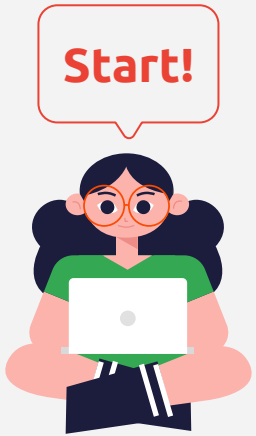
Do you want to know about Laptops

Let’s google it



Laptop Price Prediction by Regression Models

Presented by:

- Annal Ali Albeeshi
- Aesha Bakheet Aljohani

Supervisor the project:

- Mohamed Eldeeb
- kinza waqar

The Search Results Are:

1. Introduction.
2. Methodology.
3. Conclusions and further work.
4. Power Bi.

1
2
3
4

Introduction

- **1. Datasets**
  Our Two Dataset we get it from **Kaggle**.

- **2. Datasets**
  Contains of 1853 Rows & 20 Columns.

- **3. Data Cleaning**
  Dropping columns, Check for Duplicates or unnecessary data, Missing values, Outliers.

kaggle

Methodology

Pre Processing.

- Data Collation.
- Data Target.

Data Description.

- Data cleaning

EDA Analysis.

- Exploring the Data.

4- Algorithms

Methodology … 1- Pre-processing data

1. Pre-processing data

- **Data collation**

2 datasets have been selected from kaggle 1 , The first dataset is and it consists of **1303 rows** and **13 columns.**

	Unnamed: 0	Name	Processor	RAM	Operating System	Storage	Display	Warranty	Price	rating
0	0	Lenovo Ideapad S145 Core i5 10th Gen - (8 GB/1...	Intel Core i5 Processor (10th Gen)	8 GB DDR4 RAM	64 bit Windows 10 Operating System	1 TB HDD	39.62 cm (15.6 inch) Display	1 Year Onsite Warranty	â□'43,990	3.9

The second dataset 2 consists of **550 rows** and **10 columns**.

	laptop_ID	Company	Product	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price_euros
0	1	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	1339.69

- **Data cleaning**
- **Data target**

By creating a new data frame called **final_laptop.csv**, It consists of **1853 rows** and **20 columns**.

- **Data cleaning**

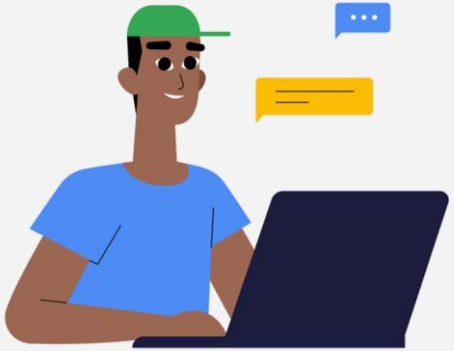
Droping colunams

Methodology … 2- Data Description

Data Description

It consists of 1853 rows and 11 columns.

1- Company- String -Laptop Manufacturer
2- Product -String -Brand and Model
3- TypeName -String -Type (Notebook, Ultrabook, Gaming, etc.)
4- Display Numeric- Screen Size
5- CPU- String -Central Processing Unit (CPU)
6- RAM Numeric - Laptop RAM
7- Memory -String- Hard Disk / SSD Memory
8- Operating System -String- Operating System
9- Weight -String- Laptop Weight
10- Price-USD -Numeric- Price
11- Warranty - String



Methodology …

- **Check for missing data (NAN Value)**

Warranty 1303 missing data, TypeName 550 missing data, Weight 550 missing data, RAM 1 missing data.

- **Check for Duplicates or unnecessary data**

Our dataset contains 36 duplicate rows , and by removing these rows we have 1817 rows, where they were previously 1853 rows.

- **Check for Outliers**

Our datasets do not contain Outliers.

