

House Price Data Analysis

by James Yigit Kasapoglu

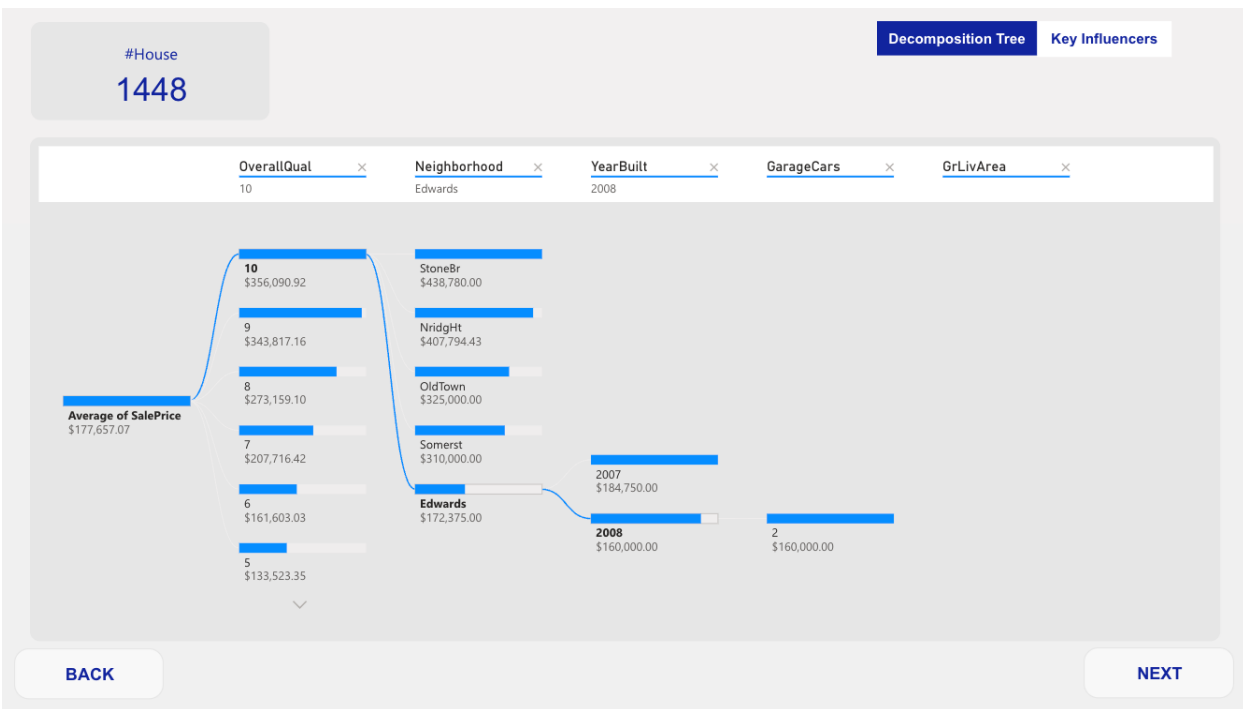
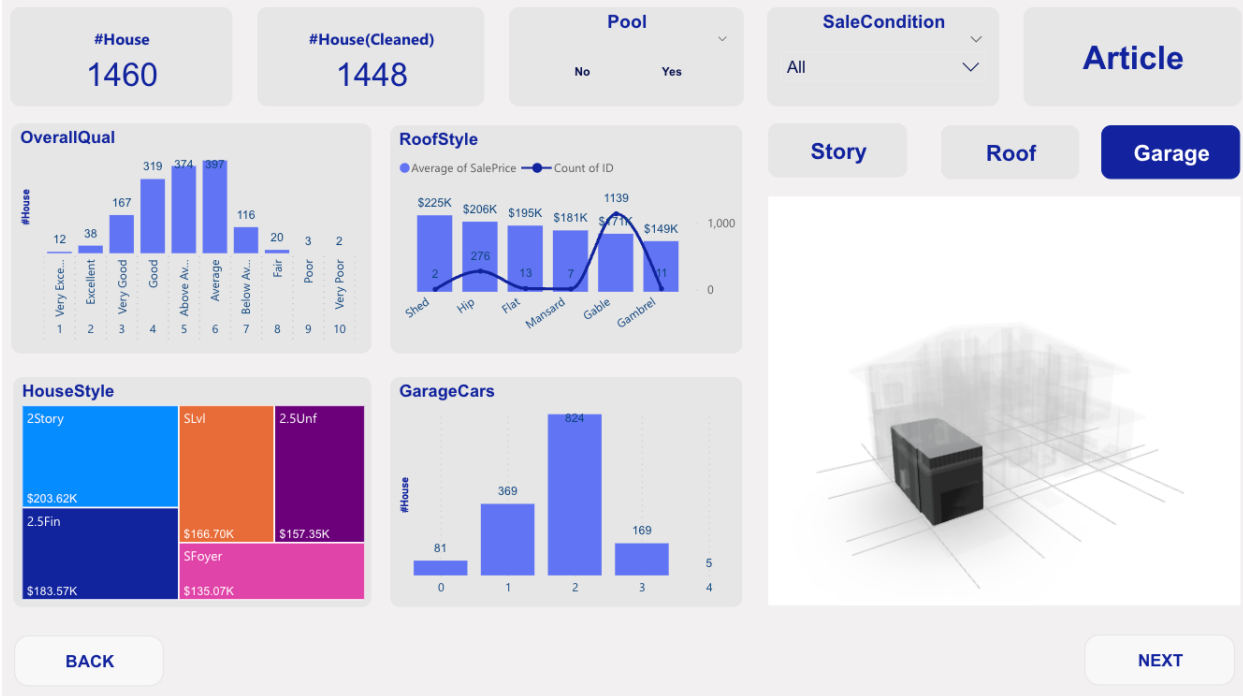
[Kaggle](#)

