

A low-angle, black and white photograph of several modern skyscrapers reaching towards a cloudy sky. The perspective creates a sense of height and scale. A large blue rectangular overlay is positioned on the left side of the image, containing white text. White geometric lines, including a right-angled triangle and a diagonal line, are visible in the top-left and bottom-right corners of the image.

# Real Estate price prediction

147\_Pun\_Pitipat



# TABLE OF CONTENTS

## **1** Goal and objective

Setting goal and objective are mandatory

## **2** Data analysis

Base on sale prices data.

## **3** Data analysis

Base on neighborhood data

## **4** Result & Recommendation

Result of randomforest and tips

# Target is to predict house price by location

## Objective

Using data science techniques to understand and forecast 'House price prediction' dataset

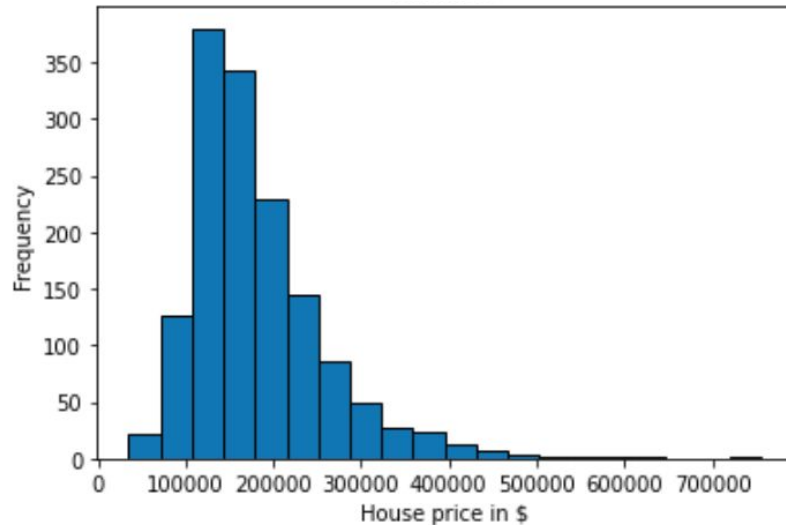
## Goal

Create model to predict the house prices

# Data analysis

## Initial analysis on sale price side

Distribution of the house price



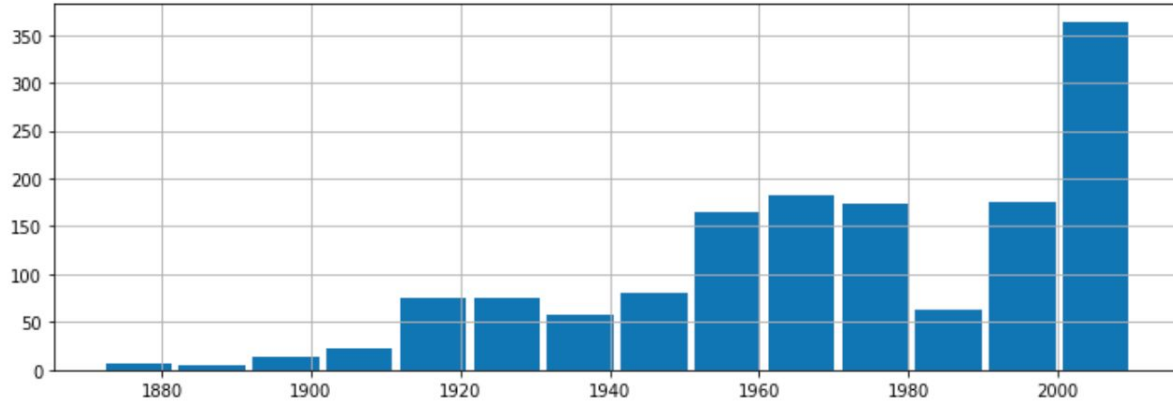
For the range of house price The average sales price is \$180,921, while median is \$163,000.

With the shape of positive skew. Where the majority of house are place between \$100,000 - \$200,000.

