Ice Breaker!

Ice Breaker!

DSIF4 Project 2

Ames Housing Data and Kaggle Challenge

Andre
Jasmeet
Wei Han

DSIF4 Project 2

Ames Housing Data and Kaggle Challenge

Andre
Jasmeet
Wei Han

Contents

Introduction & Problem Statement

- Buying / selling a property can be a daunting process
- Home owners / new buyers may feel overwhelmed by the numerous features within a property, not sure which features carry more weight in the pricing

Problem Statement:

- Based on the given dataset, determine the top 10 housing features that have the most impact on Sale Price in Iowa.
- Stakeholders:
    - Primary stakeholders: Home owners, property agents, and developers who are looking to buy/sell/develop property in Iowa.
    - Secondary stakeholders: Government body in charge of land use planning can use this model to decide what type of property to designate to a certain location type.

Dataset Overview

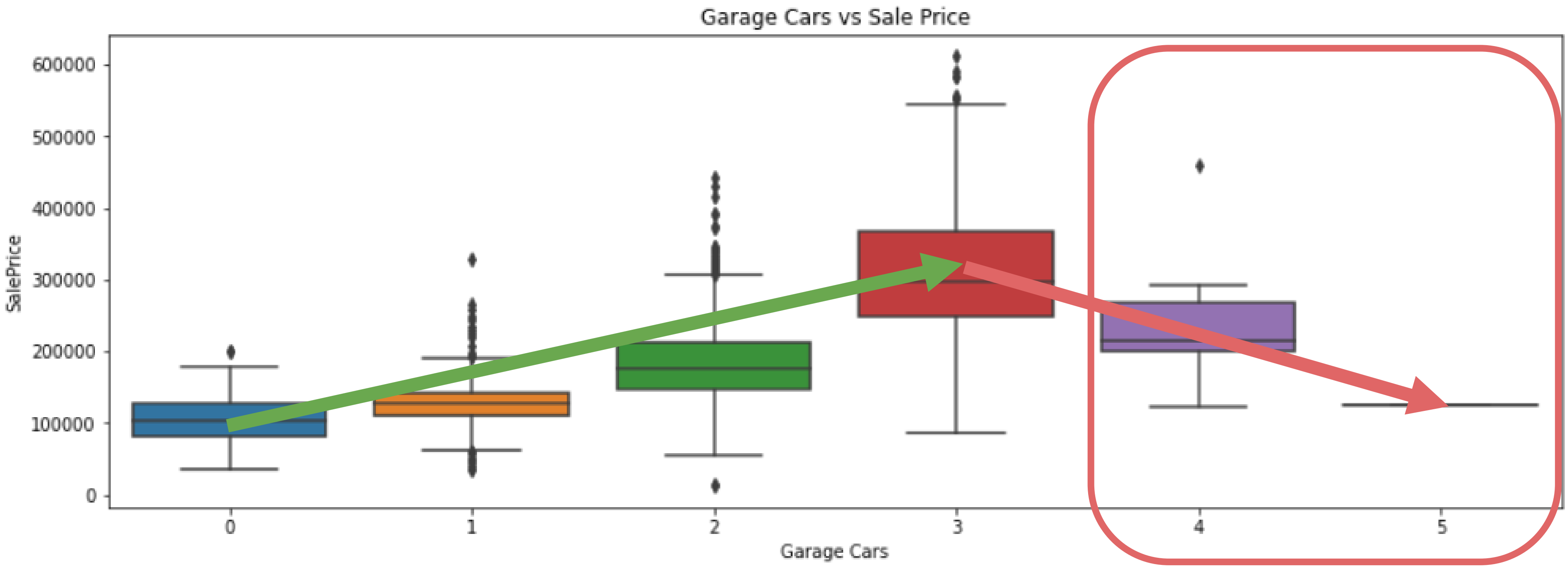
- From Ames, Iowa, USA
- Provides various aspects of housing data for residential properties sold from 2006 to 2010

Dataset Overview

- 2051 observations with 82 features
- Features include
  - Square footage
  - Condition
  - Built Quality
  - Pool
  - Materials
  - Surrounding Environment
  - …

