# Data Cleaning

## Method - Delete Rows & Columns

```
#Import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# load dataset
df = pd.read_csv(r"G:\DataSet\House Price Prediction\train.csv")

df.shape
```

```
(1460, 81)
```

```
df.head(6)
```

```
   Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape
\
0   1          60       RL         65.0     8450   Pave   NaN      Reg

1   2          20       RL         80.0     9600   Pave   NaN      Reg

2   3          60       RL         68.0    11250   Pave   NaN      IR1

3   4          70       RL         60.0     9550   Pave   NaN      IR1

4   5          60       RL         84.0    14260   Pave   NaN      IR1

5   6          50       RL         85.0    14115   Pave   NaN      IR1


  LandContour Utilities  ... PoolArea PoolQC  Fence MiscFeature
MiscVal  \
0         Lvl    AllPub  ...        0    NaN    NaN         NaN
0
1         Lvl    AllPub  ...        0    NaN    NaN         NaN
0
2         Lvl    AllPub  ...        0    NaN    NaN         NaN
0
3         Lvl    AllPub  ...        0    NaN    NaN         NaN
0
4         Lvl    AllPub  ...        0    NaN    NaN         NaN
0
5         Lvl    AllPub  ...        0    NaN  MnPrv        Shed
700
```

```
  MoSold YrSold  SaleType  SaleCondition  SalePrice
0      2   2008        WD         Normal     208500
1      5   2007        WD         Normal     181500
2      9   2008        WD         Normal     223500
3      2   2006        WD        Abnorml     140000
4     12   2008        WD         Normal     250000
5     10   2009        WD         Normal     143000

[6 rows x 81 columns]
```

```
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

df.head(6)
```

```
   Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape
\
0   1          60       RL         65.0     8450   Pave   NaN      Reg

1   2          20       RL         80.0     9600   Pave   NaN      Reg

2   3          60       RL         68.0    11250   Pave   NaN      IR1

3   4          70       RL         60.0     9550   Pave   NaN      IR1

4   5          60       RL         84.0    14260   Pave   NaN      IR1

5   6          50       RL         85.0    14115   Pave   NaN      IR1


  LandContour Utilities LotConfig LandSlope Neighborhood Condition1  \
0         Lvl    AllPub    Inside       Gtl      CollgCr       Norm
1         Lvl    AllPub       FR2       Gtl      Veenker      Feedr
2         Lvl    AllPub    Inside       Gtl      CollgCr       Norm
3         Lvl    AllPub    Corner       Gtl      Crawfor       Norm
4         Lvl    AllPub       FR2       Gtl      NoRidge       Norm
5         Lvl    AllPub    Inside       Gtl      Mitchel       Norm

  Condition2 BldgType HouseStyle  OverallQual  OverallCond  YearBuilt
\
0       Norm     1Fam     2Story            7            5       2003

1       Norm     1Fam     1Story            6            8       1976

2       Norm     1Fam     2Story            7            5       2001

3       Norm     1Fam     2Story            7            5       1915

4       Norm     1Fam     2Story            8            5       2000

5       Norm     1Fam     1.5Fin            5            5       1993
```

```
   YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType
\
