AMAZON LAPTOPS PRICE PREDICTION

Presented by
Nada alruwaythi
Nisreen alsayegh

OUTLINE

- Introduction
- Workflow
- Data
- Design
- Tools
- EDA-Analysis
- Models
- Conclusion

INTRODUCTION

Amazon represents one of the largest marketplaces on the Internet. People use its services on a daily basis to order groceries, books, laptops, and even web hosting services.

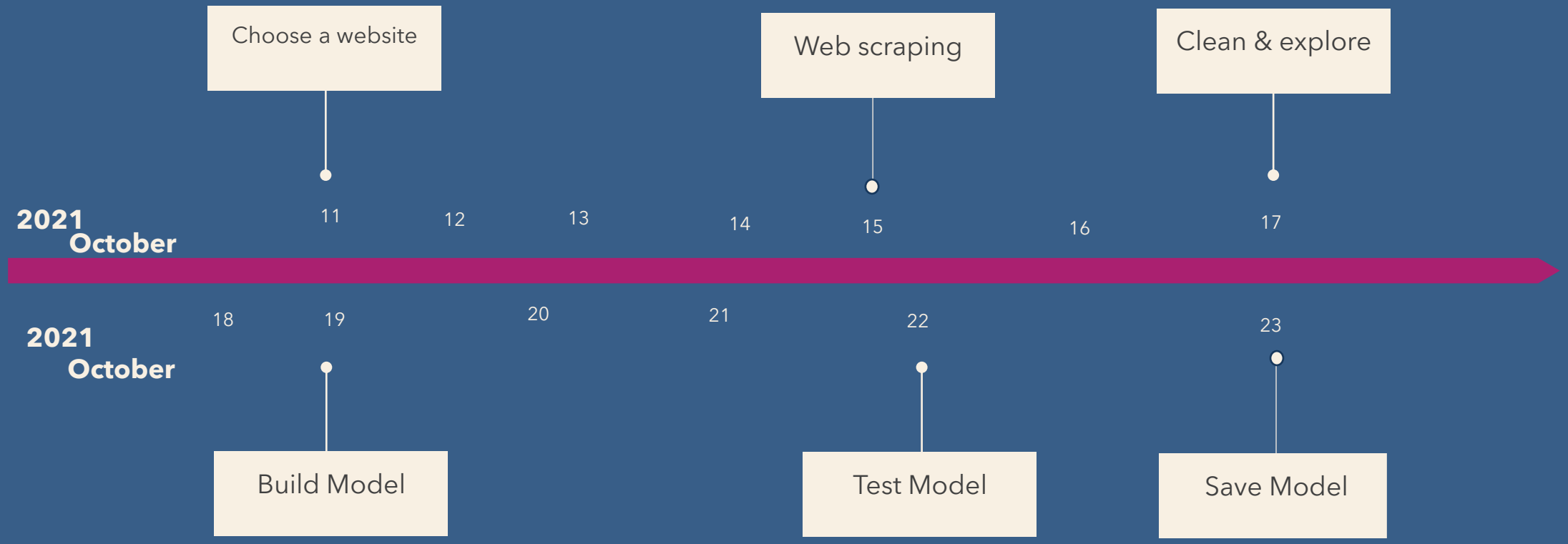
***Goals* :**

The model predicts the price of the laptop

WORKFLOW

- Web Scraping
- EDA
- Building Models
- Testing Model

TWO WEEK PLAN



SCRAPED DATA

We scrap around 500 pages from amazon.com
then we marge it with the dataset in Kaggle at the end
the total become 1747 rows and 11 columns.

Data Features:

Brand	Standing screen display size	Processor Brand	Processor Speed
RAM Size	Hard Drive Size	Hard Disk Description	Graphics Chipset Brand
Operating System	Item Weight	Price	

