Importing Libaries

```
from statsmodels.stats.outliers_influence import variance_inflation_factor
                                                                                                                                                                                                        from sklearn.model_selection import train_test_split
                                                                                                                                                                  from scipy.stats import chi2_contingency
                                                                         pd.options.display.max_columns=150
                                                                                          pd.options.display.max_columns=200
                                                                                                                                                                                                                                                            warnings.filterwarnings("ignore")
%matplotlib inline
                                 import seaborn as sns
import matplotlib.pyplot as plt
                                                                                                                                import statsmodels.api as sm
                                                                                                                                                                                      from scipy.stats import chi2
                                                                                                                                                                                                                         from sklearn import metrics
                                                                                                              from scipy import stats
import pandas as pd
                  import numpy as np
                                                                                                                                                                                                                                                import warnings
                                                                                          6
110
111
111
113
113
115
115
In [40]:
```

Retreving dataset

test=pd.read_csv(r"C:\DsTraining\five dataset for clening\house price\test.c test_sample=pd.read_csv(r"C:\DsTraining\five dataset for clening\house price train_data=pd.read_csv(r"C:\DsTraining\five dataset for clening\house price\ LotArea Street Alley LotShape LandContour Reg Reg 쮼 Reg Reg Reg Reg <u>8</u> 7 NaN data=pd.concat([train_data,test],ignore_index=True) Pave Pave Pave Pave Pave Pave Pave Pave Pave sale price predication - Jupyter Notebook 8450 0096 11250 14260 20000 10441 9550 1936 1894 9627 test=pd.merge(test,test_sample,on="Id") Id MSSubClass MSZoning LotFrontage 65.0 80.0 68.0 0.09 84.0 21.0 21.0 160.0 62.0 RM R Σ 씸 귐 꿉 귐 씸 R 20 160 160 9 20 9 9 20 85 9 2915 2915 2916 **2916** 2917 2917 2918 2918 2919 data 2914 0 In [41]: Out[41]: 12/19/22, 7:56 PM

2 2 Ξ 3 >

2 Ξ 3

>

≥

Data Preprocessing and EDA.

2919 rows × 81 columns

```
float64
                                                                                          object
                                                                                                                        object
                                                                                                                                  object
                                                                                                                                           object
object
object
object
object
object
                                                                                int64
                                                                                                              int64
                                                  Dtype
                                                                      int64
                    <class 'pandas.core.frame.DataFrame'>
                              RangeIndex: 2919 entries, 0 to 2918
                                                  Non-Null Count
                                                                                                  2433 non-null
2919 non-null
2919 non-null
                                                                                                                                                                                  2919 non-null
2919 non-null
                                                                      2919 non-null
                                                                                2919 non-null
                                                                                          2915 non-null
                                                                                                                                            2919 non-null
                                                                                                                                                                         2919 non-null
                                                                                                                                                                                                      2919 non-null
                                                                                                                                                      2919 non-null
                                                                                                                                                                non-nul]
                                        Data columns (total 81 columns):
                                                                                                                                  198 non-null
                                                                                                                                                               2917
                                                                                                                                                                                             Neighborhood
                                                                                                    LotFrontage
                                                                                                                                                      LandContour
1 data.info()
                                                                               MSSubClass
                                                                                                                                                                                                       Condition1
                                                                                                                                                                Utilities
                                                                                                                                                                          LotConfig
                                                                                                                                                                                    LandSlope
                                                                                          MSZoning
                                                                                                                                            LotShape
                                                                                                              LotArea
                                                                                                                        Street
                                                  Column
                                                                                                                                  Alley
                                                                      ы
                                                                                                                                                                         10
11
12
13
                                                                                                                                                               6
                                                                      0
                                                                                                                        2
                                                                                                                                  9
In [42]:
```

1/15

12/19/22, 7:56 PM

Step-1 find missing values.

```
r_rows=train_data.isnull().sum()
rows=train_data.shape[0]
                                                                                               0.000000
17.739726
0.000000
                                                                                                                                     ...
                                                                                      0.000000
                                                                                                                                                        0.000000
                                                                                                                                                                  0.000000
                                                                                                                                                                             0.000000
                                                                           0.000000
                                          a=r_rows/rows*100
                                                                                                                                                                            SaleCondition
                                                                                                           LotFrontage
                                  r_rows
                                                                                     MSSubClass
                                                                                                MSZoning
                                                                                                                                                                  SaleType
                                                                                                                        LotArea
                                                                                                                                            MoSold
                                                                                                                                                       YrSold
In [43]:
                                                                           Out[43]:
```

missing values, and less than 25% missing values Step-2 Drop variables having more than 25% will be fill by using appropriate central endencies

