FLIP ROBO

Housing Selling Price Prediction

Submitted by:

Aditya Maurya

ACKNOWLEDGMENT

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

INTRODUCTION

- Business Problem Framing

Describe the business problem and how this problem can be related to the real world.

Answer: The business problem is to predict the selling price of the houses. In real world many parameters like area, connectivity, quality of building materials etc. comes in to consideration. In the dataset the columns in train is Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street', 'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig', 'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType', 'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd', 'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType', 'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual', 'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1', 'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating','HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF','LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath', 'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual', 'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType', 'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual', 'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC', 'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType', 'SaleCondition', 'SalePrice']. Every column is an accessory for the complete home. Each component has its own use.

- Conceptual Background of the Domain Problem

Describe the domain related concepts that you think will be useful for better understanding of the project.

Answer: Basically we should know the people requirements as requirement of different people are different. The design and comfort of the house is one part, next comes the location of the house, House Style, heating system, garage placement in the house, number of rooms etc. There are many other accessories such as kitchen quality, Functional, Fireplaces, Fireplace, Quality etc.

- Review of Literature

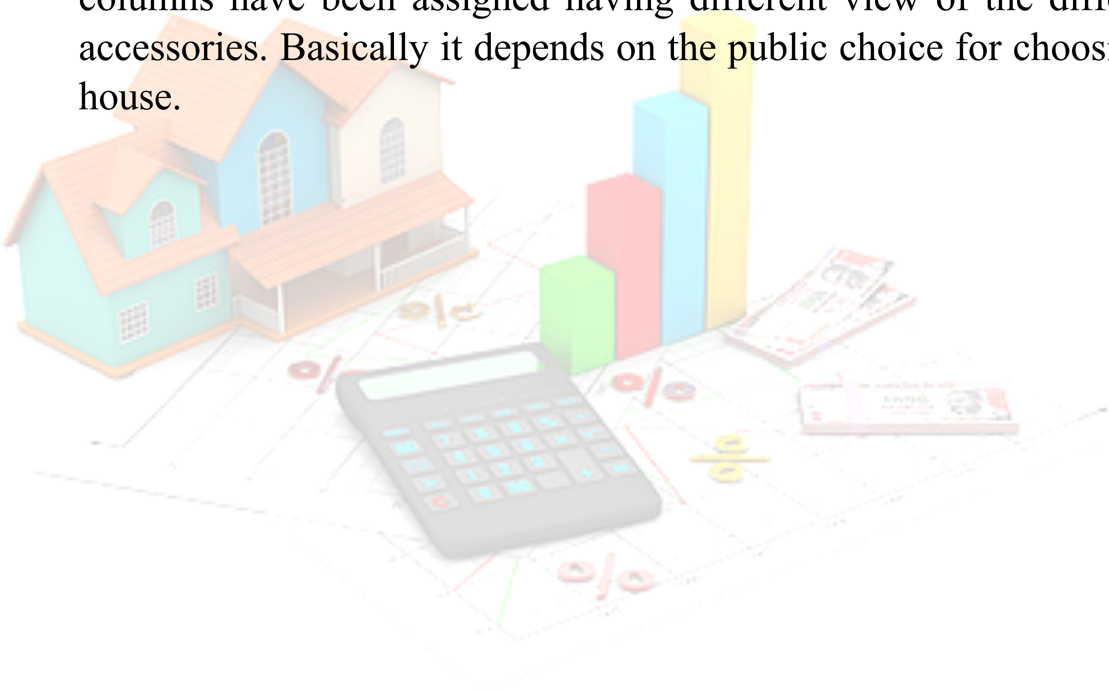
This is a comprehensive summary of the research done on the topic. The review should enumerate, describe, summarize, evaluate and clarify the research done.

Answer: As each column has a different accessory of the house it has to be analysed and prioritized based on the cost. This helps in the ordinal encoding of the machine learning model.

- Motivation for the Problem Undertaken

Describe your objective behind to make this project, this domain and what is the motivation behind.

Answer: The motivation is to build a predictive model that can predict selling price of the house including all the accessories. The total of 80 columns have been assigned having different view of the different accessories. Basically it depends on the public choice for choosing a house.



Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

Describe the mathematical, statistical and analytics modelling done during this project along with the proper justification.

- Data Sources and their formats

  What are the data sources, their origins, their formats and other details that you find necessary? They can be described here. Provide a proper data description. You can also add a snapshot of the data.

  Answer: The data can be taken by a survey of Real Estate Company, open source websites like Kaggel etc. The data is in the form of .csv file it may also be in .json or Excel files. Currently the data was provided in terms of .csv files. There is training file with 1168 columns and 81 rows. There is a test file with 292 columns and 80 rows. The below shows the column list.

```
The columns in train is Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath',
       'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual',
       'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType',
       'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual',
       'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC',
       'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType',
       'SaleCondition', 'SalePrice'],
      dtype='object')
****************************************************************************************************
The columns in test is Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath',
       'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual',
       'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType',
       'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual',
       'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC',
       'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType',
       'SaleCondition'],
      dtype='object')
```

  Finding type of the data:

```
In [14]: # Getting information on the dataset
         print ('The training dataset consists of', dt1.info())
         print('*'*100)
         print ('The testing dataset consists of',dt.info())
```

  <class 'pandas.core.frame.DataFrame'>

  RangeIndex: 1168 entries, 0 to 1167

  Data columns (total 81 columns):

```
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Id             1168 non-null   int64  
 1   MSSubClass     1168 non-null   int64  
 2   MSZoning       1168 non-null   object 
 3   LotFrontage    954 non-null    float64
 4   LotArea        1168 non-null   int64  
 5   Street         1168 non-null   object 
 6   Alley          77 non-null     object 
 7   LotShape       1168 non-null   object 
 8   LandContour    1168 non-null   object 
 9   Utilities      1168 non-null   object 
 10  LotConfig      1168 non-null   object 
 11  LandSlope      1168 non-null   object 
 12  Neighborhood   1168 non-null   object 
 13  Condition1     1168 non-null   object 
 14  Condition2     1168 non-null   object 
 15  BldgType       1168 non-null   object 
 16  HouseStyle     1168 non-null   object 
 17  OverallQual    1168 non-null   int64  
 18  OverallCond    1168 non-null   int64  
 19  YearBuilt      1168 non-null   int64  
 20  YearRemodAdd   1168 non-null   int64  
 21  RoofStyle      1168 non-null   object 
 22  RoofMatl       1168 non-null   object 
 23  Exterior1st    1168 non-null   object 
 24  Exterior2nd    1168 non-null   object 
 25  MasVnrType     1161 non-null   object 
 26  MasVnrArea     1161 non-null   float64
```

27 ExterQual 1168 non-null object
28 ExterCond 1168 non-null object
29 Foundation 1168 non-null object
30 BsmtQual 1138 non-null object
31 BsmtCond 1138 non-null object
32 BsmtExposure 1137 non-null object
33 BsmtFinType1 1138 non-null object
34 BsmtFinSF1 1168 non-null int64
35 BsmtFinType2 1137 non-null object
36 BsmtFinSF2 1168 non-null int64
37 BsmtUnfSF 1168 non-null int64
38 TotalBsmtSF 1168 non-null int64
39 Heating 1168 non-null object
40 HeatingQC 1168 non-null object
41 CentralAir 1168 non-null object
42 Electrical 1168 non-null object
43 1stFlrSF 1168 non-null int64
44 2ndFlrSF 1168 non-null int64
45 LowQualFinSF 1168 non-null int64
46 GrLivArea 1168 non-null int64
47 BsmtFullBath 1168 non-null int64
48 BsmtHalfBath 1168 non-null int64
49 FullBath 1168 non-null int64
50 HalfBath 1168 non-null int64
51 BedroomAbvGr 1168 non-null int64
52 KitchenAbvGr 1168 non-null int64
53 KitchenQual 1168 non-null object
54 TotRmsAbvGrd 1168 non-null int64
55 Functional 1168 non-null object

56 Fireplaces 1168 non-null int64
57 FireplaceQu 617 non-null object
58 GarageType 1104 non-null object
59 GarageYrBlt 1104 non-null float64
60 GarageFinish 1104 non-null object
61 GarageCars 1168 non-null int64
62 GarageArea 1168 non-null int64
63 GarageQual 1104 non-null object
64 GarageCond 1104 non-null object
65 PavedDrive 1168 non-null object
66 WoodDeckSF 1168 non-null int64
67 OpenPorchSF 1168 non-null int64
68 EnclosedPorch 1168 non-null int64
69 3SsnPorch 1168 non-null int64
70 ScreenPorch 1168 non-null int64
71 PoolArea 1168 non-null int64
72 PoolQC 7 non-null object
73 Fence 237 non-null object
74 MiscFeature 44 non-null object
75 MiscVal 1168 non-null int64
76 MoSold 1168 non-null int64
77 YrSold 1168 non-null int64
78 SaleType 1168 non-null object
79 SaleCondition 1168 non-null object
80 SalePrice 1168 non-null int64

dtypes: float64(3), int64(35), object(43)

memory usage: 739.2+ KB

The training dataset consists of None

```
*********************************************************
****************************************
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 292 entries, 0 to 291
Data columns (total 80 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Id             292 non-null    int64
 1   MSSubClass     292 non-null    int64
 2   MSZoning       292 non-null    object
 3   LotFrontage    247 non-null    float64
 4   LotArea        292 non-null    int64
 5   Street         292 non-null    object
 6   Alley          14 non-null     object
 7   LotShape       292 non-null    object
 8   LandContour    292 non-null    object
 9   Utilities      292 non-null    object
 10  LotConfig      292 non-null    object
 11  LandSlope      292 non-null    object
 12  Neighborhood   292 non-null    object
 13  Condition1     292 non-null    object
 14  Condition2     292 non-null    object
 15  BldgType       292 non-null    object
 16  HouseStyle     292 non-null    object
 17  OverallQual    292 non-null    int64
 18  OverallCond    292 non-null    int64
 19  YearBuilt      292 non-null    int64
 20  YearRemodAdd   292 non-null    int64
 21  RoofStyle      292 non-null    object
```

22	RoofMatl	292 non-null	object
23	Exterior1st	292 non-null	object
24	Exterior2nd	292 non-null	object
25	MasVnrType	291 non-null	object
26	MasVnrArea	291 non-null	float64
27	ExterQual	292 non-null	object
28	ExterCond	292 non-null	object
29	Foundation	292 non-null	object
30	BsmtQual	285 non-null	object
31	BsmtCond	285 non-null	object
32	BsmtExposure	285 non-null	object
33	BsmtFinType1	285 non-null	object
34	BsmtFinSF1	292 non-null	int64
35	BsmtFinType2	285 non-null	object
36	BsmtFinSF2	292 non-null	int64
37	BsmtUnfSF	292 non-null	int64
38	TotalBsmtSF	292 non-null	int64
39	Heating	292 non-null	object
40	HeatingQC	292 non-null	object
41	CentralAir	292 non-null	object
42	Electrical	291 non-null	object
43	1stFlrSF	292 non-null	int64
44	2ndFlrSF	292 non-null	int64
45	LowQualFinSF	292 non-null	int64
46	GrLivArea	292 non-null	int64
47	BsmtFullBath	292 non-null	int64
48	BsmtHalfBath	292 non-null	int64
49	FullBath	292 non-null	int64
50	HalfBath	292 non-null	int64

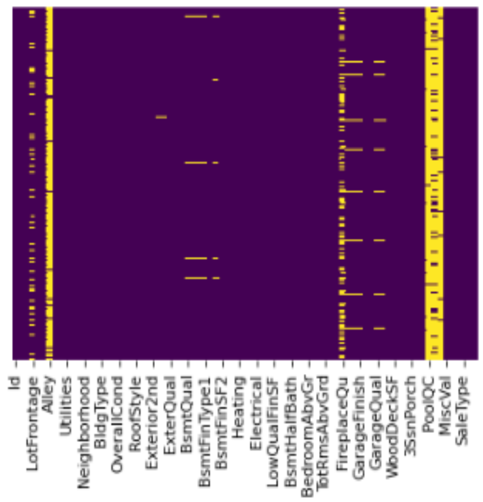
51	BedroomAbvGr	292 non-null	int64
52	KitchenAbvGr	292 non-null	int64
53	KitchenQual	292 non-null	object
54	TotRmsAbvGrd	292 non-null	int64
55	Functional	292 non-null	object
56	Fireplaces	292 non-null	int64
57	FireplaceQu	153 non-null	object
58	GarageType	275 non-null	object
59	GarageYrBlt	275 non-null	float64
60	GarageFinish	275 non-null	object
61	GarageCars	292 non-null	int64
62	GarageArea	292 non-null	int64
63	GarageQual	275 non-null	object
64	GarageCond	275 non-null	object
65	PavedDrive	292 non-null	object
66	WoodDeckSF	292 non-null	int64
67	OpenPorchSF	292 non-null	int64
68	EnclosedPorch	292 non-null	int64
69	3SsnPorch	292 non-null	int64
70	ScreenPorch	292 non-null	int64
71	PoolArea	292 non-null	int64
72	PoolQC	0 non-null	float64
73	Fence	44 non-null	object
74	MiscFeature	10 non-null	object
75	MiscVal	292 non-null	int64
76	MoSold	292 non-null	int64
77	YrSold	292 non-null	int64
78	SaleType	292 non-null	object
79	SaleCondition	292 non-null	object

dtypes: float64(4), int64(34), object(42)

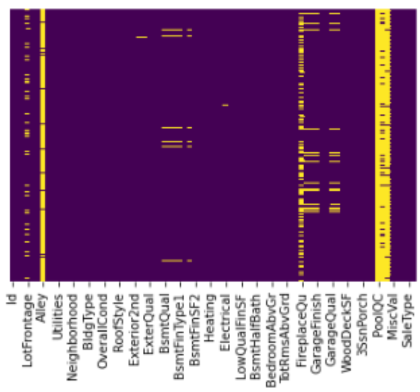
memory usage: 182.6+ KB

The testing dataset consists of NAN values:

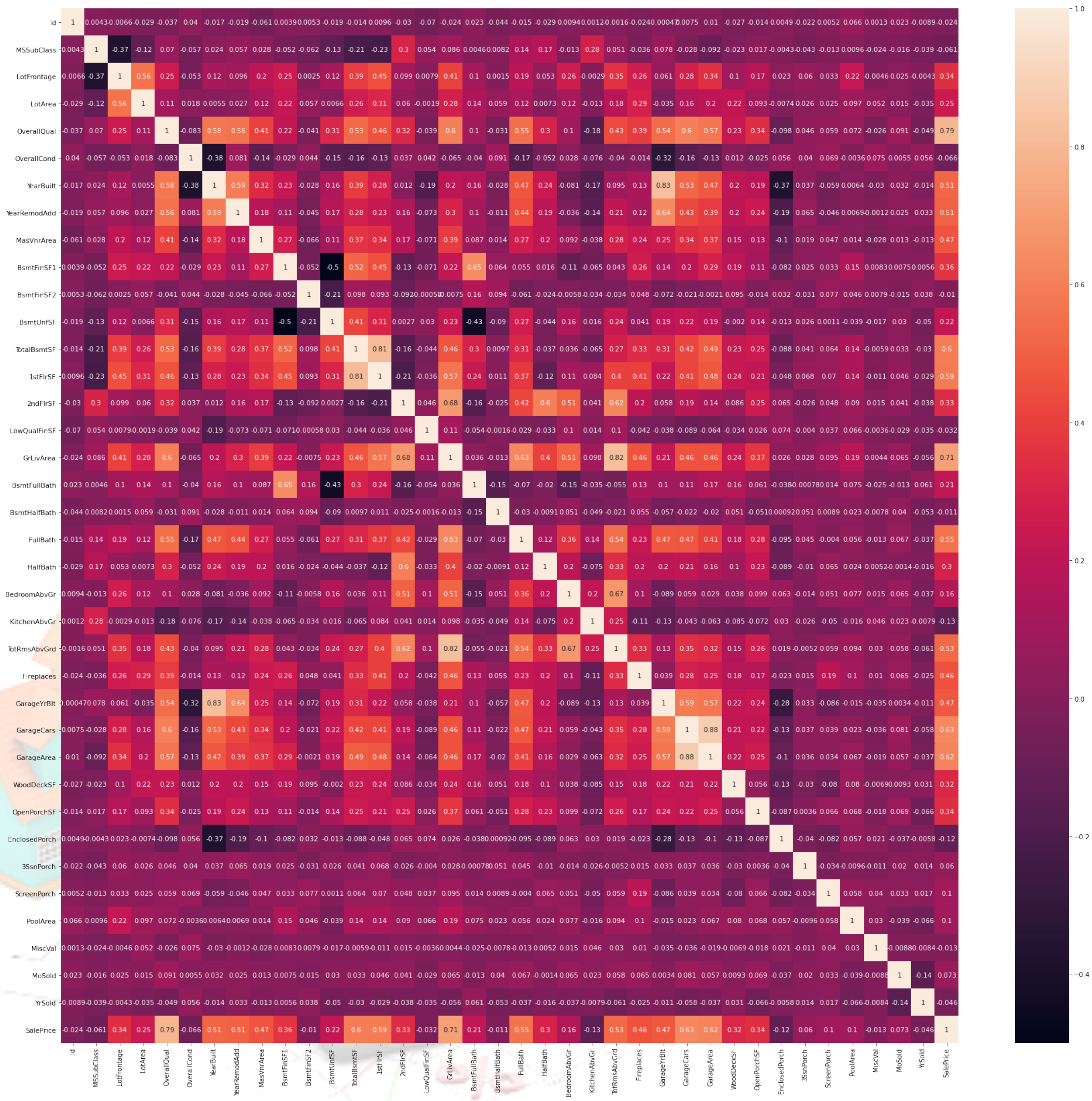
```
In [12]: sns.heatmap(dt1.isnull(), yticklabels = False, cbar = False, cmap ='viridis')
Out[12]: <AxesSubplot:>
```

