

# Project 2

## Ames Housing Data

A laptop screen is shown with a dark overlay. In the background, a line graph with a blue line and a pie chart are visible. The line graph has a label '19 av.' and a legend entry 'New Visitor'. The pie chart is partially visible. The text 'Problem statement:' is written in large white font over the screen.

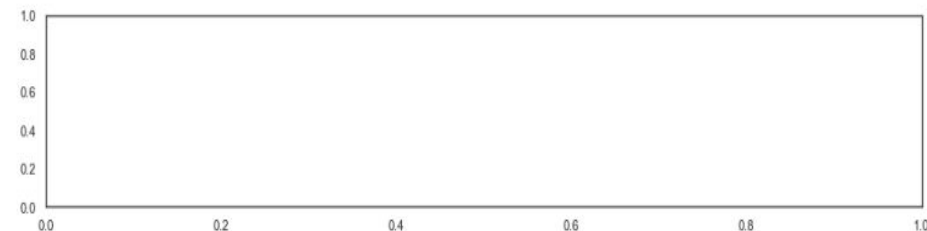
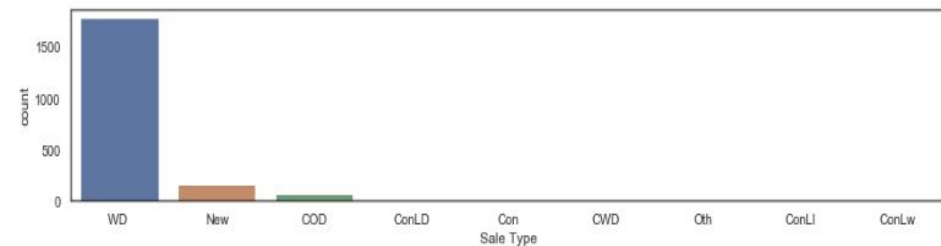
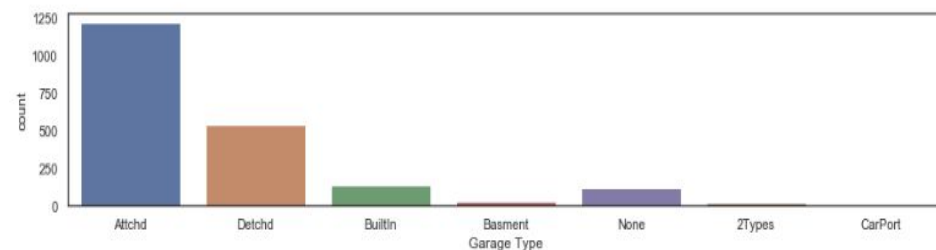
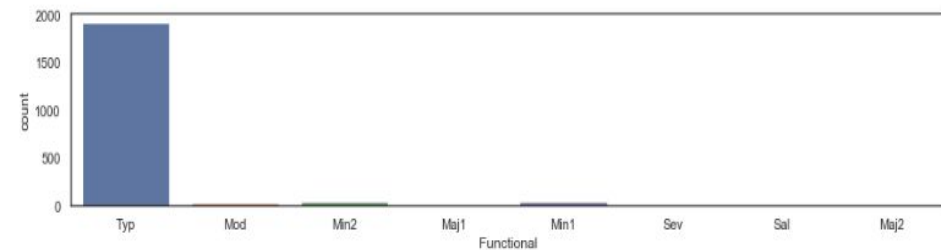
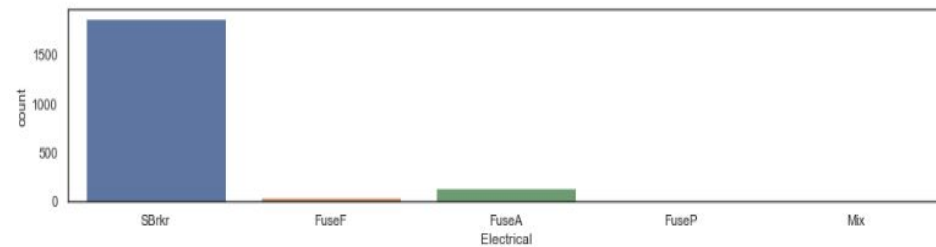
# Problem statement:

Ascertain important features that influence house prices to ensure proper valuation. The proper valuation of houses would prevent another housing bubble from taking place

# Combined Features

1. Total\_Firepl
  - a. Fireplace \*Fireplace Qu
2. Total Square Feet
  - a. Greater Living Area
  - b. Bsmt Square area
3. Total Porch Area
  - a. Open Porch SF
  - b. Enclosed Porch
  - c. 3SN Porch
  - d. Screen Porch

# Dropping some features- Lower Variance



# Final Features Column

```
X.columns
```

```
Index(['Overall Qual', 'Total SF', 'Garage Area', 'Bsmt Qual', 'Full Bath',  
      'Foundation_PConc', 'TotRms AbvGrd', 'Heating QC',  
      'Neighborhood_NridgHt', 'Garage Finish', 'Bsmt Exposure',  
      'MS Zoning_RM', 'Roof Style_Hip', 'Total_Firepl'],  
      dtype='object')
```

# R2 score

```
[ 4110... #R^2 Score for train data  
best.score(X_train,y_train)
```

```
[ 4110... 0.8074849201649046
```

```
[ 4111... #R^2 Score for test data  
best.score(X_test,y_test)
```