0          2003     Gable  CompShg     VinylSd     VinylSd    BrkFace

1          1976     Gable  CompShg     MetalSd     MetalSd       None

2          2002     Gable  CompShg     VinylSd     VinylSd    BrkFace

3          1970     Gable  CompShg     Wd Sdng     Wd Shng       None

4          2000     Gable  CompShg     VinylSd     VinylSd    BrkFace

5          1995     Gable  CompShg     VinylSd     VinylSd       None


   MasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond
BsmtExposure  \
0       196.0         Gd         TA      PConc       Gd       TA
No
1         0.0         TA         TA     CBlock       Gd       TA
Gd
2       162.0         Gd         TA      PConc       Gd       TA
Mn
3         0.0         TA         TA     BrkTil       TA       Gd
No
4       350.0         Gd         TA      PConc       Gd       TA
Av
5         0.0         TA         TA       Wood       Gd       TA
No

  BsmtFinType1  BsmtFinSF1 BsmtFinType2  BsmtFinSF2  BsmtUnfSF
TotalBsmtSF  \
0          GLQ         706          Unf           0        150
856
1          ALQ         978          Unf           0        284
1262
2          GLQ         486          Unf           0        434
920
3          ALQ         216          Unf           0        540
756
4          GLQ         655          Unf           0        490
1145
5          GLQ         732          Unf           0         64
796

  Heating HeatingQC CentralAir Electrical  1stFlrSF  2ndFlrSF
LowQualFinSF  \
0    GasA        Ex          Y      SBrkr       856       854
```

```
0
1    GasA        Ex          Y      SBrkr      1262          0
0
2    GasA        Ex          Y      SBrkr       920        866
0
3    GasA        Gd          Y      SBrkr       961        756
0
4    GasA        Ex          Y      SBrkr      1145       1053
0
5    GasA        Ex          Y      SBrkr       796        566
0

   GrLivArea  BsmtFullBath  BsmtHalfBath  FullBath  HalfBath
BedroomAbvGr  \
0       1710             1             0         2         1
3
1       1262             0             1         2         0
3
2       1786             1             0         2         1
3
3       1717             1             0         1         0
3
4       2198             1             0         2         1
4
5       1362             1             0         1         1
1

   KitchenAbvGr KitchenQual  TotRmsAbvGrd Functional  Fireplaces
FireplaceQu  \
0             1          Gd             8        Typ           0
NaN
1             1          TA             6        Typ           1
TA
2             1          Gd             6        Typ           1
TA
3             1          Gd             7        Typ           1
Gd
4             1          Gd             9        Typ           1
TA
5             1          TA             5        Typ           0
NaN

  GarageType  GarageYrBlt GarageFinish  GarageCars  GarageArea
GarageQual  \
0     Attchd       2003.0          RFn           2         548
TA
1     Attchd       1976.0          RFn           2         460
TA
2     Attchd       2001.0          RFn           2         608
TA
```

```
3     Detchd       1998.0          Unf           3         642
TA
4     Attchd       2000.0          RFn           3         836
TA
5     Attchd       1993.0          Unf           2         480