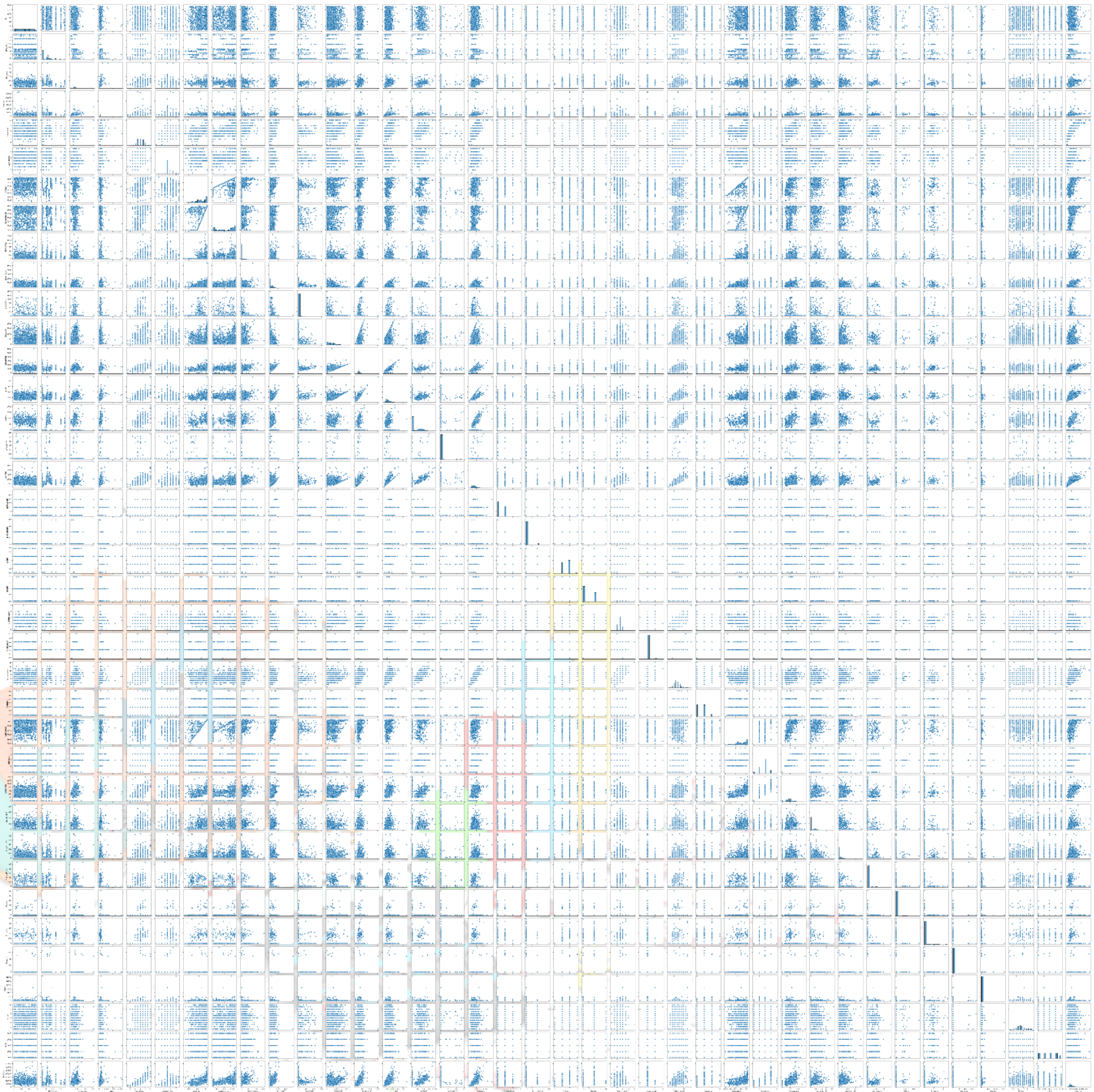



```
In [13]: sns.heatmap(dt.isnull(), yticklabels = False, cbar = False, cmap ='viridis')
Out[13]: <AxesSubplot:>
```



The correlation of the dataset is as follows:


Id
MSSubClass
LotFrontage
LotArea
OverallQual
OverallCond
YearBuilt
YearRemodAdd
MasVnrArea
BsmtFinSF1
BsmtFinSF2
BsmtUnfSF
TotalBsmtSF
1stFlrSF
2ndFlrSF
LowQualFinSF
GrLivArea
BsmtFullBath
BsmtHalfBath
FullBath
HalfBath
BedroomAbvGr
KitchenAbvGr
TotRmsAbvGrd
Fireplaces
GarageYrBlt
GarageCars
GarageArea
WoodDeckSF
OpenPorchSF
EnclosedPorch
3SsnPorch
ScreenPorch
PoolArea
MiscVal
MoSold
YrSold
SalePrice
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4



- Data Preprocessing Done

What were the steps followed for the cleaning of the data? What were the assumptions done and what were the next actions steps over that?

Answer: The BsmtQual, BsmtCond, BsmtExposure, BsmtFinType1, BsmtFinType2, FireplaceQu, GarageType, GarageFinish, GarageQual, GarageCond, Electricity have been replaced by the median. The following shows the different snapshots in the python.

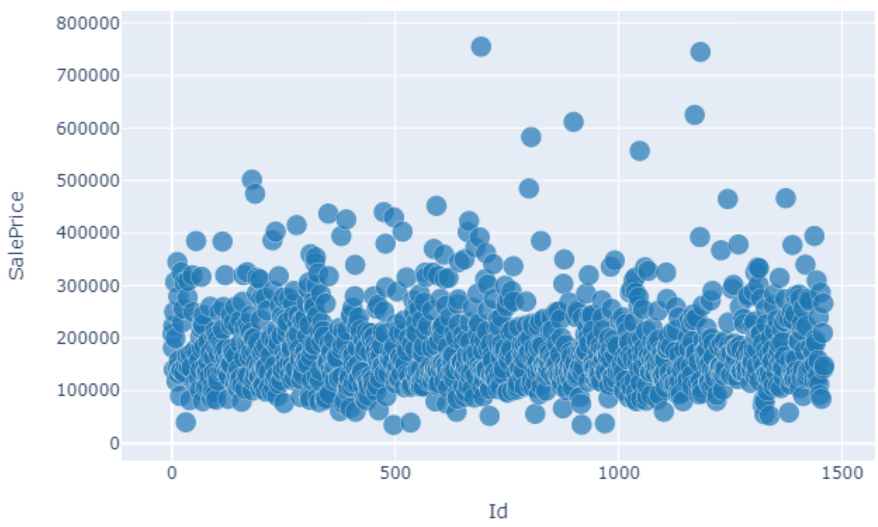
```
In [1351]: dt1['BsmtQual']=dt1['BsmtQual'].fillna(dt1['BsmtQual'].mode()[0])
            dt['BsmtQual']=dt['BsmtQual'].fillna(dt['BsmtQual'].mode()[0])

In [1352]: dt1['BsmtCond']=dt1['BsmtCond'].fillna(dt1['BsmtCond'].mode()[0])
            dt['BsmtCond']=dt['BsmtCond'].fillna(dt['BsmtCond'].mode()[0])

In [1353]: dt1['BsmtExposure']=dt1['BsmtExposure'].fillna(dt1['BsmtExposure'].mode()[0])
            dt['BsmtExposure']=dt['BsmtExposure'].fillna(dt['BsmtExposure'].mode()[0])

In [1354]: dt1['BsmtFinType1']=dt1['BsmtFinType1'].fillna(dt1['BsmtFinType1'].mode()[0])
            dt['BsmtFinType1']=dt['BsmtFinType1'].fillna(dt['BsmtFinType1'].mode()[0])

In [1355]: dt1['BsmtFinType2']=dt1['BsmtFinType2'].fillna(dt1['BsmtFinType2'].mode()[0])
            dt['BsmtFinType2']=dt['BsmtFinType2'].fillna(dt['BsmtFinType2'].mode()[0])

In [1356]: dt1['FireplaceQu']=dt1['FireplaceQu'].fillna(dt1['FireplaceQu'].mode()[0])
            dt['FireplaceQu']=dt['FireplaceQu'].fillna(dt['FireplaceQu'].mode()[0])

In [1357]: dt1['GarageType']=dt1['GarageType'].fillna(dt1['GarageType'].mode()[0])
            dt['GarageType']=dt['GarageType'].fillna(dt['GarageType'].mode()[0])

In [1358]: dt1['GarageFinish']=dt1['GarageFinish'].fillna(dt1['GarageFinish'].mode()[0])
            dt['GarageFinish']=dt['GarageFinish'].fillna(dt['GarageFinish'].mode()[0])

In [1359]: dt1['GarageQual']=dt1['GarageQual'].fillna(dt1['GarageQual'].mode()[0])
            dt['GarageQual']=dt['GarageQual'].fillna(dt['GarageQual'].mode()[0])

In [1360]: dt1['GarageCond']=dt1['GarageCond'].fillna(dt1['GarageCond'].mode()[0])
            dt['GarageCond']=dt['GarageCond'].fillna(dt['GarageCond'].mode()[0])

In [1372]: dt1.shape
Out[1372]: (1168, 81)

As more than 50% have NAN values Alley, PoolQC, Fence, MiscFeature can be neglected

In [1373]: dt1.drop(columns=['Utilities','Alley', 'PoolQC', 'Fence', 'MiscFeature'],inplace=True)

In [1374]: dt.drop(columns=['Utilities','Alley', 'PoolQC', 'Fence', 'MiscFeature'],inplace=True)

In [1375]: dt['Electrical'].fillna(dt['Electrical'].mode()[0],inplace=True)
```

- Data Inputs- Logic- Output Relationships

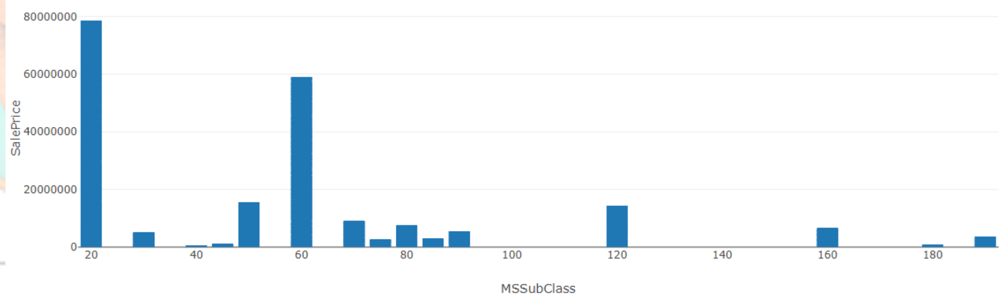
  Describe the relationship behind the data input, its format, the logic in between and the output. Describe how the input affects the output.

  Answer: For identifying the relationship between the input and output parameters bivariate analysis was performed where each column type is compared to the selling price of the house.