[Linkedin](#)

NEXT

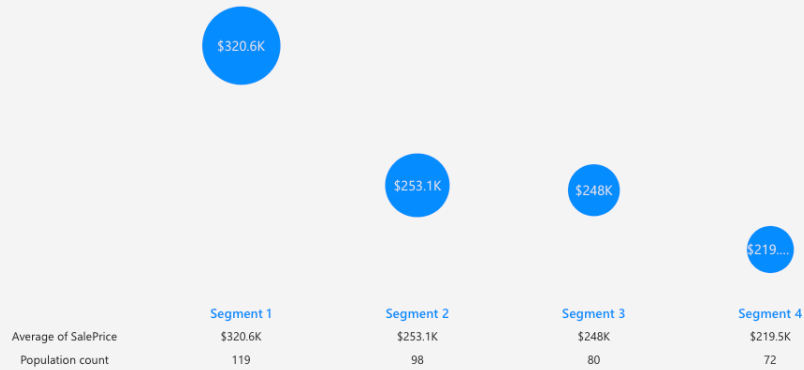






Key influencers Top segmentsWhen is SalePrice more likely to be ?

We found 4 segments and ranked them by Average of SalePrice and population size. Select a segment to see more details.



BACK

NEXT

Feature Engineering

Neighborhoodscore: This feature represents the popularity or density of each neighborhood.

HouseAge: Calculated by subtracting the year of construction from the year of building of the house.

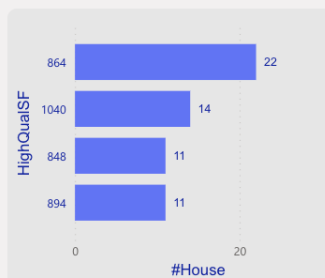
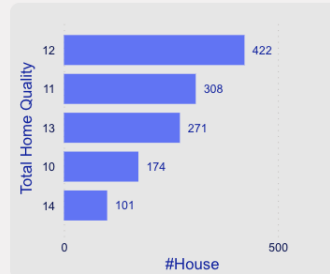
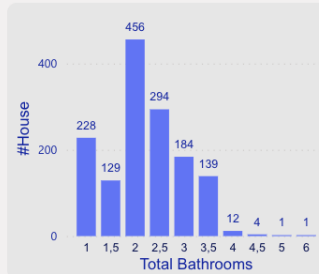
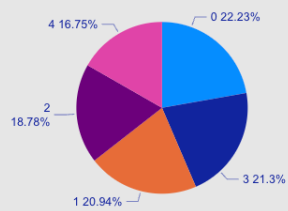
SqFtPerRoom: This property represents the ratio of the home's overall living area to the number of rooms.