TA

  GarageCond PavedDrive  WoodDeckSF  OpenPorchSF  EnclosedPorch
3SsnPorch  \
0         TA          Y           0           61              0
0
1         TA          Y         298            0              0
0
2         TA          Y           0           42              0
0
3         TA          Y           0           35            272
0
4         TA          Y         192           84              0
0
5         TA          Y          40           30              0
320

   ScreenPorch  PoolArea PoolQC  Fence MiscFeature  MiscVal  MoSold
YrSold  \
0            0         0    NaN    NaN         NaN        0       2
2008
1            0         0    NaN    NaN         NaN        0       5
2007
2            0         0    NaN    NaN         NaN        0       9
2008
3            0         0    NaN    NaN         NaN        0       2
2006
4            0         0    NaN    NaN         NaN        0      12
2008
5            0         0    NaN  MnPrv        Shed      700      10
2009

  SaleType SaleCondition  SalePrice
0       WD        Normal     208500
1       WD        Normal     181500
2       WD        Normal     223500
3       WD       Abnorml     140000
4       WD        Normal     250000
5       WD        Normal     143000

df.tail(6)

        Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley
LotShape  \
1454  1455          20       FV         62.0     7500   Pave  Pave
```

```
Reg
1455  1456          60       RL         62.0     7917   Pave   NaN
Reg
1456  1457          20       RL         85.0    13175   Pave   NaN
Reg
1457  1458          70       RL         66.0     9042   Pave   NaN
Reg
1458  1459          20       RL         68.0     9717   Pave   NaN
Reg
1459  1460          20       RL         75.0     9937   Pave   NaN
Reg

     LandContour Utilities LotConfig LandSlope Neighborhood Condition1
\
1454         Lvl    AllPub    Inside       Gtl      Somerst       Norm

1455         Lvl    AllPub    Inside       Gtl      Gilbert       Norm

1456         Lvl    AllPub    Inside       Gtl       NWAmes       Norm

1457         Lvl    AllPub    Inside       Gtl      Crawfor       Norm

1458         Lvl    AllPub    Inside       Gtl        NAmes       Norm

1459         Lvl    AllPub    Inside       Gtl      Edwards       Norm


     Condition2 BldgType HouseStyle  OverallQual  OverallCond
YearBuilt  \
1454       Norm     1Fam     1Story            7            5
2004
1455       Norm     1Fam     2Story            6            5
1999
1456       Norm     1Fam     1Story            6            6
1978
1457       Norm     1Fam     2Story            7            9
1941
1458       Norm     1Fam     1Story            5            6
1950
1459       Norm     1Fam     1Story            5            6
1965

      YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd
MasVnrType  \
1454          2005     Gable  CompShg     VinylSd     VinylSd
None
1455          2000     Gable  CompShg     VinylSd     VinylSd
None
1456          1988     Gable  CompShg     Plywood     Plywood
Stone
```

```
1457          2006     Gable  CompShg     CemntBd     CmentBd
None
1458          1996       Hip  CompShg     MetalSd     MetalSd
None
1459          1965     Gable  CompShg     HdBoard     HdBoard