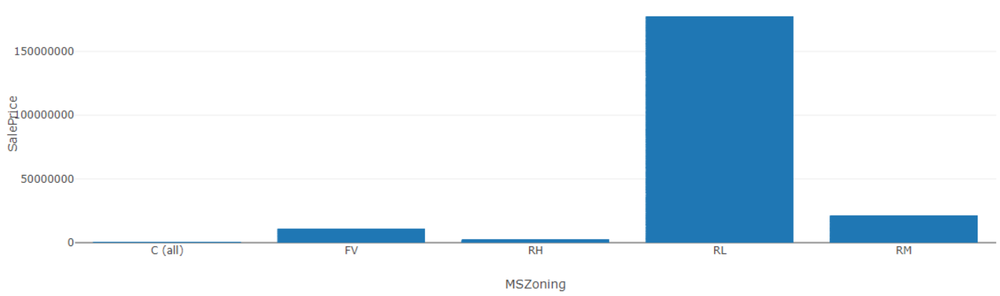
SalePrice by Id

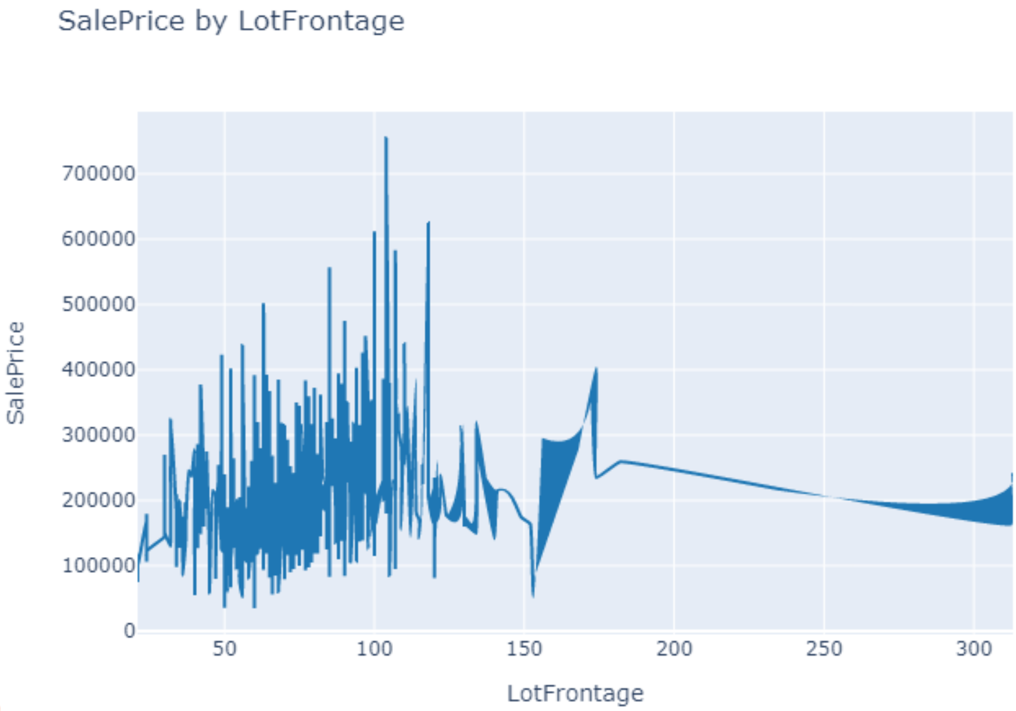


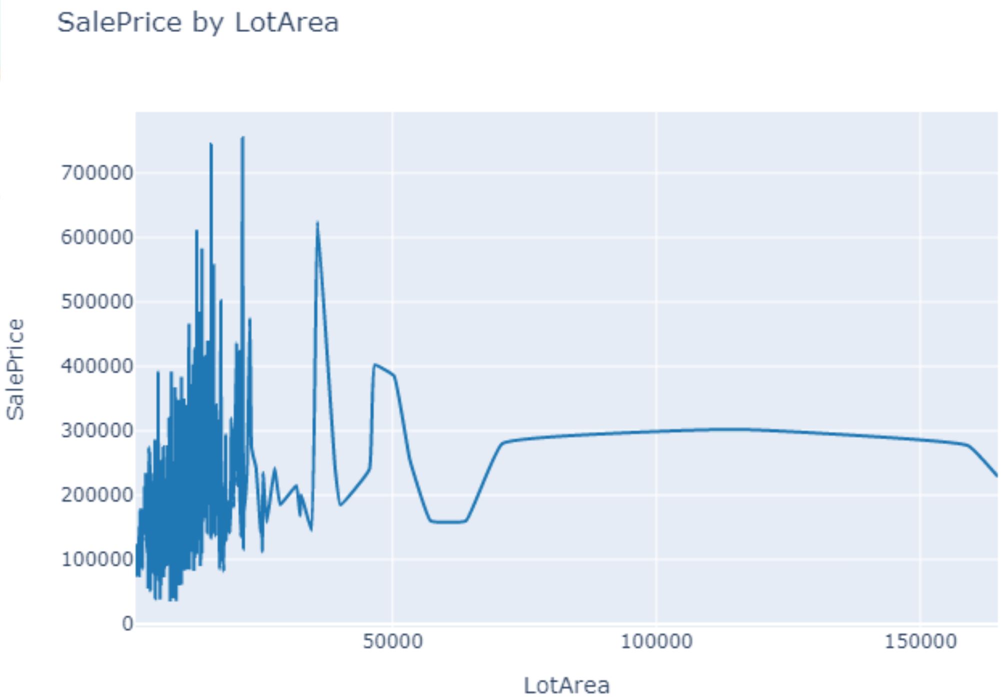
SalePrice by MSSubClass

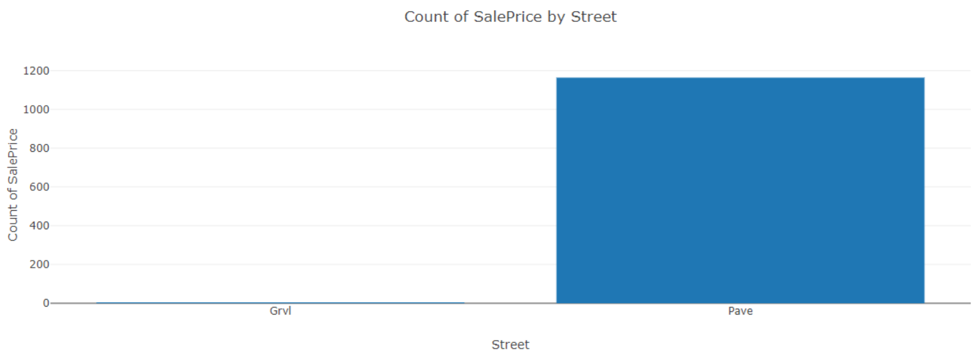


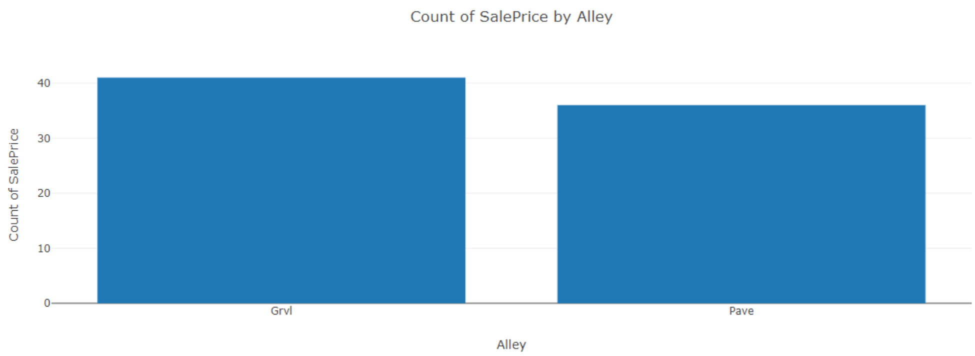
SalePrice by MSZoning

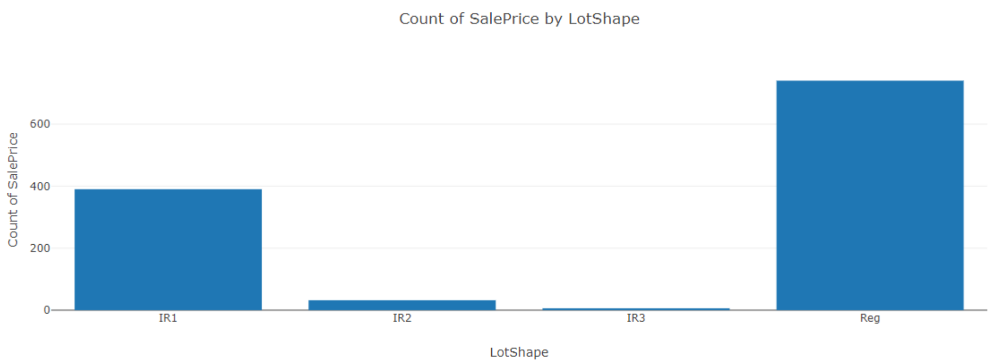


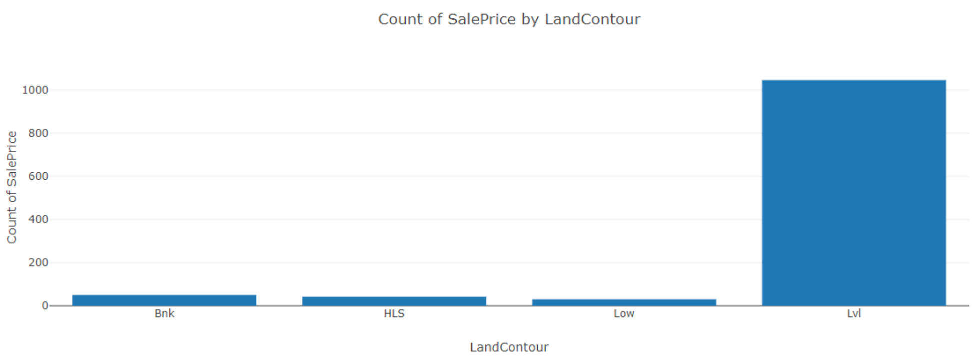

SalePrice by LotFrontage
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
50
100
150
200
250
300
LotFrontage

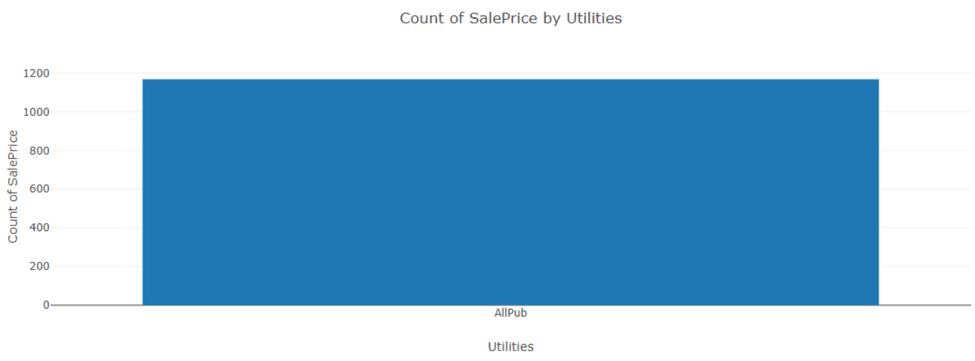

SalePrice by LotArea
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
50000
100000
150000
LotArea

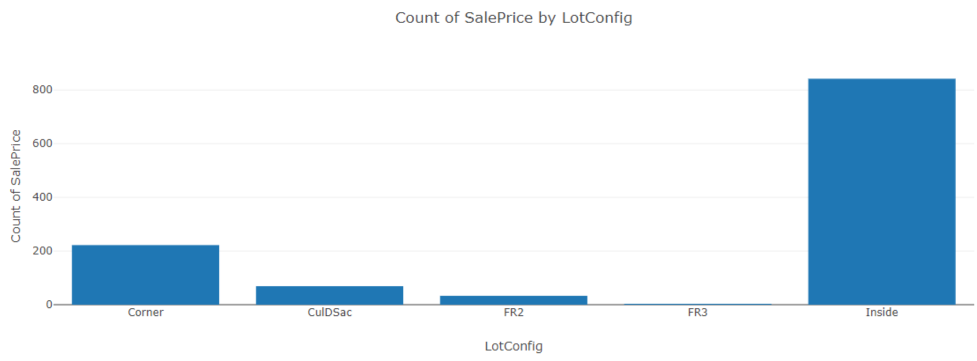

Count of SalePrice by Street
Count of SalePrice
1200
1000
800
600
400
200
0
Grvl
Pave
Street

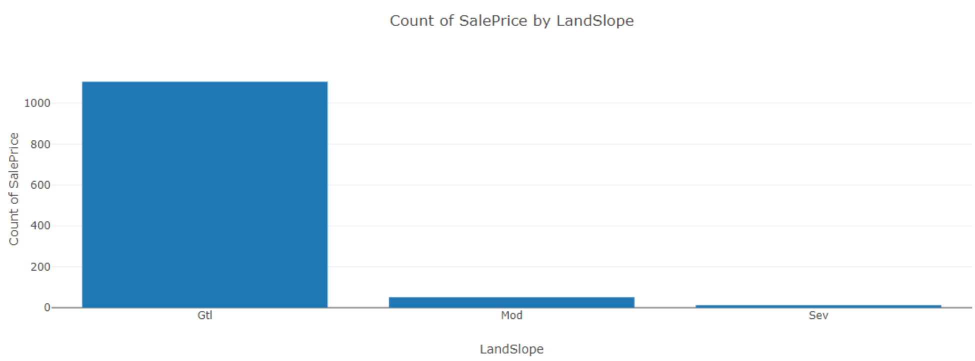

Count of SalePrice by Alley
Count of SalePrice
40
30
20
10
0
Grvl
Pave
Alley

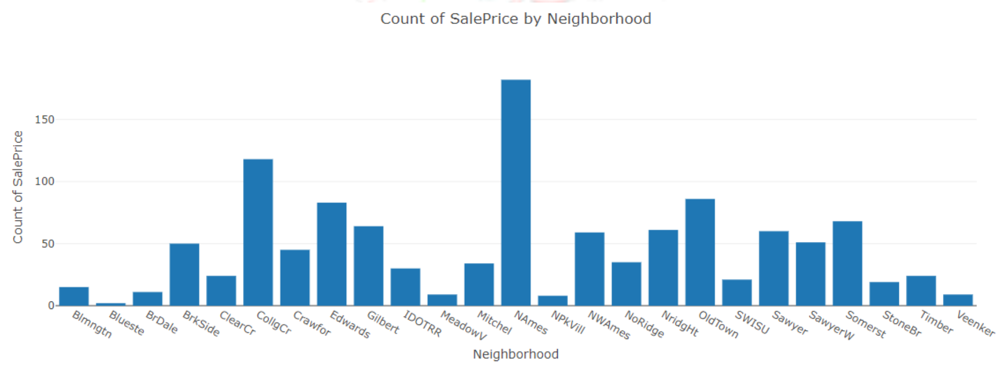

Count of SalePrice by LotShape
Count of SalePrice
600
400
200
0
IR1
IR2
IR3
Reg
LotShape

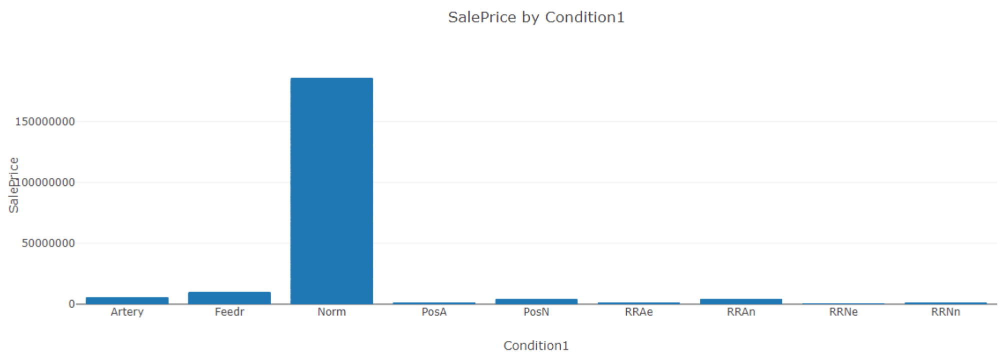

Count of SalePrice by LandContour
Count of SalePrice
1000
800
600
400
200
0
Bnk
HLS
Low
Lvl
LandContour

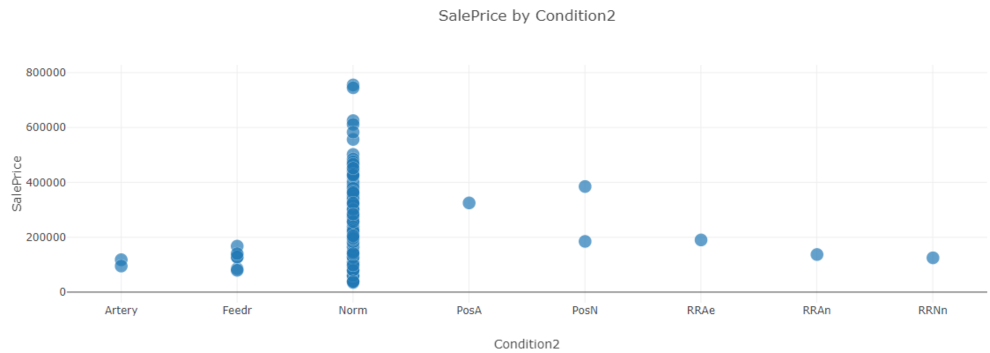

Count of SalePrice by Utilities
Count of SalePrice
0
200
400
600
800
1000
1200
AllPub
Utilities

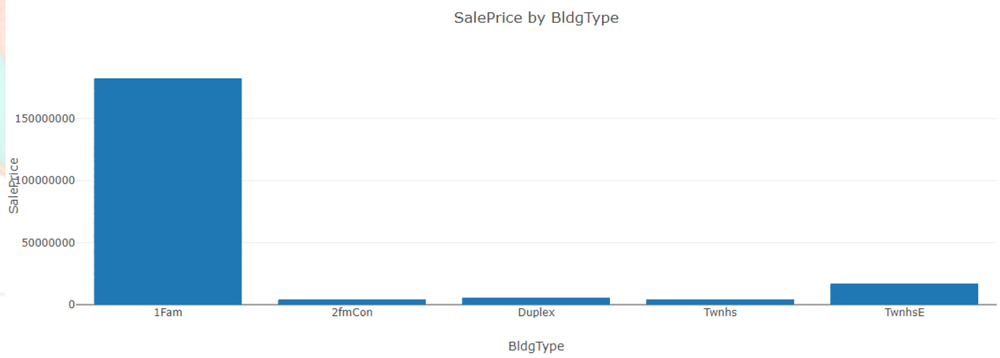

Count of SalePrice by LotConfig
Count of SalePrice
0
200
400
600
800
Corner
CulDSac
FR2
FR3
Inside
LotConfig

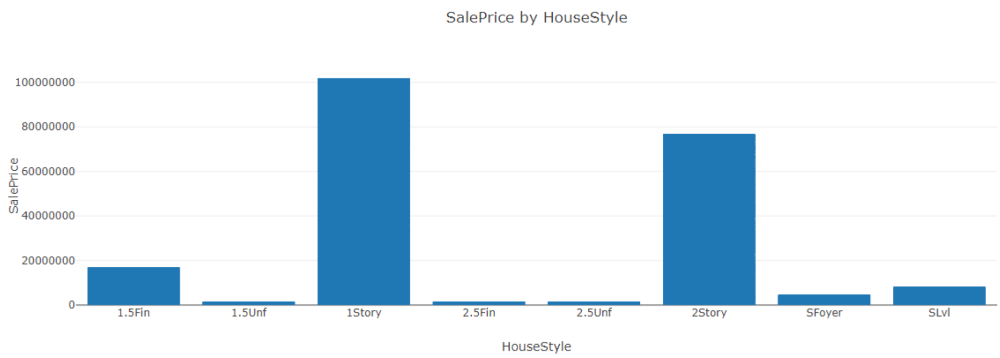

Count of SalePrice by LandSlope
Count of SalePrice
0
200
400
600
800
1000
Gtl
Mod
Sev
LandSlope

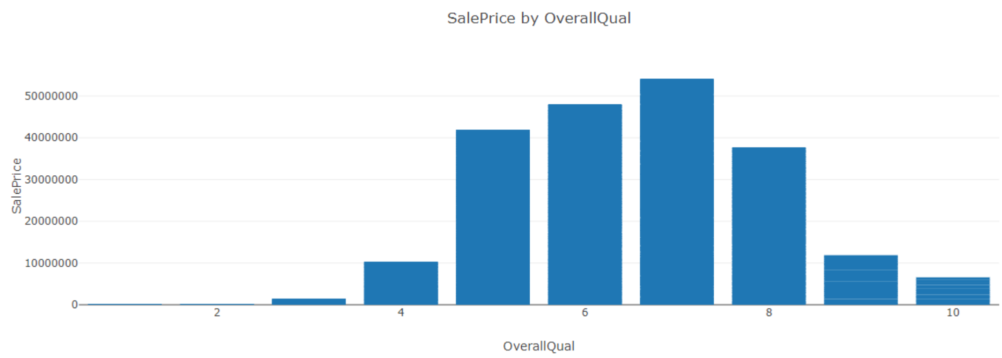

Count of SalePrice by Neighborhood
Count of SalePrice
0
50
100
150
Blmngtn
Blueste
BrDale
BrkSide
ClearCr
CollgCr
Crawfor
Edwards
Gilbert
IDOTRR
MeadowV
Mitchel
NAmes
NPkVill
NWAmes
NoRidge
NridgHt
OldTown
SWISU
Sawyer
SawyerW
Somerst
StoneBr
Timber
Veenker
Neighborhood

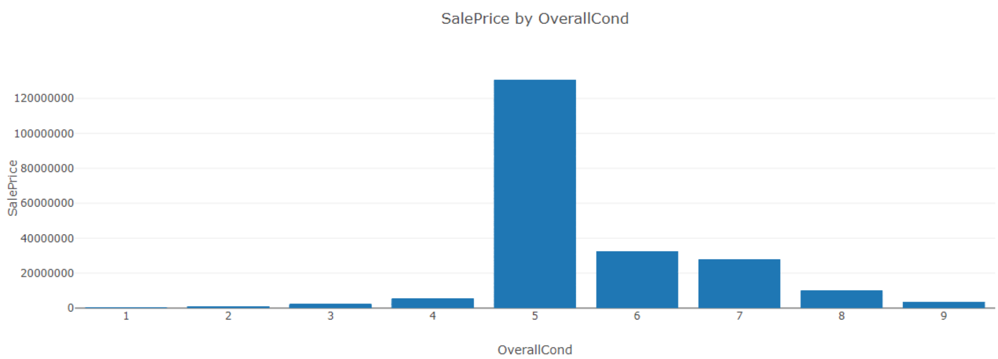

SalePrice by Condition1
SalePrice
150000000
100000000
50000000
0
Artery
Feedr
Norm
PosA
PosN
RRAe
RRAn
RRNe
RRNn
Condition1

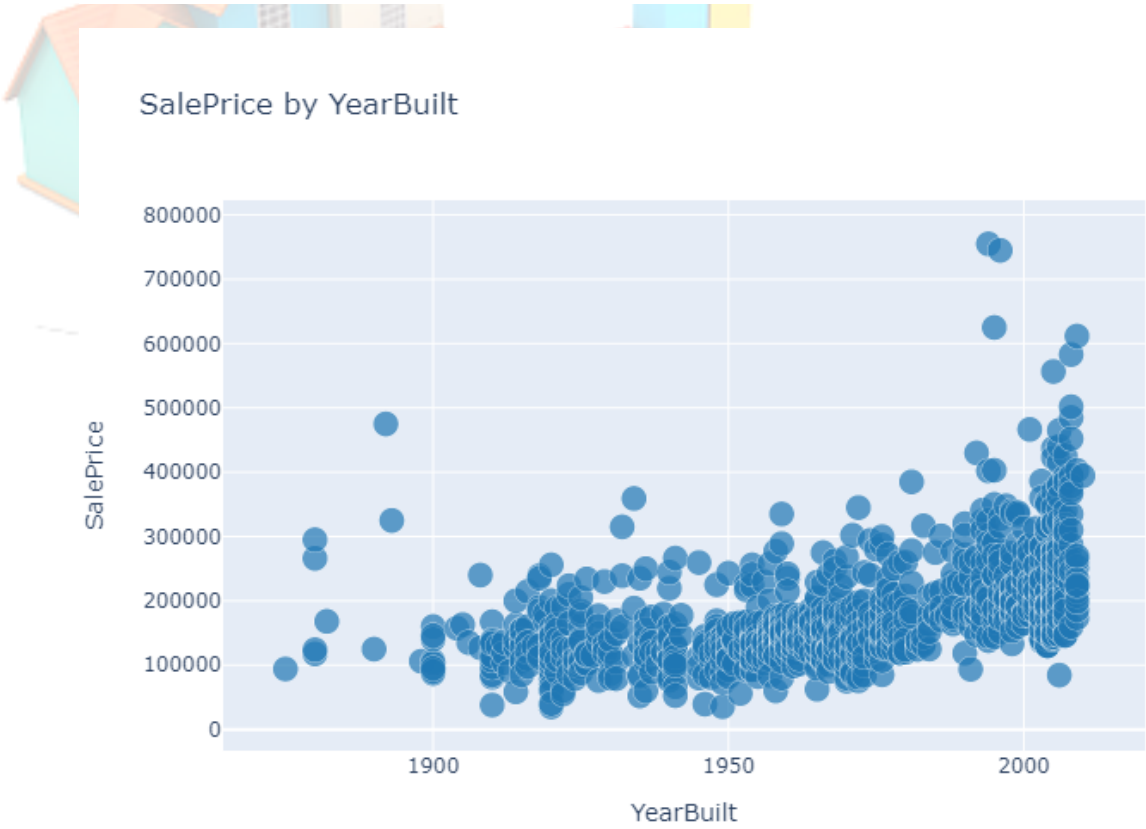

SalePrice by Condition2
SalePrice
800000
600000
400000
200000
0
Artery
Feedr
Norm
PosA
PosN
RRAe
RRAn
RRNn
Condition2