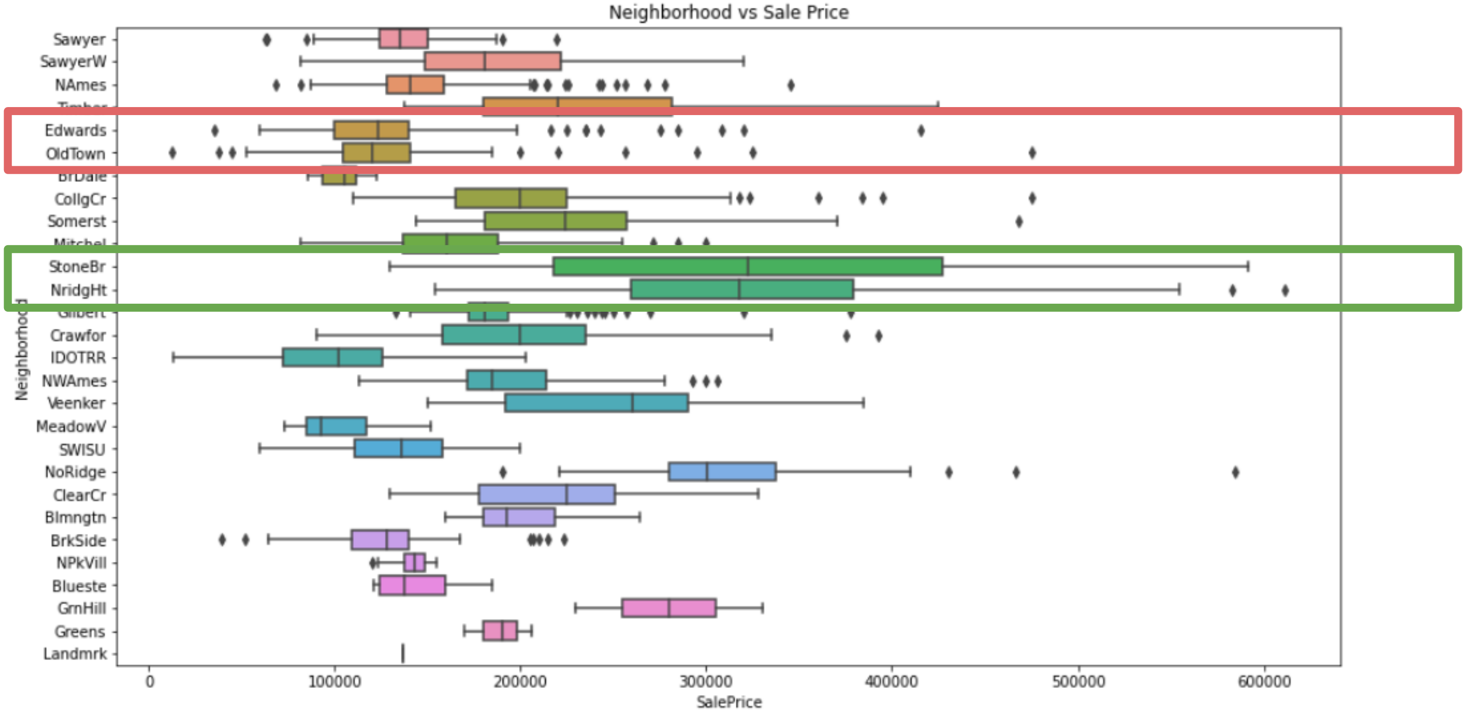
Feature Selection

More doesn’t always mean better!



Feature Selection

Certain neighborhoods have higher/lower prices.



Feature Selection

Features dropped

1	Has Pool	13	Exterior 2nd
2	Pool Area	14	Exterior 1st
3	Low Qual Fin SF	15	BsmtFin SF 2
4	Garage Cars	16	3Ssn Porch
5	Bsmt Unf SF	17	MS SubClass
6	Has Screen Porch	18	Bldg Type
7	Bsmt Half Bath	19	Lot Config
8	Yr Sold	20	Has 2nd Flr
9	Year Built	21	Mo Sold
10	Has 3Ssn Porch	22	Pool QC
11	Misc Val	23	Lot Shape
12	Alley	24	TotRms AbvGrd
13	Exterior 2nd	25	Roof Style
14	Exterior 1st	26	Fence
15	BsmtFin SF 2	27	Year Remod/Add

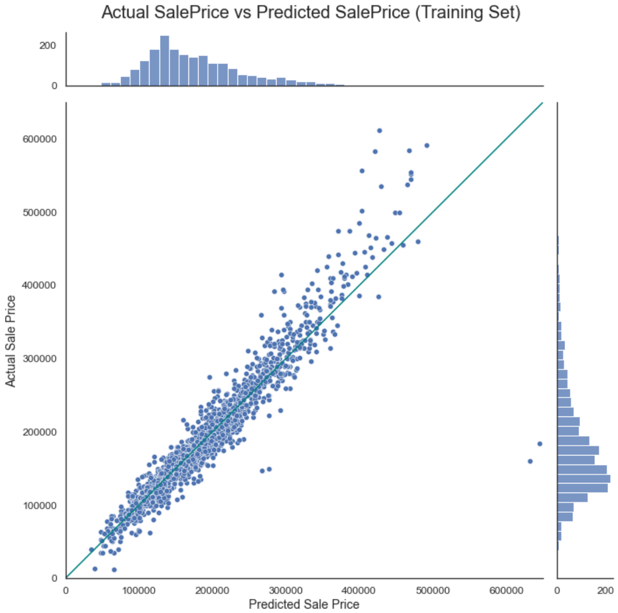
WHY?

- Too little correlation to sale price
- Too little variations between classes
- Too many of the same value
- Similar features that better explains price