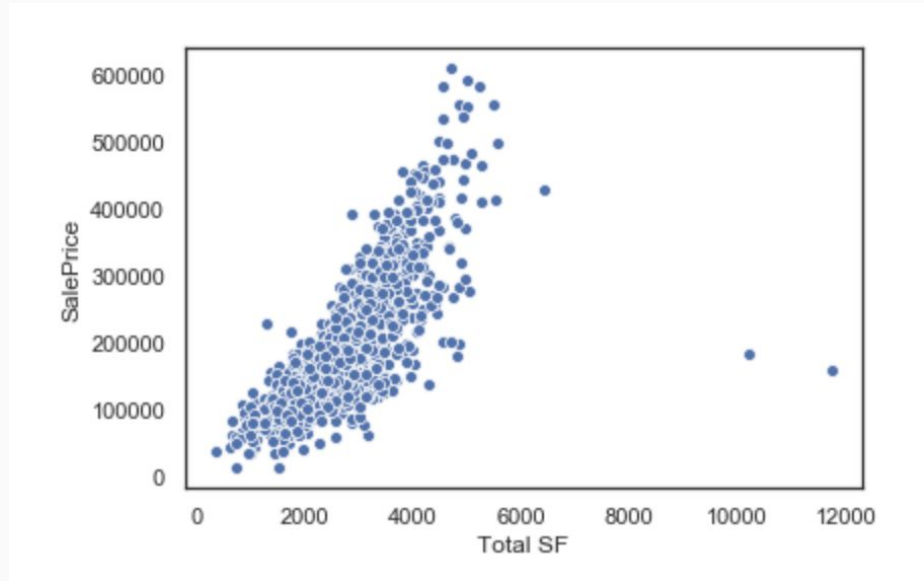
```
[ 4111... 0.8455752908422602
```

```
[ 4112... #Baselinescore for predictions  
lasso_score = cross_val_score(lasso,X_train,y_train, cv = 5)  
lasso_score.mean()
```

```
[ 4112... 0.7945773039562196
```

# Expected Findings from the predictive model

1. Overall Quality
2. Total Sq Feet of the House
  - a. Combining both Greater Liv Area & BSMT Area

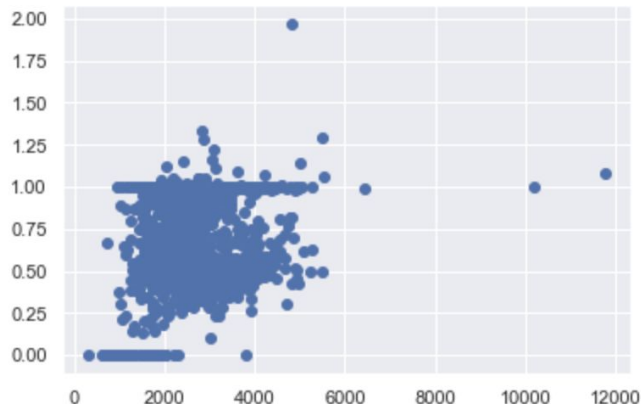


# We might lose some explanatory power by using the total Square Feet data

```
[624]: #We can add up some values that are related with each other  
train['Total SF'] = train['Total Bsmt SF'] + train['Gr Liv Area']  
train['SF_ratio'] = train['Total Bsmt SF'] / train['Gr Liv Area']
```

```
[626]: plt.scatter(x='Total SF', y='SF_ratio', data=train)
```

```
[626]: <matplotlib.collections.PathCollection at 0x1a3b69a2b0>
```





# Interesting finds through the predictive model

1. **Garage Area**
2. **Fireplace Quality**
3. **Basement Exposure**



Why are Garage  
Areas and  
Fireplace quality  
essential for the  
houses in Ames?



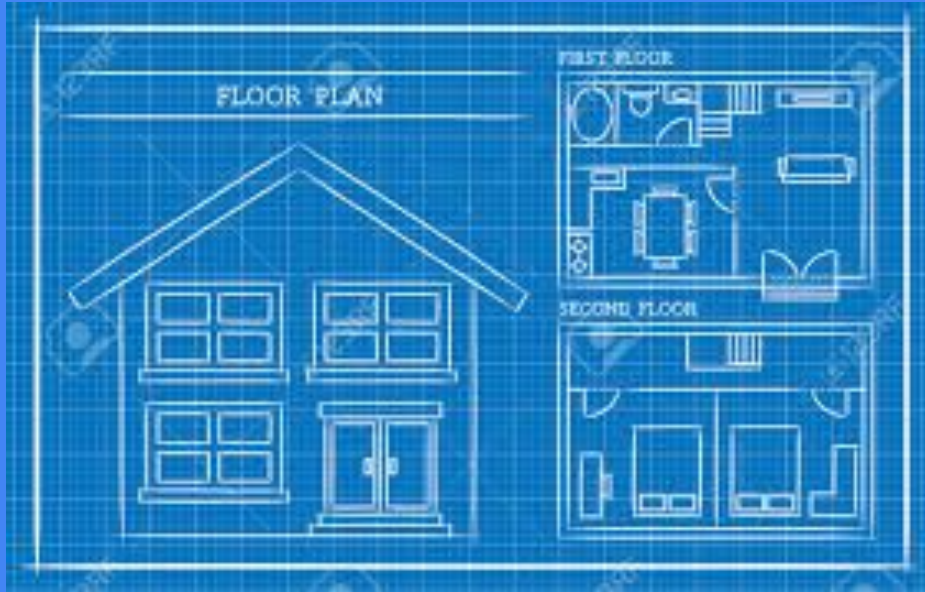
<https://www.thrillist.com/travel/nation/states-with-the-worst-winters-worst-us-states-for-winter>



- **Why are these factors important?**
  - Value house prices accurately
    - Prevent another housing bubble from occurring
  - People are more aware of the house prices



# Things to note:



- *Some general features in this model can be used in predicting other houses from a different city.*
- *The other features that are city specific*
  - *Winter prone- Harsh Winter*
  - *Major cities have higher number of apartments*
    - *Different features will be used to predict the prices*