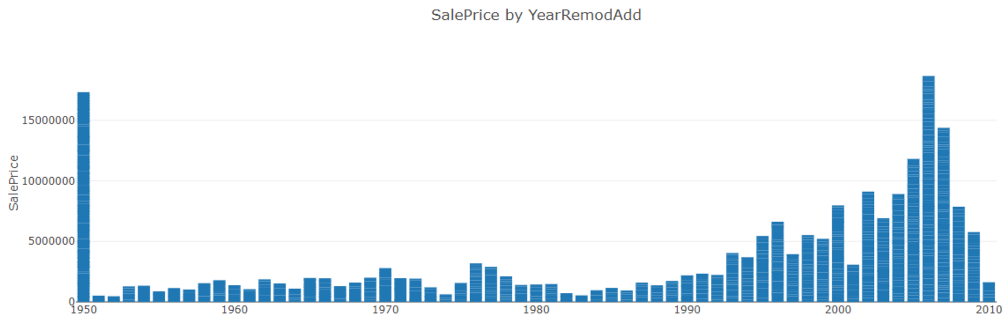

SalePrice by BldgType
SalePrice
150000000
100000000
50000000
0
1Fam
2fmCon
Duplex
Twnhs
TwnhsE
BldgType

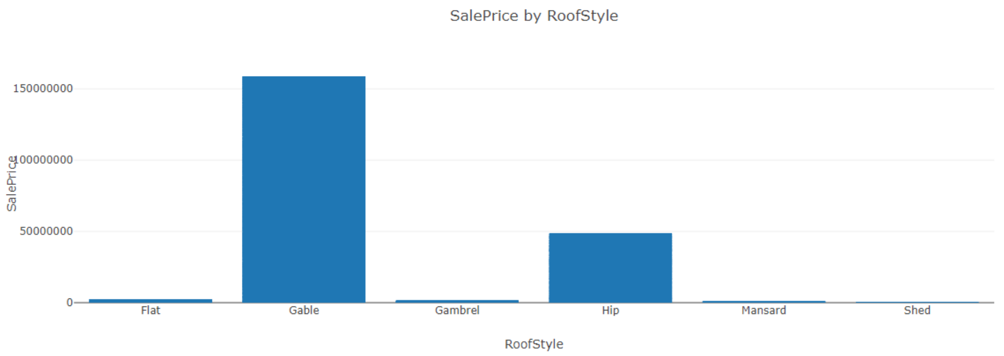

SalePrice by HouseStyle
SalePrice
100000000
80000000
60000000
40000000
20000000
0
1.5Fin
1.5Unf
1Story
2.5Fin
2.5Unf
2Story
SFoyer
SLvl
HouseStyle

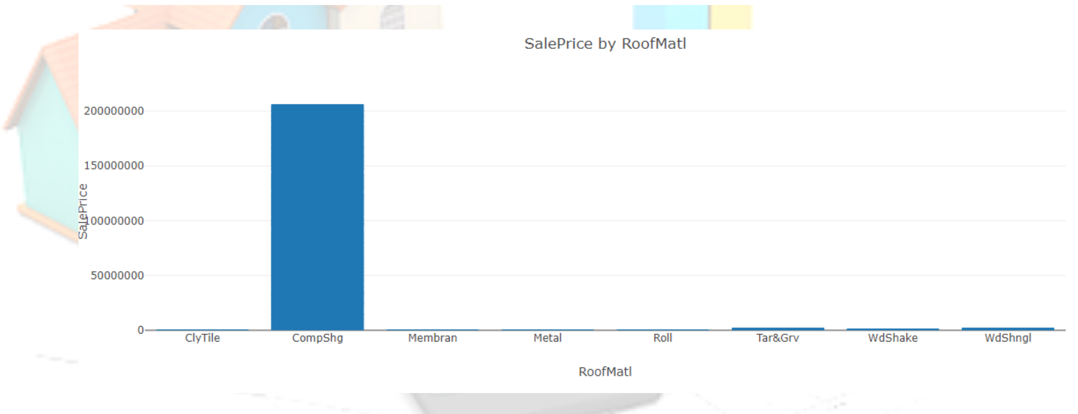

SalePrice by OverallQual
SalePrice
50000000
40000000
30000000
20000000
10000000
0
2
4
6
8
10
OverallQual

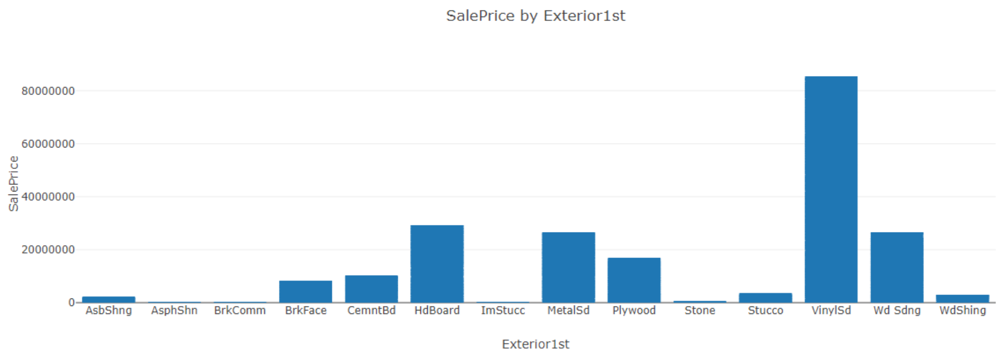

SalePrice by OverallCond
SalePrice
120000000
100000000
80000000
60000000
40000000
20000000
0
1
2
3
4
5
6
7
8
9
OverallCond

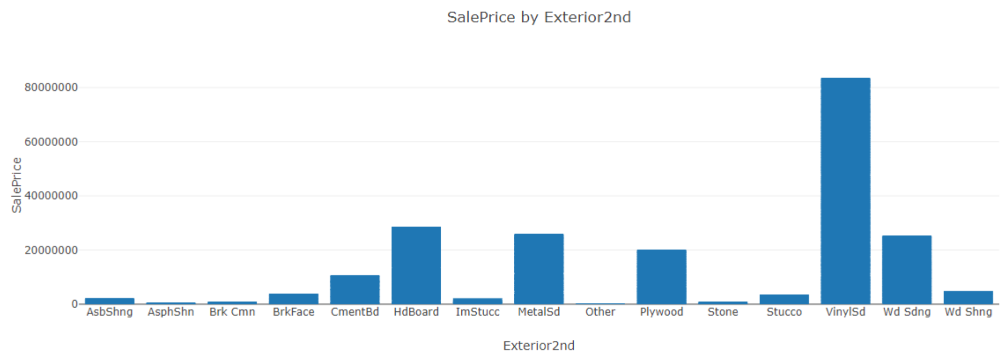

SalePrice by YearBuilt
SalePrice
800000
700000
600000
500000
400000
300000
200000
100000
0
1900
1950
2000
YearBuilt

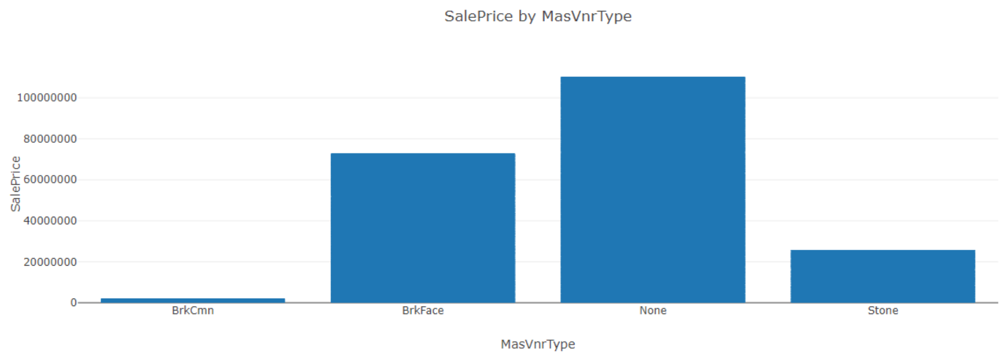

SalePrice by YearRemodAdd
SalePrice
0
5000000
10000000
15000000
1950
1960
1970
1980
1990
2000
2010

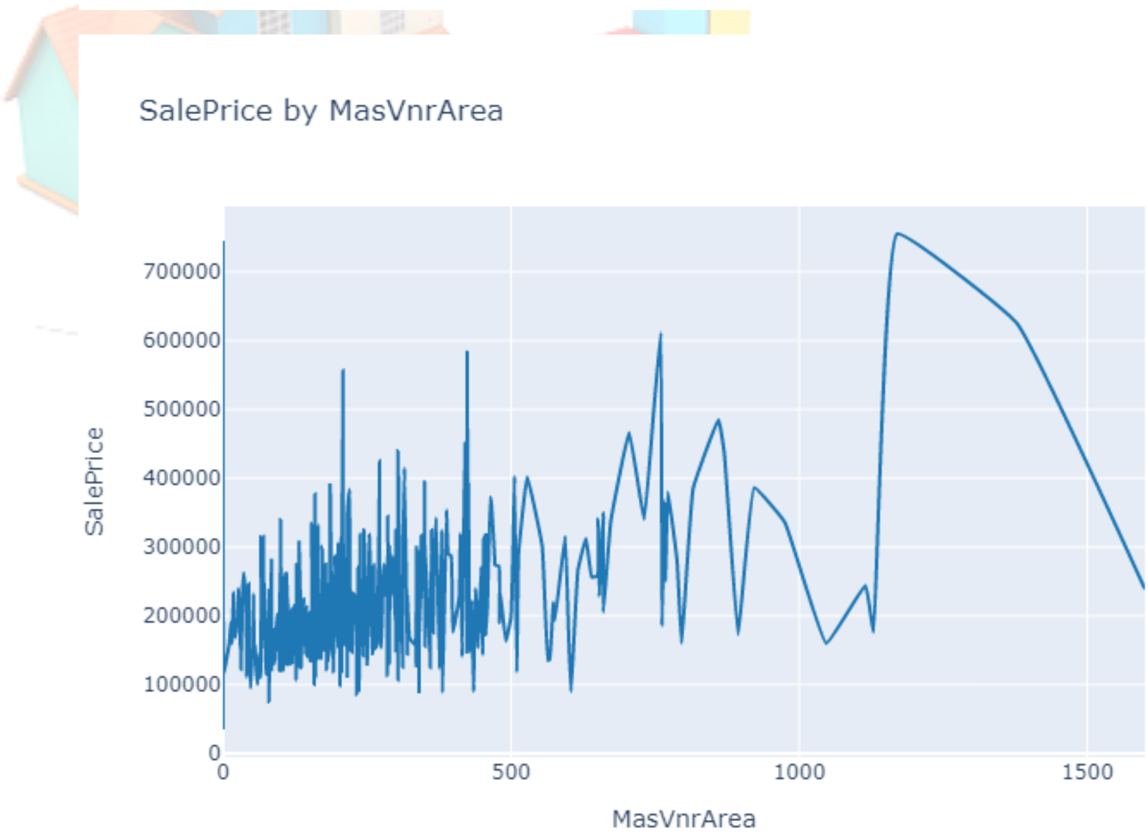

SalePrice by RoofStyle
SalePrice
0
50000000
100000000
150000000
Flat
Gable
Gambrel
Hip
Mansard
Shed
RoofStyle

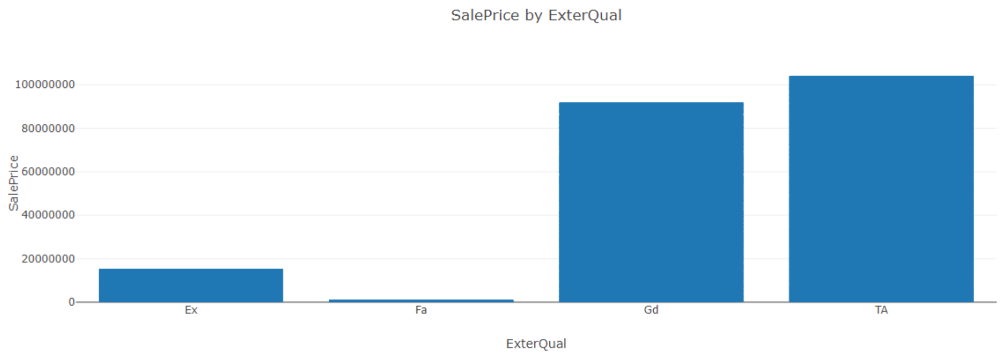

SalePrice by RoofMatl
SalePrice
0
50000000
100000000
150000000
200000000
ClyTile
CompShg
Membran
Metal
Roll
Tar&Grv
WdShake
WdShngl
RoofMatl

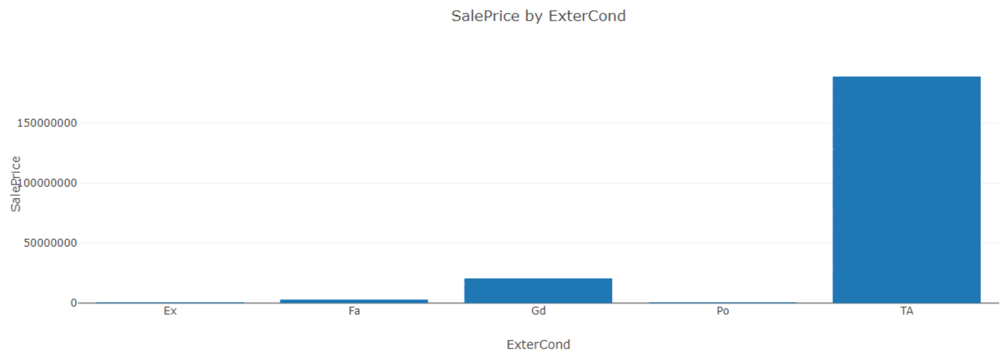

SalePrice by Exterior1st
SalePrice
0
20000000
40000000
60000000
80000000
AsbShng
AsphShn
BrkComm
BrkFace
CemntBd
HdBoard
ImStucc
MetalSd
Plywood
Stone
Stucco
VinylSd
Wd Sdng
WdShing
Exterior1st

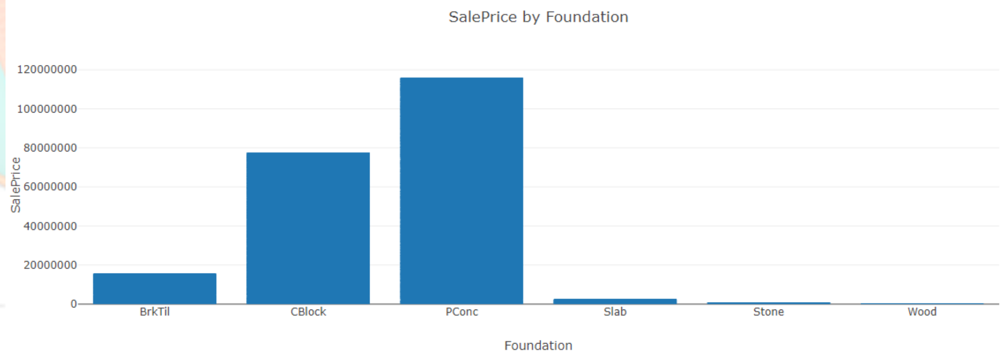

SalePrice by Exterior2nd
SalePrice
80000000
60000000
40000000
20000000
0
AsbShng
AsphShn
Brk Cmn
BrkFace
CmentBd
HdBoard
ImStucc
MetalSd
Other
Plywood
Stone
Stucco
VinylSd
Wd Sdng
Wd Shng
Exterior2nd

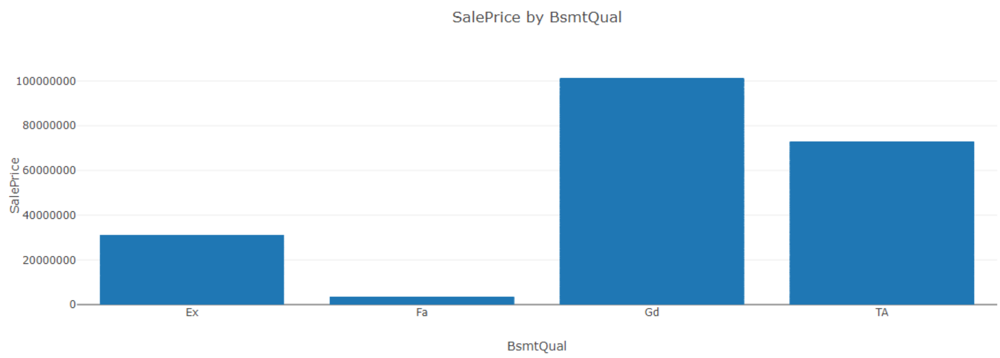

SalePrice by MasVnrType
SalePrice
100000000
80000000
60000000
40000000
20000000
0
BrkCmn
BrkFace
None
Stone
MasVnrType

