AMES REAL ESTATE

Ames, IA real estate data analysis

Presented by Cristina Sahoo

Problem Statement

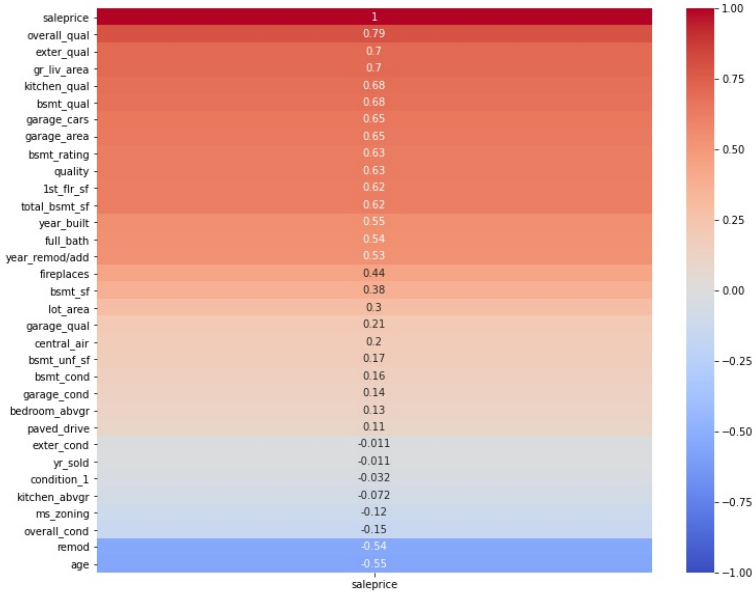
The Ames housing dataset contains detailed information about homes that were sold between 2006 and 2010, along with their sale price. Our real estate agents want to be able to advise their selling clients on the best sale price for their home. This project aims to analyze the Ames housing dataset and create a model that predicts the sale price of a home.

Data source:
https://www.kaggle.com/prevek18/ames-housing-dataset

Why is this important?

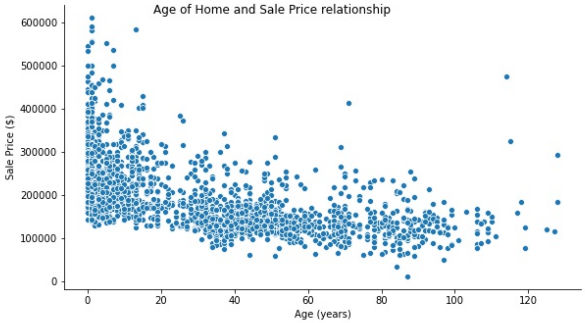
As was shown by the 2008 financial crisis, it is essential for the economy that the amount of a bank's real estate lending matches the actual value of the real estate. Getting a good estimate of the price of a house is hard even for the most seasoned real estate agents.

Select appropriate features to predict the sale price of a home



A positive correlation is a relationship between two variables in which both variables move in the same direction.

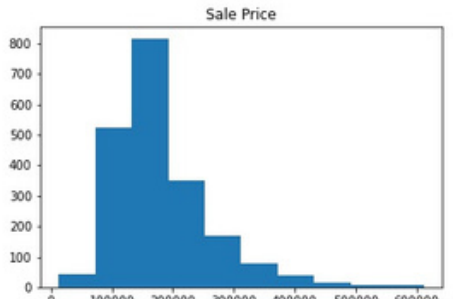
A negative correlation is a relationship between two variables in which one variable increases as the other decreases, and vice versa



Normalize the target variable

The target (saleprice) is right/positive skewed.

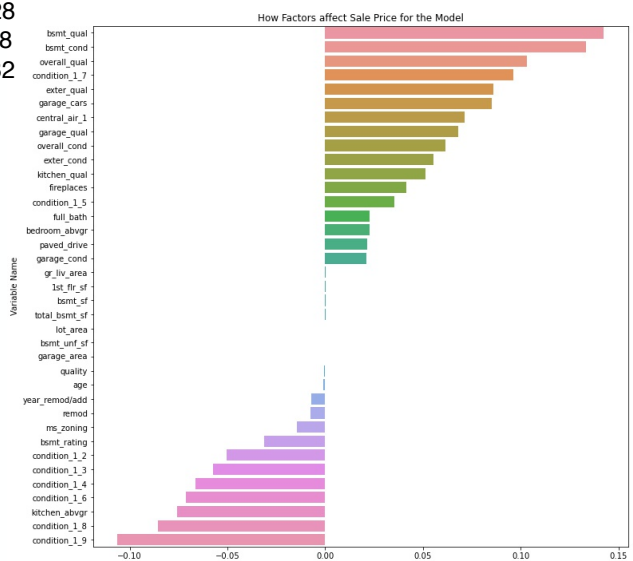
By applying the natural log to the values of the saleprice column, we get a distribution that is closer to normal. By doing this, we can improve our model.



Evaluate the model

Train score 0.8552046088188228
Test score 0.8879463265423998
Baseline 0.8060271761299932

16	kitchen_qual	0.051063
17	fireplaces	0.041384
32	condition_1_5	0.035164
13	full_bath	0.022798
14	bedroom_abvgr	0.022689
22	paved_drive	0.021415
21	garage_cond	0.020858
12	gr_liv_area	0.000149
11	1st_flr_sf	0.000059
27	bsmt_sf	0.000042
10	total_bsmt_sf	0.000022
1	lot_area	0.000003
9	bsmt_unf_sf	-0.000020
19	garage_area	-0.000035
24	quality	-0.000426
23	age	-0.001200
4	year_remod/add	-0.007287



Conclusions

For 1 point increase in the quality of the kitchen, holding all other features constant, we can expect, on average, a 5% increase in the price of a home.

If the home has a fireplace, the increase in the price is estimated at about 4%.

For 1 year increase in the age of a home, we can expect a decrease of about 0.12% in the price.

Next steps:

Further evaluate and tune the model for a more accurate prediction. Consider adding/engineering more features. Additional research is needed into specific neighborhoods and other factors that influenced home prices, such as the market crash in 2008.

Resources:

https://www.opendoor.com/w/blog/factors-that-influence-home-value

https://peltarion.com/knowledge-center/documentation/tutorials/predict-real-estate-prices

https://www.kaggle.com/jepsds/feature-selection-using-selectkbest?utm_campaign=News&utm_medium=Community&utm_source=DataCamp.com

Questions...