None

      MasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond  \
1454         0.0        Gd        TA      PConc       Gd       TA
1455         0.0        TA        TA      PConc       Gd       TA
1456       119.0        TA        TA     CBlock       Gd       TA
1457         0.0        Ex        Gd      Stone       TA       Gd
1458         0.0        TA        TA     CBlock       TA       TA
1459         0.0        Gd        TA     CBlock       TA       TA

     BsmtExposure BsmtFinType1  BsmtFinSF1 BsmtFinType2  BsmtFinSF2  \
1454           No          GLQ         410          Unf           0
1455           No          Unf           0          Unf           0
1456           No          ALQ         790          Rec         163
1457           No          GLQ         275          Unf           0
1458           Mn          GLQ          49          Rec        1029
1459           No          BLQ         830          LwQ         290

      BsmtUnfSF  TotalBsmtSF Heating HeatingQC CentralAir
Electrical  \
1454        811         1221    GasA        Ex          Y      SBrkr

1455        953          953    GasA        Ex          Y      SBrkr

1456        589         1542    GasA        TA          Y      SBrkr

1457        877         1152    GasA        Ex          Y      SBrkr

1458          0         1078    GasA        Gd          Y      FuseA

1459        136         1256    GasA        Gd          Y      SBrkr


      1stFlrSF  2ndFlrSF  LowQualFinSF  GrLivArea  BsmtFullBath
BsmtHalfBath  \
1454      1221         0             0       1221             1
0
1455       953       694             0       1647             0
0
1456      2073         0             0       2073             1
0
1457      1188      1152             0       2340             0
0
1458      1078         0             0       1078             1
0
```

```
1459      1256         0              0       1256              1
0

      FullBath  HalfBath  BedroomAbvGr  KitchenAbvGr KitchenQual  \
1454         2         0             2             1          Gd
1455         2         1             3             1          TA
1456         2         0             3             1          TA
1457         2         0             4             1          Gd
1458         1         0             2             1          Gd
1459         1         1             3             1          TA

      TotRmsAbvGrd Functional  Fireplaces FireplaceQu GarageType
GarageYrBlt  \
1454             6        Typ           0         NaN     Attchd
2004.0
1455             7        Typ           1          TA     Attchd
1999.0
1456             7       Min1           2          TA     Attchd
1978.0
1457             9        Typ           2          Gd     Attchd
1941.0
1458             5        Typ           0         NaN     Attchd
1950.0
1459             6        Typ           0         NaN     Attchd
1965.0

     GarageFinish  GarageCars  GarageArea GarageQual GarageCond
PavedDrive  \
1454          RFn           2         400         TA         TA
Y
1455          RFn           2         460         TA         TA
Y
1456          Unf           2         500         TA         TA
Y
1457          RFn           1         252         TA         TA
Y
