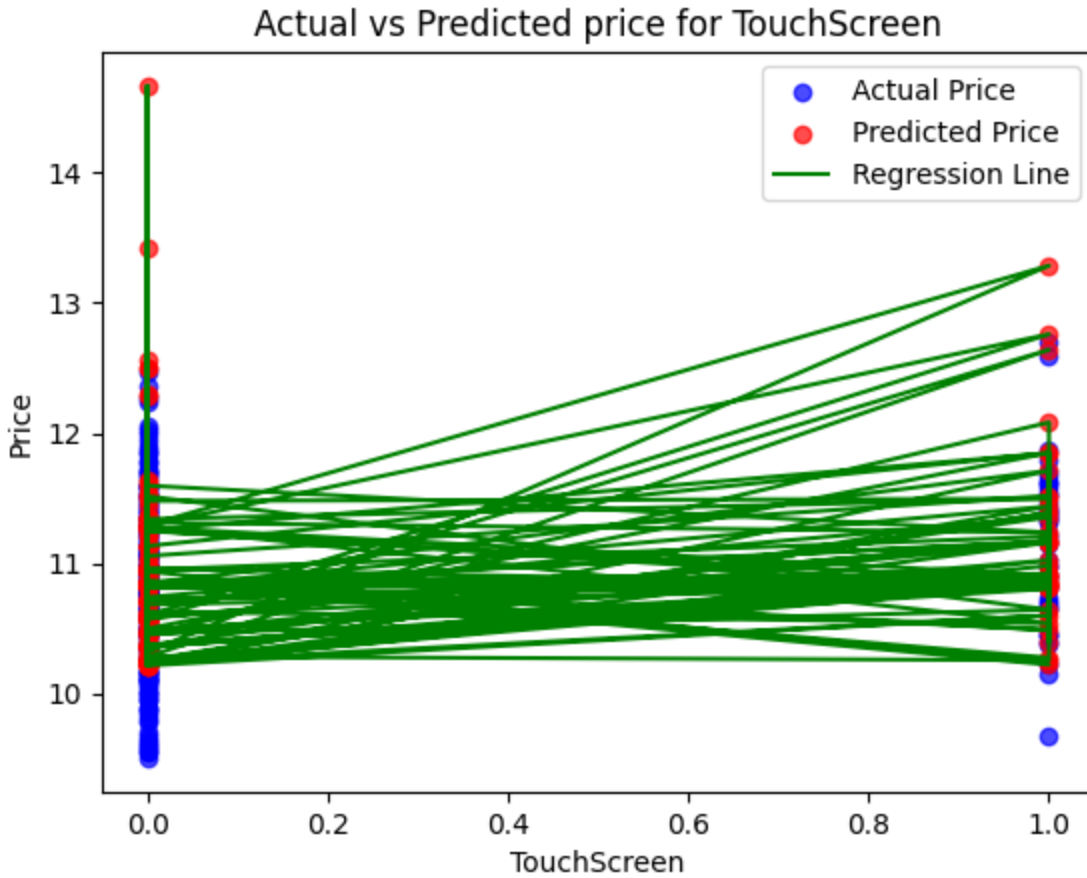
Box Plot:



Price distribution shown in box plot. This box plot showing the distribution of the "Price" variable. The box represents the interquartile range (IQR), where the middle 50% of data lies, with the line inside the box indicating the median. The "whiskers" extend to the smallest and largest data points within 1.5 times the IQR. The point above the upper whisker is an outlier, representing a price significantly higher than the rest of the data.

Graph Plot:





References:

- Zulkifley, N.H., Rahman, S.A., Ubaidullah, N.H. and Ibrahim, I., 2020. House price prediction using a machine learning model: a survey of literature. *International Journal of Modern Education and Computer Science*, 12(6), pp.46-54.
- Mamoudan, M.M., Mohammadnazari, Z., Ostadi, A. and Esfahbodi, A., 2024. Food products pricing theory with application of machine learning and game theory approach. *International Journal of Production Research*, 62(15), pp.5489-5509.