DATA CLEANING

EXPLORING DATA

Explore data after merging both datasets

CLEANING DATA

By removing Null and Duplicate values

Before :row1681,columns11

After dummies:
row 1681 ,columns 62

TOOLS

BEAUTIFULSOUP

Collect data from website

PANDAS, MITO

Explore & clean data

MATPLOTLIB, SEABORN

Visualise data and models

EXCEL

Download data as csv file

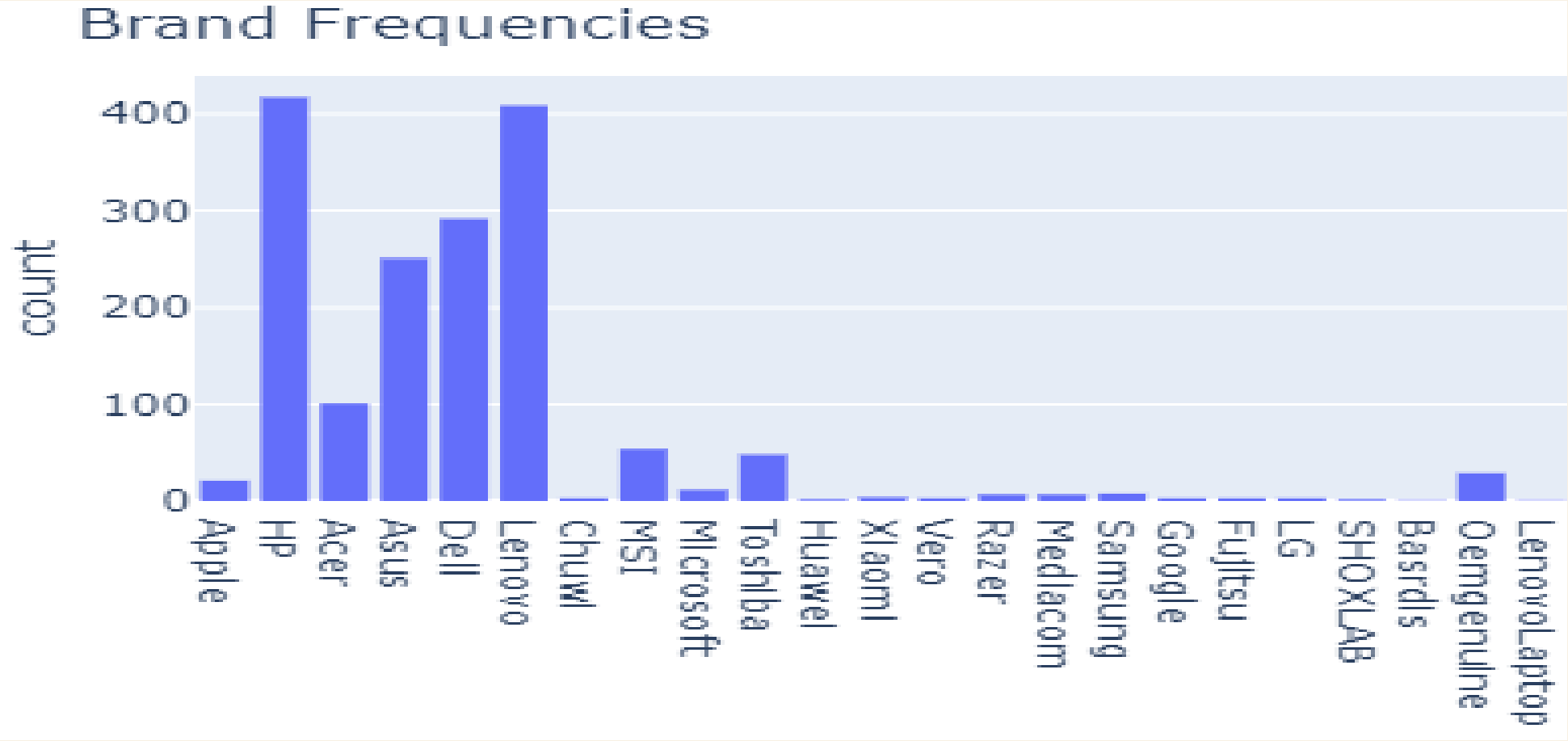
SKLEARN

Models training

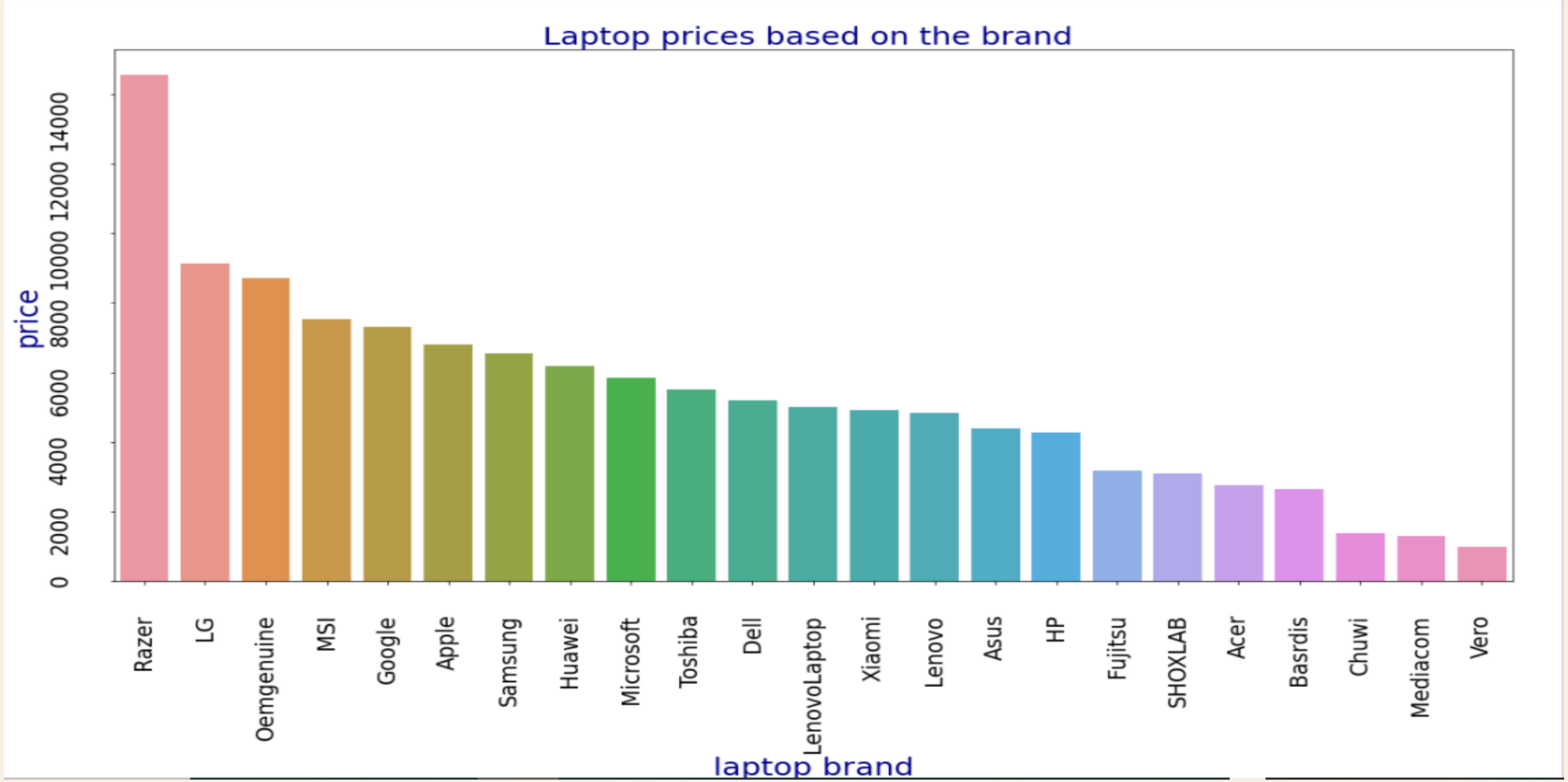
FLASK

build a web application

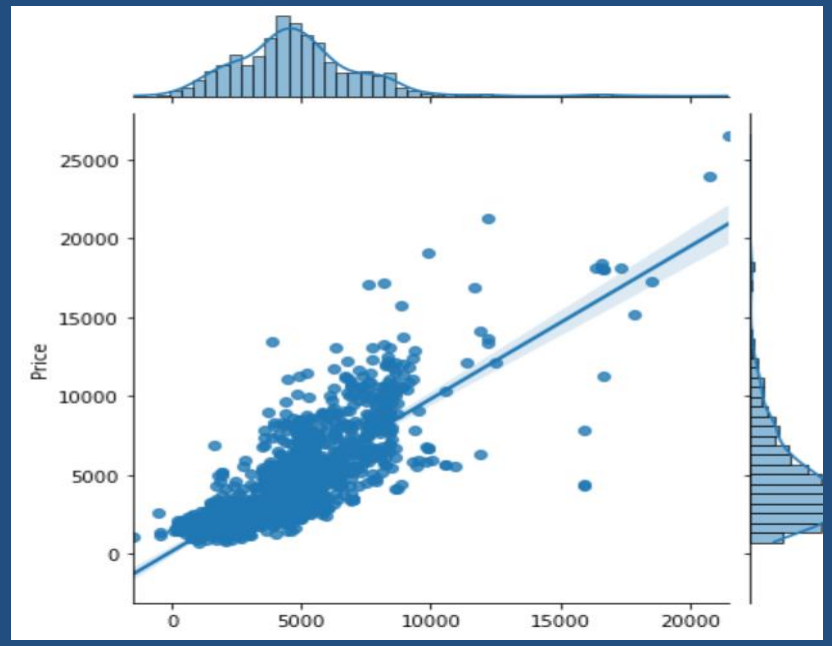
EDA - ANALYSIS



EDA - ANALYSIS



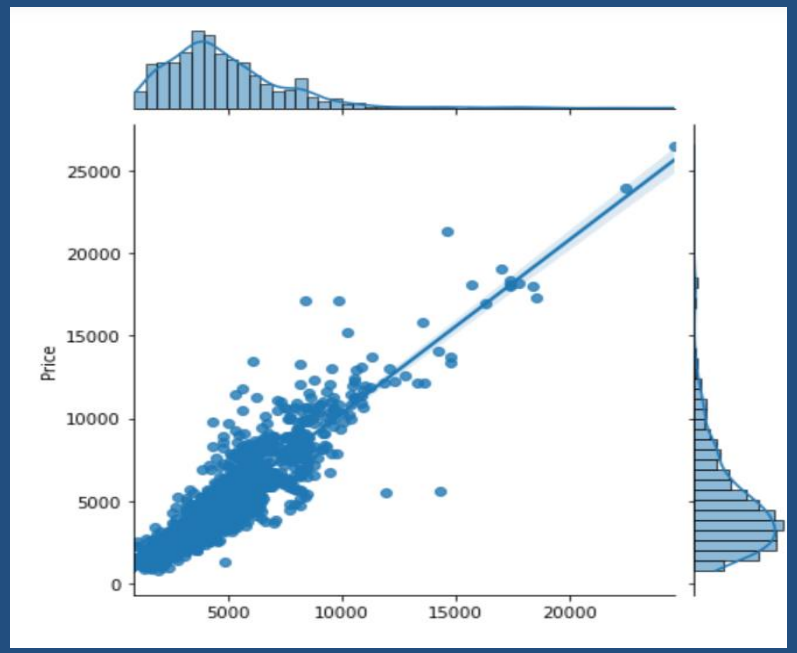
MODELS



Linear Regression

Linear Regression train R^2:0.646

