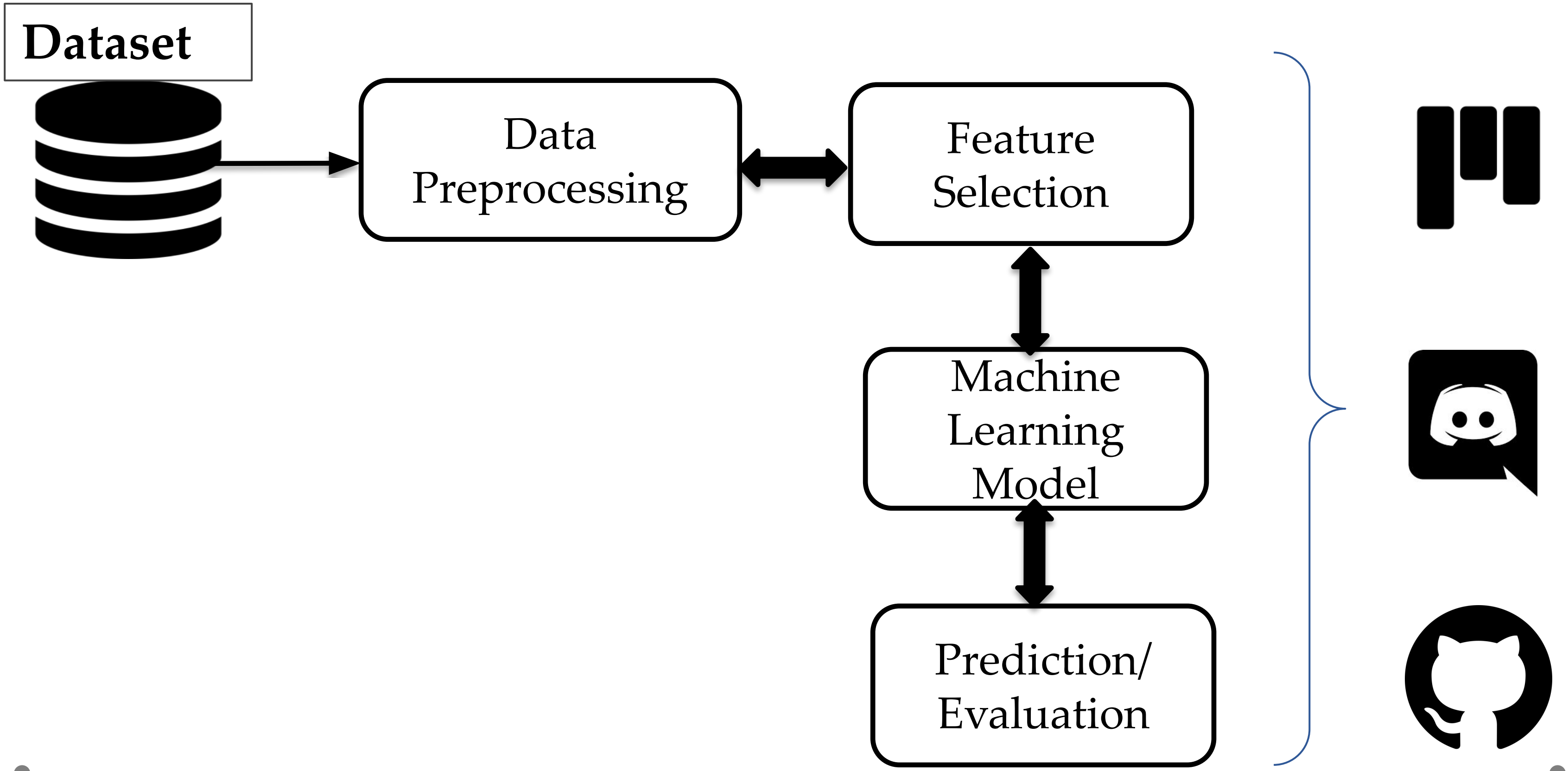
House Price Prediction

Francesco Mariottini
Joachim Kotek
Ankita Haldia
Manasa Noolu

Project Management

- The main objective of the project is to predict the Belgium's House Price using Linear Regression Model.