0.000000

SalePrice

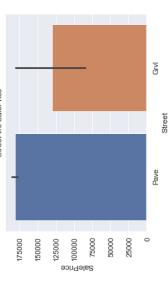
Length: 81, dtype: float64

```
data["BsmtFullBath"]=data["BsmtFullBath"].fillna(data["BsmtFullBath"].mean()
data["BsmtHalfBath"]=data["BsmtHalfBath"].fillna(data["BsmtHalfBath"].mean()
1 | data=data.drop(columns=["Alley","FireplaceQu","PoolQC","Fence","MiscFeature"
                                                                                                                                                                                                           data["GarageYPBIt"]=data["GarageYrBIt"].fillna(data["GarageYrBIt"]_mean())
data["MasVnrArea"]=data["MasVnrArea"].fillna(data["MasVnrArea"].mean())
data["BsmtFinSF1"]=data["BsmtFinSF1"].fillna(data["BsmtFinSF1"].mean())
data["BsmtUnfSF"]=data["BsmtUnfSF"].fillna(data["BsmtUnfSF"].mean())
                                                                                                                                                                                                                                                                                                                                                                                                                  data["TotalBsmtSF"]=data["TotalBsmtSF"].fillna(data["TotalBsmtSF"].mean())
data["BsmtFinSF2"]=data["BsmtFinSF2"].fillna(data["BsmtFinSF2"].mean())
                                                                                                                                                                         data["LotFrontage"]=data["LotFrontage"].fillna(data["LotFrontage"].mean())
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 data["GarageCars"]=data["GarageCars"].fillna(data["GarageCars"].mean())
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               data["GarageArea"]=data["GarageArea"].fillna(data["GarageArea"].mean())
                                                                                                                               #numerical data we used to fill mean.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                         In [44]:
                                                                                                                               In [45]:
```

```
data["Exterior1st"]=data["Exterior1st"].fillna(data["Exterior1st"].mode()[0]
                                                                                                                                                                                                                                                                                                                                                                                                                                                      data["BsmtExposure"]=data["BsmtExposure"].fillna(data["BsmtExposure"].mode()
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       data["BsmtFinType1"]=data["BsmtFinType1"].fillna(data["BsmtFinType1"].mode()
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                data["GarageFinish"]=data["GarageFinish"].fillna(data["GarageFinish"].mode()
                                                                                                                                                                                                                                             data["Exterior2nd"]=data["Exterior2nd"].fillna(data["Exterior2nd"].mode()[0]
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        data["BsmtFinType2"]=data["BsmtFinType2"].fillna(data["BsmtFinType2"].mode()
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         data["KitchenQual"]=data["KitchenQual"].fillna(data["KitchenQual"].mode()[0
                                            data["Electrical"]=data["Electrical"].fillna(data["Electrical"].mode()[0])
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              data["GarageCond"]=data["GarageCond"].fillna(data["GarageCond"].mode()[0])
data["SaleType"]=data["SaleType"].fillna(data["SaleType"].mode()[0])
                                                                                                                                                                                                                                                                                                  data["MasVnrType"]=data["MasVnrType"].fillna(data["MasVnrType"].mode()[0])
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          data["Functional"]=data["Functional"].fillna(data["Functional"].mode()[0])
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            data["GarageType"]=data["GarageType"].fillna(data["GarageType"].mode()[0])
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             data["GarageQual"]=data["GarageQual"].fillna(data["GarageQual"].mode()[0])
                                                                                                                                               data["Utilities"]=data["Utilities"].fillna(data["Utilities"].mode()[0])
                                                                                                                                                                                                                                                                                                                                                        data["BsmtQual"]=data["BsmtQual"].fillna(data["BsmtQual"].mode()[0])
                                                                                                                                                                                                                                                                                                                                                                                                     data["BsmtCond"]=data["BsmtCond"].fillna(data["BsmtCond"].mode()[0])
                                                                                                  data["MSZoning"]=data["MSZoning"].fillna(data["MSZoning"].mode()[0])
    #chracteries data we used to mode or so on.
                                                                                                                                                                                                                                                                                                                                                                                                                                                      10
11
11
11
12
13
14
17
17
17
18
In [46]:
```