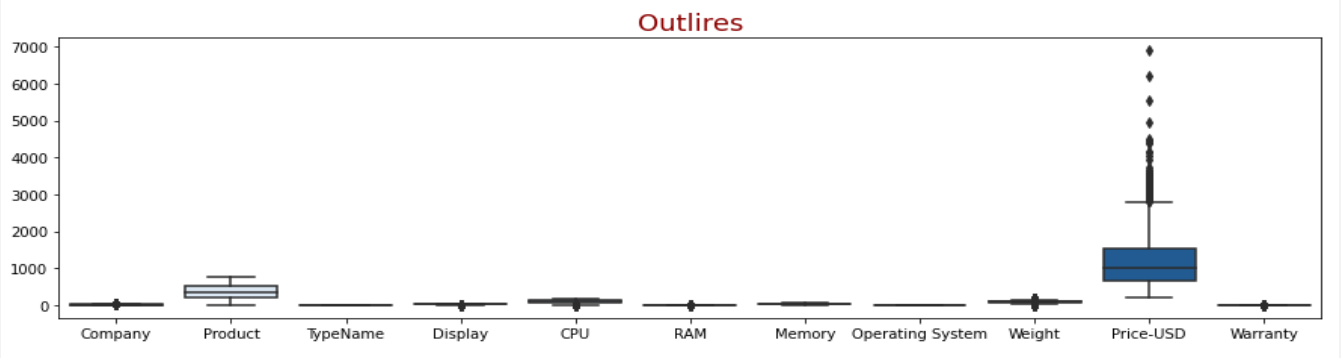


Fig 1: Check for Outliers

3- EDA Analysis

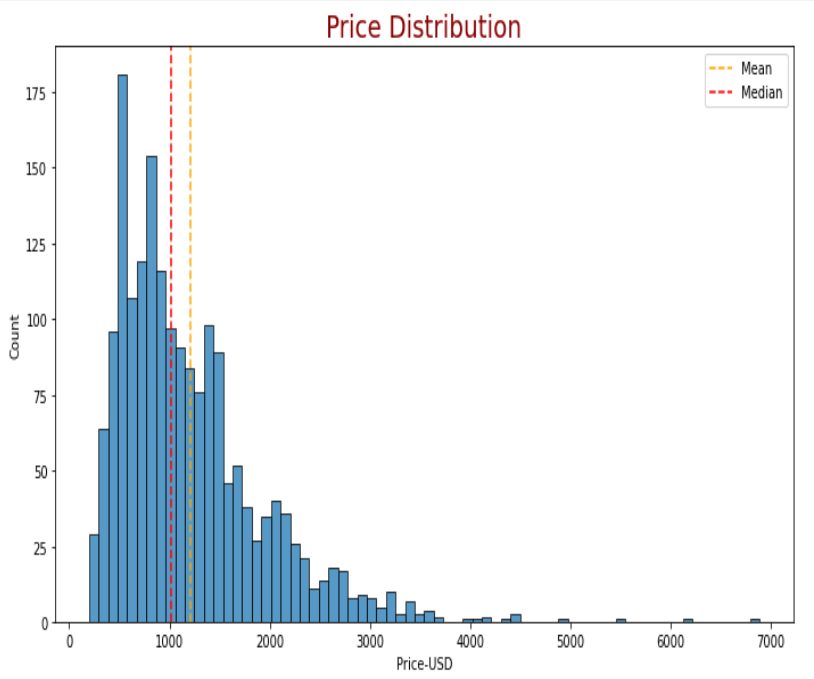


Fig 2: Analysis for price

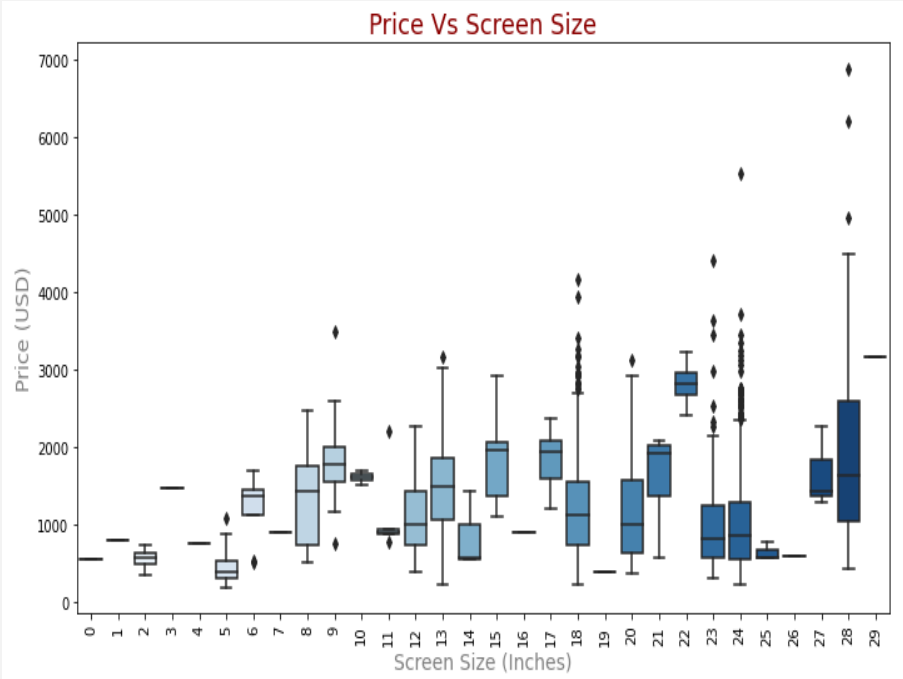


Fig 3: Analysis for price vs screen size

Methodology … Algorithms

- **Regression models**

1- Random Forest Regression model.
2- Linear Regression model.

- **Models building steps :**

1. Label Encoder
2. Train model
3. Model evalulation
4. Experiments
5. Comparison of the performance of the models

Algorithms … Linear Regression model

Linear Regression model

Linear regression is one of models machine learning-supervised learning.