Data preprocessing

- Null facades_number' <- median by 'property_subtype'
- "Soft" outliers detection & removal by Tukey fences

Original dataset (10607 records)

Column	Outliers count	Outliers [%]	First Outlier
price	994	9.37	988000
rooms_number	246	2.32	8.0
area	686	6.47	410.0

Apartments joined with Statbel dataset (423 records)

Column	Outliers count	Outliers [%]	First Outlier
price	29	6.86	1549000.0
rooms_number	21	4.96	14.0
area	24	5.67	716.0

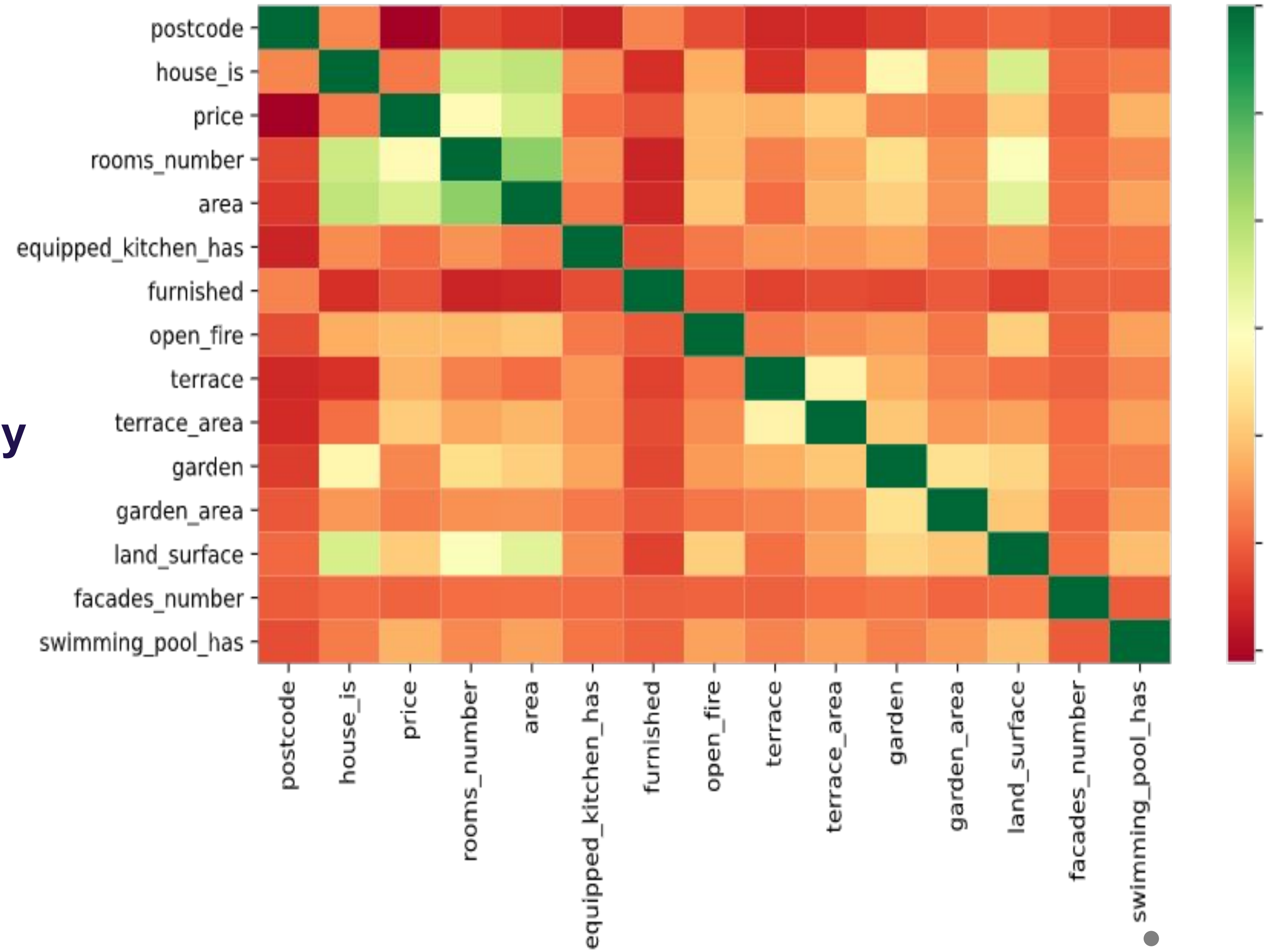
Feature Selection