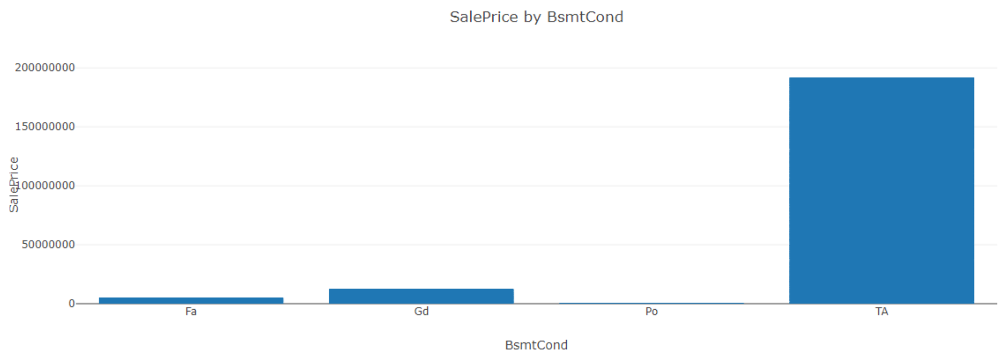

SalePrice by MasVnrArea
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
500
1000
1500
MasVnrArea

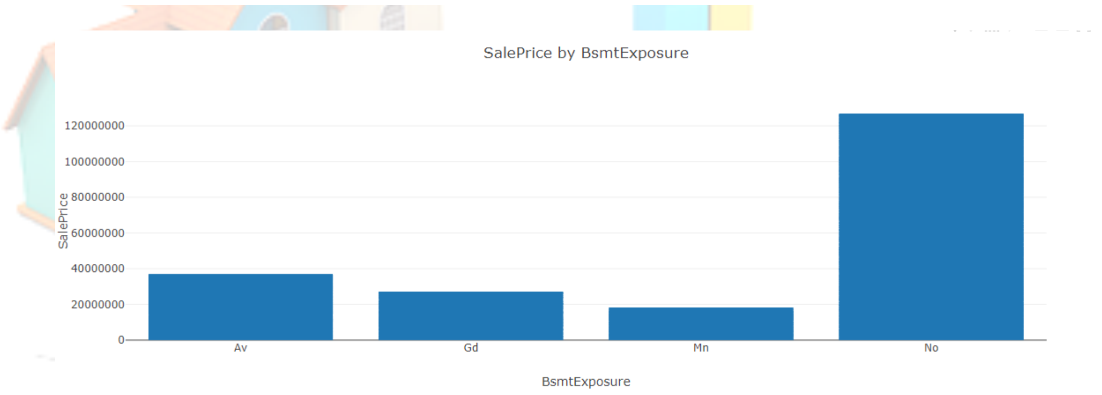

SalePrice by ExterQual
100000000
80000000
60000000
40000000
20000000
0
SalePrice
Ex
Fa
Gd
TA
ExterQual

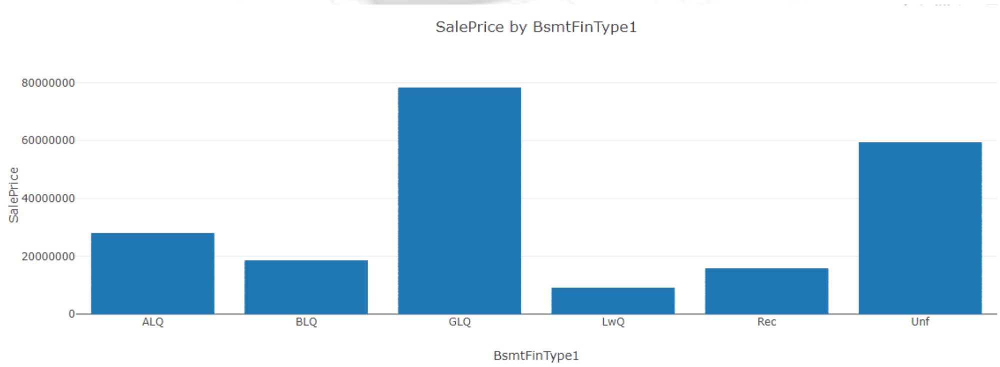

SalePrice by ExterCond
150000000
100000000
50000000
0
SalePrice
Ex
Fa
Gd
Po
TA
ExterCond

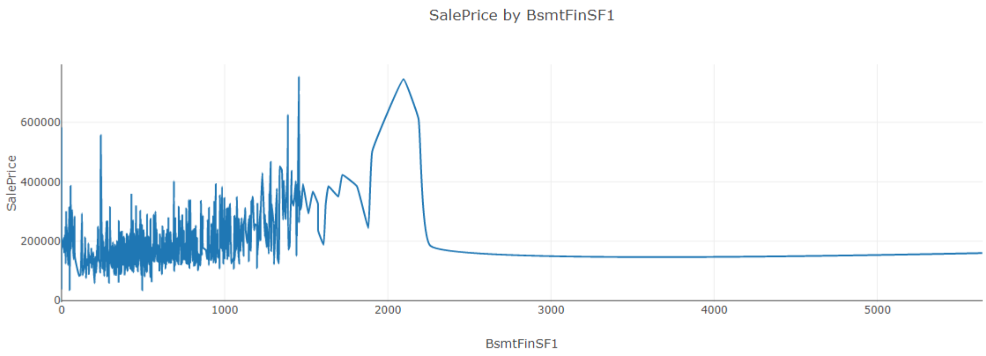

SalePrice by Foundation
120000000
100000000
80000000
60000000
40000000
20000000
0
SalePrice
BrkTil
CBlock
PConc
Slab
Stone
Wood
Foundation

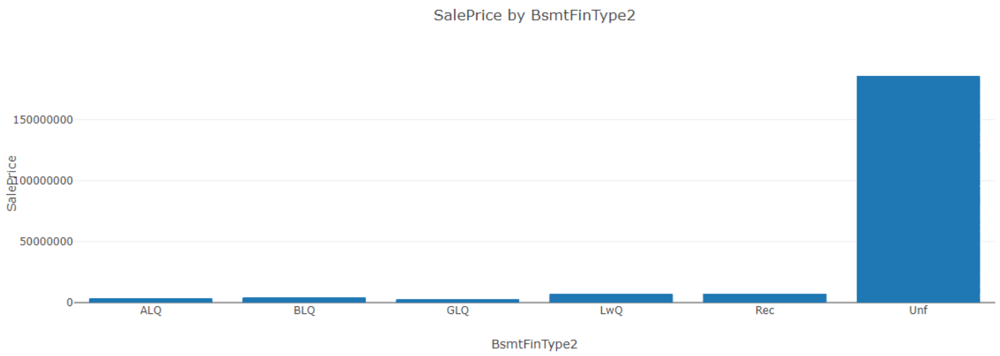

SalePrice by BsmtQual
SalePrice
100000000
80000000
60000000
40000000
20000000
0
Ex
Fa
Gd
TA
BsmtQual

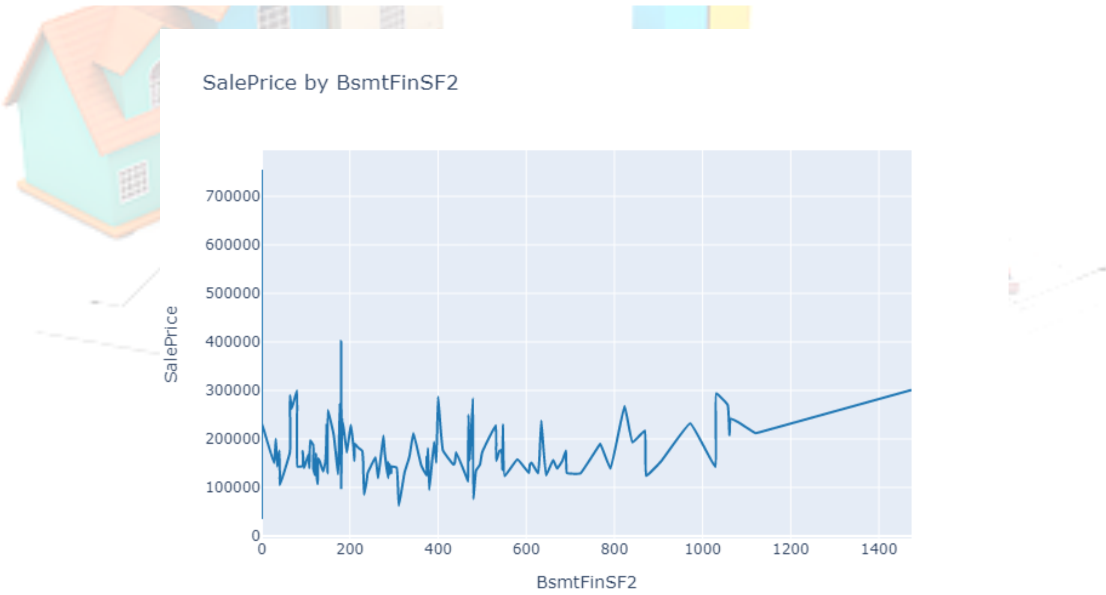

SalePrice by BsmtCond
SalePrice
200000000
150000000
100000000
50000000
0
Fa
Gd
Po
TA
BsmtCond

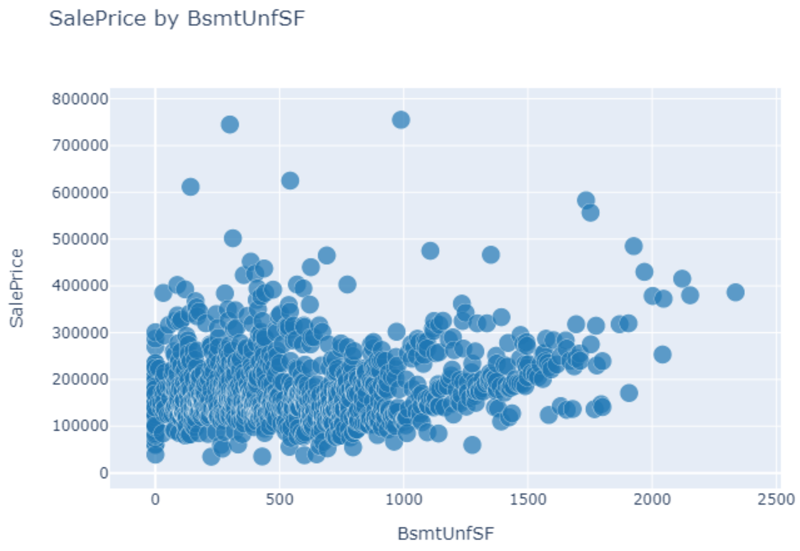

SalePrice by BsmtExposure
SalePrice
120000000
100000000
80000000
60000000
40000000
20000000
0
Av
Gd
Mn
No
BsmtExposure

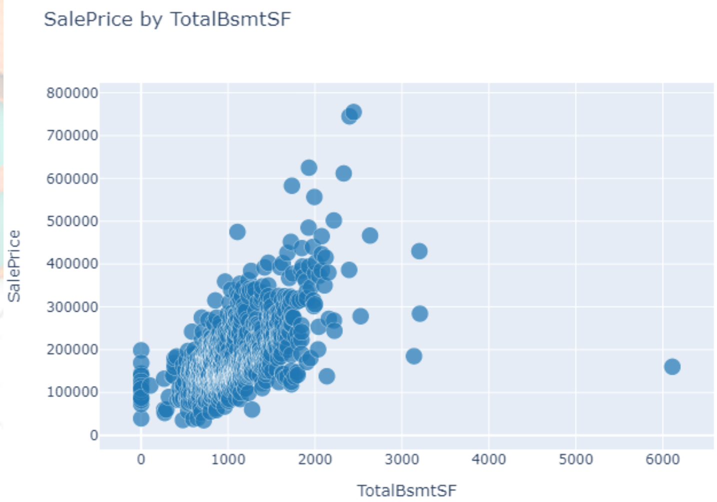

SalePrice by BsmtFinType1
SalePrice
80000000
60000000
40000000
20000000
0
ALQ
BLQ
GLQ
LwQ
Rec
Unf
BsmtFinType1

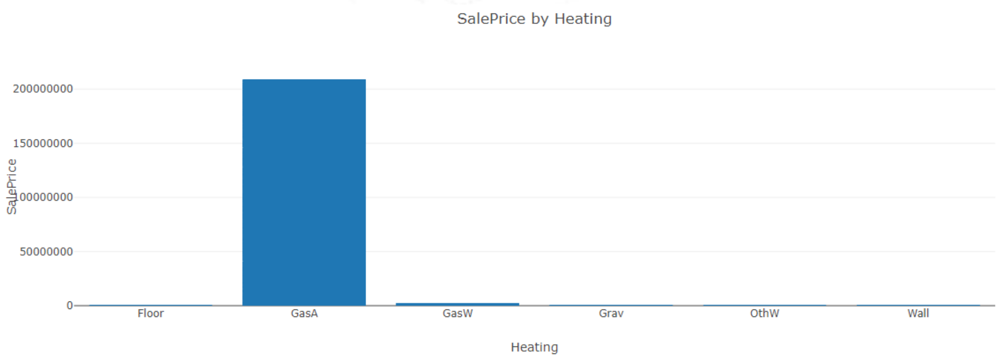

SalePrice by BsmtFinSF1
SalePrice
600000
400000
200000
0
0
1000
2000
3000
4000
5000
BsmtFinSF1

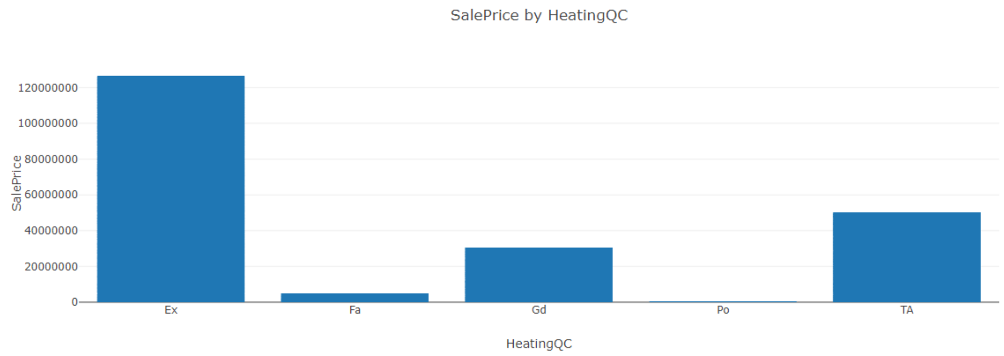

SalePrice by BsmtFinType2
SalePrice
150000000
100000000
50000000
0
ALQ
BLQ
GLQ
LwQ
Rec
Unf
BsmtFinType2

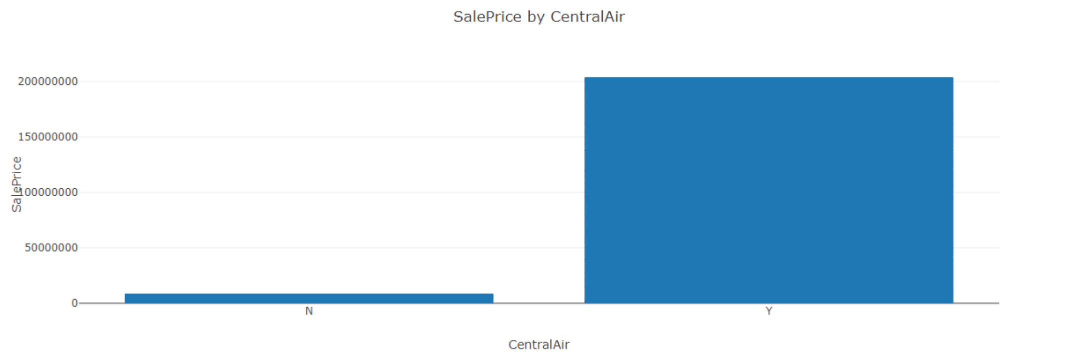

SalePrice by BsmtFinSF2
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
200
400
600
800
1000
1200
1400
BsmtFinSF2

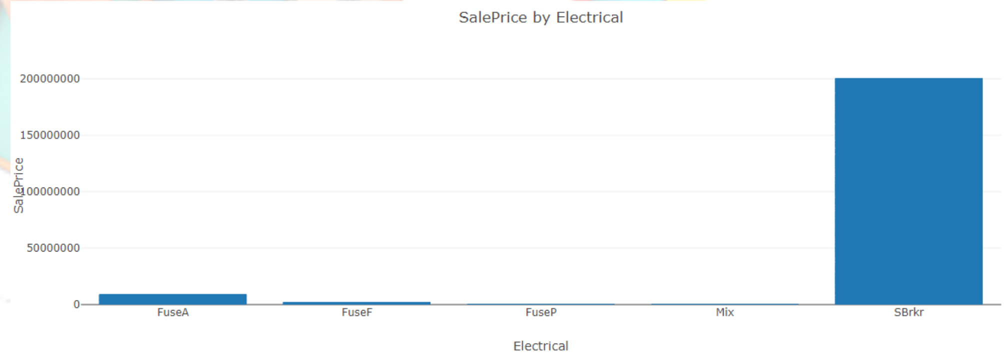

SalePrice by BsmtUnfSF
800000
700000
600000
500000
400000
300000
200000
100000
0
SalePrice
0
500
1000
1500
2000
2500
BsmtUnfSF