What is meant by linear regression? Linear regression aims at finding a linear relationship between two continuous variables. By searching for an statistical relationship but not a deterministic relationship.
Regression: Predict the real-valued output for each individual, based on input data.

Model building steps :

1. Label Encoder
2. Train model
3. Model evalulation
4. Experiments
5. Comparison of the performance of the models

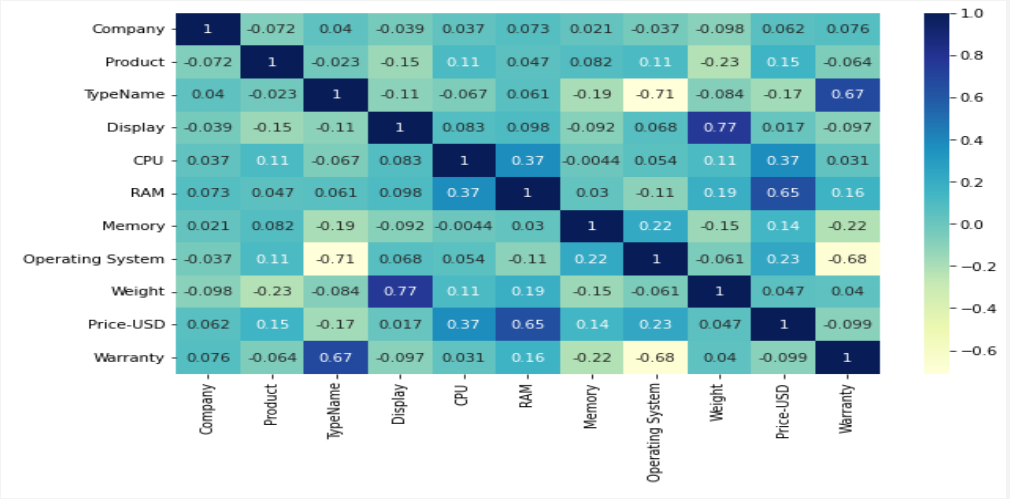


Fig 4: Features Correlating

Linear Regression model … Cont

Model building steps :

1. Label Encoder
2. Train model

The model is trained by dividing the dataset into two sets: training set 80% and testing set 20%
Through the training set, the model is trained, and through the test set, the prediction is calculated by ***score***.