# Data analysis

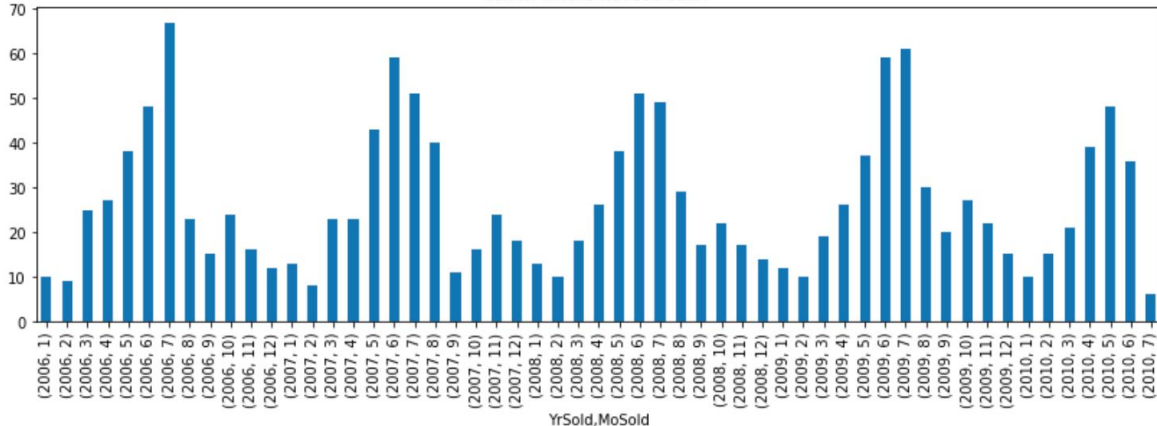
## Initial analysis on sale price side

When were the houses built?



Using 'YearBuild' feature to find out that majority of house are build on 20s.

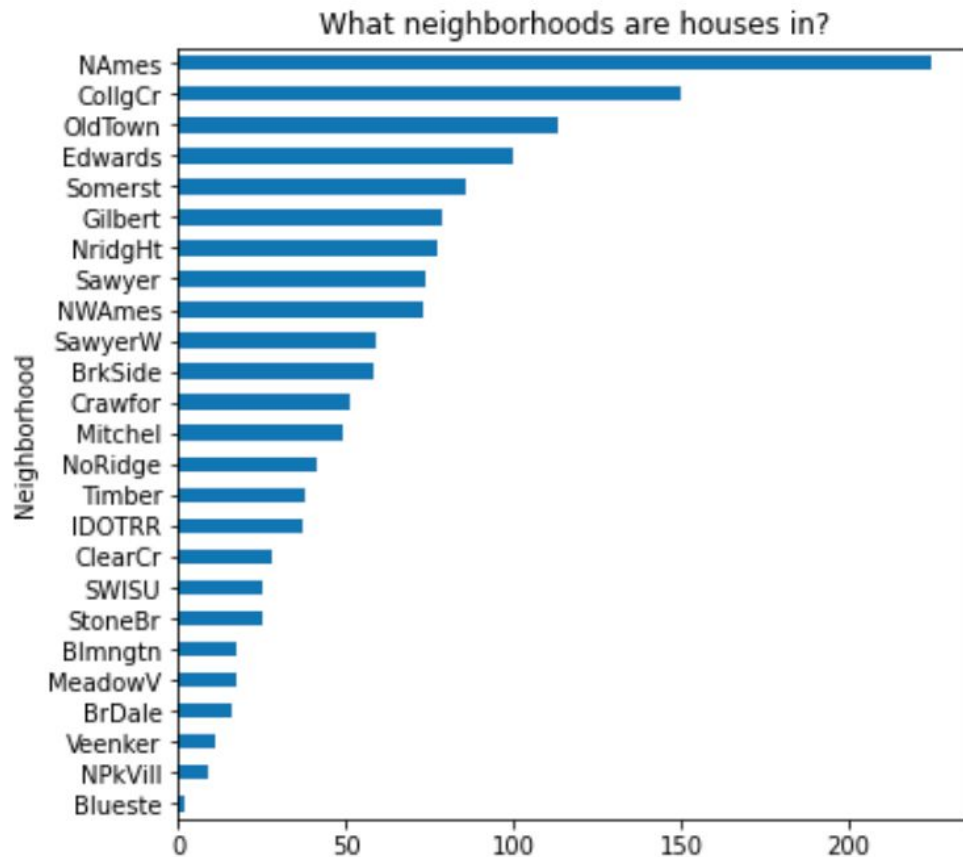
When where houses sold?