1458          Unf           1         240         TA         TA
Y
1459          Fin           1         276         TA         TA
Y

      WoodDeckSF  OpenPorchSF  EnclosedPorch  3SsnPorch
ScreenPorch  \
1454           0          113              0          0            0

1455           0           40              0          0            0

1456         349            0              0          0            0

1457           0           60              0          0            0
```

```
1458         366            0            112          0            0

1459         736           68              0          0            0


      PoolArea PoolQC  Fence MiscFeature  MiscVal  MoSold  YrSold
SaleType  \
1454         0    NaN    NaN          NaN        0      10    2009
WD
1455         0    NaN    NaN          NaN        0       8    2007
WD
1456         0    NaN  MnPrv          NaN        0       2    2010
WD
1457         0    NaN  GdPrv         Shed     2500       5    2010
WD
1458         0    NaN    NaN          NaN        0       4    2010
WD
1459         0    NaN    NaN          NaN        0       6    2008
WD

     SaleCondition  SalePrice
1454        Normal     185000
1455        Normal     175000
1456        Normal     210000
1457        Normal     266500
1458        Normal     142125
1459        Normal     147500

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
Id               1460 non-null int64
MSSubClass       1460 non-null int64
MSZoning         1460 non-null object
LotFrontage      1201 non-null float64
LotArea          1460 non-null int64
Street           1460 non-null object
Alley            91 non-null object
LotShape         1460 non-null object
LandContour      1460 non-null object
Utilities        1460 non-null object
LotConfig        1460 non-null object
LandSlope        1460 non-null object
Neighborhood     1460 non-null object
Condition1       1460 non-null object
Condition2       1460 non-null object
BldgType         1460 non-null object
```

```
HouseStyle        1460 non-null object
OverallQual       1460 non-null int64
OverallCond       1460 non-null int64
YearBuilt         1460 non-null int64
YearRemodAdd      1460 non-null int64
RoofStyle         1460 non-null object
RoofMatl          1460 non-null object
Exterior1st       1460 non-null object
Exterior2nd       1460 non-null object
MasVnrType        1452 non-null object
MasVnrArea        1452 non-null float64
ExterQual         1460 non-null object
ExterCond         1460 non-null object
Foundation        1460 non-null object
BsmtQual          1423 non-null object
BsmtCond          1423 non-null object
BsmtExposure      1422 non-null object
BsmtFinType1      1423 non-null object
BsmtFinSF1        1460 non-null int64
BsmtFinType2      1422 non-null object
BsmtFinSF2        1460 non-null int64
BsmtUnfSF         1460 non-null int64
TotalBsmtSF       1460 non-null int64
Heating           1460 non-null object
HeatingQC         1460 non-null object
CentralAir        1460 non-null object
Electrical        1459 non-null object
1stFlrSF          1460 non-null int64
2ndFlrSF          1460 non-null int64
LowQualFinSF      1460 non-null int64
GrLivArea         1460 non-null int64
BsmtFullBath      1460 non-null int64
