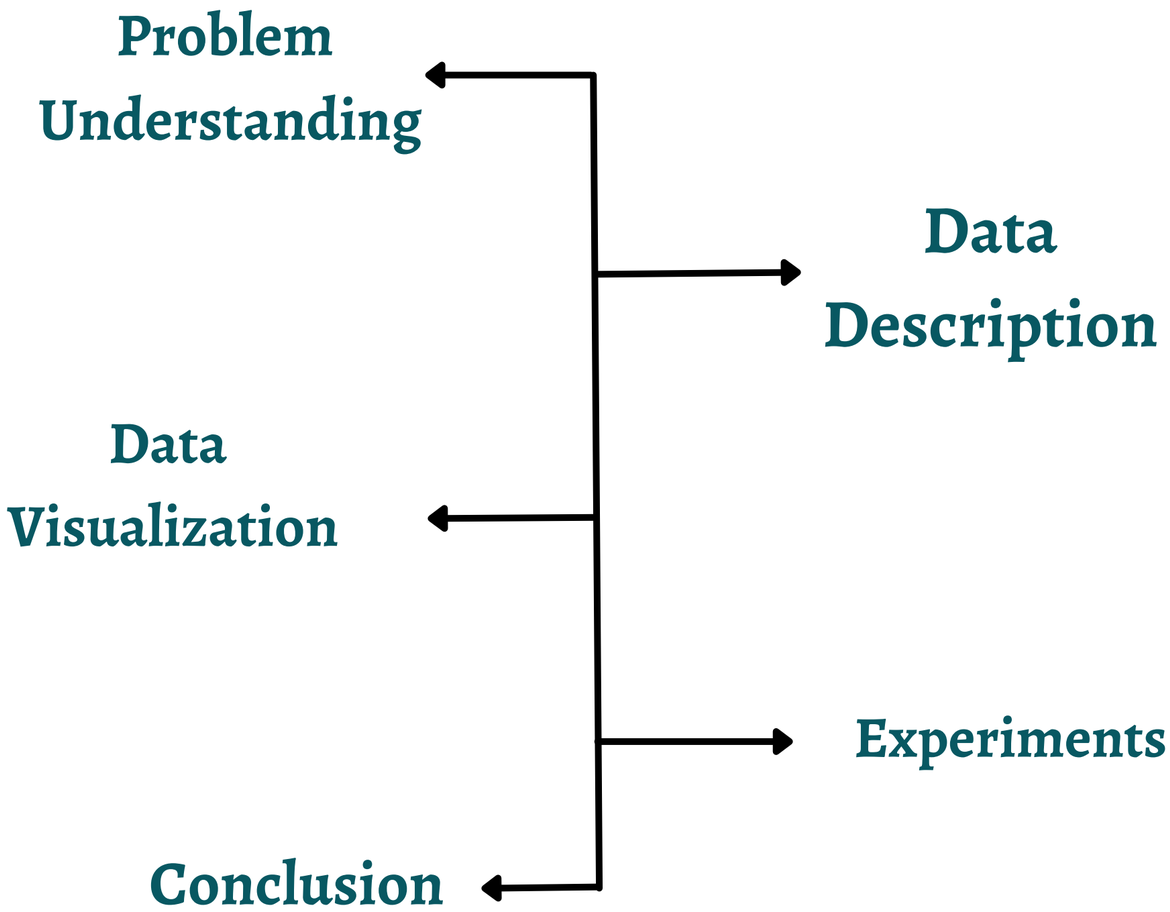
HOTELS PRICE PREDICTION IN SAUDI ARABIA

Methodology



Methodology

Problem Understanding

- ## Overview

One of Saudi Arabia 2030 vision key initiatives is “Quality of life”. This initiative comes with the aim of diversifying and enriching the tourism and entertainment experience in the Kingdom.