SCORE	Training set 80%	Testing set 20%
	0.54	0.50

3. Model evalulation **on testing data**

- Mean Squared Error (MSE) = 283926
- Mean Absolute Error (MAE) = 361
- Root Mean Square Error (RMSE) = 532
- R^2 Score = 0.50

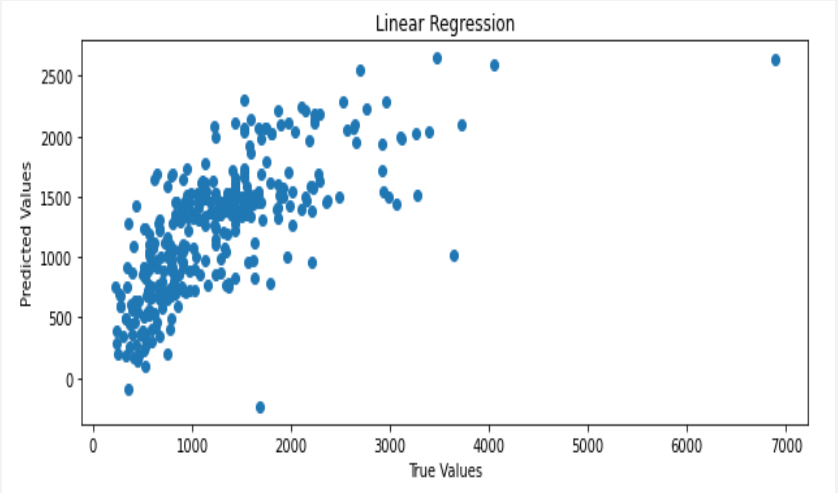


Fig 5: Predicted Values for LR

Linear Regression model … Cont

Model building steps:

Experiments

1: select some of the columns
2: Polynomial Feature
3: adding interaction terms

score	select some of the columns	Polynomial Feature	adding interaction terms
Train R^2	0.44	0.56	0.57
Validation R^2	0.38	0.53	0.53

Algorithms … Random Forest Regression model

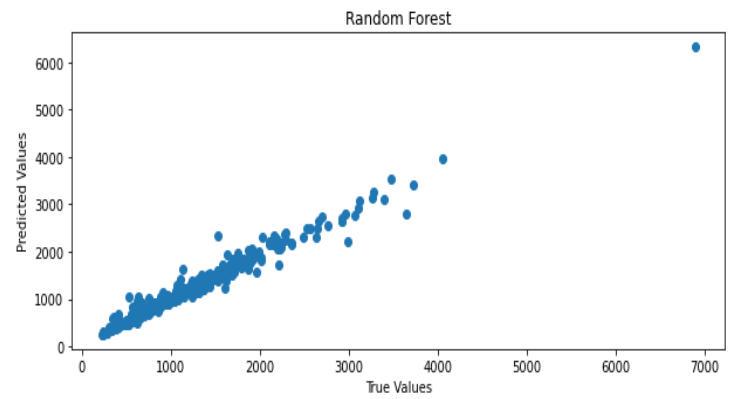


Fig 6: Predicted Values for RFR

	Feature Importance
RAM	0.459986
Weight	0.118541
Product	0.112920
CPU	0.088912
Operating System	0.058137

Model building steps:

1. Label Encoder
2. Train model

Similar to the previous model LR.

SCORE	Training set 80%	Testing set 20%
	0.97	0.96

3. Model evalulation on testing data

- Mean Squared Error (MSE) = 83
- Mean Absolute Error (MAE) = 19389
- Root Mean Square Error (RMSE) = 139
- R^2 Score = 0.96