BsmtHalfBath      1460 non-null int64
FullBath          1460 non-null int64
HalfBath          1460 non-null int64
BedroomAbvGr      1460 non-null int64
KitchenAbvGr      1460 non-null int64
KitchenQual       1460 non-null object
TotRmsAbvGrd      1460 non-null int64
Functional        1460 non-null object
Fireplaces        1460 non-null int64
FireplaceQu       770 non-null object
GarageType        1379 non-null object
GarageYrBlt       1379 non-null float64
GarageFinish      1379 non-null object
GarageCars        1460 non-null int64
GarageArea        1460 non-null int64
GarageQual        1379 non-null object
GarageCond        1379 non-null object
```

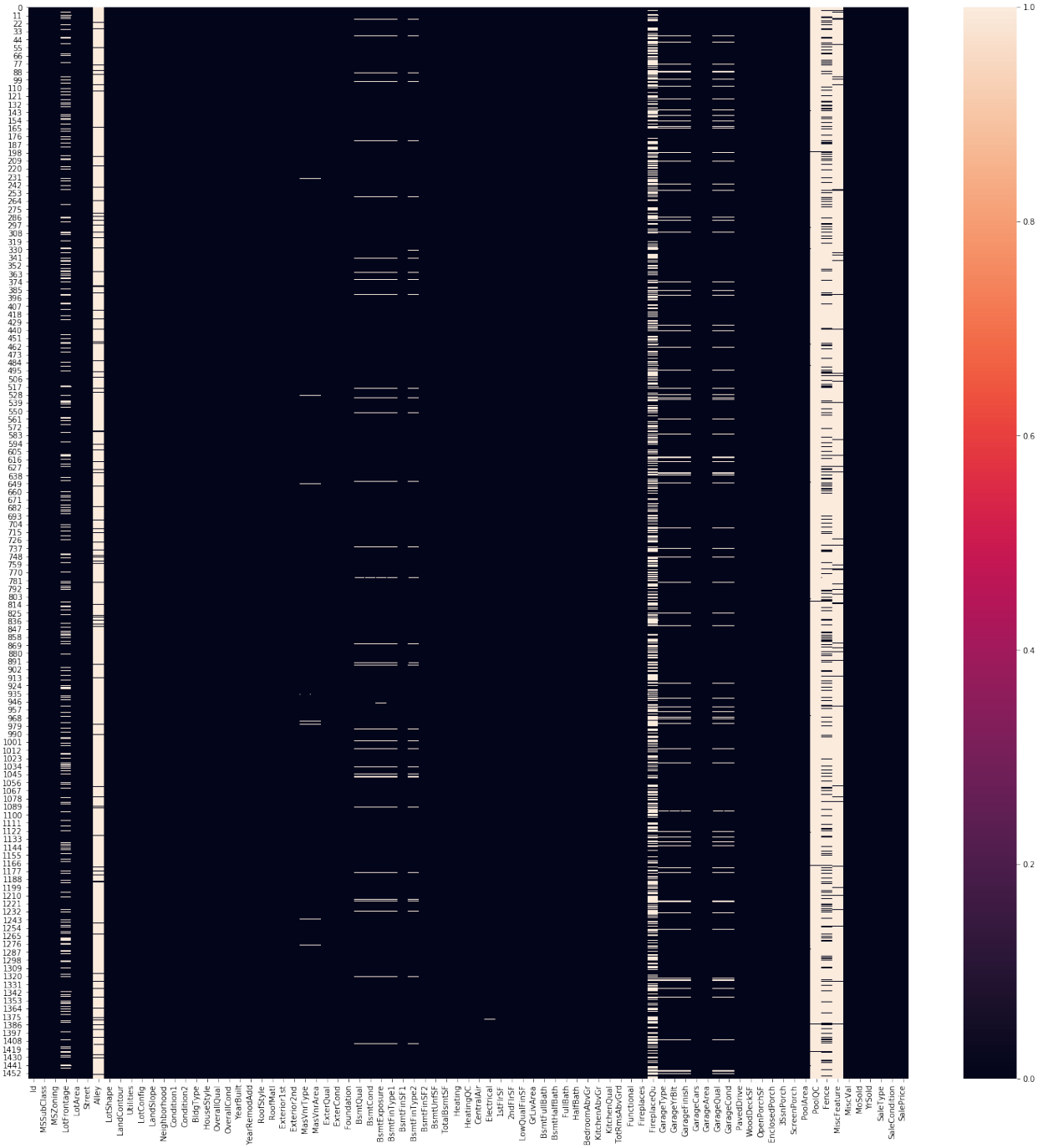
```
PavedDrive        1460 non-null object
WoodDeckSF        1460 non-null int64
OpenPorchSF       1460 non-null int64
EnclosedPorch     1460 non-null int64
3SsnPorch         1460 non-null int64
ScreenPorch       1460 non-null int64
PoolArea          1460 non-null int64
PoolQC            7 non-null object
Fence             281 non-null object
MiscFeature       54 non-null object
MiscVal           1460 non-null int64
MoSold            1460 non-null int64
YrSold            1460 non-null int64
SaleType          1460 non-null object
SaleCondition     1460 non-null object
SalePrice         1460 non-null int64
dtypes: float64(3), int64(35), object(43)
memory usage: 924.0+ KB
```

```
df.isnull().sum()
```

```
Id                   0
MSSubClass           0
MSZoning             0
LotFrontage        259
LotArea              0
Street               0
Alley             1369
LotShape             0
LandContour          0
Utilities            0
LotConfig            0
LandSlope            0
Neighborhood         0
Condition1           0
Condition2           0
BldgType             0
HouseStyle           0
OverallQual          0
OverallCond          0
YearBuilt            0
YearRemodAdd         0
RoofStyle            0
RoofMatl             0
Exterior1st          0
Exterior2nd          0
MasVnrType           8
MasVnrArea           8
ExterQual            0
ExterCond            0
```

```
Foundation          0
BsmtQual           37
BsmtCond           37
BsmtExposure       38
BsmtFinType1       37
BsmtFinSF1          0
BsmtFinType2       38
BsmtFinSF2          0
BsmtUnfSF           0
TotalBsmtSF         0
Heating             0
HeatingQC           0
CentralAir          0
Electrical          1
1stFlrSF            0
2ndFlrSF            0
LowQualFinSF        0
GrLivArea           0
BsmtFullBath        0
BsmtHalfBath        0
FullBath            0
HalfBath            0
BedroomAbvGr        0
KitchenAbvGr        0
KitchenQual         0
TotRmsAbvGrd        0
Functional          0
Fireplaces          0
FireplaceQu       690
GarageType         81
GarageYrBlt        81
GarageFinish       81
GarageCars          0
GarageArea          0
GarageQual         81
GarageCond         81
PavedDrive          0
WoodDeckSF          0
OpenPorchSF         0
EnclosedPorch       0
3SsnPorch           0
ScreenPorch         0
PoolArea            0
PoolQC           1453
Fence            1179
MiscFeature      1406
MiscVal             0
MoSold              0
YrSold              0
```

```
SaleType             0
SaleCondition        0
SalePrice            0
dtype: int64

