Manvith S Rao

MIT - CCE-24

(test dataset)

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

test_df=pd.read_csv('test.csv')

test_df.shape

(1459, 80)

test_df.head()

     Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley
LotShape  \
0  1461          20       RH         80.0    11622   Pave   NaN
Reg
1  1462          20       RL         81.0    14267   Pave   NaN
IR1
2  1463          60       RL         74.0    13830   Pave   NaN
IR1
3  1464          60       RL         78.0     9978   Pave   NaN
IR1
4  1465         120       RL         43.0     5005   Pave   NaN
IR1

  LandContour Utilities  ... ScreenPorch PoolArea PoolQC  Fence
MiscFeature  \
0         Lvl    AllPub  ...         120        0    NaN  MnPrv
NaN
1         Lvl    AllPub  ...           0        0    NaN    NaN
Gar2
2         Lvl    AllPub  ...           0        0    NaN  MnPrv
NaN
3         Lvl    AllPub  ...           0        0    NaN    NaN
NaN
4         HLS    AllPub  ...         144        0    NaN    NaN
NaN

  MiscVal MoSold  YrSold  SaleType  SaleCondition
0       0      6    2010        WD         Normal
1   12500      6    2010        WD         Normal
```

```
2       0      3    2010         WD          Normal
3       0      6    2010         WD          Normal
4       0      1    2010         WD          Normal

[5 rows x 80 columns]
```

#Check Null Values

```
test_df.isnull().sum()
```

```
Id                  0
MSSubClass          0
MSZoning            4
LotFrontage       227
LotArea             0
                 ...
MiscVal             0
MoSold              0
YrSold              0
SaleType            1
SaleCondition       0
Length: 80, dtype: int64
```

#Fill Missing Values

```
test_df['LotFrontage']=test_df['LotFrontage'].fillna(test_df['LotFrontage'].mean())

test_df['MSZoning']=test_df['MSZoning'].fillna(test_df['MSZoning'].mode()[0])

test_df.shape
```

```
(1459, 80)
```

```
test_df.drop(['Alley'],axis=1,inplace=True)

test_df.shape
```

```
(1459, 79)
```

```
test_df['BsmtCond']=test_df['BsmtCond'].fillna(test_df['BsmtCond'].mode()[0])
test_df['BsmtQual']=test_df['BsmtQual'].fillna(test_df['BsmtQual'].mode()[0])

test_df['FireplaceQu']=test_df['FireplaceQu'].fillna(test_df['FireplaceQu'].mode()[0])
test_df['GarageType']=test_df['GarageType'].fillna(test_df['GarageType'].mode()[0])
```

```
test_df.drop(['GarageYrBlt'],axis=1,inplace=True)
```

```
test_df.shape
```

```
(1459, 78)
```

```
test_df['GarageFinish']=test_df['GarageFinish'].fillna(test_df['Garage
Finish'].mode()[0])
test_df['GarageQual']=test_df['GarageQual'].fillna(test_df['GarageQual
'].mode()[0])
test_df['GarageCond']=test_df['GarageCond'].fillna(test_df['GarageCond
'].mode()[0])
```