Group 'Year sold' and 'Month sold' data to see trend of house sold. Peak time during June-July and drowse on early of the year.

As we are interesting in predicting sales price with location, we then focus on location side.

# Data analysis

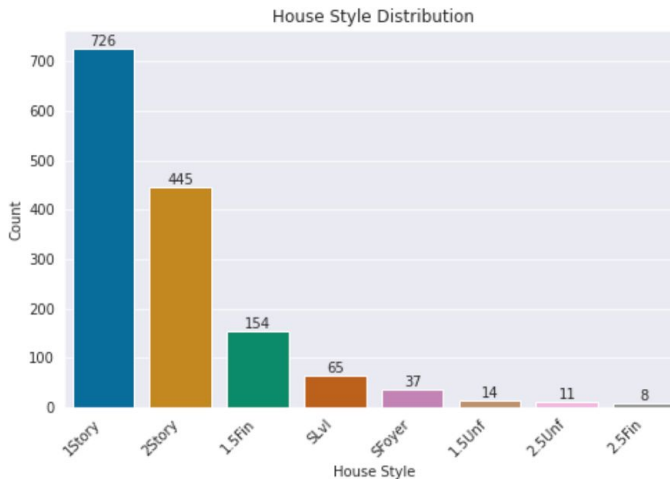
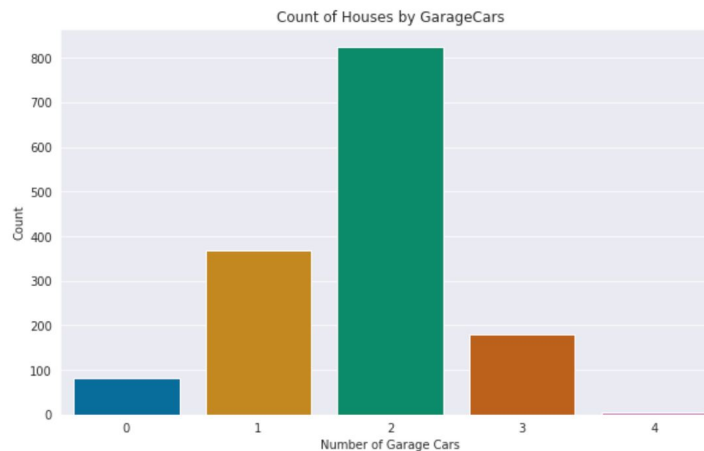
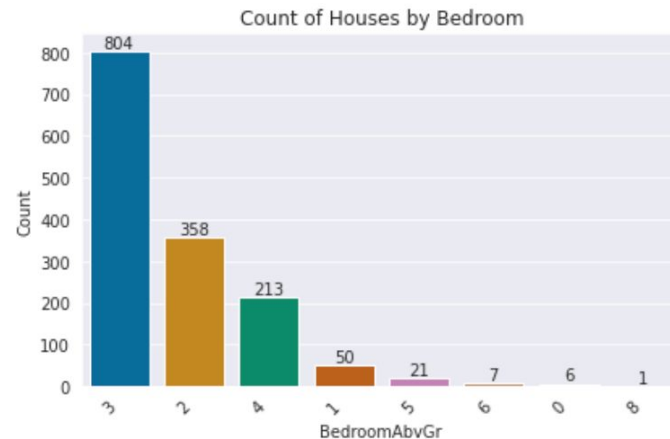
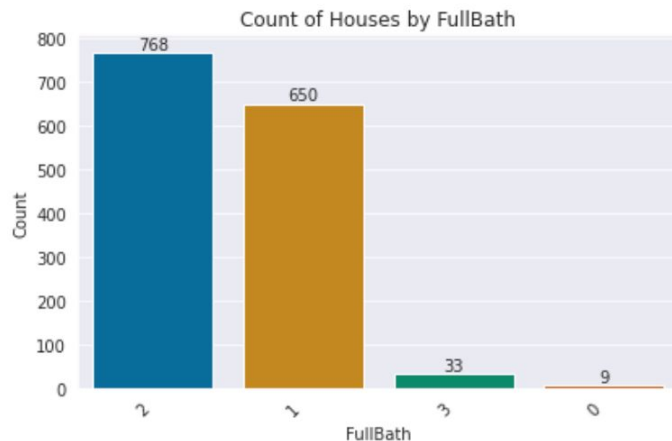


