House Prices Prediction

Abdulrahman Almegren – Yasir Albahlal

content:

- Data
- Data Cleaning
- EDA
- Models comparison
- Final model

Data:

01 About 30000 house scraped from Aqar website

02 11 columns

03 PROJECT GOALS

The objective of this project is to predict Riyadh house prices.

04 9 Features and 1 predict (price)

Data Cleaning:

Drop or fix nulls

Change some nulls with 0 and drop others

Delete duplicate rows

7500 duplicated row

Rename columns and change datatypes

Rename Arabic columns and change objects to numerical

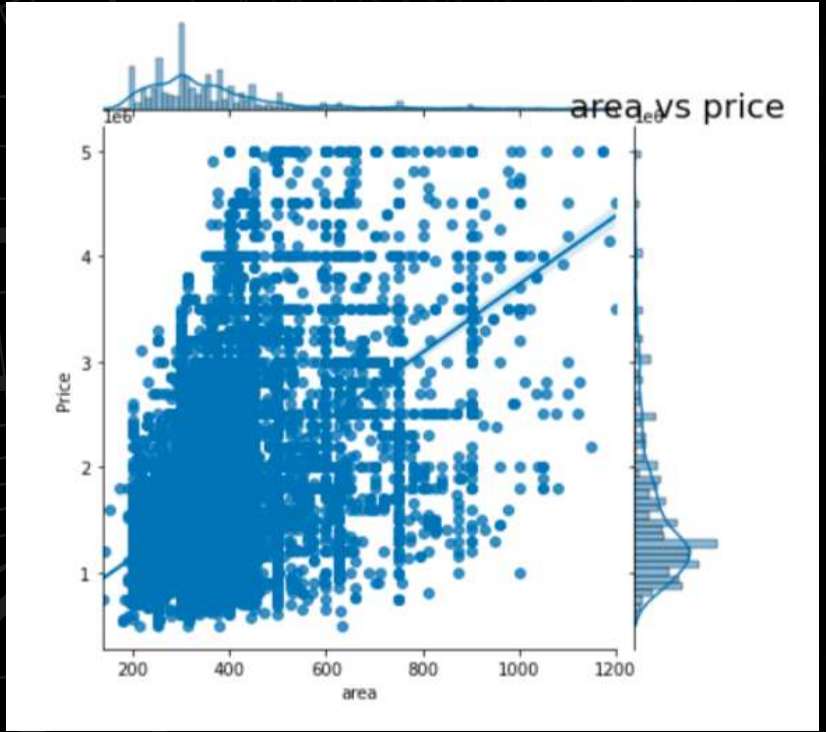
Drop outliers

Limit prices, area , apartments and street width

EDA

Features correlation

Price VS Regions and area



Comparing regression models:

Regression name	Train	Validation
Linear Regression	0.745	0.761
Polynomial(D=2)	0.800	0.778
Lasso	0.745	0.761
Ridge	0.745	0.761
LassoCV with poly features	0.792	0.785

LassoCV with Polyfeatures AND Conclusion

Based on the train and validation score we choose LassoCV with polyfeatures and apply the test on this regression.

Best model is LassoCV with polynomial features.

Regression name	Train	Validation	Test
LassoCV with polyfeatures	0.792	0.785	0.822

THANKS

Does anyone have any questions?