Steps :-

1. Correlation Matrix

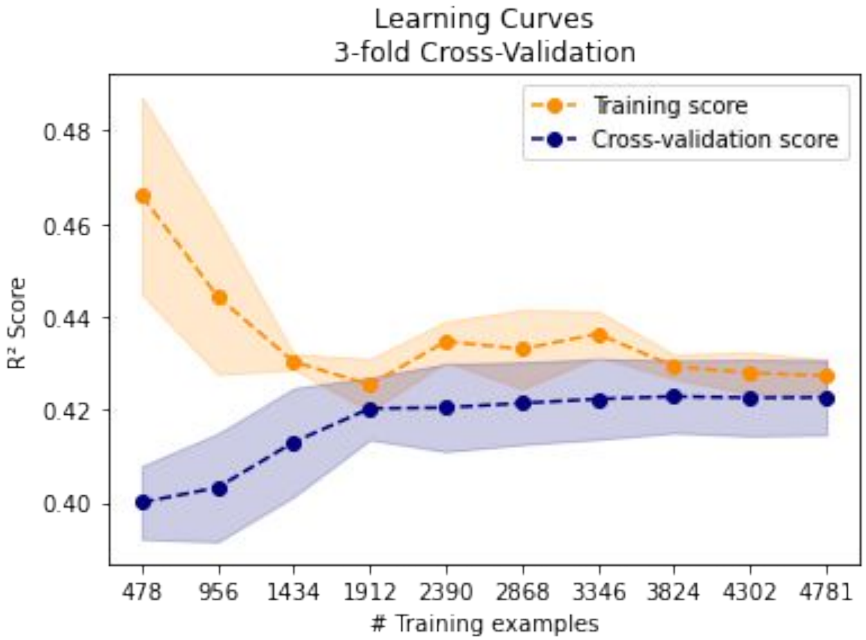
2. Chi Square Contingency

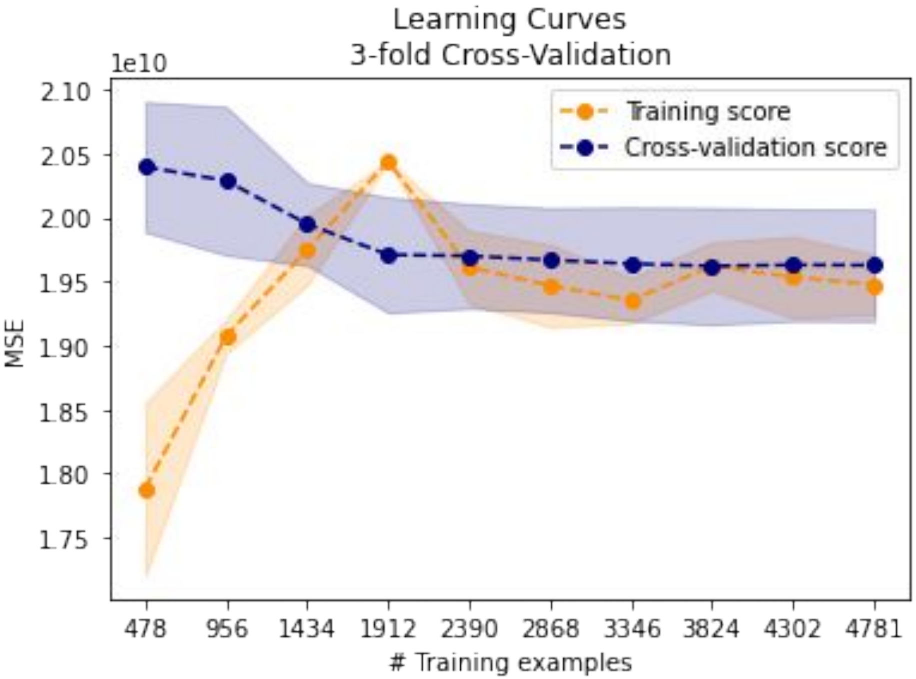
3. One Hot Encoding



Machine Learning Model

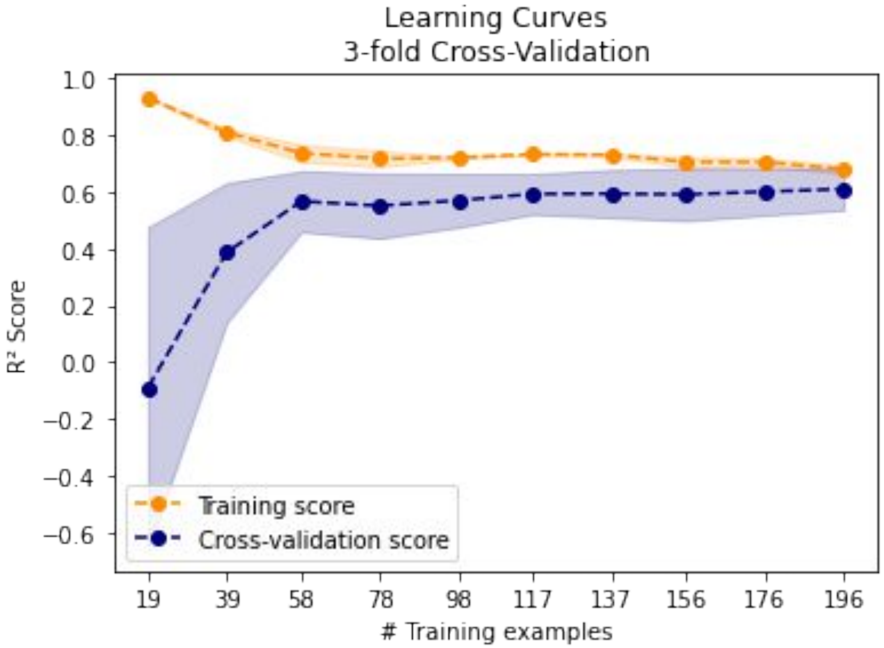
Whole Dataset