Total_Home_Quality: This feature aims to better reflect the market value of the home by presenting the overall quality and state of maintenance of the home.

Total_Bathrooms: This feature provides this information by calculating the total of full and half bathrooms.

HighQualSF: This feature represents the totality of the home's high-quality living space (the sum of the first floor and second floor areas).

HouseAge



SqFtPerRoom

156.86

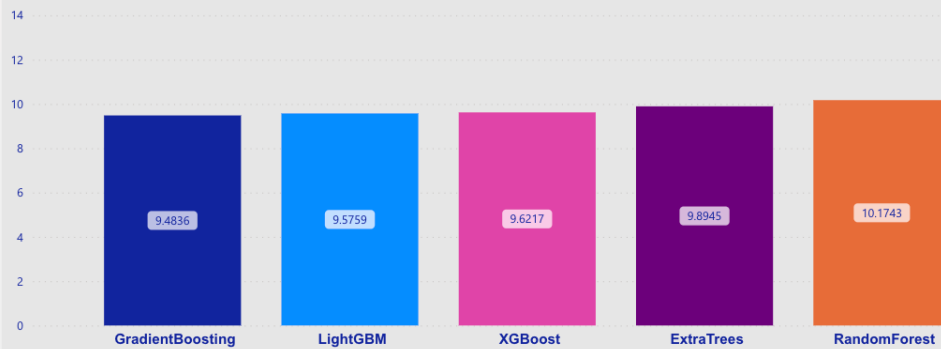
14,57 M^2

BACK

NEXT

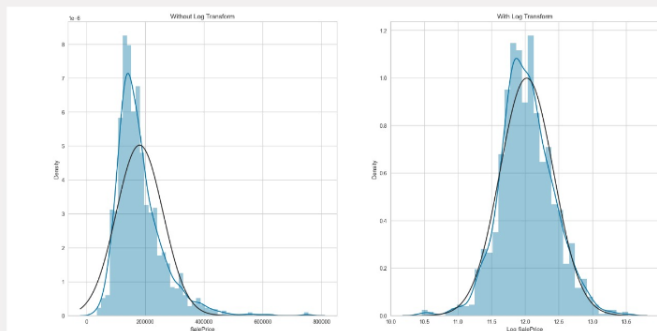
MAPE

Sort 1 ● 2 ● 3 ● 4 ● 5



BACK

NEXT



Before

After

Feature	Skew	Absolute Skew	Skewed
0 LotFrontage	2.019251	2.019251	True
1 LotArea	12.195142	12.195142	True
2 OverallQual	0.216721	0.216721	False
3 OverallCond	0.692355	0.692355	True
4 YearBuilt	-0.612831	0.612831	True
5 YearRemodAdd	-0.503044	0.503044	True

Feature	Skew	Absolute Skew	Skewed
0 LotFrontage	-0.789832	0.789832	True
1 LotArea	-0.137263	0.137263	False
2 OverallQual	0.216721	0.216721	False
3 OverallCond	-0.253754	0.253754	False
4 YearBuilt	-0.640470	0.640470	True
5 YearRemodAdd	-0.510480	0.510480	True

BACK

NEXT

What changed?

- 32 skewed columns reduced to 19
- Logarithmic Transformation implemented to target and features
- DataFrame encoded with One Hot Encoder
- Implemented standardization with Scikit-learn's StandardScaler class
- ML models compared with PyCarat's compare_models
- Implemented KFold cross validation
- Final prediction made with Bayesian Ridge