EDA

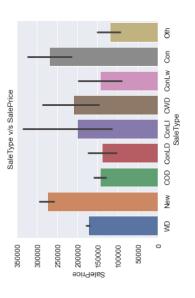
```
_{
m 1} #In this graph we see that in street "Pave(Paved)" is more higher as compare
                             sns.barplot(y="SalePrice",x="Street",data=train_data)
plt.title("Street v/s SalePrice")
                                                                                                                                                                                        Street v/s SalePrice
                                                                                                           plt.show()
   In [47]:
```



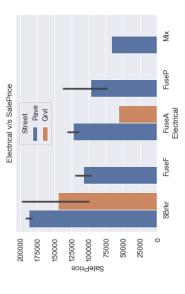
localhost:8888/notebooks/projects/sale price predication ipynb#Importing-Libaries

In [48]:

12/19/22, 7:56 PM



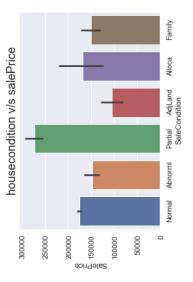
#we observed that in electrical(SBrkr) "street" price rate is very high.
sns.barplot(data=train_data,x="Electrical",y="SalePrice",hue="Street")
s plt.title("Electrical v/s SalePrice")
4 plt.show() In [49]:



12/19/22, 7:56 PM

sale price predication - Jupyter Notebook

1 # We observed that in salecondition (partial) saleprice is very high.
2 sns.barplot(data=train_data, y="SalePrice", x="SaleCondition")
3 plt.title("housecondition v/s salePrice", size=18)
4 plt.show() In [50]:

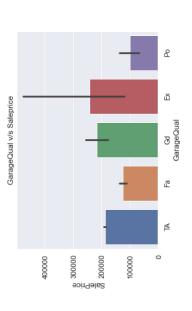


1 #In this graph were we observed saleprice is Less than privious(2009) year.
2 sns.lineplot(data=train_data, x="YrSold", y="SalePrice") In [51]:

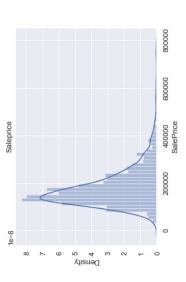


localhost:8888/notebooks/projects/sale price predication.ipynb#Importing-Libaries

In [52]: 1 #if your Garage quality is excellent and good so the sale price of the house
2 sns.barplot(x="GarageQual",y="SalePrice",data=train_data)
3 plt.title("GarageQual v/s Saleprice")
4 plt.show()













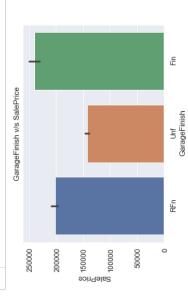
듐

New COD ConLD ConLI CWD ConLw Con

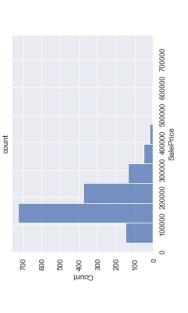
Ø

200

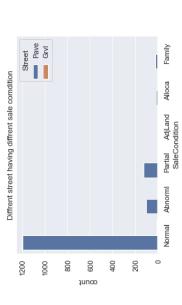
400



#we obseved that maximumm pepole buy houses in the range of 2Lakh. sns.histplot(train_data["SalePrice"],kde=False,bins=10) plt.title("count") plt.show() In [56]:







12/19/22, 7:56 PM

sale price predication - Jupyter Notebook

```
-0.25
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          -0.50
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            -0.75
                                                                                                                                                                                              1.00
                                                                                                                                                                                                                                                                   0.75
                                                                                                                                                                                                                                                                                                                                     0.50
                                                                                                                                                                                                                                                                                                                                                                                                          0.25
                                                                                                                                                                                                                                                                                                                                                                                                                                                                            - 0.00
                                                                  sns.heatmap(data.corr(),vmin=1,vmax=-1,annot=True,cmap="Spectral")
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          0.00743.7e-05
                                                                                                                                                                                                            0089-0.041-0.0028-0.0250.00015-0.015-0.035-0.0071-0.01-0.022-0.047-0.022-0.014-0.00820.0064-0.012
                                                                                                                                                                                                                                                                              0.1 0.021 0.016 0.054 0.094 0.069 0.0042 0.3
                                                                                                                                                                                                                                                                                                              0.069 0.071 0.044 0.044 -0.017 0.034 -0.0063-0.066
                                                                                                                                                                                                                                                                                                                                           0.14 -0.11 0.014 0.065 0.0045 0.0450.00012 0.35
                                                                                                                                                                                                                                                                                                                                                                             0.081 -0.068 0.027 0.053 0.044-0.0046-0.0036 0.15
                                                                                                                                                                                                                                                                                                                                                                                                                0.035-0.0097 0.027 0.042 0.067 0.037 0.023 -0.013
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            0.0094 0.048 0.064 0.077 0.034 0.24
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              1.033 -0.064 0.093 0.0087 -0.021 -0.071
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           0.03 -0.00640.000790.027 0.03
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           0.026 0.0071 0.029 0.082
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           0.012 -0.042 0.081
                                                                                                                                                                                                                                                                                                                                                                                                                                             0.18 -0.082 -0.023 0.036 0.0014 0.027 -0.0011 0.23
                                                                                                                                                                                                                                                                                                                                                                                                                                                                            0.099 0.0085 0.032 0.35
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            -0.12 -0.0039-0.052 0.094 0.057 0.018
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          0.0064-0.00120.0042-0.00630.000130.0036 0.023 -0.0011 0.032 0.018 0.034 -0.021 0.027 0.029 -0.042 0.007
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          0.077 0.00870.00079.0071 0.012
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           0.094 0.064 0.093 -0.0065 0.026
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           0.022 -0.049 0.054 0.044 0.065 0.053 0.042 0.036 0.17 -0.052 0.048 -0.064 -0.03
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             0.03
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          0.071
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           0.044 0.014 0.027 0.027 -0.023 0.019 -0.00390.0094-0.033
                                     num_col=data.select_dtypes(exclude='object')
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            -0.11 -0.068-0.0097-0.0820.00097-0.12 -0.06
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       0.24
                                                                                                                                                                                                                                                                                                                                                                                                                                             0.12
                                                                                                                                                                                                                                                                                                                                                                                                                                                                            0.23
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          0.0082 -0.029 0.069 0.034 0.045 -0.0046 0.037 0.027 0.0085 0.057
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             0.24
sns.set(rc={"figure.figsize":(15,8)})
                                                                                                                                                                                                                                                                                                                                                                                                          0.058 0.039
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          0.35
                                                                                                                                                                                                                                                                              0.13 0.13 0.026 0.034 0.26
                                                                                                                                                                                                                                                                                                              0.042 0.084 -0.089 -0.031
                                                                                                                                                                                                                                                                                                                                           0.14 0.015 0.19 0.27
                                                                                                                                                                                                                                                                                                                                                                                -0.033 0.17
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               0.14 0.081 -0.035 0.18 0.16
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           -0.017 0.0045 0.044 0.067 0.0014 0.099
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            0.19 0.052 0.12
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          -0.013 0.23
                                                                                                                                                                                                                                                                                                                                                                                                                                                                            0.17 0.039
                                                                                                                                                                                                                                                                                                                                                                                                                                             0.19 -0.033 -0.058
                                                                                                                                                                                                                                                                                                                                                                                                                0.015
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             0.35
                                                                                                                                                                                                                                                                                                                                                                             BsmtFullBath 0.000150.0099 0.13 -0.042
                                                                                                                                                                                                                                                                                                                                                                                                                                                                            0.035 -0.055 0.26 -0.031
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            0.02
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             0.066
                                                                                                                                                                                                                                                                                                                                                                                                                                             -0.015 0.18 0.034 -0.089
                                                                                                                                                                                                                                                                                                                                                 -0.025 0.0054 0.13
                                                                                                                                                                                                                                                                                                                                                                                                                0.01 -0.0019 0.026
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               0.0071-0.018 0.16
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              0.022 -0.021 0.021
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           -0.047 -0.038 0.016
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           0.014 -0.0031 0.094
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          -0.012 -0.088 0.3
                                                                                                                                                                                                                                                                                                                 0.0028-0.066-0.036
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            0.01 -0.016 0.1
                                                                                                                                                                                                                                                                              0.041 -0.2
                                                                                                 plt.show()
                                                                                                                                                                                                                                                                              LotArea
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           reenPorch
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           PoolArea
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          MiscVal
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       MoSold
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          SalePrice
                                                                                                                                                                                                                                                                                                                                                 MasVnrArea
                                                                                                                                                                                                                                                                                                                                                                                                                3smtHalfBath
                                                                                                                                                                                                                                                                                                                 OverallCond
                                                                                                                                                                                                                                            MSSubClass
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            oodDeckSF
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               penPorchSF
  In [60]:
```