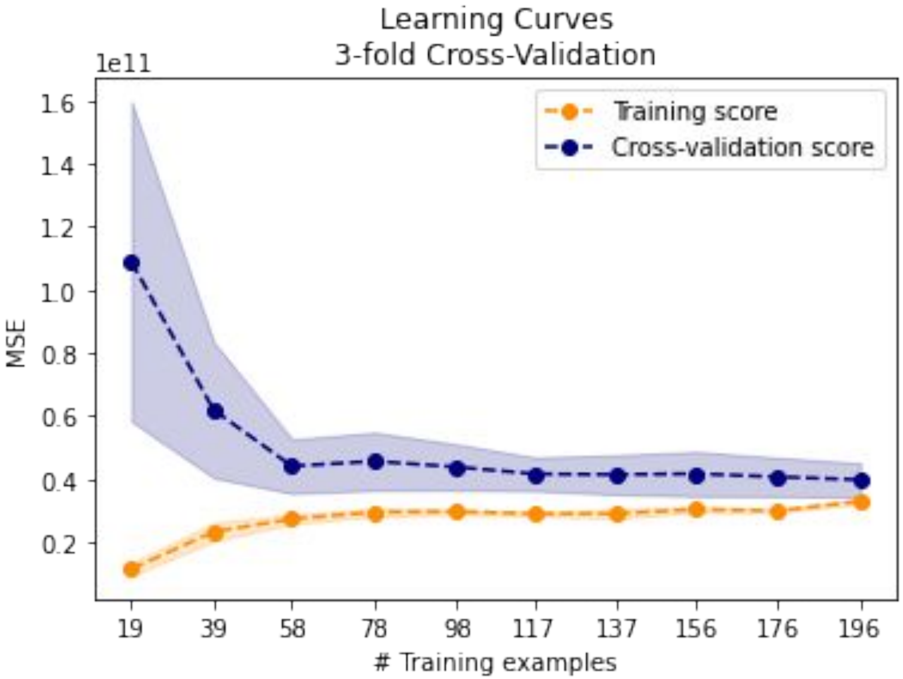


Machine Learning Model

Appartments Only

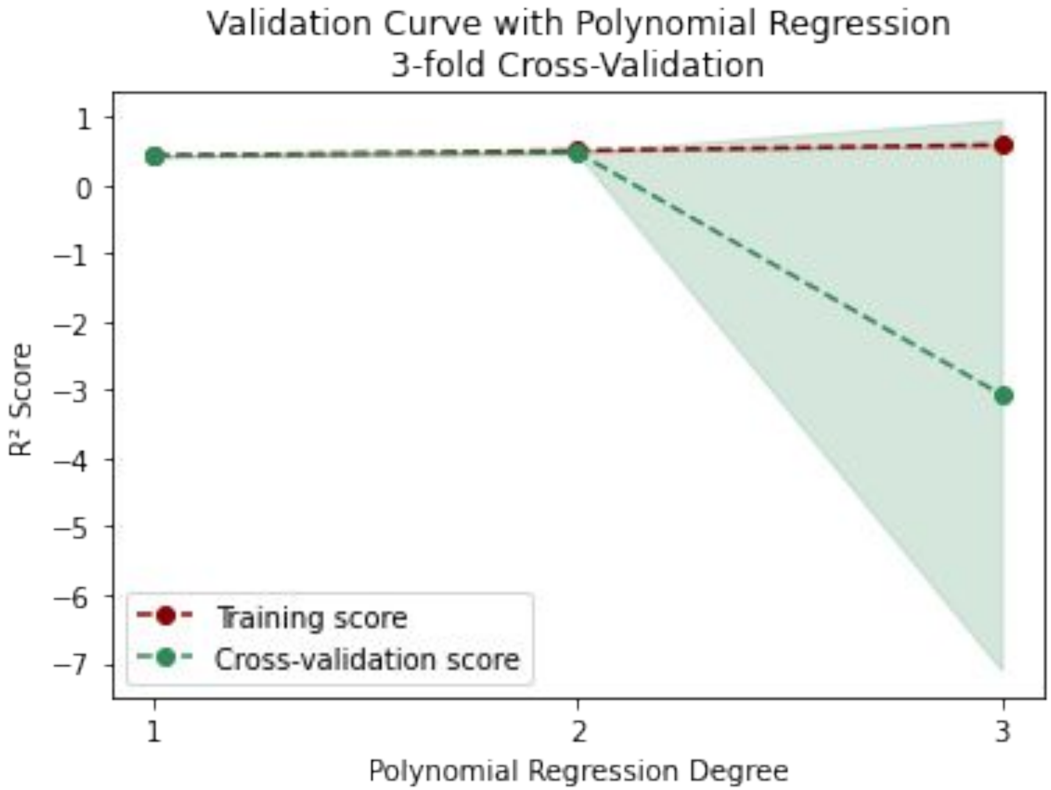


Learning Curves
3-fold Cross-Validation



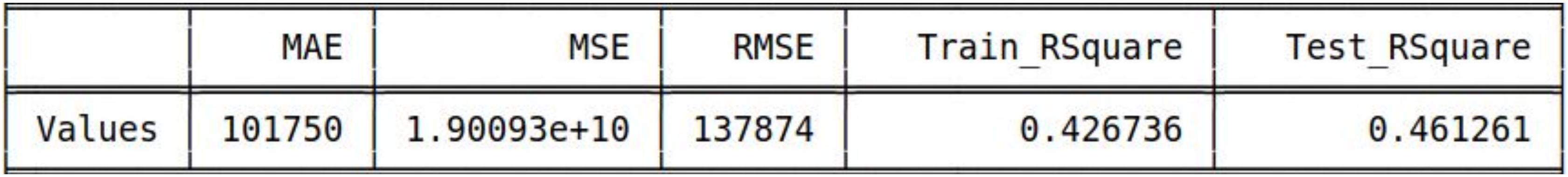
Learning Curves
3-fold Cross-Validation