plt.figure(figsize=(25,25))
sns.heatmap(df.isnull())

<matplotlib.axes._subplots.AxesSubplot at 0x1c4c7c8af60>
```



```
null_var = df.isnull().sum()/df.shape[0] *100
null_var
```

```
Id                0.000000
MSSubClass        0.000000
MSZoning          0.000000
LotFrontage      17.739726
LotArea           0.000000
Street            0.000000
Alley            93.767123
LotShape          0.000000
LandContour       0.000000
Utilities         0.000000
LotConfig         0.000000
LandSlope         0.000000
Neighborhood      0.000000
Condition1        0.000000
Condition2        0.000000
BldgType          0.000000
HouseStyle        0.000000
OverallQual       0.000000
OverallCond       0.000000
YearBuilt         0.000000
YearRemodAdd      0.000000
RoofStyle         0.000000
RoofMatl          0.000000
Exterior1st       0.000000
Exterior2nd       0.000000
MasVnrType        0.547945
MasVnrArea        0.547945
ExterQual         0.000000
ExterCond         0.000000
Foundation        0.000000
BsmtQual          2.534247
BsmtCond          2.534247
BsmtExposure      2.602740
BsmtFinType1      2.534247
BsmtFinSF1        0.000000
BsmtFinType2      2.602740
BsmtFinSF2        0.000000
BsmtUnfSF         0.000000
TotalBsmtSF       0.000000
Heating           0.000000
HeatingQC         0.000000
CentralAir        0.000000
Electrical        0.068493
1stFlrSF          0.000000
2ndFlrSF          0.000000
LowQualFinSF      0.000000
GrLivArea         0.000000
```

```
BsmtFullBath      0.000000
BsmtHalfBath      0.000000
FullBath          0.000000
HalfBath          0.000000
BedroomAbvGr      0.000000
KitchenAbvGr      0.000000
KitchenQual       0.000000
TotRmsAbvGrd      0.000000
Functional        0.000000
Fireplaces        0.000000
FireplaceQu      47.260274
GarageType        5.547945
GarageYrBlt       5.547945
GarageFinish      5.547945
GarageCars        0.000000
GarageArea        0.000000
GarageQual        5.547945
GarageCond        5.547945
PavedDrive        0.000000
WoodDeckSF        0.000000
OpenPorchSF       0.000000
EnclosedPorch     0.000000
3SsnPorch         0.000000
ScreenPorch       0.000000
PoolArea          0.000000
PoolQC           99.520548
Fence            80.753425
MiscFeature      96.301370
MiscVal           0.000000
MoSold            0.000000
YrSold            0.000000
SaleType          0.000000
SaleCondition     0.000000
SalePrice         0.000000
dtype: float64
```

```
drop_columns  = null_var[null_var >17].keys()
drop_columns
```

```
Index(['LotFrontage', 'Alley', 'FireplaceQu', 'PoolQC', 'Fence',
       'MiscFeature'],
      dtype='object')
```

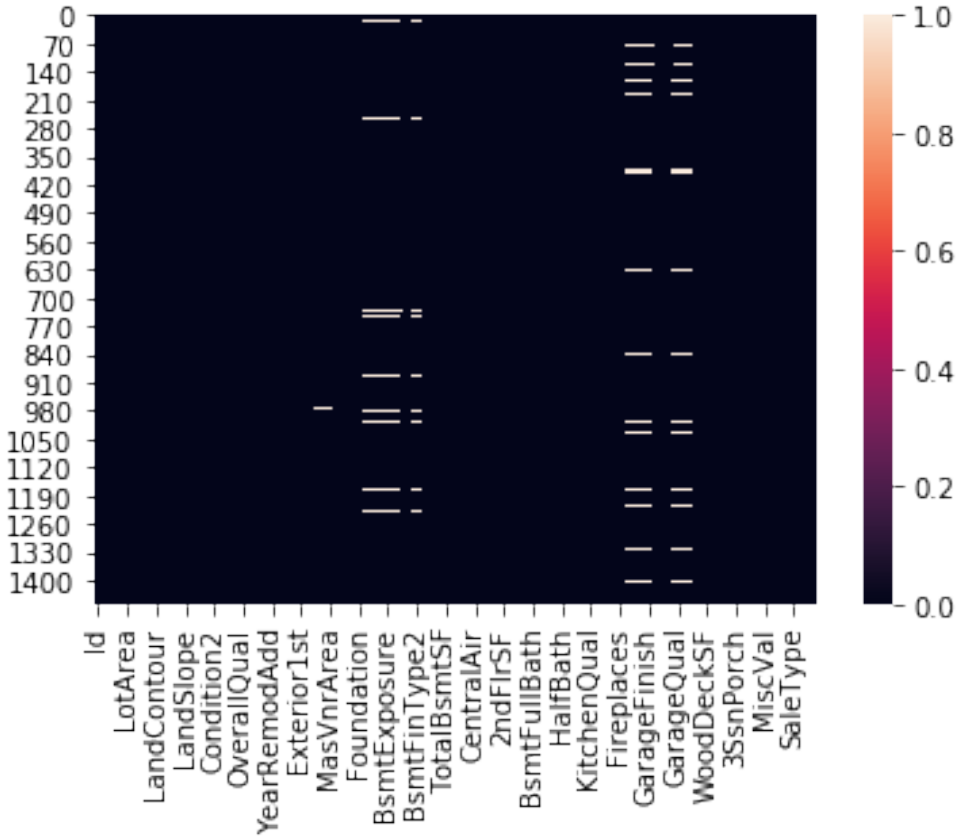
```
df2_drop_clm=df.drop(columns=drop_columns)
```

```
df2_drop_clm.shape
```

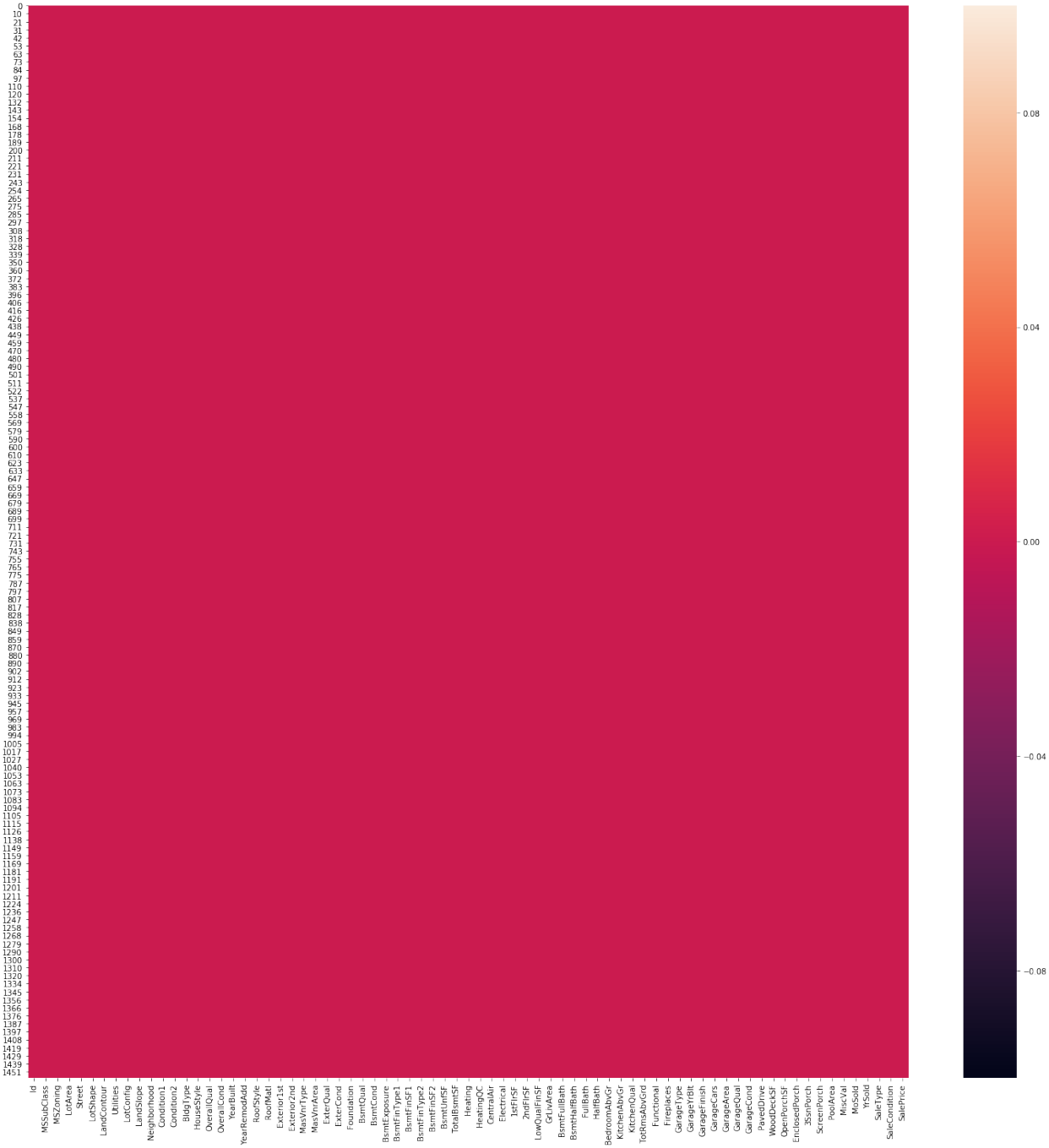
```
(1460, 75)
```

```
sns.heatmap(df2_drop_clm.isnull())
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c4c819a630>
```




```
df3_drop_rows = df2_drop_clm.dropna()

df3_drop_rows.shape

(1338, 75)

plt.figure(figsize=(25,25))
sns.heatmap(df3_drop_rows.isnull())

<matplotlib.axes._subplots.AxesSubplot at 0x1c4c8377550>
```




```
df3_drop_rows.isnull().sum().sum()
```

```
0
```

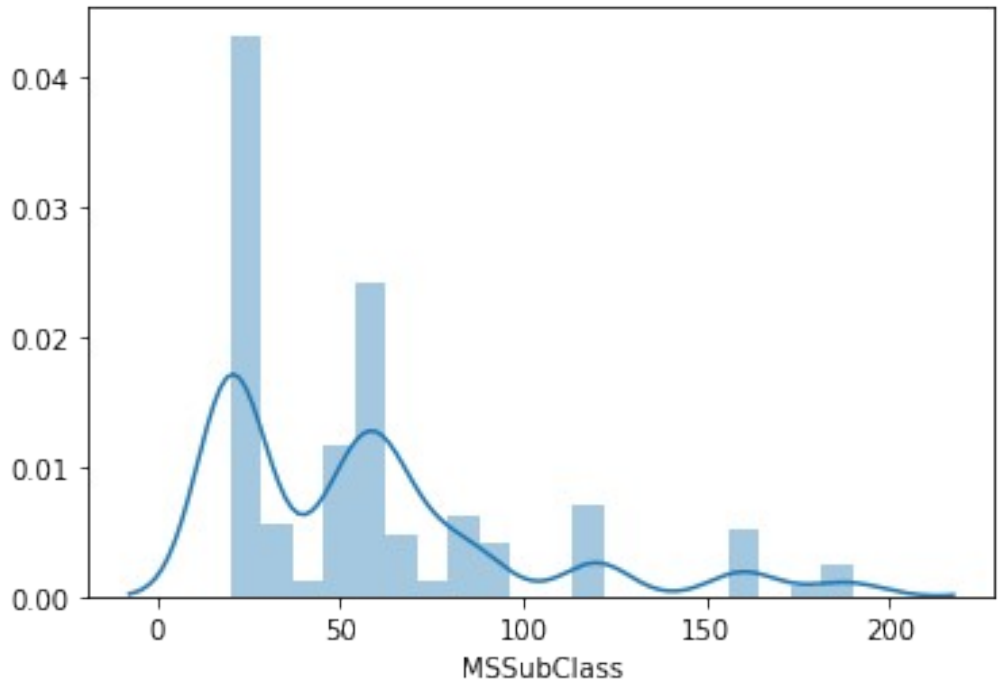
```
df3_drop_rows.select_dtypes(include=['int64','float64']).columns
```

```