Where the top 5 neighborhoods are NAmes, CollgCr, OldTown, Edwards and Somerst.

Where NAmes have the highest household number among all neighborhood with over 200 house. And approximately 150 household at CollgCr.

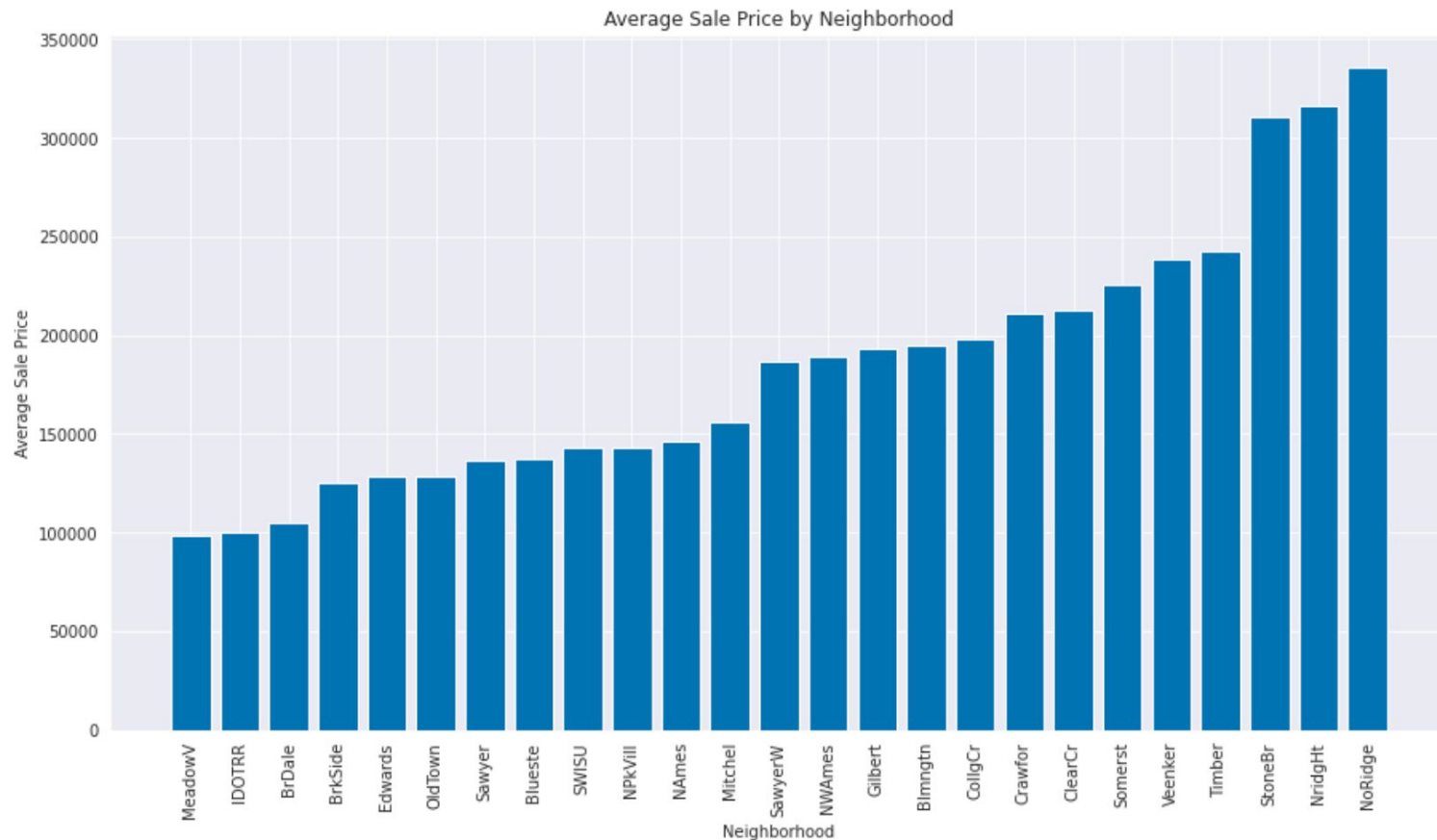
As we are interesting in predicting sales price with location, we then focus on location side.