Machine Learning Model

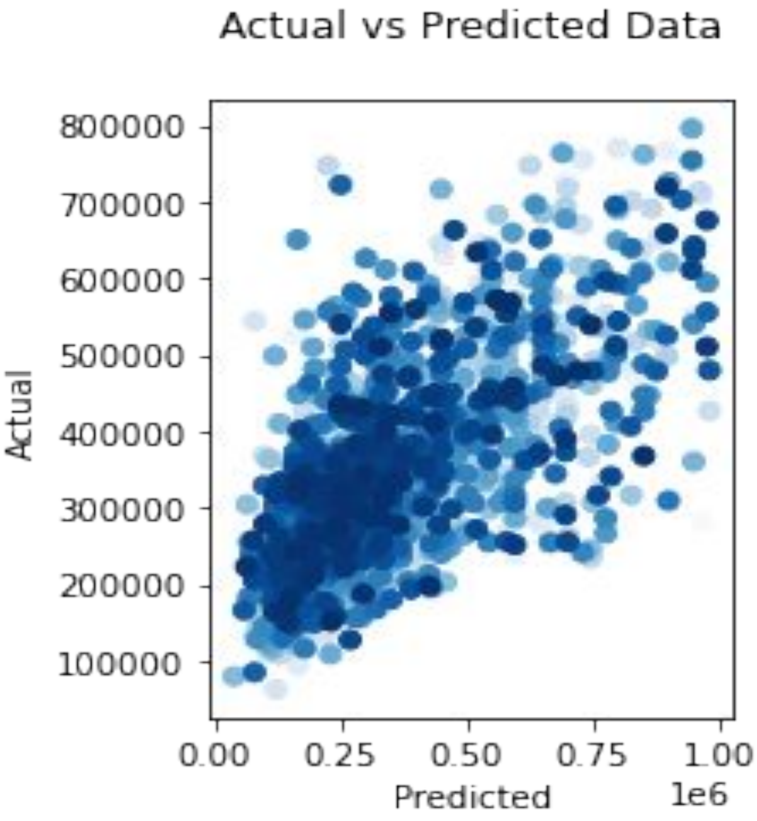


Evaluation

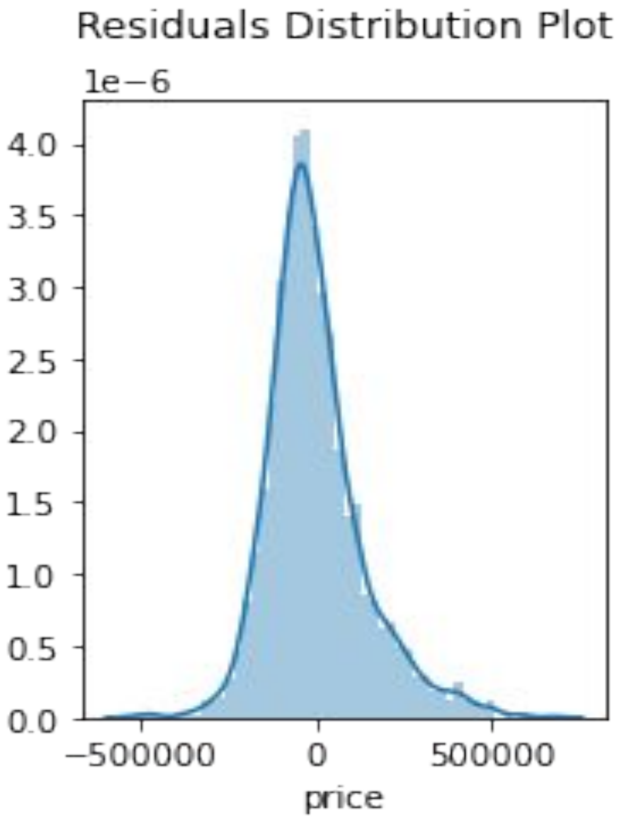
Regression Metrics on Whole Dataset



	MAE	MSE	RMSE	Train_RSquare	Test_RSquare
Values	101750	1.90093e+10	137874	0.426736	0.461261