Multicollinary(using VIF method)

```
data=data.drop(columns=["BsmtFinSF1","LotFrontage","GarageArea","BedroomAbvG
                                                                             vif_data["VIF"]=[variance_inflation_factor(num_col.values,i)
for i in range(len(num_col.columns))]
                                                                                                                                                                                                                                                                                                                                 #variable having range 5 to 10 showa high multicolinarity.
                                    vif_data["feature"]=num_col.columns
                                                                                                                                                                                                                                                                                                #if vif<10 we drop that variable
vif_data=pd.DataFrame()
                                                                                                                                                                                                           vif_data.round(1)
In [59]:
```

Outlier Detection (Extream end outliers will be removed.)

```
1 data=data.drop(columns=['LotArea','MasVnrArea','WoodDeckSF','OpenPorchSF','E
In [21]:
```

9/15

<class 'pandas.core.frame.DataFrame'> RangeIndex: 2919 entries, 0 to 2918

1 data.info()

In [22]:

12/19/22, 7:56 PM

Non-Null Count

Data columns (total 50 columns):

Column ы

2919 non-null

2919 non-null 2919 non-null 2919 non-null 2919 non-null 2919 non-null 2919 non-null 2919 non-null 2919 non-null 2919 non-null 2919 non-null

MSSubClass

MSZoning LotShape

Street

object object

int64

12/19/22, 7:56 PM

Fill appropriate outlier using z_score method

```
1 | upper_limit=data['BsmtFullBath'].mean()+3*data['BsmtFullBath'].std()
                                    lower_limit=data['MSSubClass'].mean()-3*data['MSSubClass'].std()
data['MSSubClass'] = np.where(
   data['MSSubClass']>upper_limit,
1 | upper_limit=data['MSSubClass'].mean()+3*data['MSSubClass'].std()
                                                                                                                                       data['MSSubClass']<0,
                                                                                                                                                                             data['MSSubClass']
                                                                                              upper_limit,
                                                                                                                     np.where(
                                                                               5
6
7
7
8
8
8
11
11
12
12
In [61]:
                                                                                                                                                                                                                                                                          In [62]:
```

```
lower_limit=data['BsmtFullBath'].mean()-3*data['BsmtFullBath'].std()
data['BsmtFullBath'] = np.where(
                                  data['BsmtFullBath']>upper_limit,
                                                                                            data['BsmtFullBath']<0,</pre>
                                                                                                                                   data['BsmtFullBath']
                                                      upper_limit,
                                                                             np.where(
                                                                                                                                   110
```

object object object object object object object object object

2919 non-null

2919 non-null 2919 non-null 2919 non-null

Exterior1st Exterior2nd

MasVnrType

ExterQual ExterCond

2919 non-null 2919 non-null 2919 non-null

int64

Overallcond

RoofStyle

RoofMat1

HouseStyle

BldgType

2919 non-null 2919 non-null 2919 non-null 2919 non-null 2919 non-null

Neighborhood

Condition1 Condition2

_andContour

Utilities

LotConfig LandSlope

```
1 upper_limit=data['BsmtHalfBath'].mean()+3*data['BsmtHalfBath'].std()
                                       lower_limit=data['BsmtHalfBath'].mean()-3*data['BsmtHalfBath'].std()
                                                                                 data['BsmtHalfBath']>upper_limit,
                                                                data['BsmtHalfBath'] = np.where(
                                                                                                                                                  data['BsmtHalfBath']<0,
                                                                                                                                                                                            data['BsmtHalfBath']
                                                                                                        upper_limit,
                                                                                                                             np.where(
                                                                                                                                                8
9
11
12
12
In [63]:
```

float64 float64

BsmtFullBath BsmtHalfBath

CentralAir

HeatingQC

Heating

Electrical

object object object object object object

KitchenQual

HalfBath

Functional Fireplaces GarageType GarageQual GarageCond PavedDrive

int64

int64

object

2919 non-null 2919 non-null 2919 non-null

GarageFinish

int64 int64 int64

2919 non-null 2919 non-null 2919 non-null

3SsnPorch

PoolArea

MiscVal

object

object object object object object object

> 2919 non-null 2919 non-null

BsmtExposure BsmtFinType2

BsmtFinType1

2919 non-null 2919 non-null

Foundation

BsmtQual BsmtCond localhost:8888/notebooks/projects/sale price predication.ipynb#Importing-Libaries