```
test_df.drop(['PoolQC','Fence','MiscFeature'],axis=1,inplace=True)
```

```
test_df.shape
```

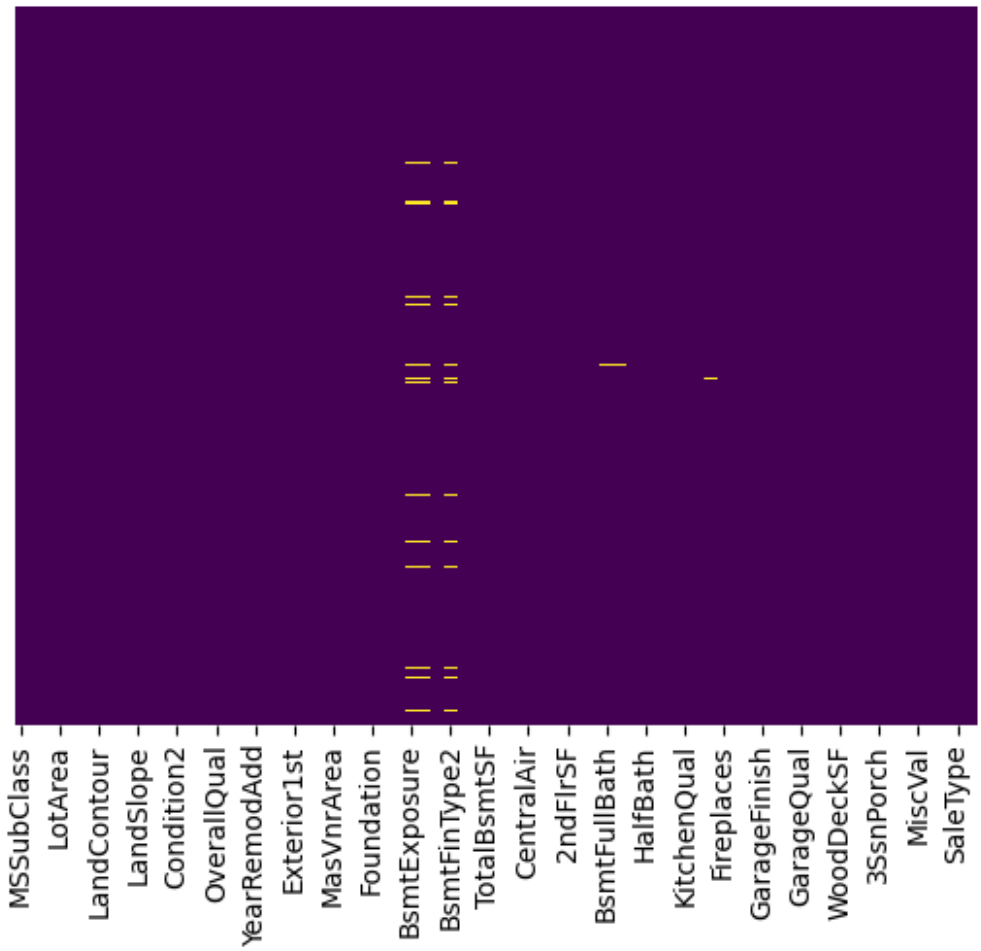
```
(1459, 75)
```

```
test_df.drop(['Id'],axis=1,inplace=True)
```

```
test_df['MasVnrType']=test_df['MasVnrType'].fillna(test_df['MasVnrType
'].mode()[0])
test_df['MasVnrArea']=test_df['MasVnrArea'].fillna(test_df['MasVnrArea
'].mode()[0])
```

```
sns.heatmap(test_df.isnull(),yticklabels=False,cbar=False,cmap='viridi
s')
```

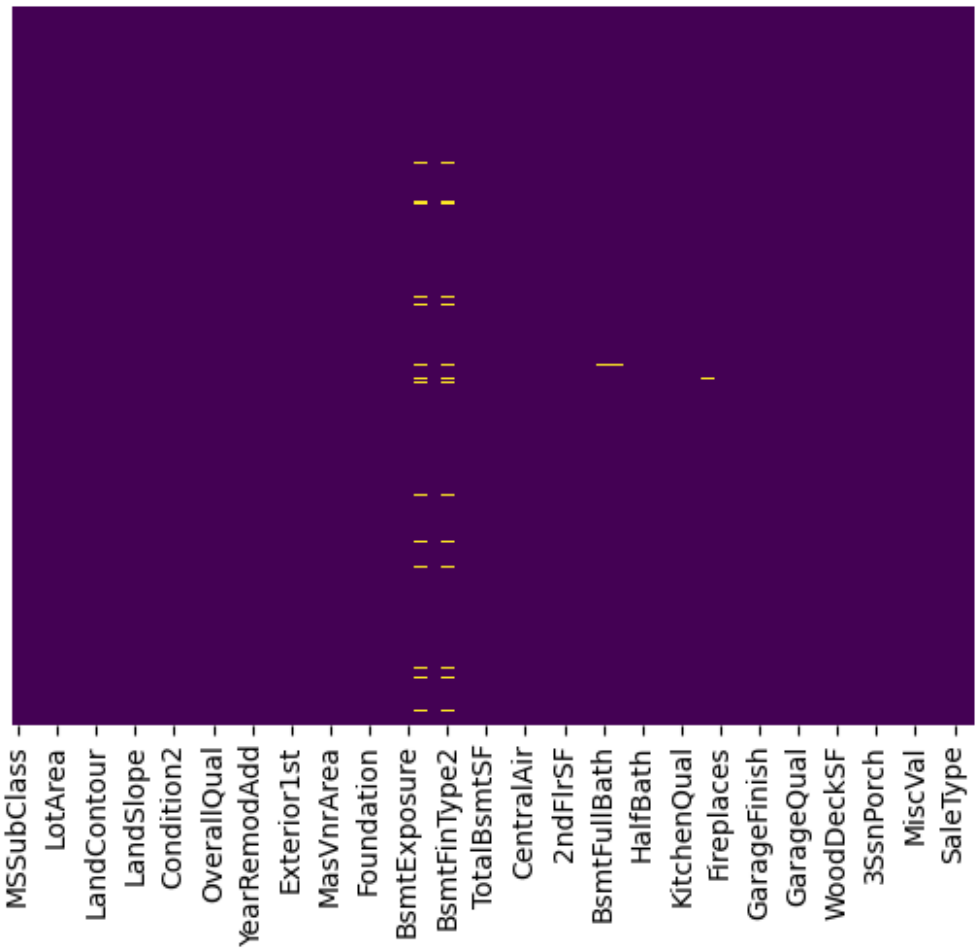
```
<Axes: >
```