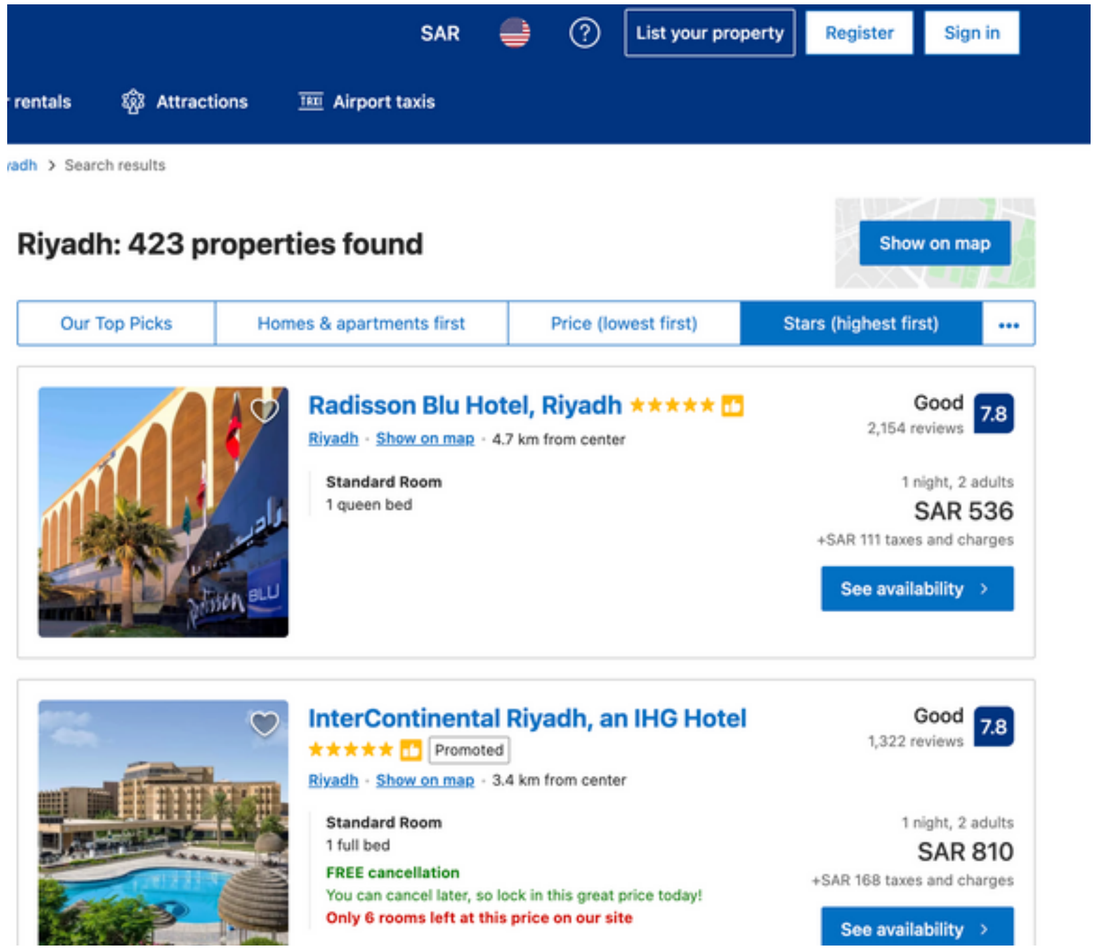
- ## Problem Statment

In this time of the year, hotels and resorts are in high demand in the Kingdom. In this project we build a model to predict the prices in riyals, given the rating score, and the number of reviews.