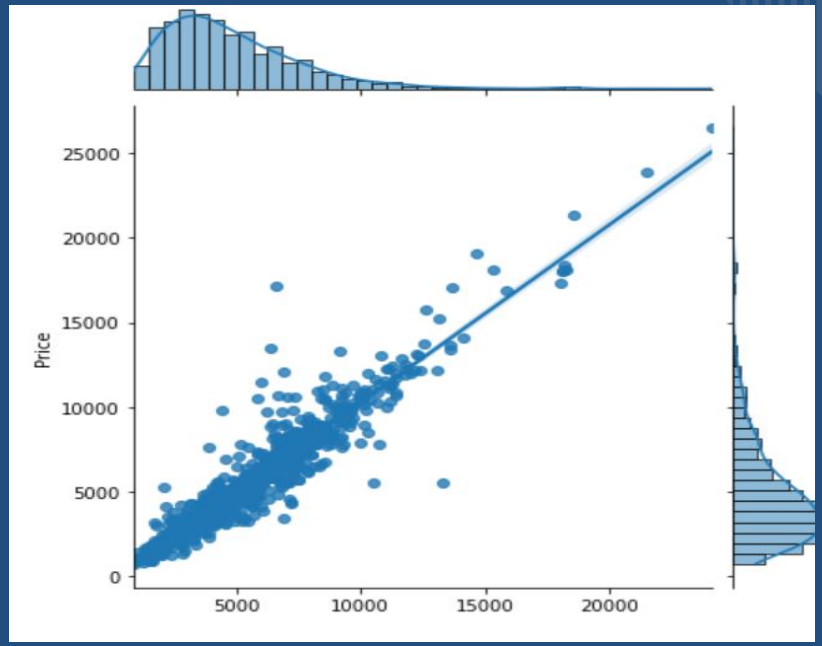
Linear Regression validation R^2: 0.582



Gradient Boosting Regressor

Gradient Boosting Regressor train R^2: 0.865

Gradient Boosting Regressor validation R^2: 0.710



Forest Regressor

Forest Regressor train R^2:0.954

Forest Regressor validation R^2: 0.738

THE BEST MODELS

Random Forest Regressor Train R^2: 0.954
Random Forest Regressor Validation R^2: 0.738
**After apply feature engineering : ** 0.948

Random Forest Regressor Test R^2 : 0.834

Gradient Boosting Regressor Train R^2: 0.861
Gradient Boosting Regressor Validation R^2: 0.706

**After apply feature engineering : ** 0.865

Gradient Boosting Regressor Test R^2 : 0.810

SUMMARY

Linear Regression	Forest Regression	Gradient Boosting	Polynomial	Lasso	Ridge	ElasticNet	
0.646	0.954	0.862	0.865	0.626	0.626	0.626	Training
0.582	0.738	0.710	-26*10^12	0.534	0.534	0.534	validation
	0.948	0.865					Feature engineering-Training
	0.723	0.710					Feature engineering-validation
	0.834	0.803		0.543	0.545	0.559	Testing

CONCLUSION

After building different models Random Forest Regressor has the highest r^2 in both validation stage and testing stage removing unneeded columns has increased r^2 in Gradient Boosting Regressor but not in Random Forest Regressor .

FUTURE WORK

COLLECT AND SCRAP MORE DATA -

- SCRAP DATA FORM OTHER WEBSITE

- EXPLORE DIFFERENT MODELS

THANK YOU

Nada alruwaythi
Nisreen alsayegh