SalePrice by TotalBsmtSF
800000
700000
600000
500000
400000
300000
200000
100000
0
SalePrice
0
1000
2000
3000
4000
5000
6000
TotalBsmtSF


SalePrice by Heating
200000000
150000000
100000000
50000000
0
SalePrice
Floor
GasA
GasW
Grav
OthW
Wall
Heating

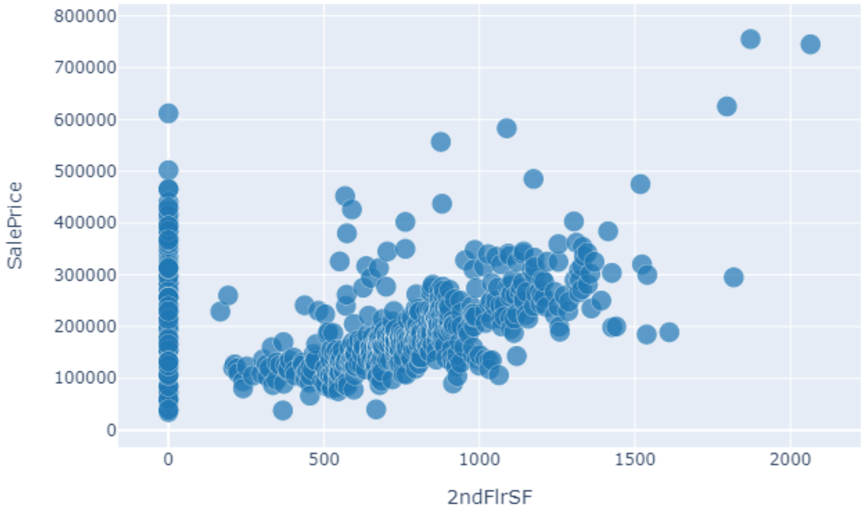

SalePrice by HeatingQC
120000000
100000000
80000000
60000000
40000000
20000000
0
SalePrice
Ex
Fa
Gd
Po
TA
HeatingQC


SalePrice by CentralAir
200000000
150000000
100000000
50000000
0
SalePrice
N
Y
CentralAir

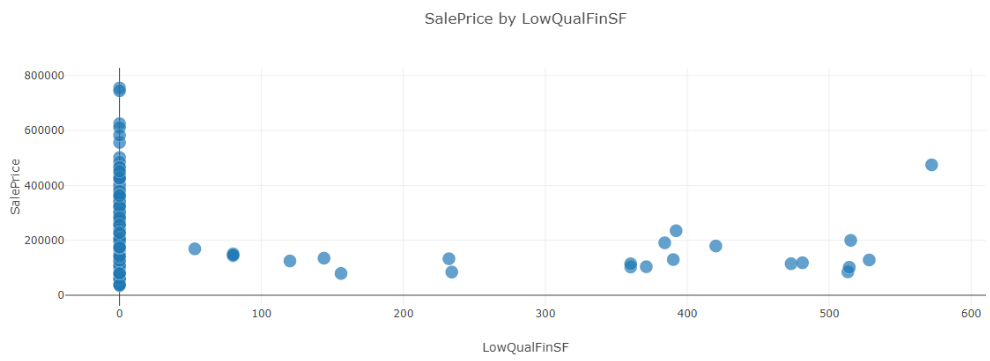

SalePrice by Electrical
200000000
150000000
100000000
50000000
0
SalePrice
FuseA
FuseF
FuseP
Mix
SBrkr
Electrical


SalePrice by 1stFlrSF
600000
400000
200000
0
SalePrice
500
1000
1500
2000
2500
3000
3500
4000
4500
1stFlrSF

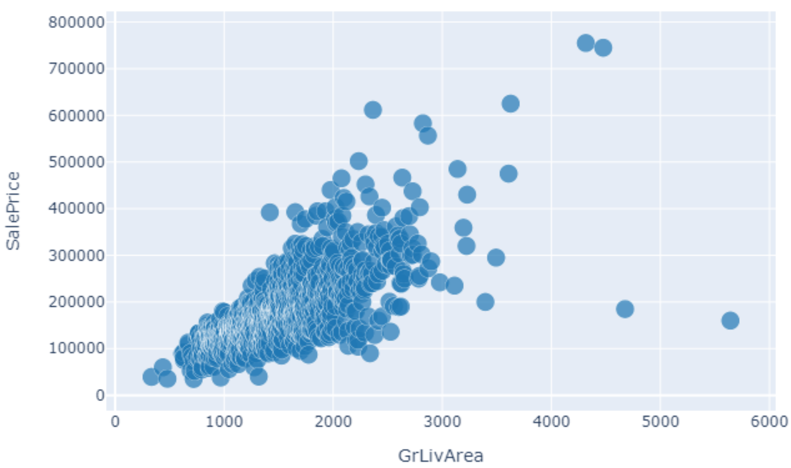
SalePrice by 2ndFlrSF

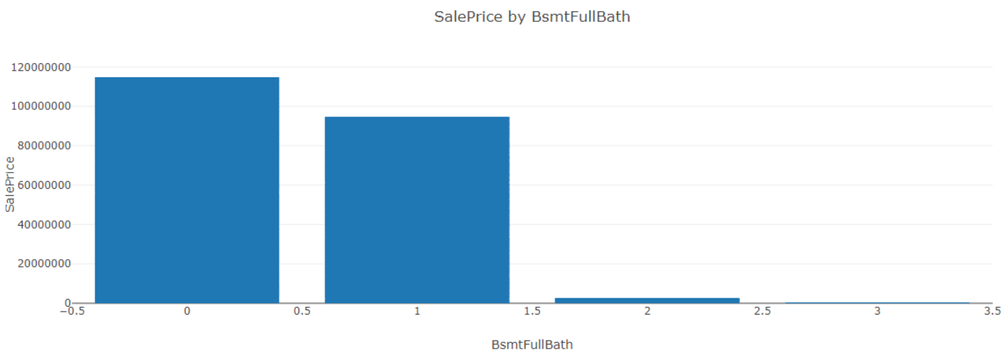


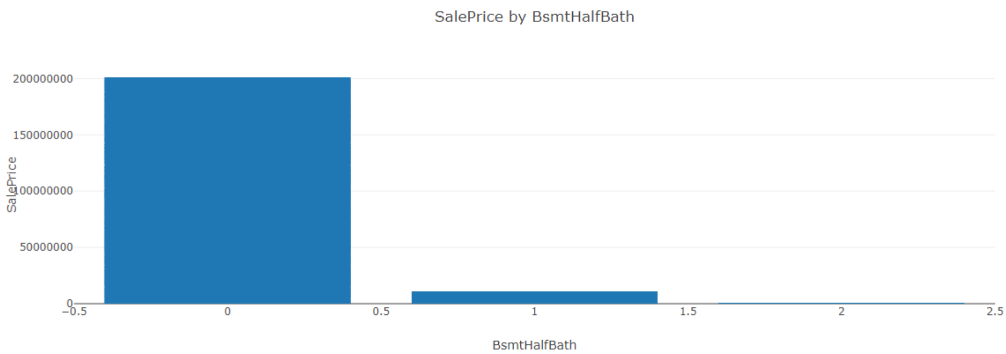
SalePrice by LowQualFinSF

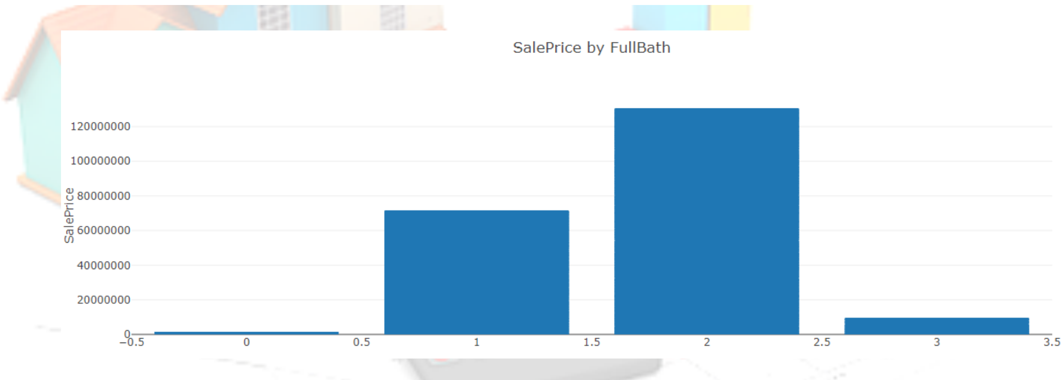


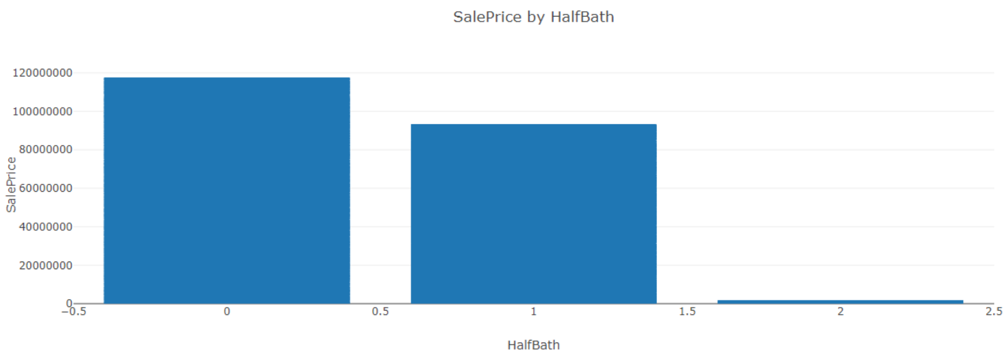
SalePrice by GrLivArea

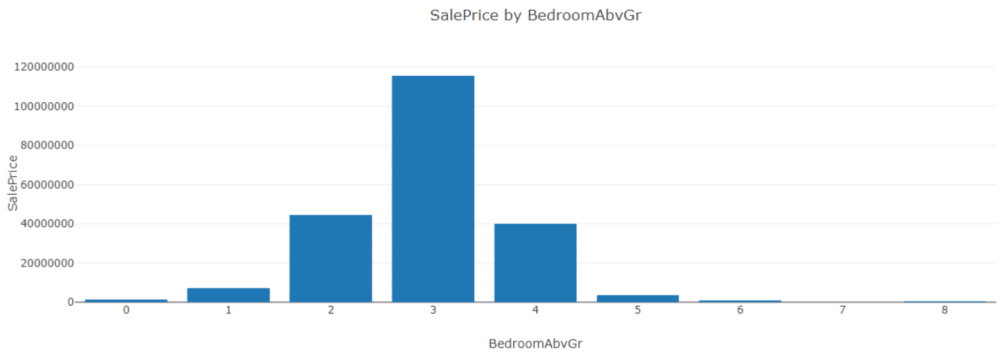


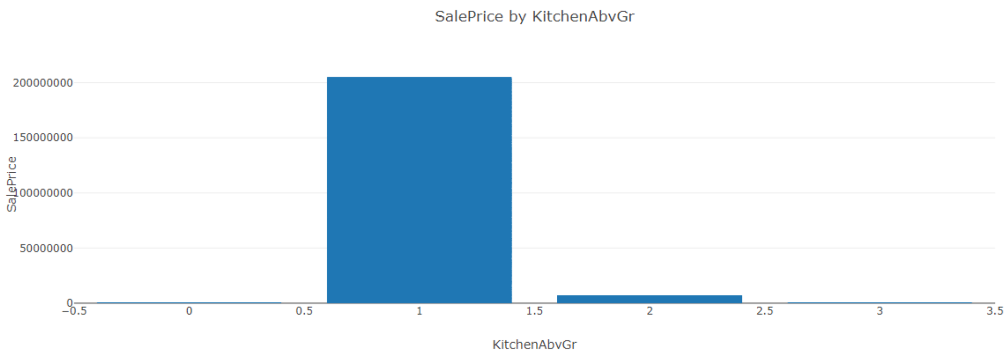

SalePrice by BsmtFullBath
SalePrice
120000000
100000000
80000000
60000000
40000000
20000000
0
−0.5
0
0.5
1
1.5
2
2.5
3
3.5
BsmtFullBath

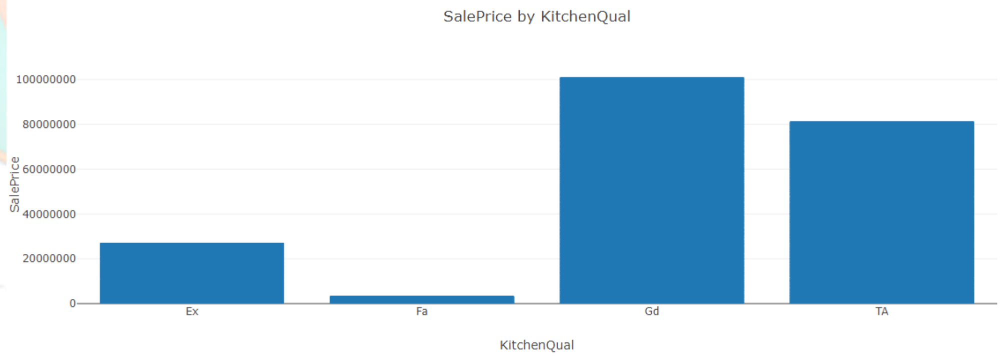

SalePrice by BsmtHalfBath
SalePrice
200000000
150000000
100000000
50000000
0
−0.5
0
0.5
1
1.5
2
2.5
BsmtHalfBath

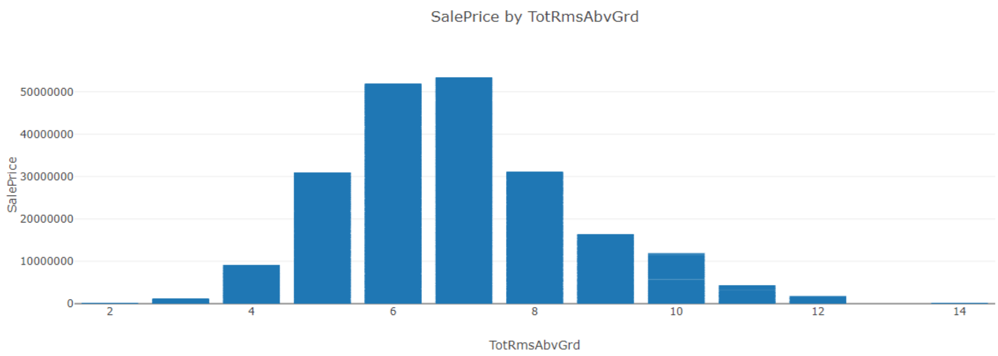

SalePrice by FullBath
SalePrice
120000000
100000000
80000000
60000000
40000000
20000000
0
−0.5
0
0.5
1
1.5
2
2.5
3
3.5

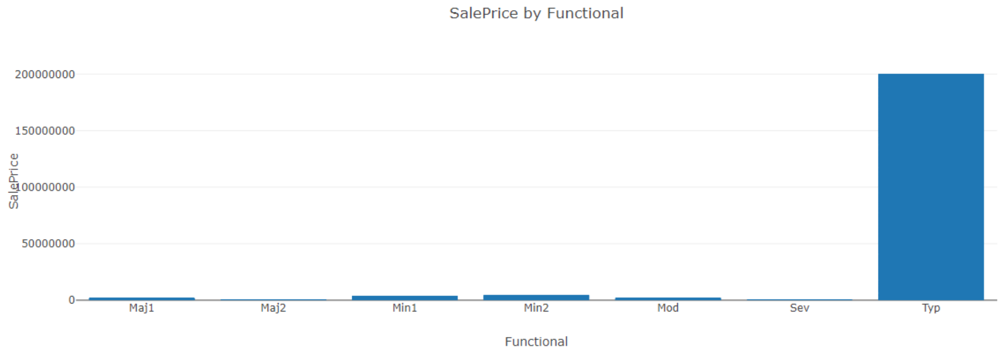

SalePrice by HalfBath
SalePrice
120000000
100000000
80000000
60000000
40000000
20000000
0
−0.5
0
0.5
1
1.5
2
2.5
HalfBath

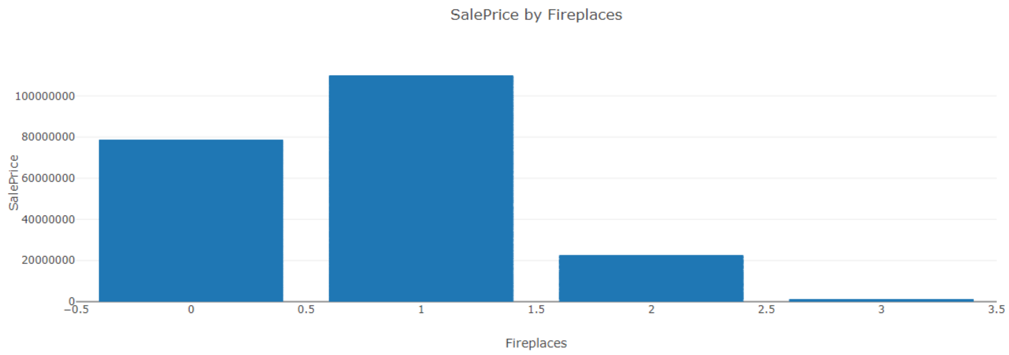

SalePrice by BedroomAbvGr
120000000
100000000
80000000
60000000
40000000
20000000
0
SalePrice
0
1
2
3
4
5
6
7
8
BedroomAbvGr