# Data analysis



Now we plot a graph of Sales price and neighborhood

# Data analysis





# Result & Recommendation

## Result

- As a result RMSE(Root mean squared error): 0.321

## Trend

**Sale Price:** 100,000 - 200,000 \$

**Neighborhood:** NAmes, CollgCr, OldTown, Sawyer, Edwards, Gilbert, NWAmes

**House Style:** 1 story

**Full Bath:** 1 Full bath

**Bedroom:** 3 Bedrooms

**Garsage:** 2 cars

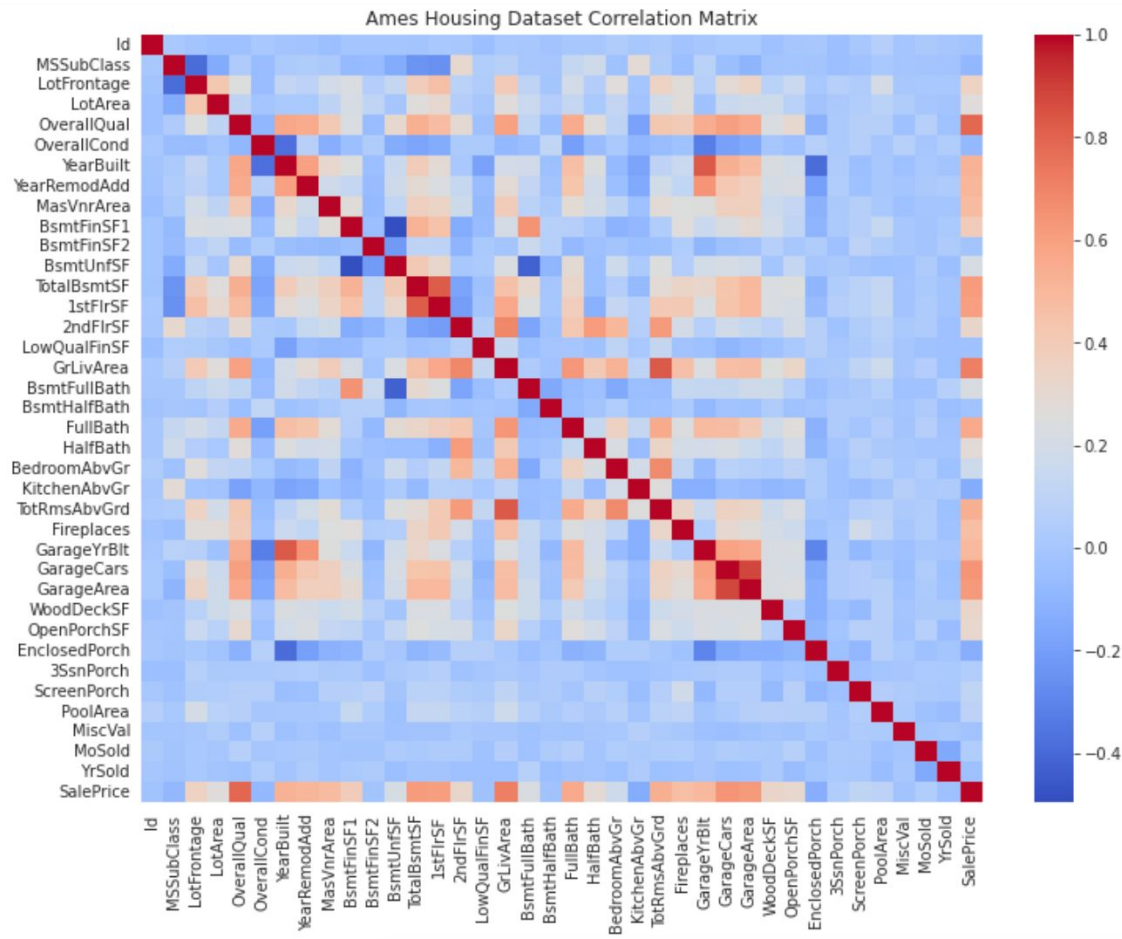
## Tips:

- For an investor or any buyer, other factors also important:
  - Time of house sold usually drop on every 2nd quarter of the year (Buyer might get a deal off)
  - Overall quality and facilities provide are also very potential when buying or even selling



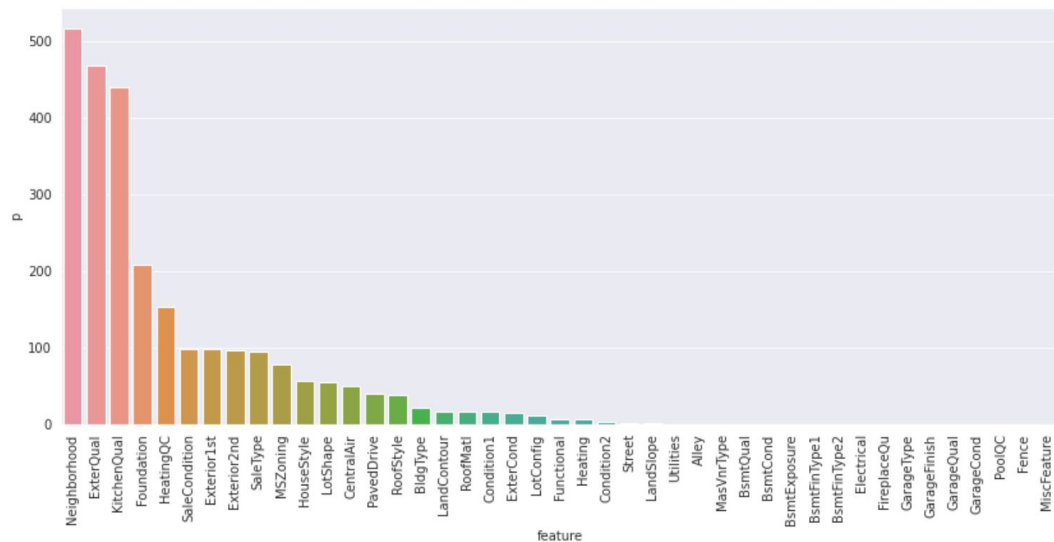
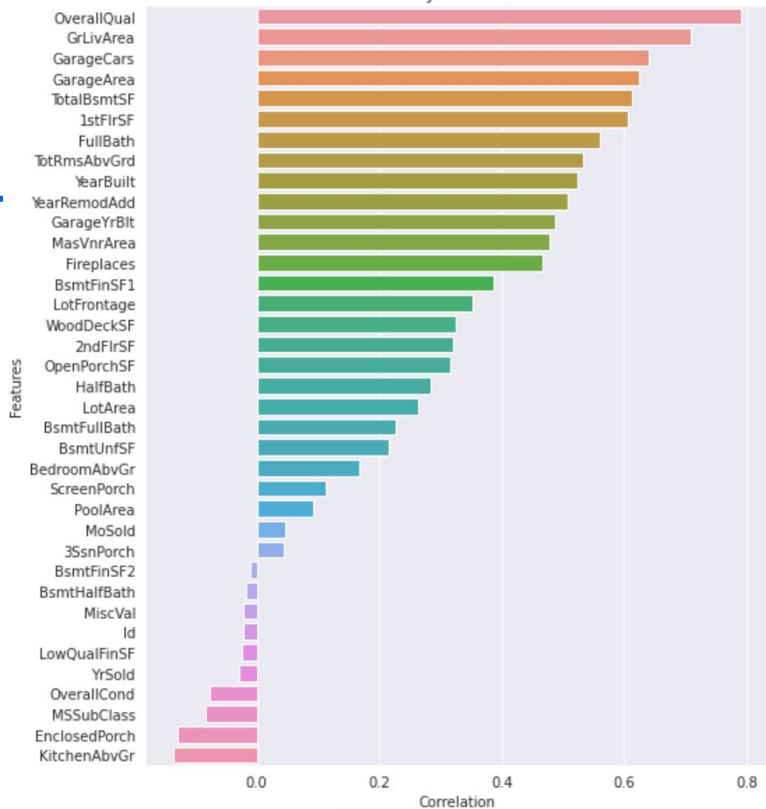
**Thankyou for  
listening**