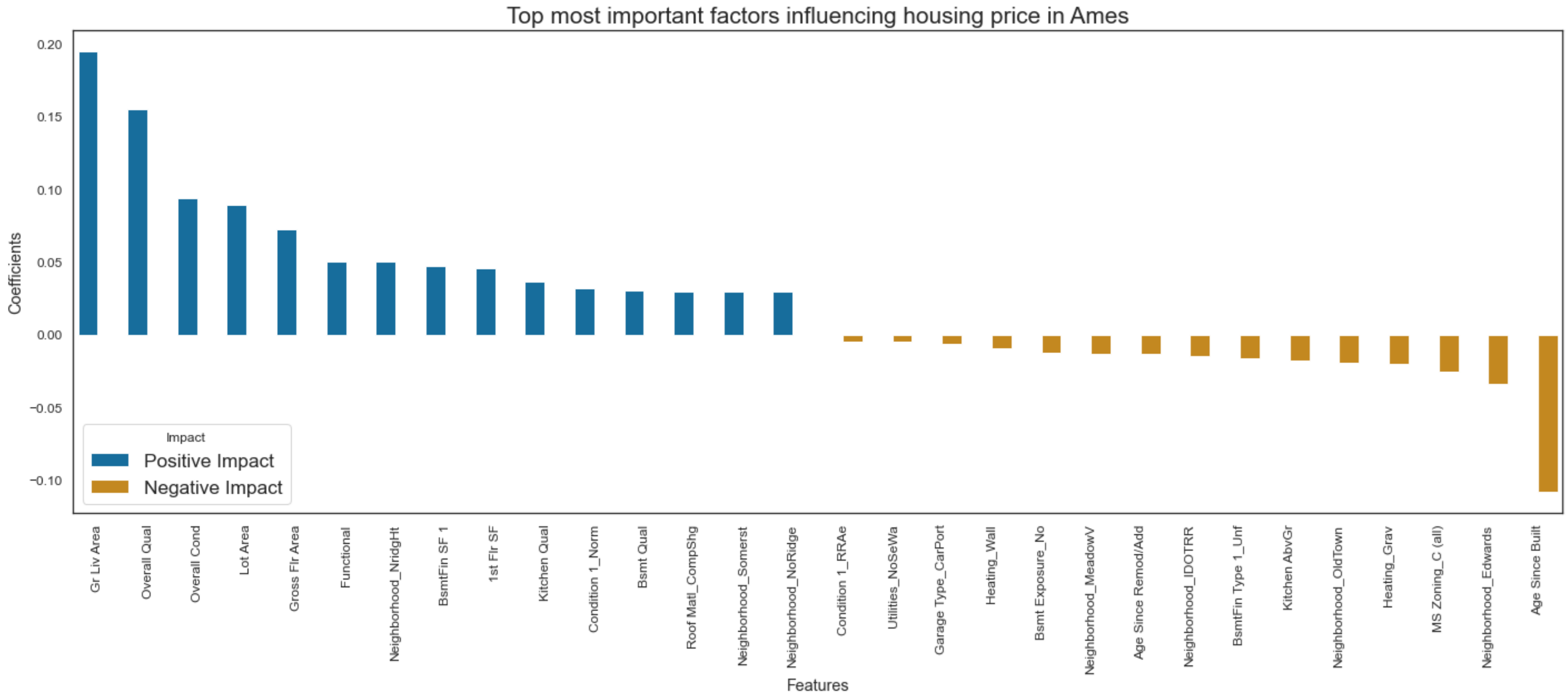
Model Evaluation

Adjusted R-square = 0.916

Our model explains 91.6% of variability in Sale Price