Comparison of the performance of the models

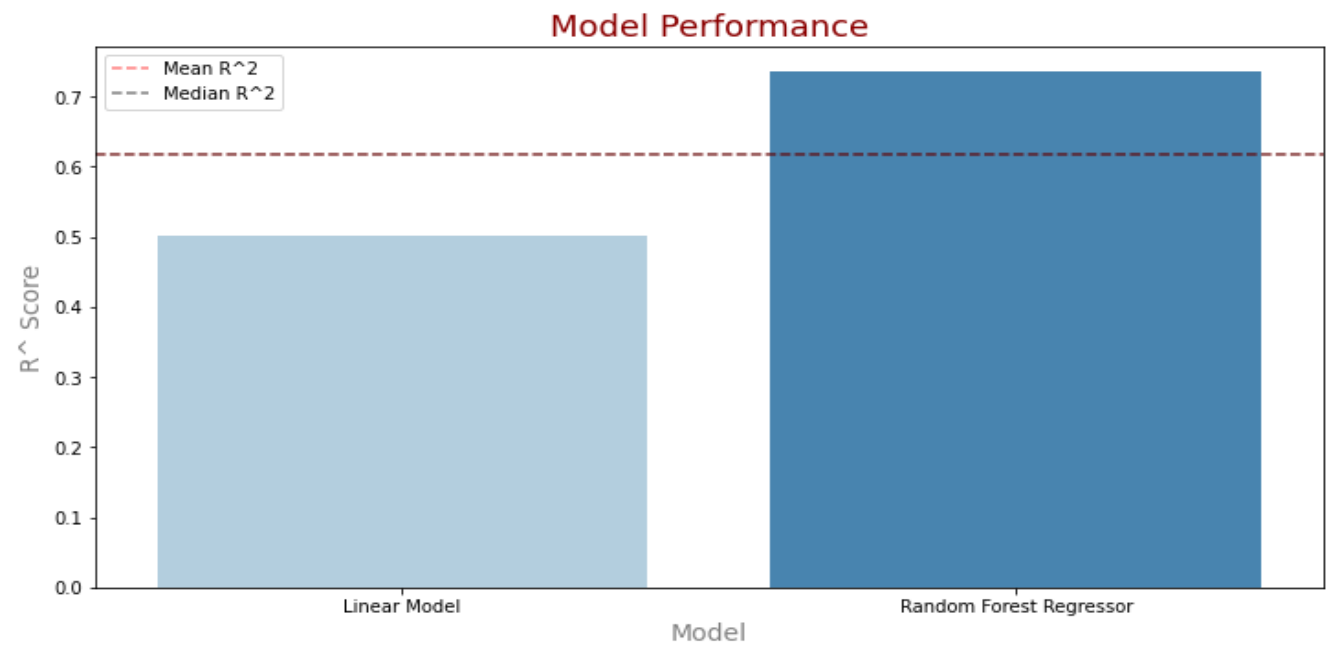


Fig 7: Model Performance

Conclusion and Future work

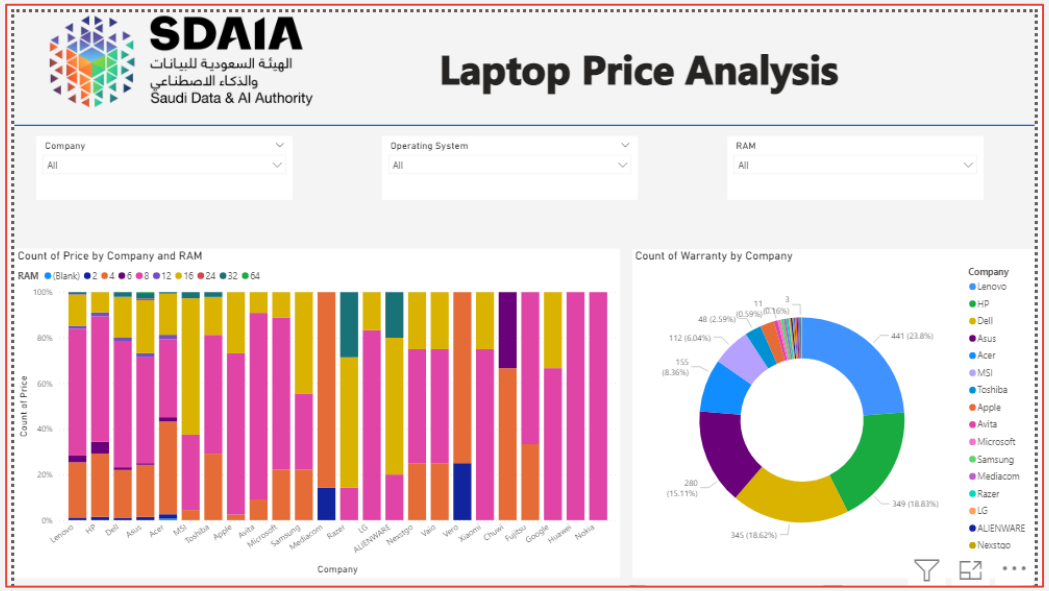
Conclusion:

We note that the prediction results for the Linear Regression model are low even after conducting experiments, as there is a slight increase in the results, but in general the results are not satisfactory or high, like the Random Forest Regression model where the prediction results were in training = 97 and testing = 96.

Future work

Scrap data form website.
Explore different models.

let's see Analysis in Power Bi



Thanks

Do you have any questions?