# Reference



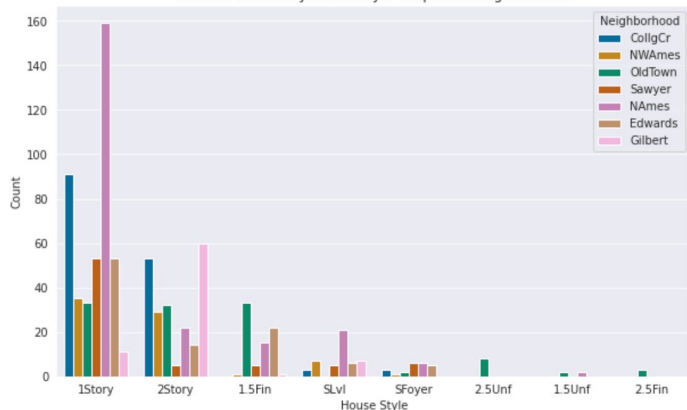
# Reference

Features sorted by correlation to SalePrice

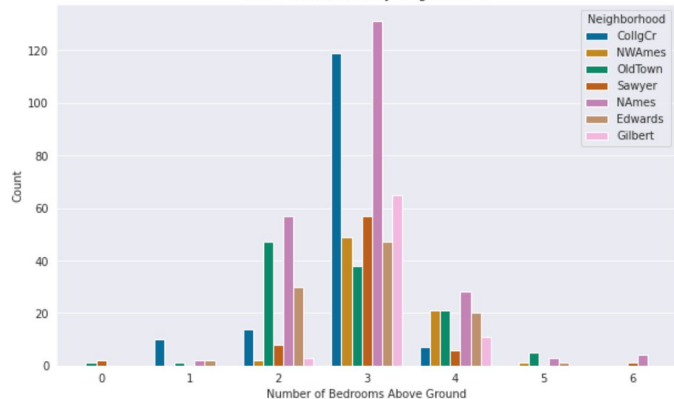


# Reference

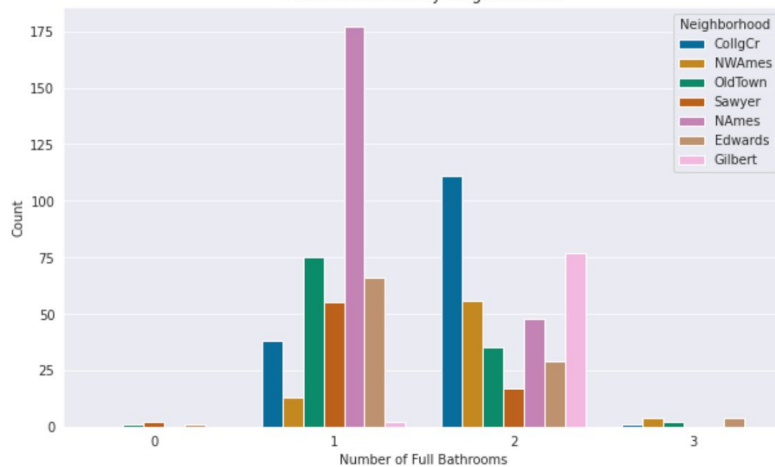
Count of Houses by House Style in Specific Neighborhoods



Count of Bedrooms by Neighborhood



Count of FullBath by Neighborhoods



Count of Garage Cars by Neighborhood

