

King County Home Prediction

Contributor: Prabagini Thambinathan

Overview

- This project focuses on predicting the housing price with various factors affecting the price.
- It involves in investigating house sales in the King County area and building a multiple linear regression model to predict the sale price.



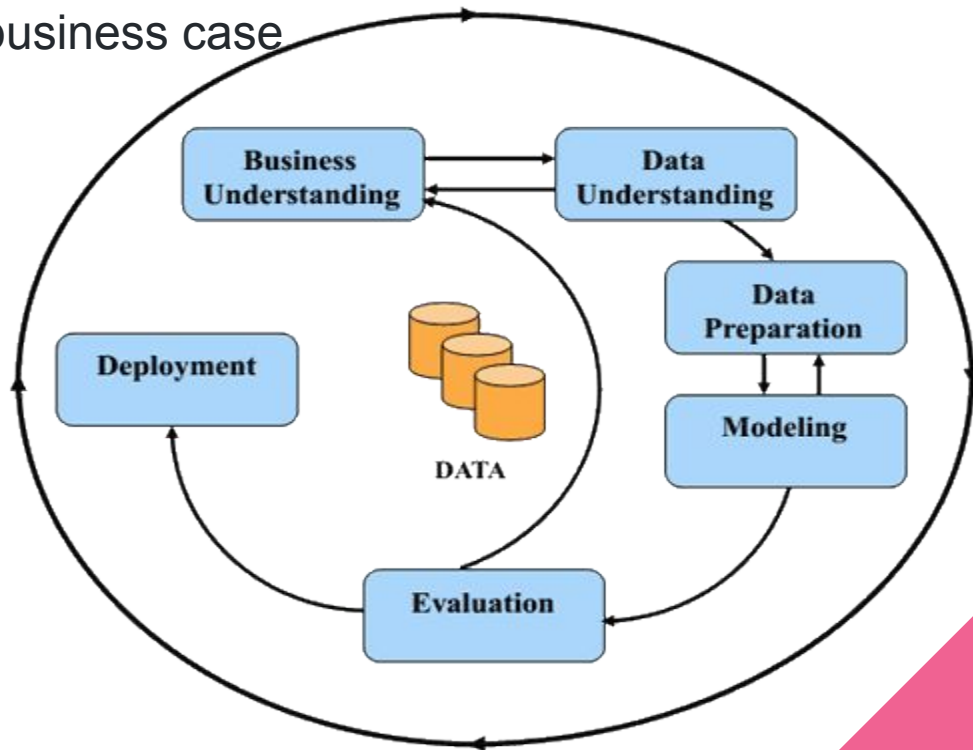
Data

- Real Property Sales (EXTR_RPSale.csv)
 - Residential Building (EXTR_ResBldg.csv)
 - Parcel (EXTR_Parcel.csv)
- Used these table for my model analysis.



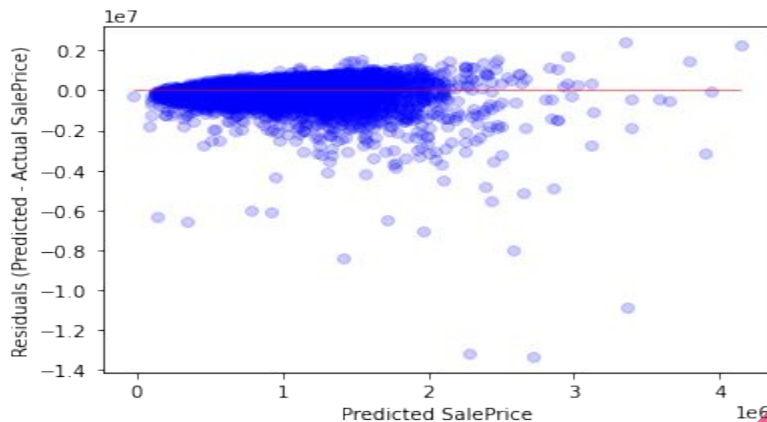
Methods

- By using CRISP-DM method, iterate the modeling process with feature engineering, model fit, evaluation, and repeating until the model proves to be useful to my business case



Baseline Model

- My first simple regression model tells that about 39.8% variance of Saleprice. I have used only one feature.
- p-value for the Intercept is 0.001, which is considered statistically significant(<0.05).
- Coefficient for SqFtTotLiving has a p-value of 0.000 which is less than (0.05) and it is also statistically significant.
- This model predicts a saleprice of -27,650 USD when square feet to living is 0 sqft.
- With an increase in 1 sqft, the saleprice is expected to increase by 382.8579 USD.



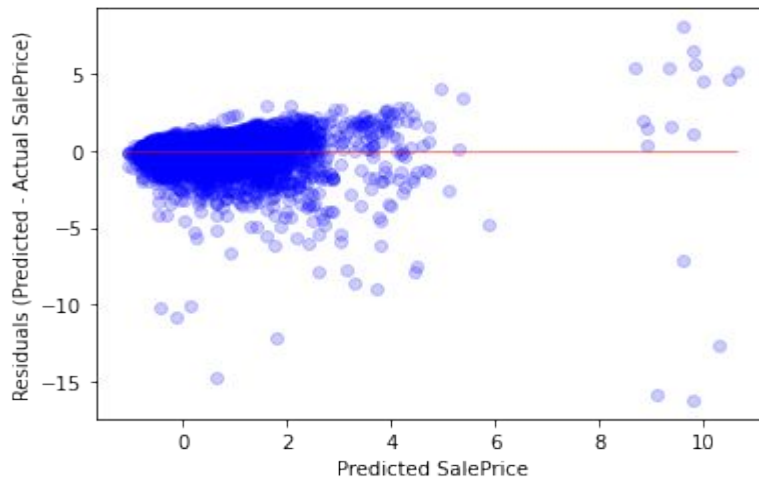
Transformation

- Log Transformed both target and predictors
- Dropped outliers from targeted variable
Saleprice
- Performed squareroot scaling
- Added categorical features



Final Model

- In the model 7 added categorical features to the predictor variables.
- Adding categorical features such as 'HeatSystem', 'HeatSource', 'Bedrooms', 'BldgGrade', Model 7 tells that about 53.2% variance of SalePrice.



Conclusion

- The most important factor in housing is the square footage of the house.
- After cleaning and modifying the given data, Model 7 has given the satisfactory result and considering it as the final model with 53.2% variance of Saleprice.





Thank You!

Email: mail4prabagini@gmail.com

GitHub: [@prabagini](#)