12/19/22, 7:56 PM

```
upper_limit=data['Fireplaces'].mean()+3*data['Fireplaces'].std()
                               lower_limit=data['Fireplaces'].mean()-3*data['Fireplaces'].std()
                                                data['Fireplaces'] = np.where(
   data['Fireplaces']>upper_limit,
                                                                                                               data['Fireplaces']<0,
                                                                                                                                              data['Fireplaces']
                                                                                 upper limit,
                                                                                                  np.where(
                                                                    In [64]:
```

```
1 | upper_limit=data['3SsnPorch'].mean()+3*data['3SsnPorch'].std()
                                lower_limit=data['3SsnPorch'].mean()-3*data['3SsnPorch'].std()
data['3SsnPorch'] = np.where(
                                                                      data['3SsnPorch']>upper_limit,
                                                                                                                     data['3SsnPorch']<0,
                                                                                                                                                      data['3SsnPorch']
                                                                                     upper_limit,
                                                                                                      np.where(
                                                                      5
6
7
7
11
11
12
12
  In [65]:
```

```
____upper_limit=data['PoolArea'].mean()+3*data['PoolArea'].std()
                                   lower_limit=data['PoolArea'].mean()-3*data['PoolArea'].std()
data['PoolArea'] = np.where(
    data['PoolArea']>upper_limit,
                                                                                                                 data['PoolArea']<0,
                                                                                                                                           data['PoolArea']
                                                                                upper_limit,
np.where(
                                                                   5
6
7
7
8
8
8
9
10
11
11
12
      In [66]:
```

```
1 | upper_limit=data['MiscVal'].mean()+3*data['MiscVal'].std()
                                lower_limit=data['MiscVal'].mean()-3*data['MiscVal'].std()
                                                              data['MiscVal']>upper_limit,
                                                 data['MiscVal'] = np.where(
                                                                                                                data['MiscVal']<0,
                                                                                                                                               data['MiscVal']
                                                                                upper_limit,
                                                                                                 np.where(
                                                                 5
6
7
7
8
8
8
9
10
11
11
12
  In [67]:
```

```
In [ ]:
```

All object converted into integers.

```
object_list=data.select_dtypes(include=['object']).columns
1 | from sklearn.preprocessing import LabelEncoder
2 | le=LabelEncoder()
                                                                                                 data[i]=le.fit_transform(data[i])
                                                                                   for i in object_list:
                          le=LabelEncoder()
    In [68]:
```

Train-Test-Spliting

```
1 x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_stat
1 x=data.iloc[:,:-1].values
2 y=data.iloc[:,-1].values
  In [69]:
                                                                                                              In [70]:
```

Linear Regression

```
reg.fit(x_train,y_train)
print(f'The score of Linear regression is {reg.score(x_test,y_test)*100} %')
1 from sklearn.linear_model import LinearRegression
                                    reg=LinearRegression()
      In [71]:
```

The score of Linear regression is 39.260968207579374 %

```
In [72]: 1 y_pred=reg.predict(x_test)
```

XGBOOST Modeling.

```
xg_reg.fit(x_train,y_train)
print(f'The score of xgboost is {xg_reg.score(x_test,y_test)*100} %')
                            xgb.XGBRegressor().get_params()
                                                      xg_reg=xgb.XGBRegressor()
1 import xgboost as xgb
In [73]:
```

The score of xgboost is 72.00224981007838

```
In [74]: 1 xgb_preds = xg_reg.predict(x_test)
```

Decision Tree

sale price predication - Jupyter Notebook

12/19/22, 7:56 PM

```
In [75]: 1 from sklearn.tree import DecisionTreeRegressor
2 regr=DecisionTreeRegressor()
3 regr.fit(x_train,y_train)
4 print(f'The score of Decision Tree is {reg.score(x_test,y_test)*100})%')
```