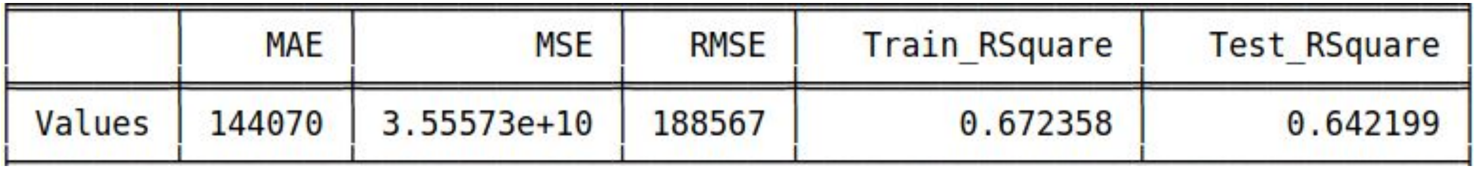
Actual vs Predicted Data



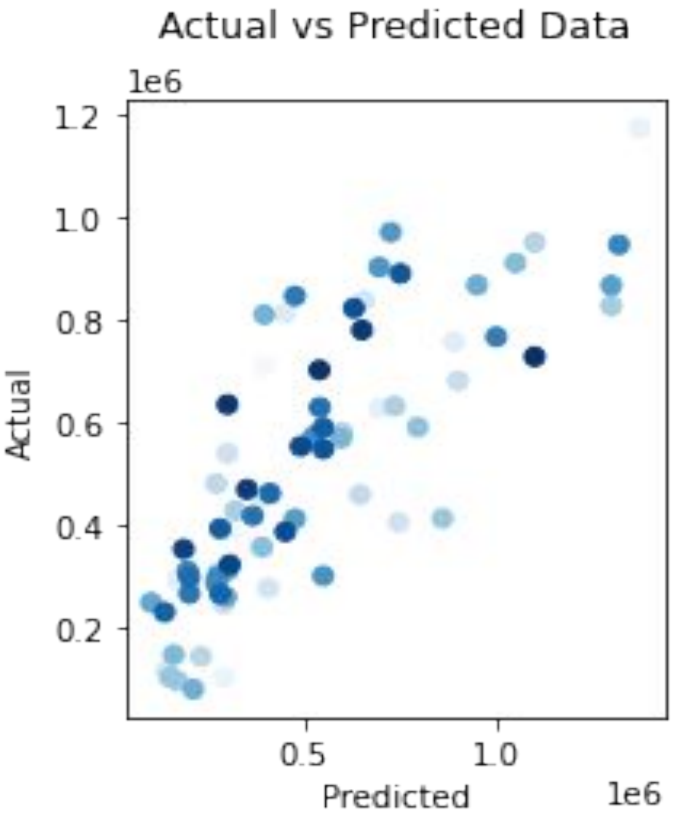
Residuals Distribution Plot

Evaluation

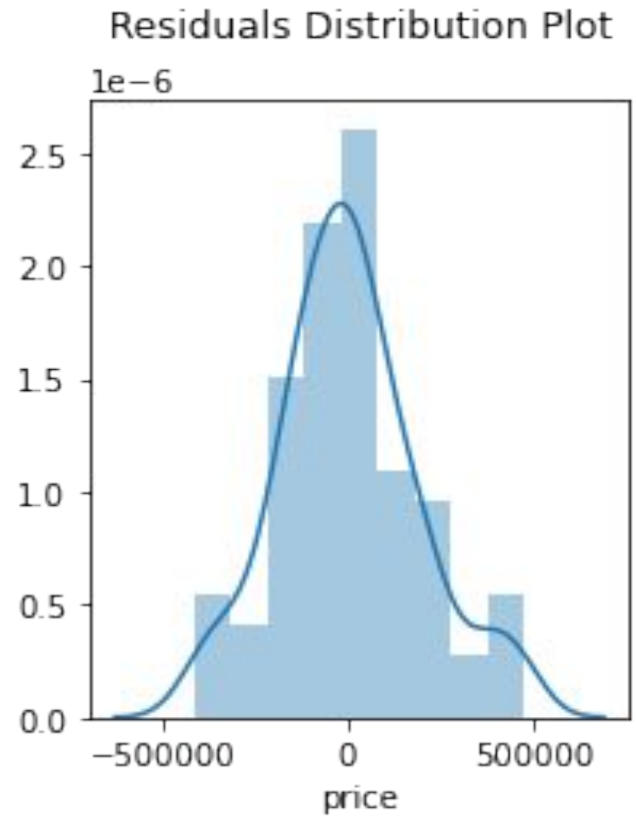
Regression Metrics with Apartments



	MAE	MSE	RMSE	Train_RSquare	Test_RSquare
Values	144070	3.55573e+10	188567	0.672358	0.642199



Actual vs Predicted Data



Residuals Distribution Plot

Observations

- Complete dataset with more data and variables (new features) could have also helped.
- Better correlation found using a subset but no time to explore.
- Feature selection plays vital role in increasing the accuracy.

Queries???

Thank You