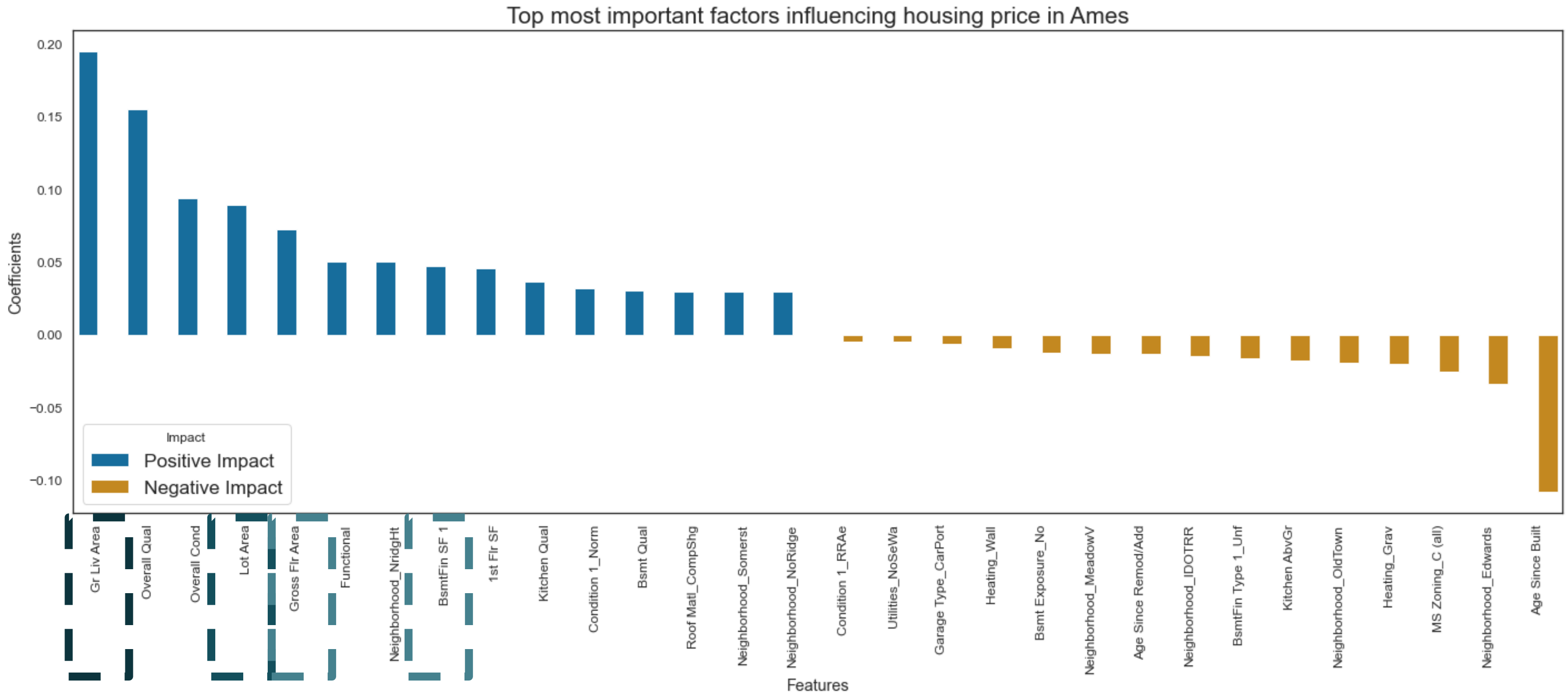


Conclusion & Recommendation



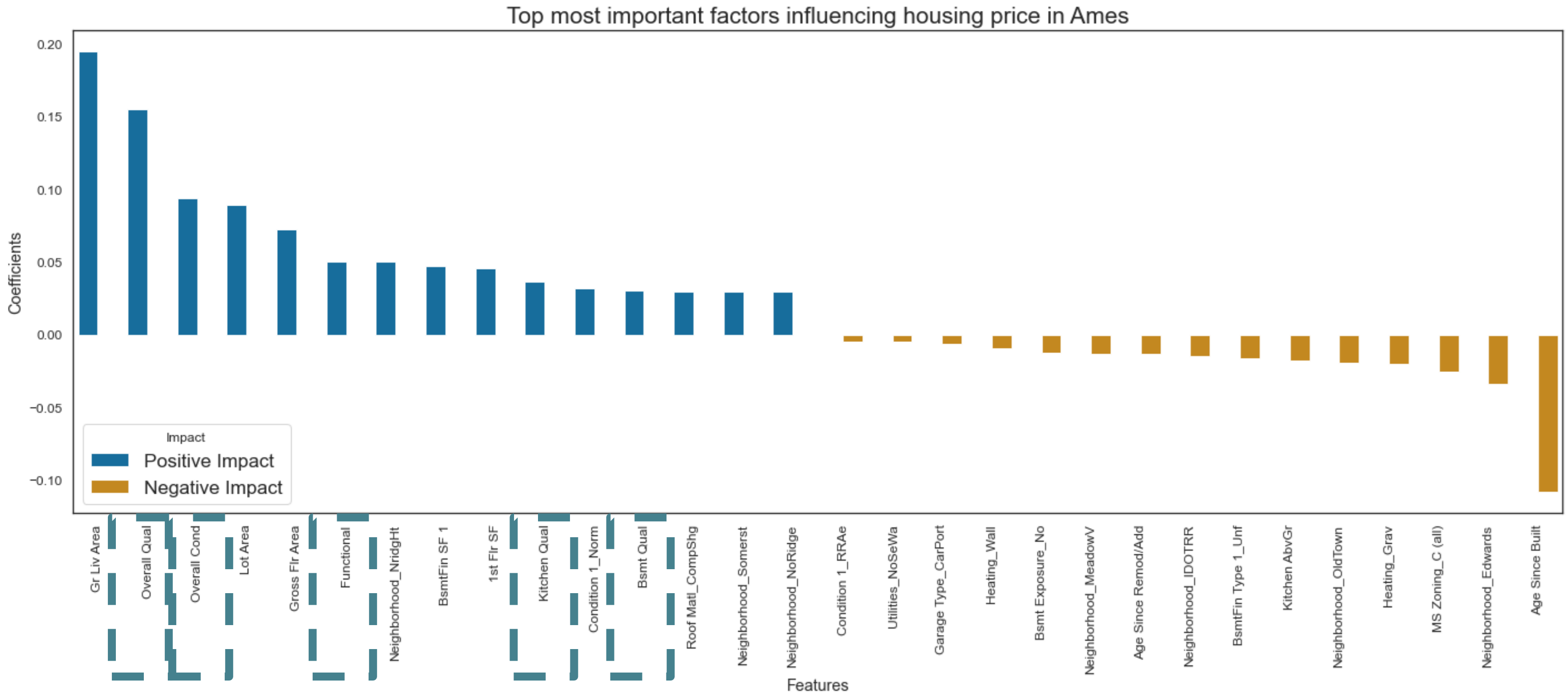
Conclusion & Recommendation

Above grade Area > Lot Area > Gross Floor Area > Basement Area



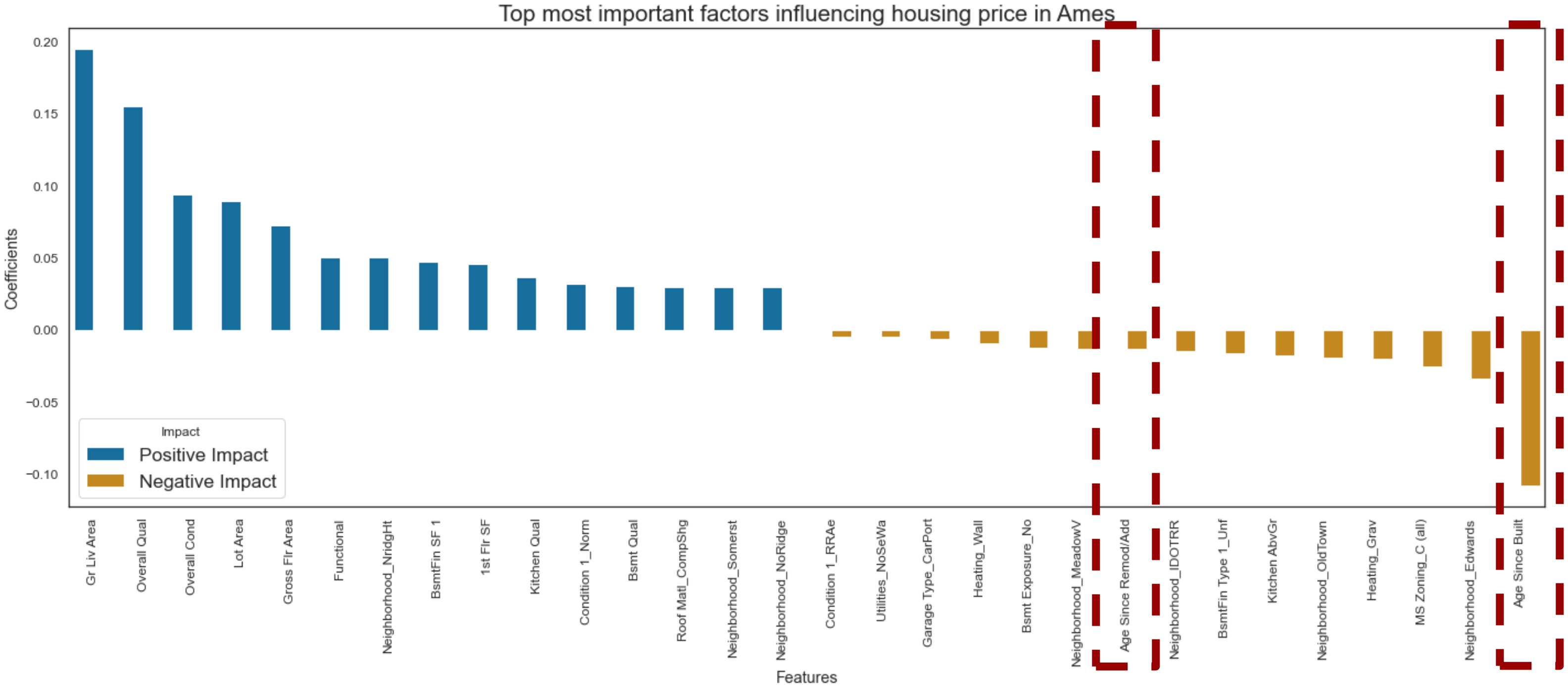
Conclusion & Recommendation

Hire a professional to access the housing quality!