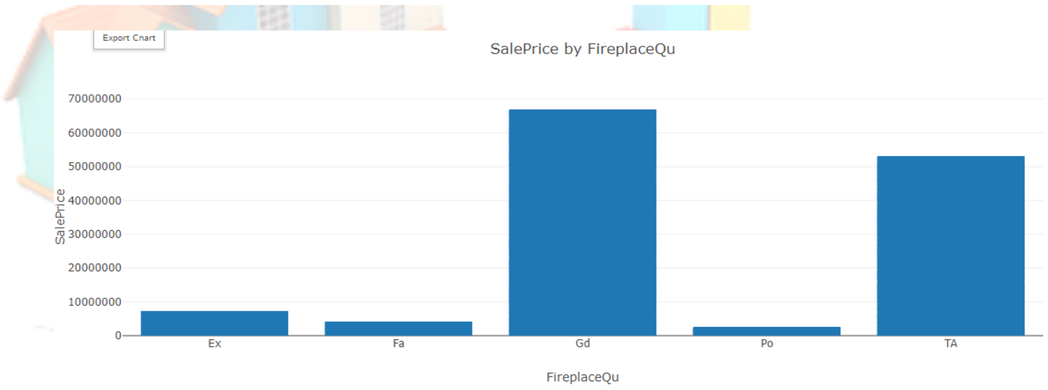

SalePrice by KitchenAbvGr
200000000
150000000
100000000
50000000
0
SalePrice
−0.5
0
0.5
1
1.5
2
2.5
3
3.5
KitchenAbvGr

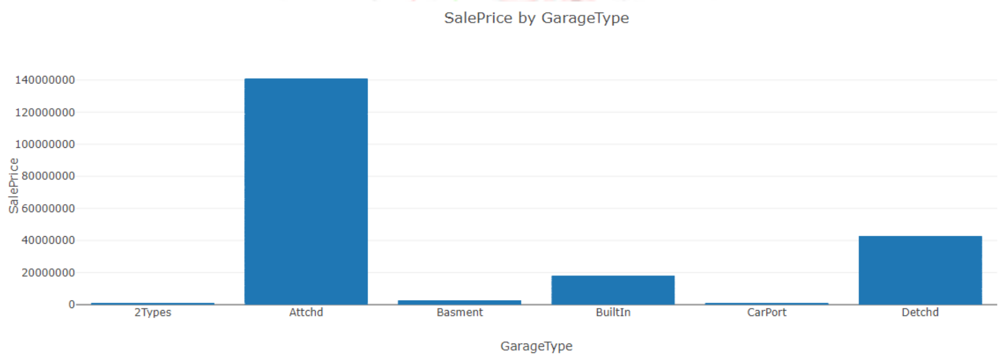

SalePrice by KitchenQual
100000000
80000000
60000000
40000000
20000000
0
SalePrice
Ex
Fa
Gd
TA
KitchenQual

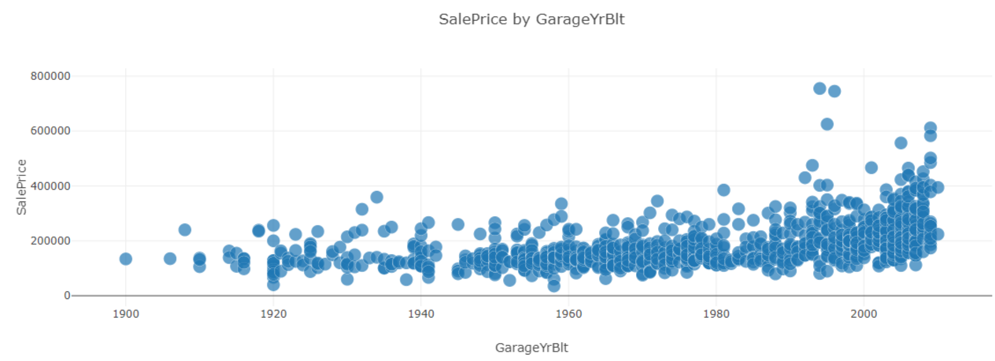

SalePrice by TotRmsAbvGrd
50000000
40000000
30000000
20000000
10000000
0
SalePrice
2
4
6
8
10
12
14
TotRmsAbvGrd

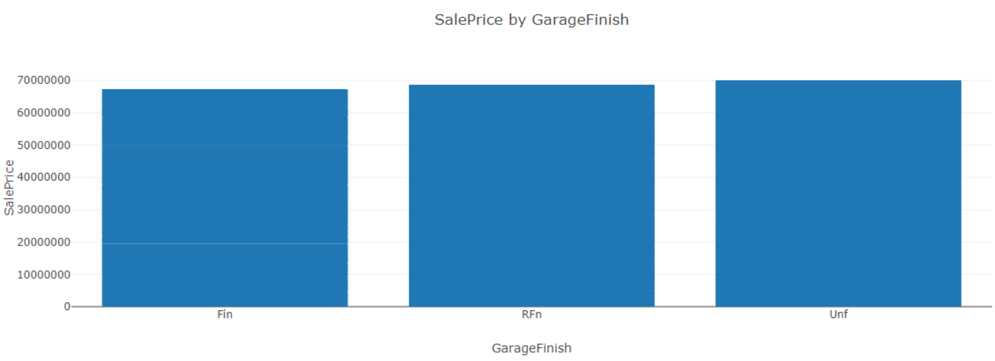

SalePrice by Functional
200000000
150000000
100000000
50000000
0
SalePrice
Maj1
Maj2
Min1
Min2
Mod
Sev
Typ
Functional

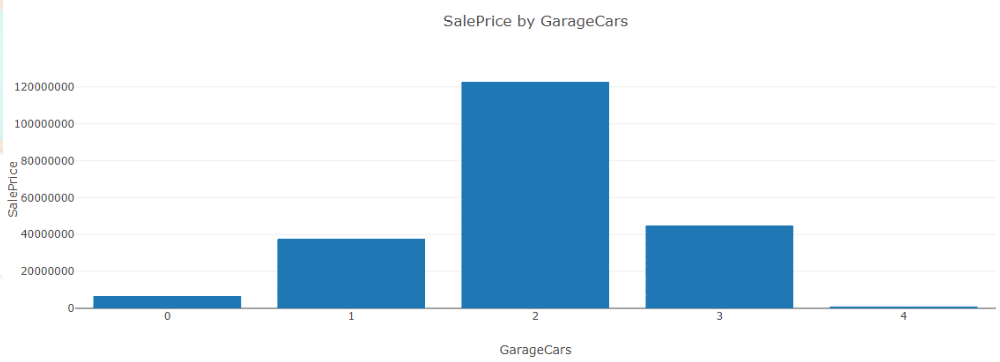

SalePrice by Fireplaces
100000000
80000000
60000000
40000000
20000000
0
SalePrice
−0.5
0
0.5
1
1.5
2
2.5
3
3.5
Fireplaces

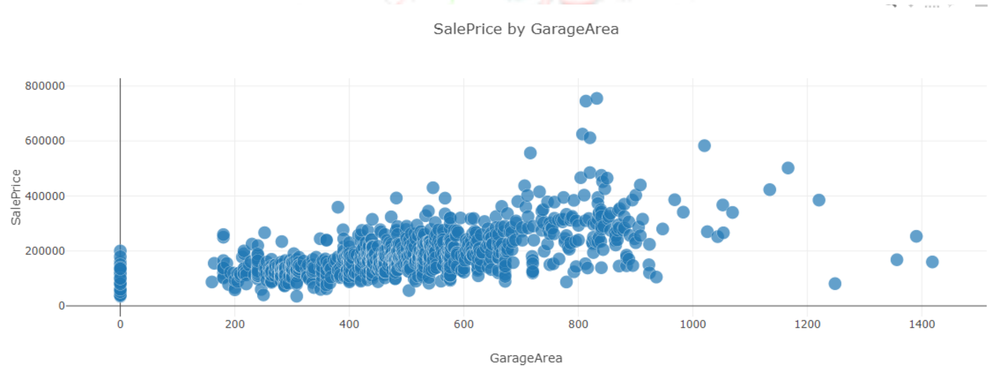

Export Chart
SalePrice by FireplaceQu
70000000
60000000
50000000
40000000
30000000
20000000
10000000
0
SalePrice
Ex
Fa
Gd
Po
TA
FireplaceQu

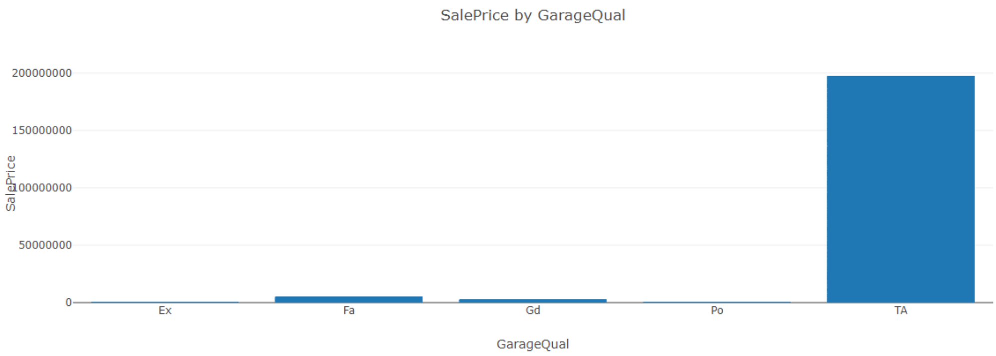

SalePrice by GarageType
140000000
120000000
100000000
80000000
60000000
40000000
20000000
0
SalePrice
2Types
Attchd
Basment
BuiltIn
CarPort
Detchd
GarageType

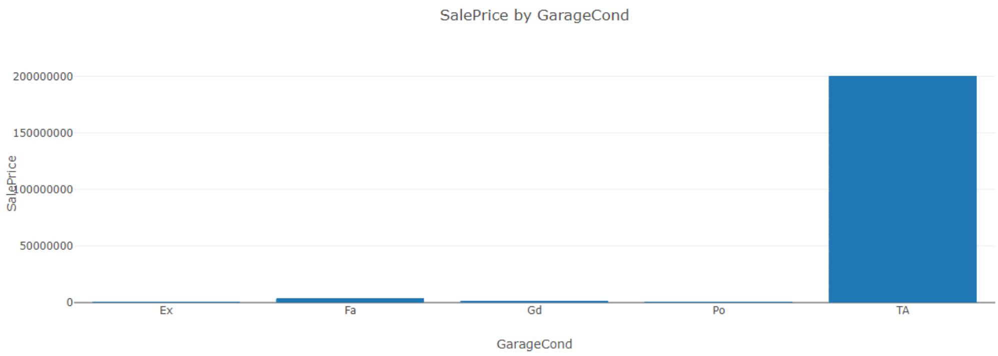

SalePrice by GarageYrBlt
SalePrice
800000
600000
400000
200000
0
1900
1920
1940
1960
1980
2000
GarageYrBlt

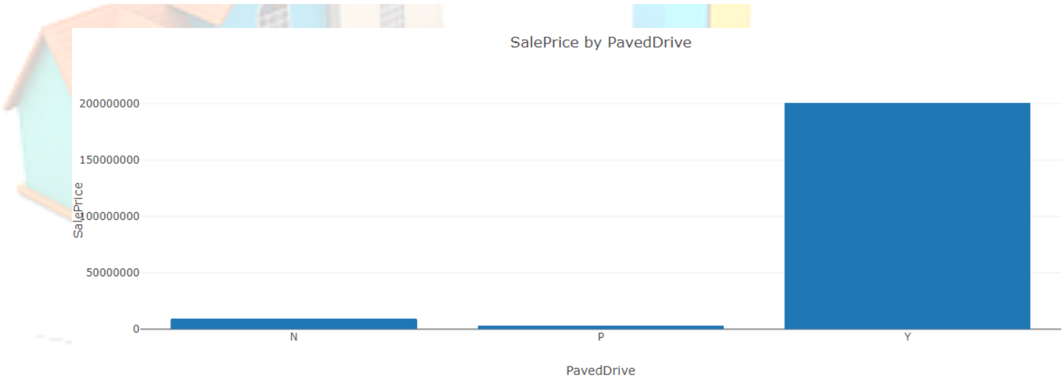

SalePrice by GarageFinish
SalePrice
70000000
60000000
50000000
40000000
30000000
20000000
10000000
0
Fin
RFn
Unf
GarageFinish

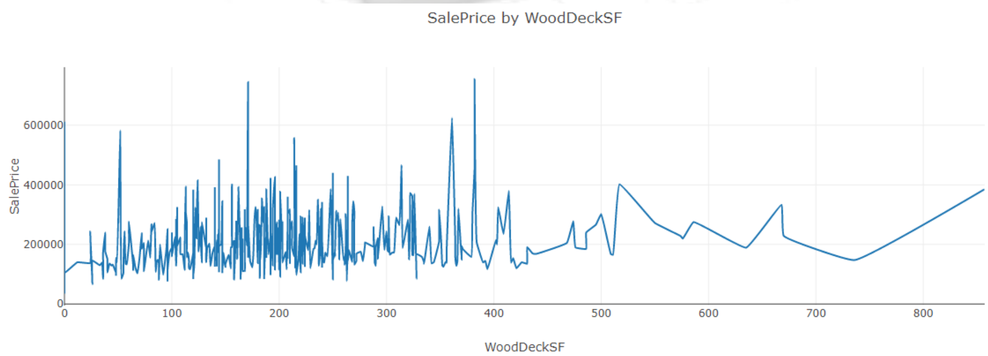

SalePrice by GarageCars
SalePrice
120000000
100000000
80000000
60000000
40000000
20000000
0
0
1
2
3
4
GarageCars

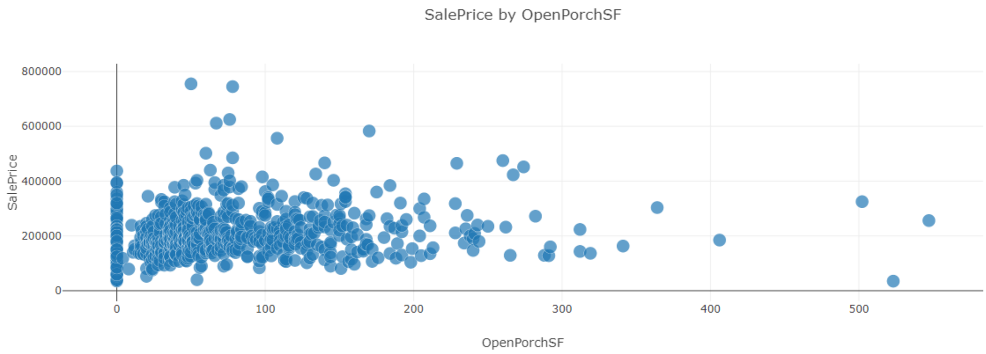

SalePrice by GarageArea
SalePrice
800000
600000
400000
200000
0
0
200
400
600
800
1000
1200
1400
GarageArea

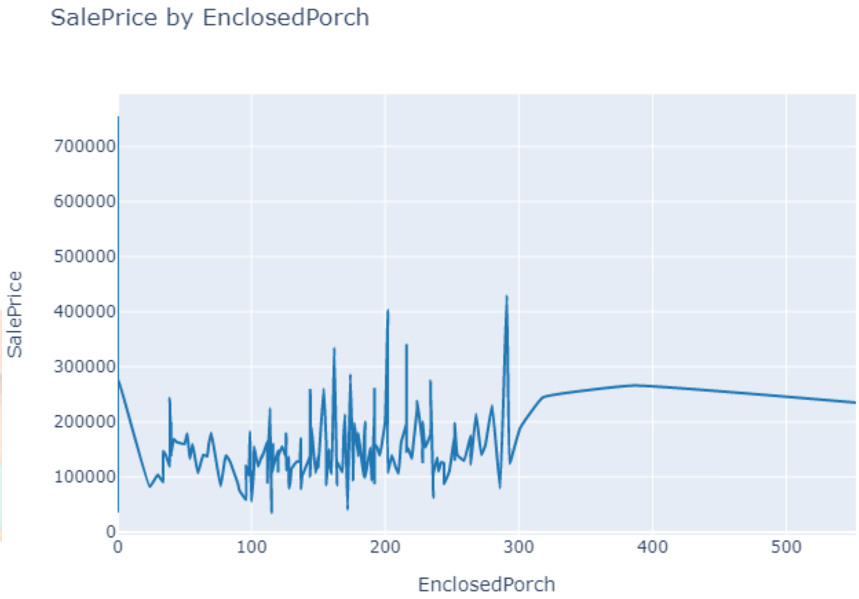

SalePrice by GarageQual
200000000
150000000
100000000
50000000
0
SalePrice
Ex
Fa
Gd
Po
TA
GarageQual

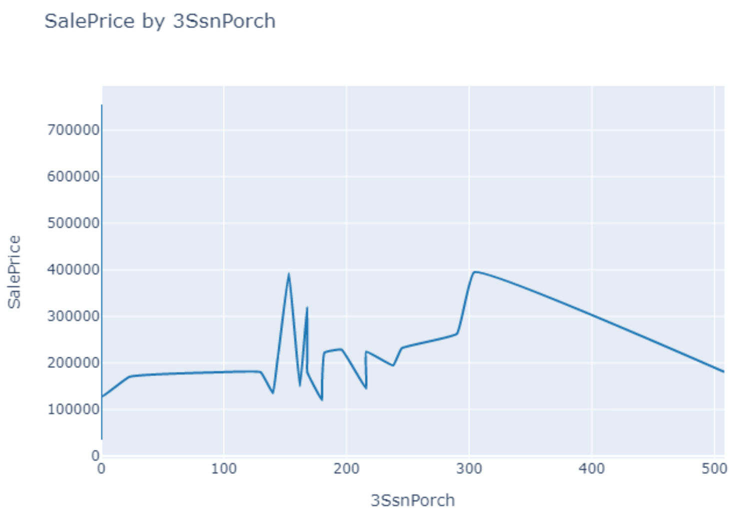

SalePrice by GarageCond
200000000
150000000
100000000
50000000
0
SalePrice
Ex
Fa
Gd
Po
TA
GarageCond