Methodology

Data Description

They data was token by web-scraping (booking) website and combine it with data from kaggle. it is about data of 1224 hotels in Saudi Arabia in the current month (December 2021),



Methodology

Data validation

Saudi
Welcome to Arabia

MISSING VALUES

Handling Some Missing Values

DUPLICATS

Handling Duplicates Rows

OUTLIARES

Checking Outliers

Methodology

Data Description

Saudi
Welcome to Arabia

Features

Hotel Name

Rating

Rating Title

Number of Ratings

Room Type

Place

Target

Price

Methodology

Data Visualization

Saudi Welcome to Arabia

Heatmap Correlation

Baseline Model

Train Score : 0.096

Validation Score : 0.130

Experiments

1

Dummy Variables

Train Score

0.1248

Validation Score

0.1024

2

Count Value for Location

Train Score

0.1671

Validation Score

0.1325

3

polynomial Equ

Train Score

0.1694

Validation Score

0.1329

4

Features Engineering (Adding Interaction Terms)

Train Score

0.225

Validation Score

0.175

5

Box Cox

Train Score

0.6436

Validation Score

0.629

6

Lasso

$\lambda = 2$

Train Score

0.640

Validation Score

0.631

7

Ridge

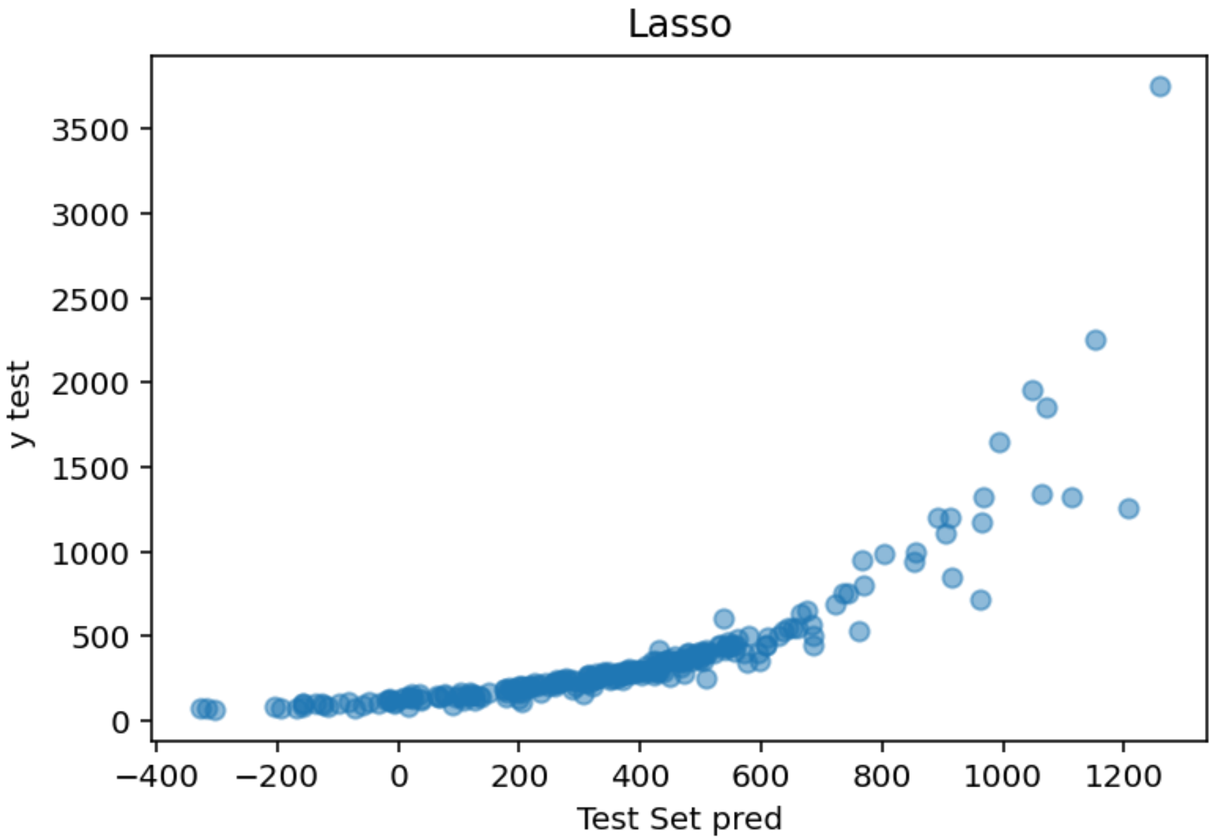
$\lambda = 6$

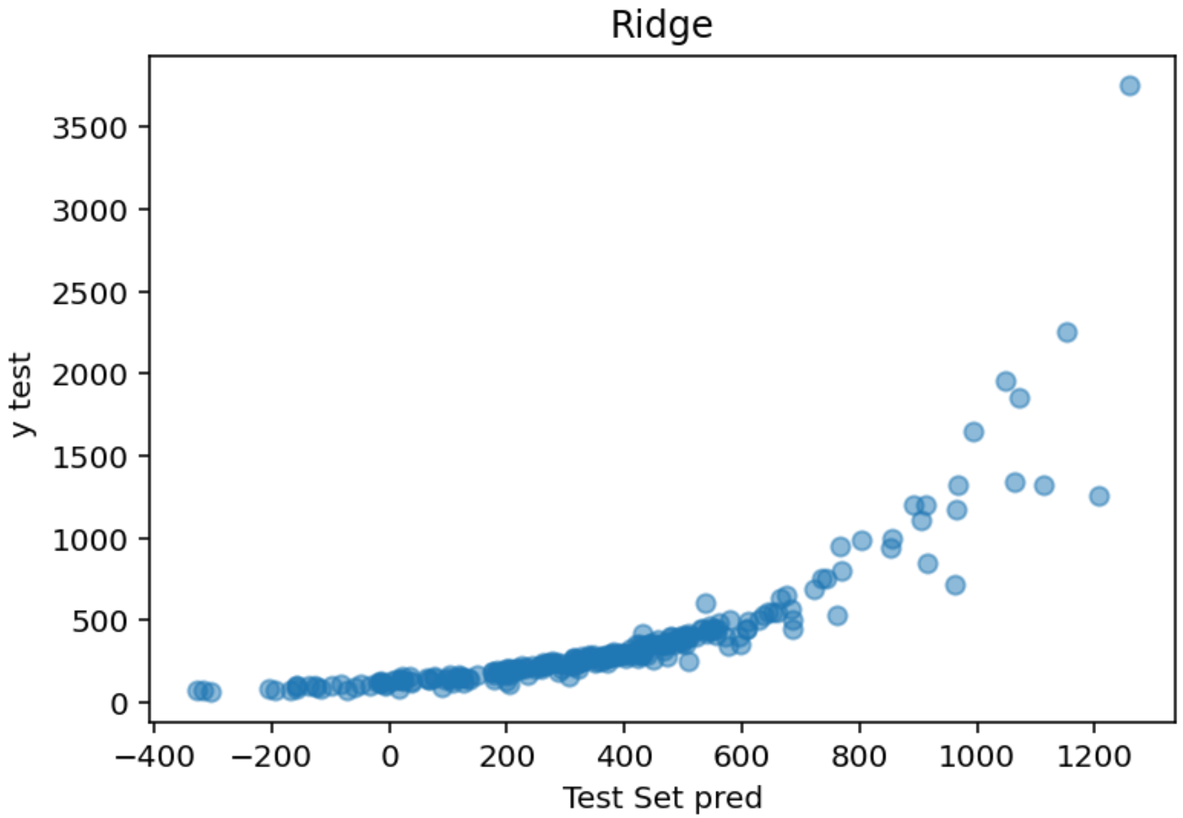
Train Score

0.643

Validation Score

0.631





SSE = 1.316065e+07

Conclusion

	Baseline	Dummy	Count	Polynomial Equ	Group-by price mean per location	Adding Interaction Terms	Box-Cos	LASSO	RIDGE
Train Score	0.096	0.124	0.1671	0.1694	0.169	0.225	0.643	0.64	0.64
Validation Score	0.1308	0.102	0.1325	0.1329	0.132	0.175	0.629	0.63	0.63

Best Model Train Score = 0.64

Best Model Validation Score = 0.63

Best Model Test Score = 0.63

Thank You ..

Raghad Albarrak

Maryam Aljasham

Welcome to Arabia