Index(['Id', 'MSSubClass', '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'],
      dtype='object')
```

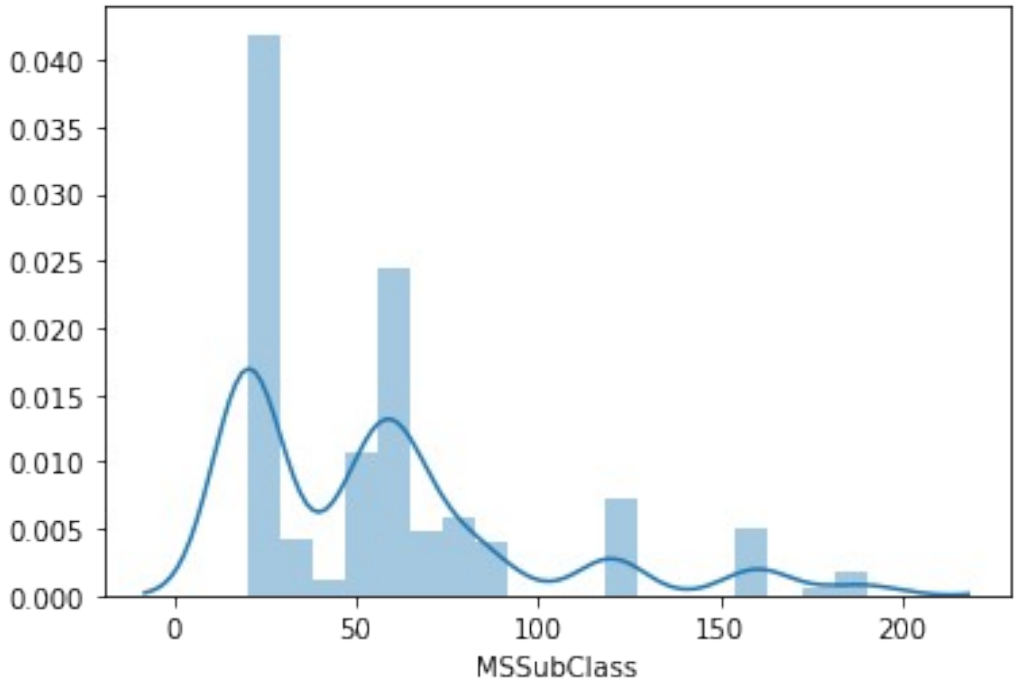
```
sns.distplot(df['MSSubClass'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c4ca6797b8>
```

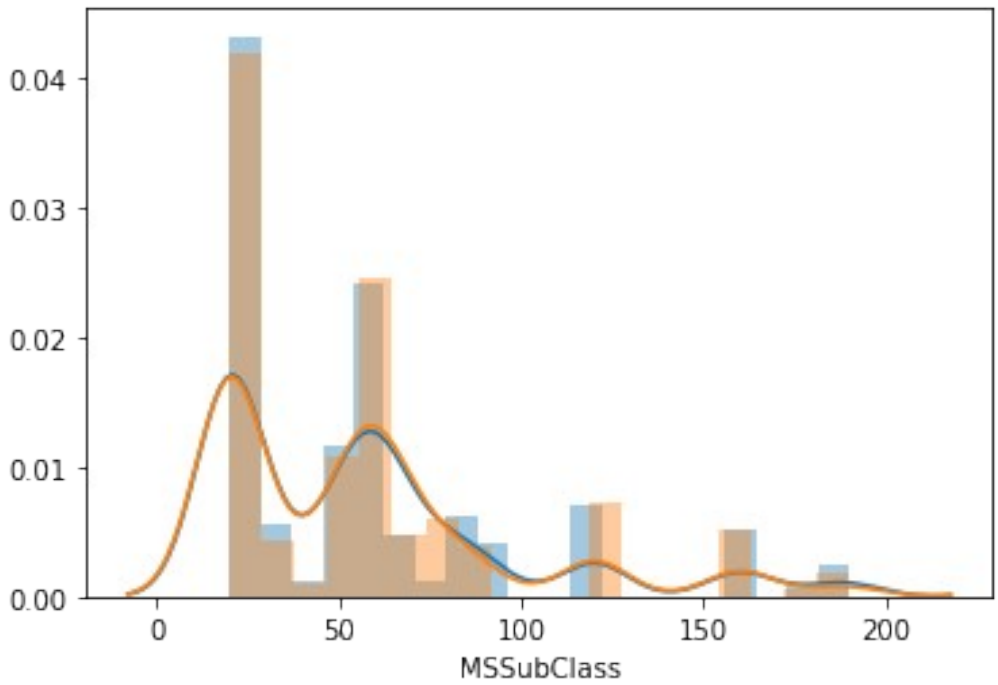



```
sns.distplot(df3_drop_rows['MSSubClass'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c4c93cc908>
```

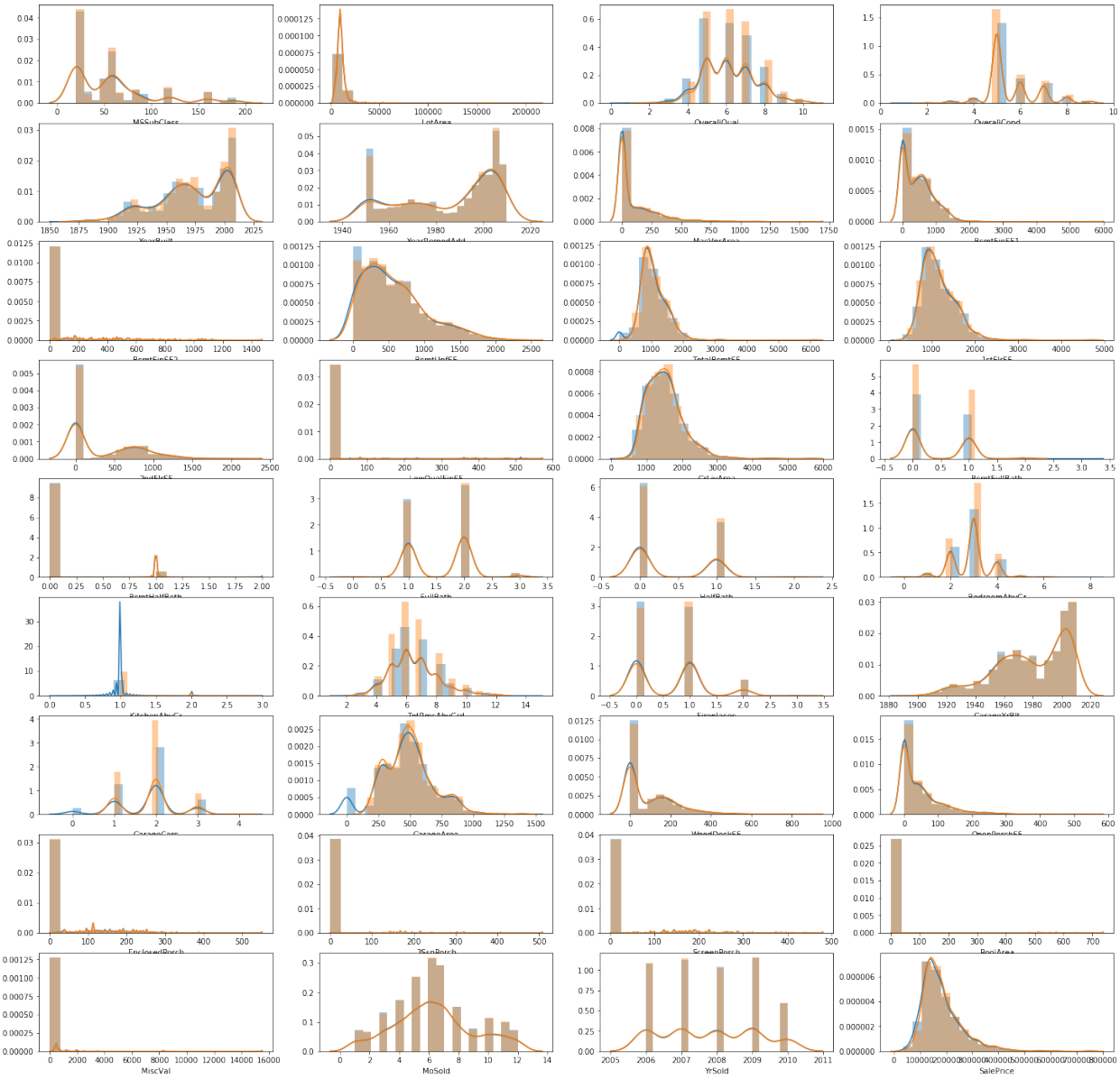



```
sns.distplot(df['MSSubClass'])
sns.distplot(df3_drop_rows['MSSubClass'])

<matplotlib.axes._subplots.AxesSubplot at 0x1c4ca89e668>
```



```
num_var = ['MSSubClass', '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']

plt.figure(figsize=(25,25))
for i,var in enumerate(num_var):
    plt.subplot(9,4,i+1)
    sns.distplot(df[var], bins=20)
    sns.distplot(df3_drop_rows[var], bins=20)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\numpy\lib\
histograms.py:839: RuntimeWarning: invalid value encountered in
greater_equal
  keep = (tmp_a >= first_edge)
C:\ProgramData\Anaconda3\lib\site-packages\numpy\lib\
histograms.py:840: RuntimeWarning: invalid value encountered in
less_equal
  keep &= (tmp_a <= last_edge)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\nonparametric\
kde.py:447: RuntimeWarning: invalid value encountered in greater
  X = X[np.logical_and(X > clip[0], X < clip[1])] # won't work for two
columns.
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\nonparametric\
kde.py:447: RuntimeWarning: invalid value encountered in less
  X = X[np.logical_and(X > clip[0], X < clip[1])] # won't work for two
columns.
```




```
df3_drop_rows.select_dtypes(include=['object']).columns

Index(['MSZoning', 'Street', 'LotShape', 'LandContour', 'Utilities',
       'LotConfig', 'LandSlope', 'Neighborhood', 'Condition1',
'Condition2',
       'BldgType', 'HouseStyle', 'RoofStyle', 'RoofMatl',
'Exterior1st',
       'Exterior2nd', 'MasVnrType', 'ExterQual', 'ExterCond',
'Foundation',
       'BsmtQual', 'BsmtCond', 'BsmtExposure', 'BsmtFinType1',
'BsmtFinType2',
       'Heating', 'HeatingQC', 'CentralAir', 'Electrical',
'KitchenQual',
       'Functional', 'GarageType', 'GarageFinish', 'GarageQual',
```

```
'GarageCond',
       'PavedDrive', 'SaleType', 'SaleCondition'],
      dtype='object')
```

```
pd.concat([df['MSZoning'].value_counts()/df.shape[0] * 100,

df3_drop_rows['MSZoning'].value_counts()/df3_drop_rows.shape[0] *
100], axis=1,
         keys=['MSZoning_org', 'MSZoning_clean'])
```

```
         MSZoning_org  MSZoning_clean
RL          78.835616       79.671151
RM          14.931507       14.275037
FV           4.452055        4.633782
RH           1.095890        0.822123
C (all)      0.684932        0.597907
```

```
def cat_var_dist(var):
    return pd.concat([df[var].value_counts()/df.shape[0] * 100,
          df3_drop_rows[var].value_counts()/df3_drop_rows.shape[0] *
100], axis=1,
         keys=[var+'_org', var+'clean'])


cat_var_dist('MSZoning')
```

```
         MSZoning_org  MSZoningclean
RL          78.835616      79.671151
RM          14.931507      14.275037
FV           4.452055       4.633782
RH           1.095890       0.822123
C (all)      0.684932       0.597907
```

```
print("Thank You.....-:)")
```

```
Thank You.....-:)
```