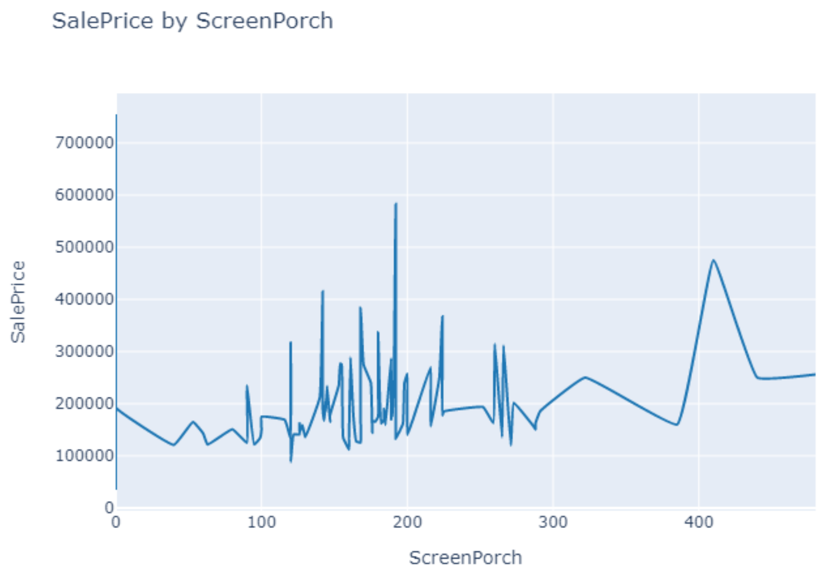

SalePrice by PavedDrive
200000000
150000000
100000000
50000000
0
SalePrice
N
P
Y
PavedDrive

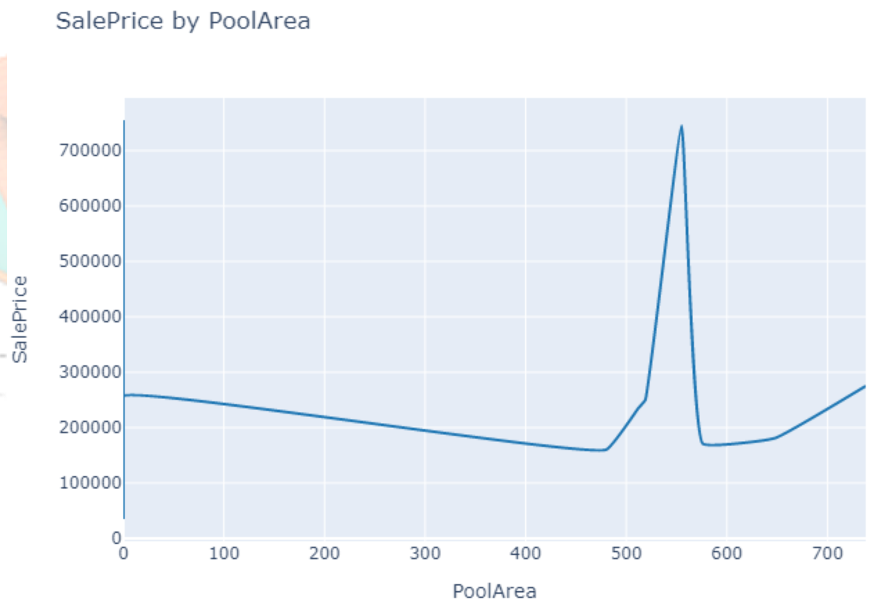

SalePrice by WoodDeckSF
600000
400000
200000
0
SalePrice
0
100
200
300
400
500
600
700
800
WoodDeckSF

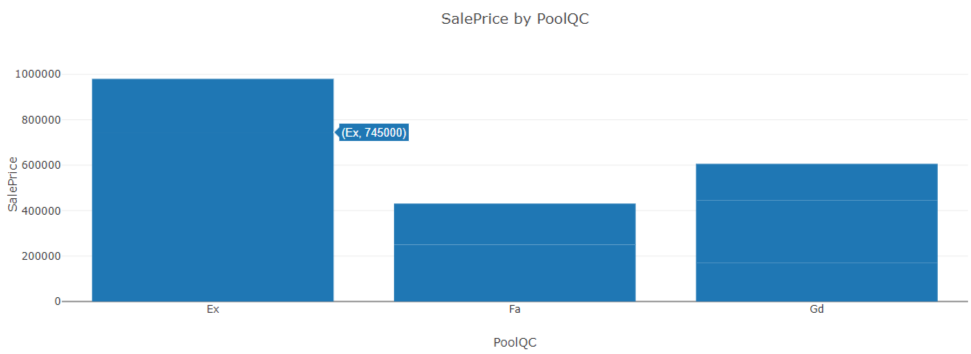

SalePrice by OpenPorchSF
800000
600000
400000
200000
0
SalePrice
0
100
200
300
400
500
OpenPorchSF

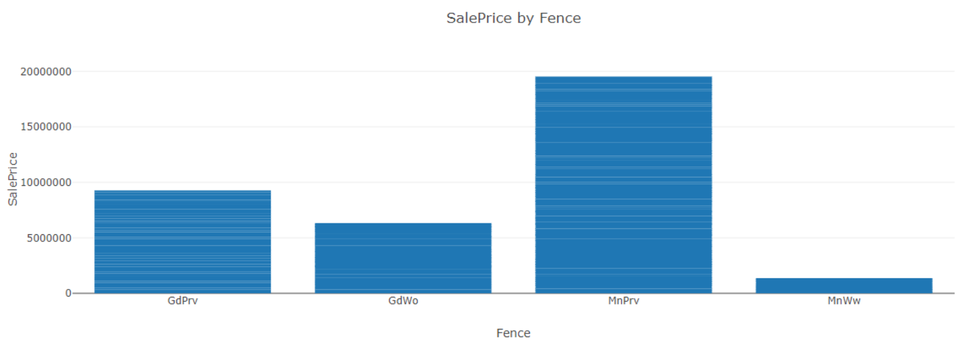

SalePrice by EnclosedPorch
700000
600000
500000
400000
300000
200000
100000
0
SalePrice
0
100
200
300
400
500
EnclosedPorch

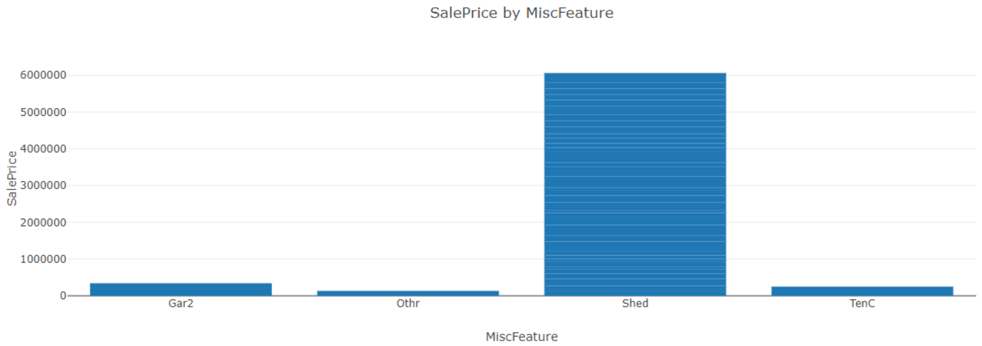

SalePrice by 3SsnPorch
700000
600000
500000
400000
300000
200000
100000
0
SalePrice
0
100
200
300
400
500
3SsnPorch

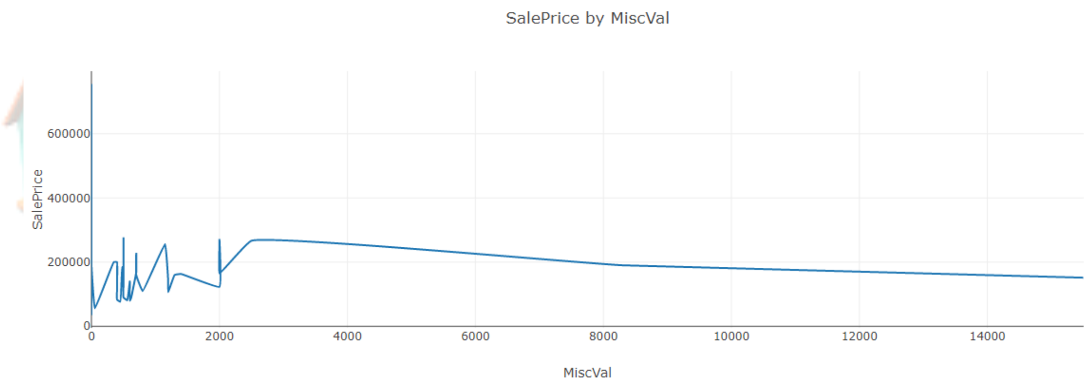

SalePrice by ScreenPorch
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
100
200
300
400
ScreenPorch

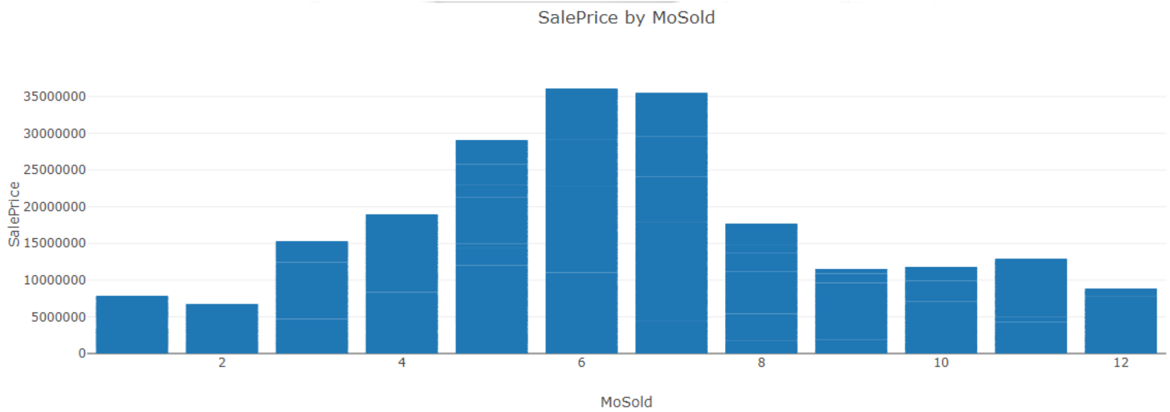

SalePrice by PoolArea
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
100
200
300
400
500
600
700
PoolArea

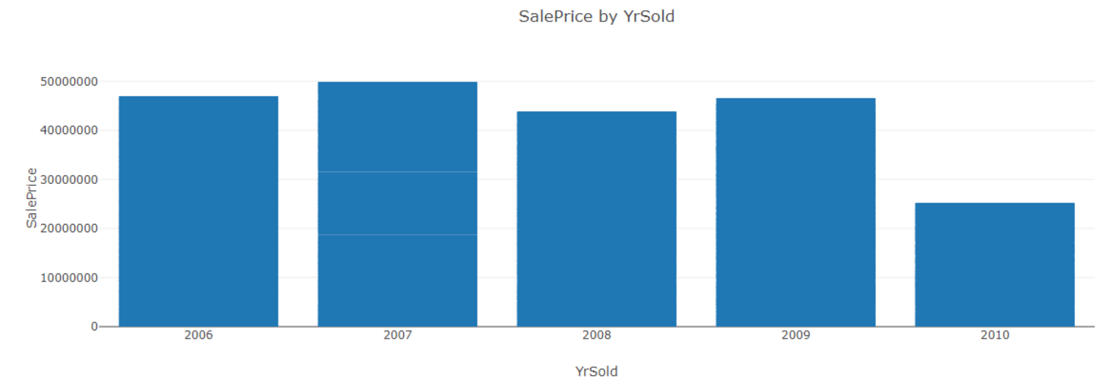

SalePrice by PoolQC
SalePrice
1000000
800000
600000
400000
200000
0
(Ex, 745000)
Ex
Fa
Gd
PoolQC

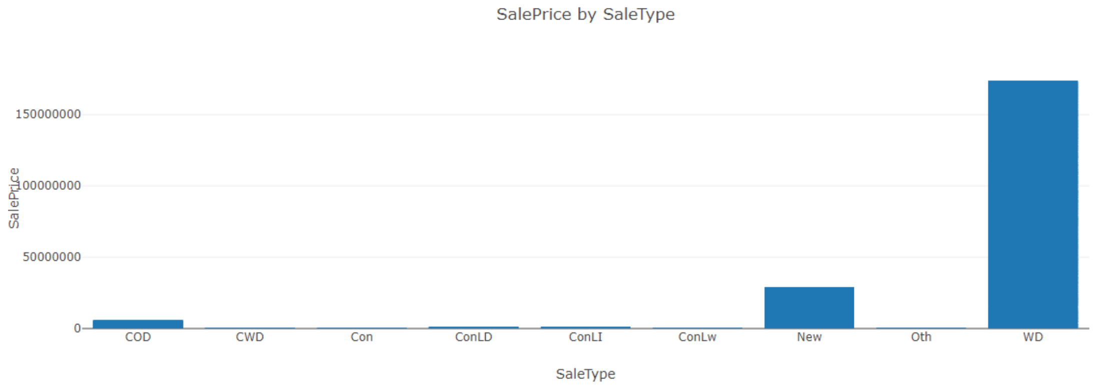

SalePrice by Fence
20000000
15000000
10000000
5000000
0
SalePrice
GdPrv
GdWo
MnPrv
MnWw
Fence

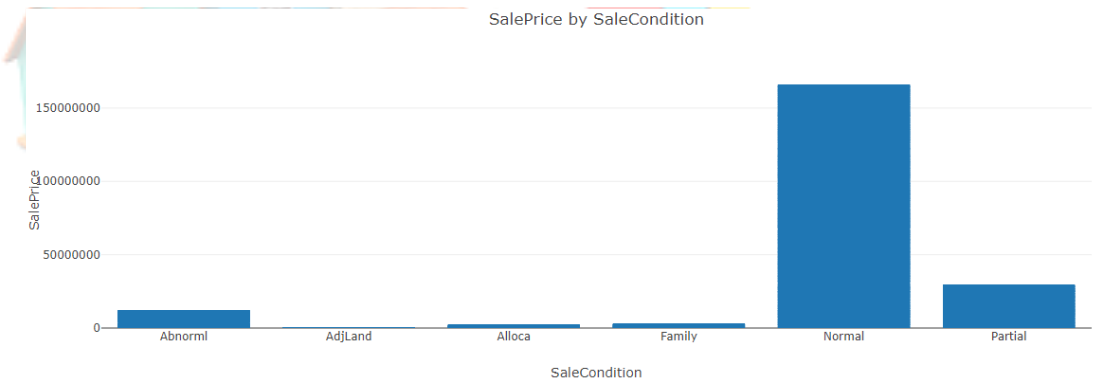

SalePrice by MiscFeature
6000000
5000000
4000000
3000000
2000000
1000000
0
SalePrice
Gar2
Othr
Shed
TenC
MiscFeature


SalePrice by MiscVal
600000
400000
200000
0
SalePrice
0
2000
4000
6000
8000
10000
12000
14000
MiscVal


SalePrice by MoSold
35000000
30000000
25000000
20000000
15000000
10000000
5000000
0
SalePrice
2
4
6
8
10
12
MoSold







- State the set of assumptions (if any) related to the problem under consideration

  Here, you can describe any presumptions taken by you.

  Answer: The columns having more than 50% NAN values have been neglected.

```
In [1372]: dt1.shape
Out[1372]: (1168, 81)
```

As more than 50% have NAN values Alley, PoolQC, Fence, MiscFeature can be neglected

```
In [1373]: dt1.drop(columns=['Utilities','Alley', 'PoolQC', 'Fence', 'MiscFeature'],inplace=True)

In [1374]: dt.drop(columns=['Utilities','Alley', 'PoolQC', 'Fence', 'MiscFeature'],inplace=True)

In [1375]: dt['Electrical'].fillna(dt['Electrical'].mode()[0],inplace=True)
```

- Hardware and Software Requirements and Tools Used

  Listing down the hardware and software requirements along with the tools, libraries and packages used. Describe all the software tools used along with a detailed description of tasks done with those tools.

  The different libraries and packages used are:

  1. Pandas, 2. Numpy, 3. Matplotlib, 4. Sklearn and 5.Dtale etc.

  Pandas: for importing the dataset

  Matplotlib and Dtale: For graphing

  Sklearn: Modelling

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  Describe the approaches you followed, both statistical and analytical, for solving of this problem.

- Testing of Identified Approaches (Algorithms)

  Listing down all the algorithms used for the training and testing.

- Run and Evaluate selected models

  Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

- Key Metrics for success in solving problem under consideration

  What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

- Visualizations

Mention all the plots made along with their pictures and what were the inferences and observations obtained from those. Describe them in detail.

If different platforms were used, mention that as well.

- Interpretation of the Results

Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

CONCLUSION

- Key Findings and Conclusions of the Study

Describe the key findings, inferences, observations from the whole problem.

Answer: Form the visualization we can see that different groups having like RL, Pave, Gravel in alley, level of the land, location of the property, garage location etc. decides the price of the house.

- Learning Outcomes of the Study in respect of Data Science

List down your learnings obtained about the power of visualization, data cleaning and various algorithms used. You can describe which algorithm works best in which situation and what challenges you faced while working on this project and how did you overcome that.

Answer: The 80 columns are all important when purchasing a house. The visualization by bar and line graphs of bivariate analysis gives a clear picture of different types of accessories of house to the selling price. Data cleaning is very important as the data many contain NAN and junk entries which doesn't given any information. When 80 columns are provided the greatest challenge is to understand the column which gives picture to provide an ordinal encoding in model selection. This was the greatest challenge.

- Limitations of this work and Scope for Future Work

What are the limitations of this solution provided, the future scope? What all steps/techniques can be followed to further extend this study and improve the results.

Answer: The solution provided only has the accuracy of 82 approximately by using neural networks this accuracy can be increased