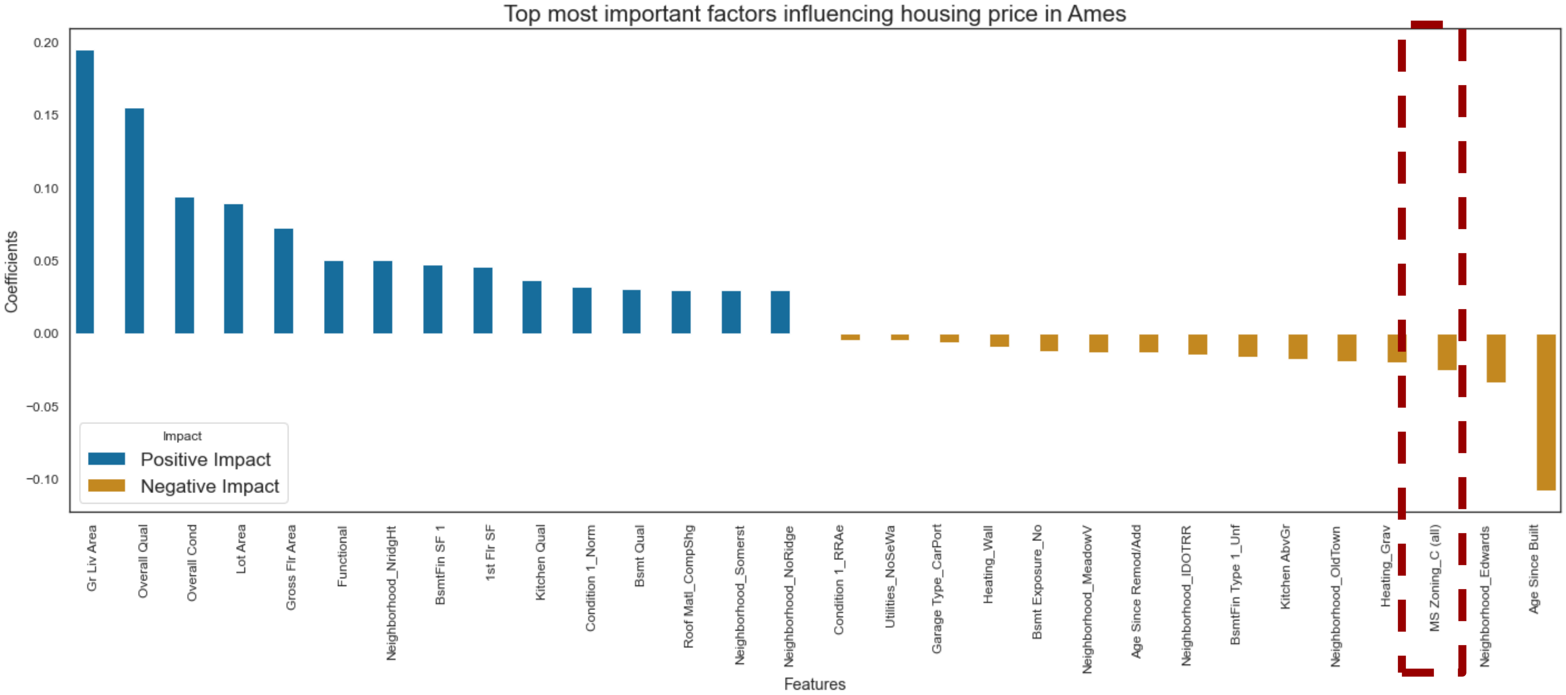
Conclusion & Recommendation

Try not to choose old houses unless you want to remodel it



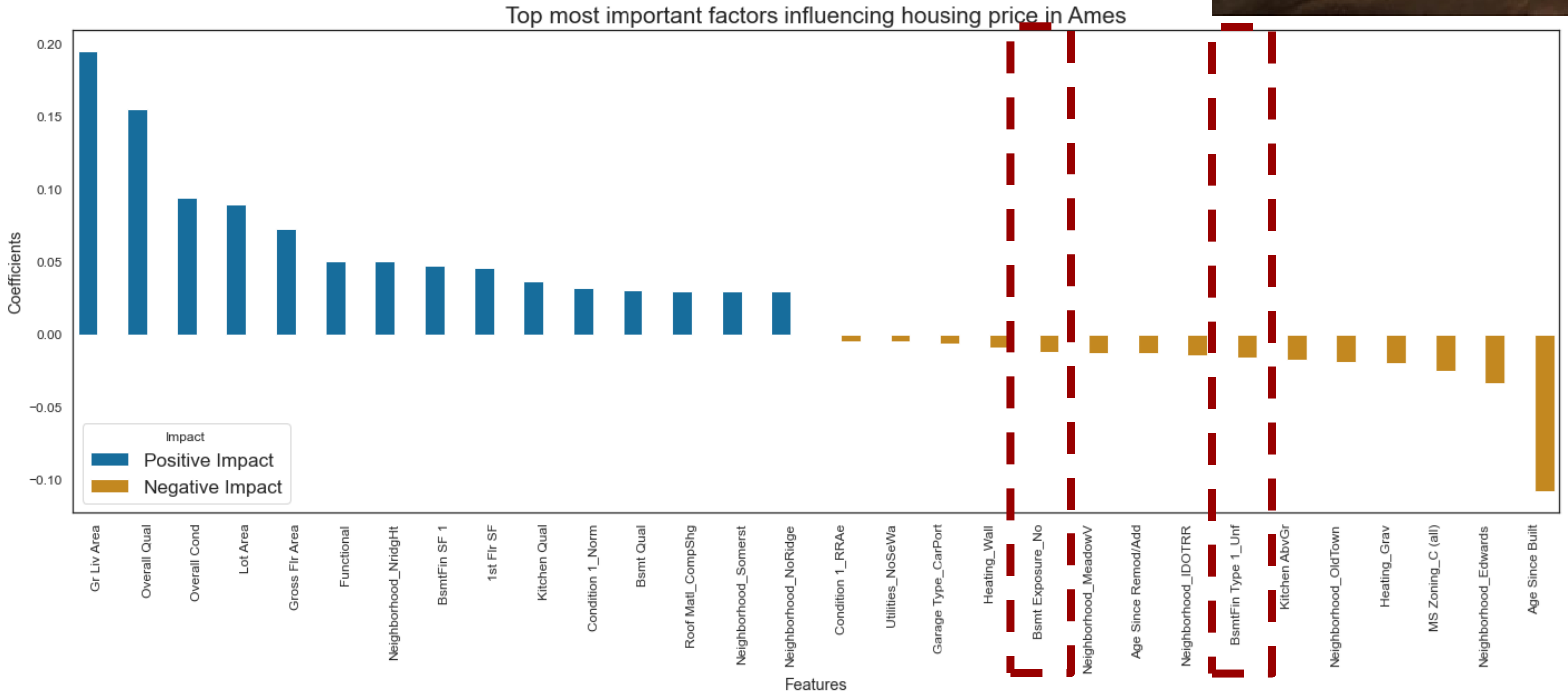
Conclusion & Recommendation

Avoid commercial zones



