Housing Price Modeling

Presented by Waya Piyapanopas

About

In this project, I will be evaluating different housing features like "bathroom quality, kitchen size, and roof styles" to optimize a model which predict house prices. Through the qualitatively impact features the homeowners who wishes to sell their house will be able to identify which features of their house should or shouldn't be mentioned to their real estate agent to get the most value out of the pricing.

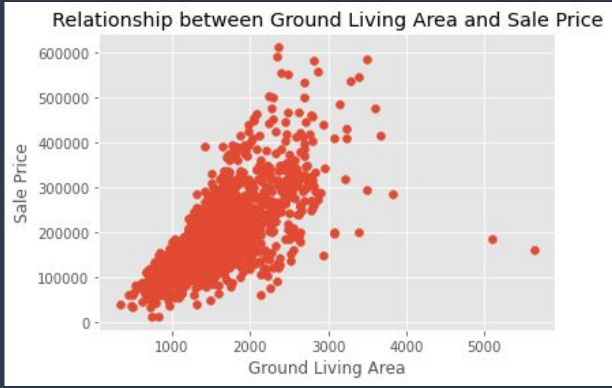
The Data that will be used to create models:

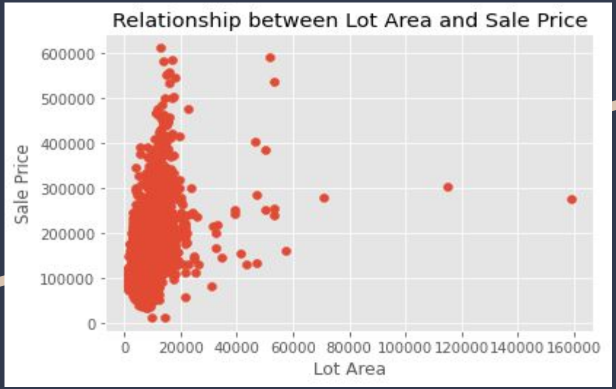
The data given consisted of 2 datasets with roughly 80 unique features that are both numerical and categorical. Only one of the 2 datasets contain the SalePrice column, which contains the price of the house. That dataset will be used as a model to predict the price of the house in the second dataset.

	Id	PID	MS SubClass	MS Zoning	Lot Frontage	Lot Area	Street	Alley
0	109	533352170	60	RL	NaN	13517	Pave	NaN
1	544	531379050	60	RL	43.0	11492	Pave	NaN
2	153	535304180	20	RL	68.0	7922	Pave	NaN
3	318	916386060	60	RL	73.0	9802	Pave	NaN
4	255	906425045	50	RL	82.0	14235	Pave	NaN

5 rows × 81 columns

Process:





The process will start off with assessing null values within the data, and outliers that may negatively impact the quality of our models. Creating additional features, such as age and total living space, that may increase the accuracy of our model's prediction LinearRegression, Lasso, Elastic, and Ridge models will be created and evaluated in this project.

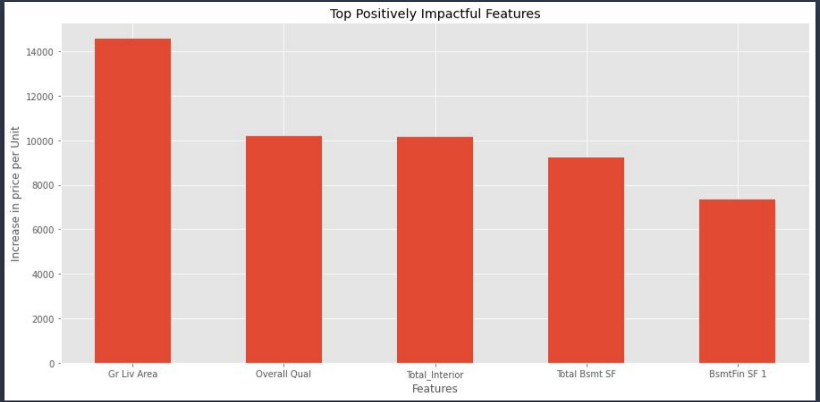
Top Features That Increases House Price:

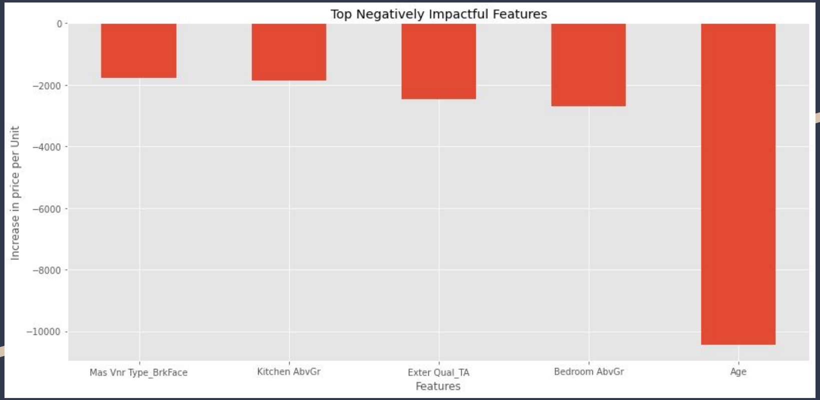
	Coef	Col
0	14544.715331	Gr Liv Area
2	10194.323699	Overall Qual
1	10159.767415	Total_Interior
3	9245.659186	Total Bsmt SF
5	7350.246985	BsmtFin SF 1

Top Features That Decreases House Price:

	Coef	Col
43	-1763.197288	Mas Vnr Type_BrkFace
30	-1868.051720	Kitchen AbvGr
23	-2462.269939	Exter Qual_TA
26	-2688.299831	Bedroom AbvGr
4	-10437.679207	Age

Using lasso and Elastic models, we can determine the quality of each features. Ones that are negatively impacting our models' quality will be zeroed out. Once the quality features were defined we can create an optimized model, in which did not contain too high or too low variance. Additionally, it perform decently well on unseen data.





Evaluation of the features/ Business Recommendation

The features that relates to sizing (size of the living space, lot, garage, etc.), quality and condition (in excellent quality), location (neighborhood), and age has the highest impact on the price of the house Age being the most negatively impacting the price of the house. With an increase in only 1 year of age, the price of the house would roughly drop by USD 10,000.

As a homeowner looking to improve the price of a house, making changes to the quality of the interior takes priority. For example, ensuring that the quality of the kitchen is excellent will roughly drive the price of your house by USD 5600.

Location:

The NeighborHoods that increase the price of the house ordering from highest increase to lowest increase:

Neighborhood_NridgHt,

Neighborhood_StoneBr,

Neighborhood_NoRidge,

Neighborhood_Somerst,

Neighborhood_Crawfor,

and Neighborhood_NWAmes (This neighborhood has negative impact to price)

Things that should be taken into consideration

Because Neighborhood helps a lot with the quality of my model, the model is not generalize universally. However, this model would perform okay without the Neighboorhood features on other cities. At the base, the price of houses would start at 181, 350 Most Impactful features (in order): Ground Living Area, Age, Overall Qual, Total Interior space, Total Basement space, Overall Condition.

Conclusion

This Ridge model that was created had decent accuracy on unseen data. The features that positively influence the price of the house would include any feature that deals with area, quality, and location. In the other hand, low quality/condition of any certain features will pull the price of the house down. High age being the most negatively impactful.

THANK YOU FOR YOUR ATTENTION