PHONE PRICE

REGRESSION

INTRODUCTION

- With the technological development, mobile phones have become a necessity for the majority of people.

- This linear regression project aims to predict the prices of the phones in the future based on the most important characteristics that affect its value in the market.

PROBLEM STATEMENT

- What is the effect of features on phone prices ?

OBJECTIVE

- Predict mobile phone prices
- Based on specific phones features

DATA SET

- The data to be tested in this project are scraped from 91mobiles.com
- Before cleaning: 2658 rows and (108) columns.
- After cleaning: 2473 rows and (41) columns.

TOOLS

- Pandas
- Numpy
- Sklearn
- Statsmodels
- Seaborn
- Matplotlib
- SciPy
- Selenium & BeautifulSoup
- Pickle

PROJECT WORKFLOW

Web Scraping

- 2658 rows
- 108 columns

EDA

- Drop columns
- Rename
- Data Cleaning
- Outliers

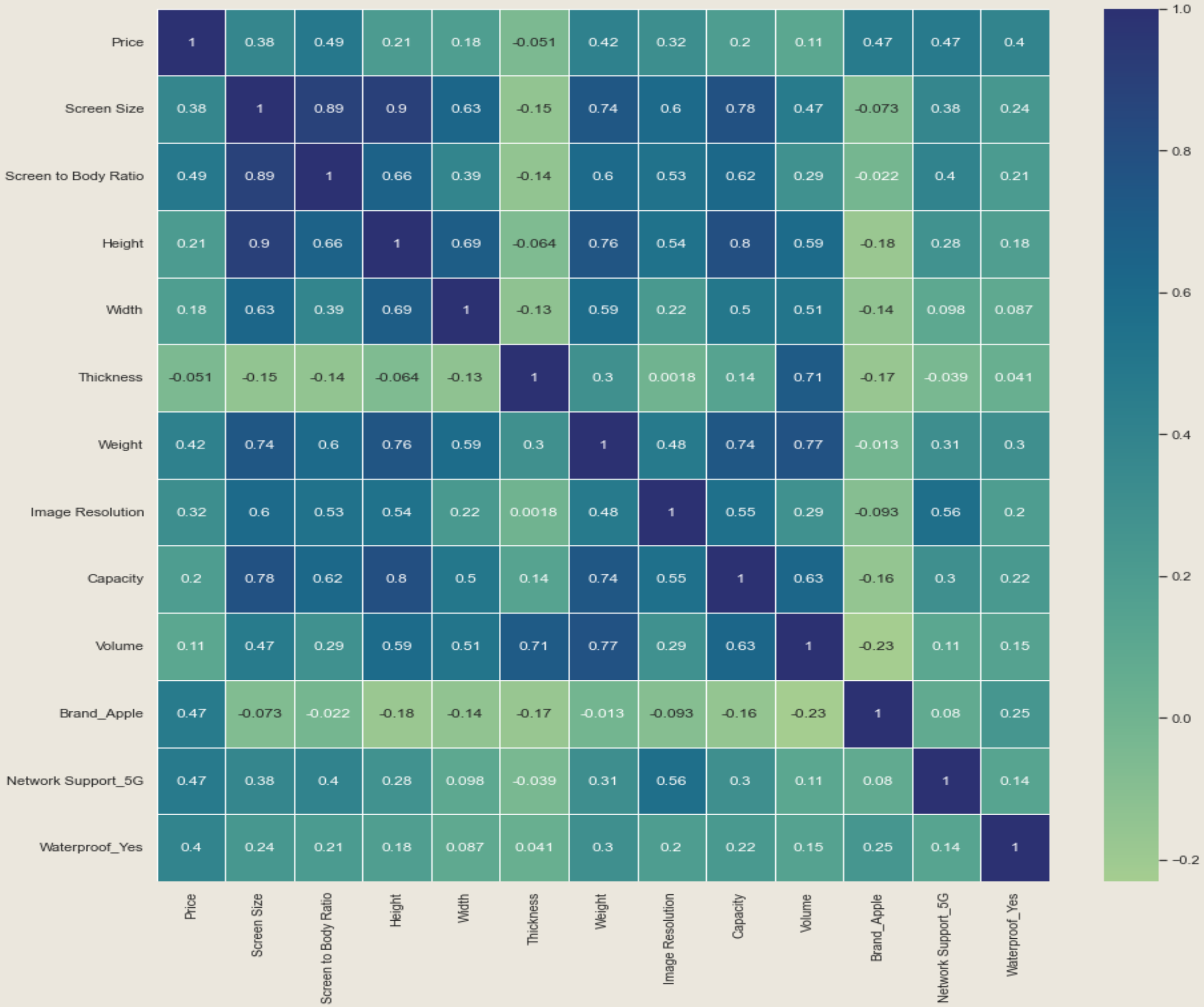
Regression

- Feature Engineering
- Cross-validation
- Scaling
- Training Models

CORRELATIONS WITH TARGET

Highest correlated features with target:

- Brand
- Screen Size
- Screen to Body Ratio
- 5G Network support
- Waterproof



REGRESSION MODELS

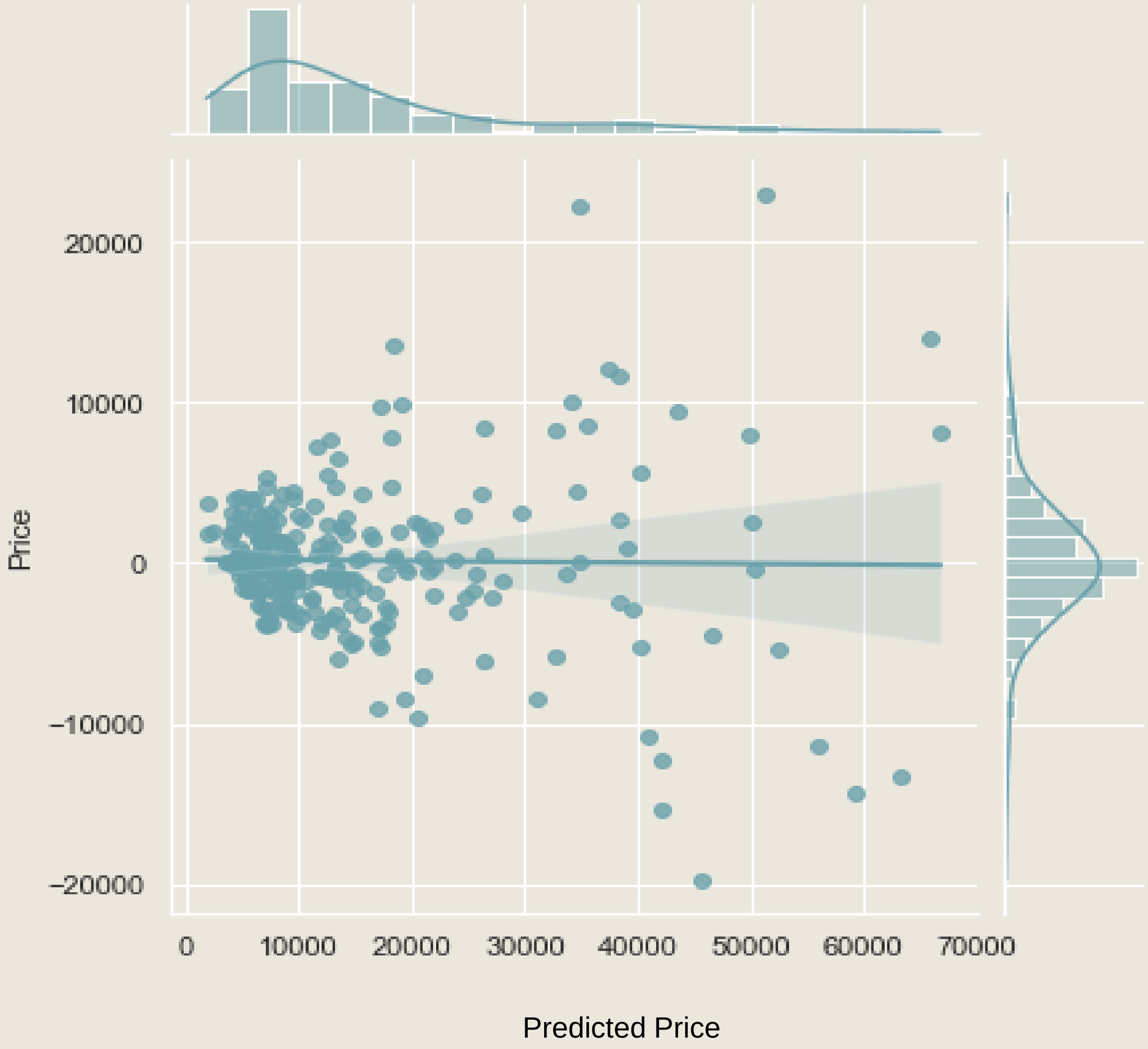
MODELS	TEST SCORES	**LOSS FUNCTION (MAE)**
Simple Linear (CV)	0.8708464488562737	3277.46
Lasso (CV)	0.8738624622859356	3202.47
Ridge (CV)	0.873863131914 1522	3203.46
ElasticNet (CV)	0.8746795574928716	3183.75
Polynomial (Degree 2)	0.5277092107850869	4087.94
Random Forest	0.7757199361551717	4098.03
Gradient Boosting Regressor	0.8768546692714321	3222.66

CONCLUSION

- As seen in the previous table, the best model in terms of highest score and lowest loss is the **ElasticNet.**

- The following regression plot represents model target vs predicted model target.

- Even though the data was cleaned and scaled, the model did not capture all the patterns in the data.



Thank you for listening