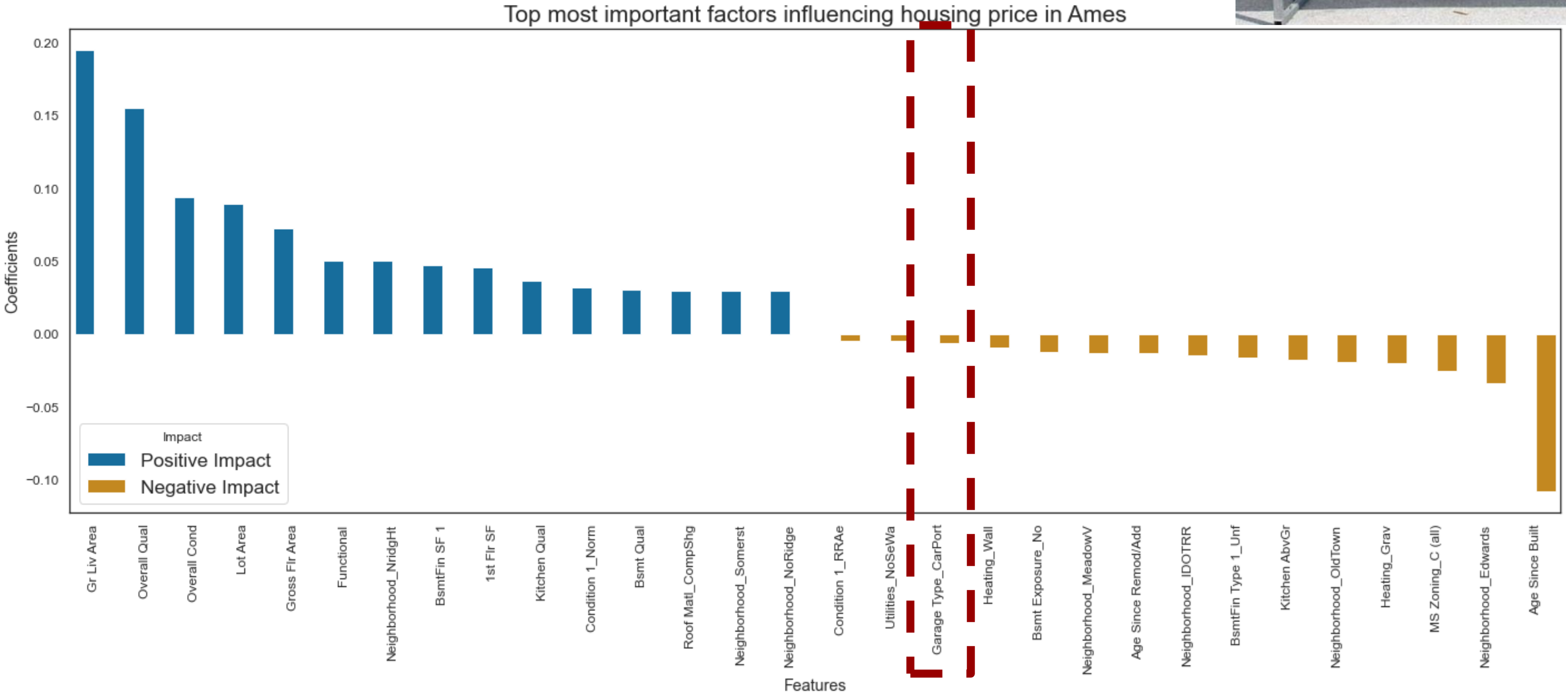
Conclusion & Recommendation

Some basement characteristics to avoid.



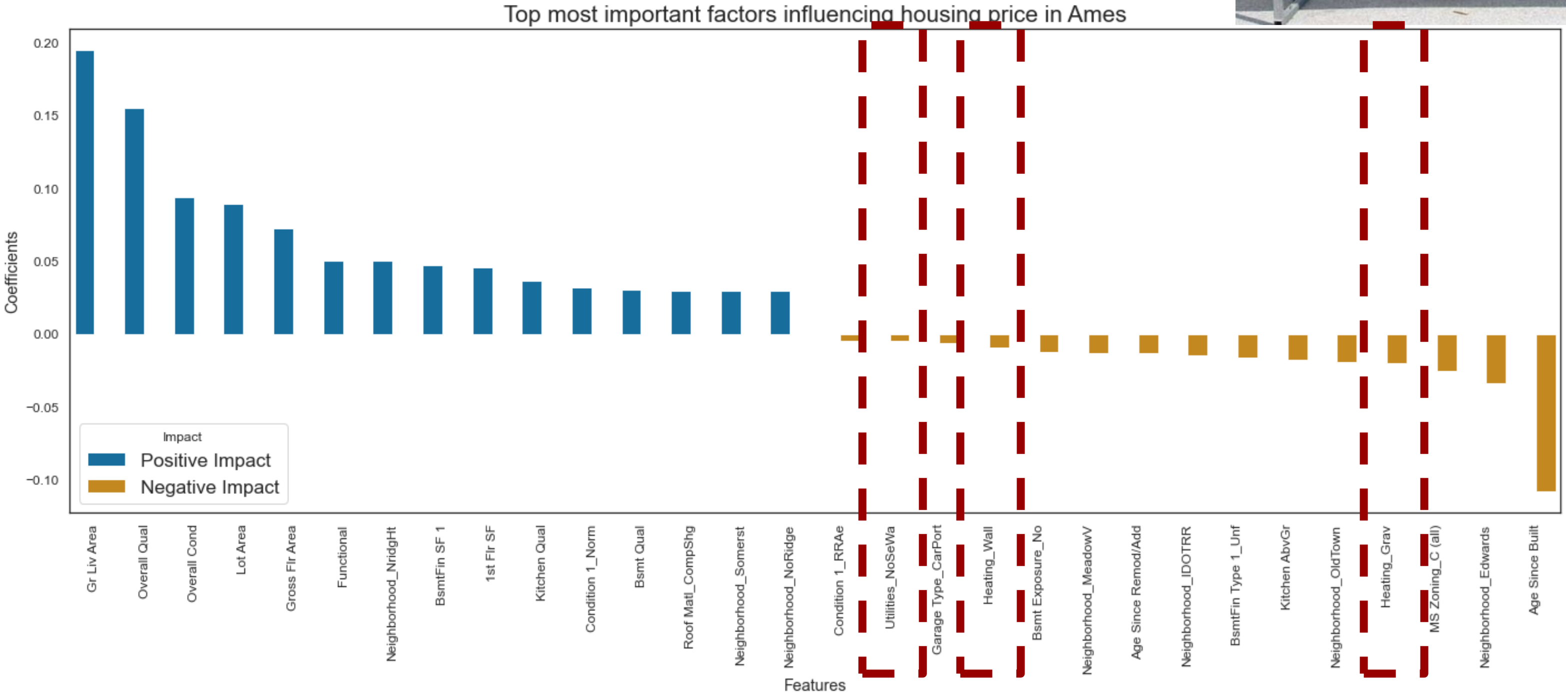
Conclusion & Recommendation

Avoid carport garage.

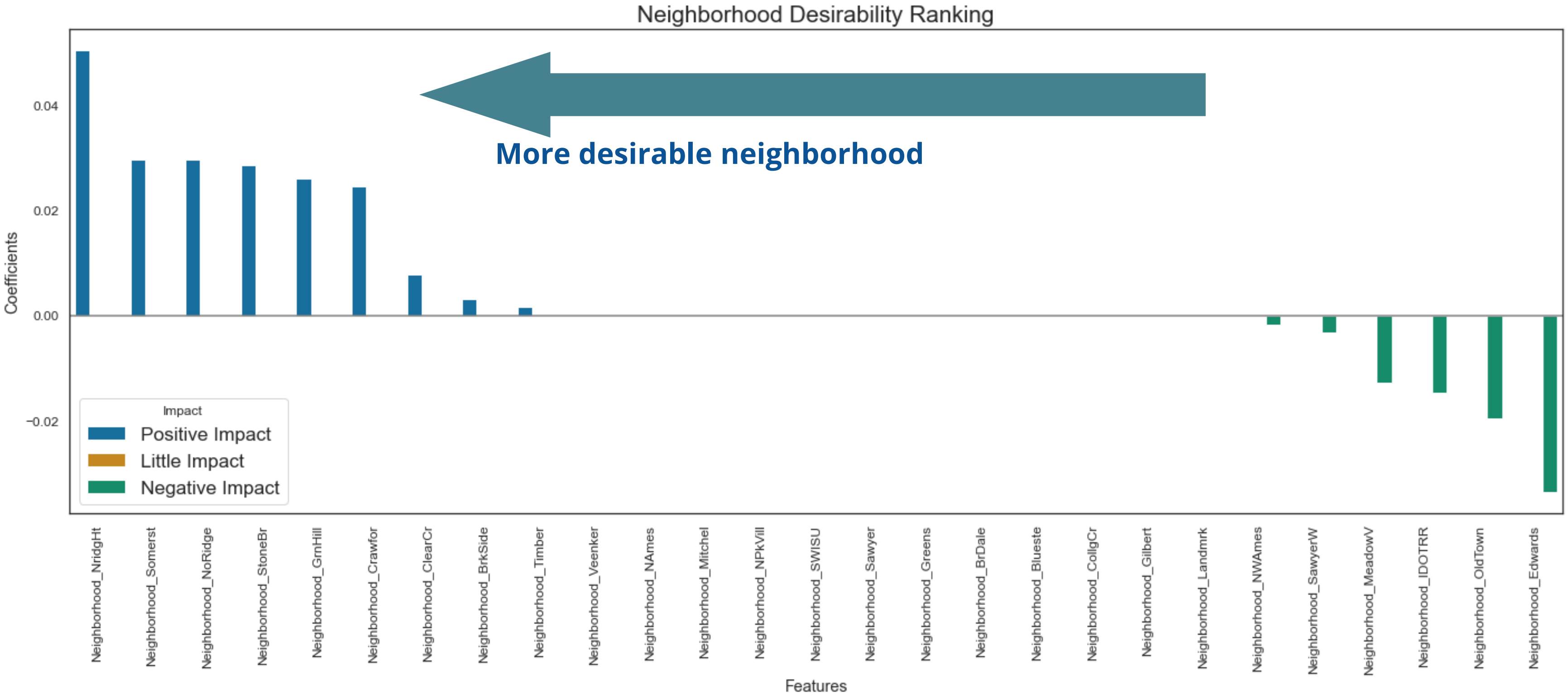


Conclusion & Recommendation

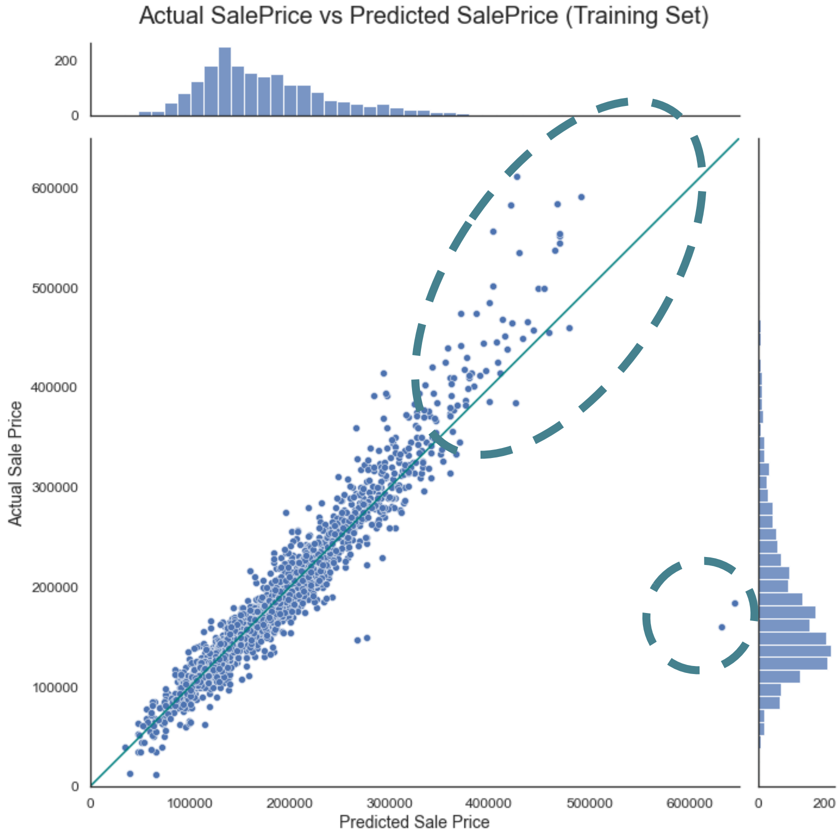
Pay attention to utilities and heating.



Conclusion & Recommendation



Limitations



Tends to "Underpredict" for high value properties.

2 properties are far off the prediction

Future Steps

- High networth properties would require a separate dataset with more entries
- Collect more features which would also affect the Sale Price of a property
    - proximity to amenities
    - transport networks
    - network coverage, etc
- Missing data could also be made compulsory to answer, and also have an algorithm to determine the relevance of the given answer.

Q&A