```
test_df['BsmtExposure']=test_df['BsmtExposure'].fillna(test_df['BsmtExposure'].mode()[0])

sns.heatmap(test_df.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

```
<Axes: >
```




```
test_df['BsmtFinType2']=test_df['BsmtFinType2'].fillna(test_df['BsmtFi
nType2'].mode()[0])

test_df.loc[:, test_df.isnull().any()].head()
```

```
  Utilities Exterior1st Exterior2nd BsmtFinType1  BsmtFinSF1
BsmtFinSF2  \
0    AllPub     VinylSd     VinylSd          Rec       468.0
144.0
1    AllPub     Wd Sdng     Wd Sdng          ALQ       923.0
0.0
2    AllPub     VinylSd     VinylSd          GLQ       791.0
0.0
3    AllPub     VinylSd     VinylSd          GLQ       602.0
0.0
4    AllPub     HdBoard     HdBoard          ALQ       263.0
0.0

   BsmtUnfSF  TotalBsmtSF  BsmtFullBath  BsmtHalfBath KitchenQual
Functional  \
```

```
0      270.0         882.0            0.0            0.0           TA
Typ
1      406.0        1329.0            0.0            0.0           Gd
Typ
2      137.0         928.0            0.0            0.0           TA
Typ
3      324.0         926.0            0.0            0.0           Gd
Typ
4     1017.0        1280.0            0.0            0.0           Gd
Typ
```

```
   GarageCars  GarageArea SaleType
0         1.0       730.0       WD
1         1.0       312.0       WD
2         2.0       482.0       WD
3         2.0       470.0       WD
4         2.0       506.0       WD
```

```
test_df['Utilities']=test_df['Utilities'].fillna(test_df['Utilities'].
mode()[0])
test_df['Exterior1st']=test_df['Exterior1st'].fillna(test_df['Exterior
1st'].mode()[0])
test_df['Exterior2nd']=test_df['Exterior2nd'].fillna(test_df['Exterior
2nd'].mode()[0])
test_df['BsmtFinType1']=test_df['BsmtFinType1'].fillna(test_df['BsmtFi
nType1'].mode()[0])
test_df['BsmtFinSF1']=test_df['BsmtFinSF1'].fillna(test_df['BsmtFinSF1
'].mean())
test_df['BsmtFinSF2']=test_df['BsmtFinSF2'].fillna(test_df['BsmtFinSF2
'].mean())
test_df['BsmtUnfSF']=test_df['BsmtUnfSF'].fillna(test_df['BsmtUnfSF'].
mean())
test_df['TotalBsmtSF']=test_df['TotalBsmtSF'].fillna(test_df['TotalBsm
tSF'].mean())

test_df['BsmtFullBath']=test_df['BsmtFullBath'].fillna(test_df['BsmtFu
llBath'].mode()[0])
test_df['BsmtHalfBath']=test_df['BsmtHalfBath'].fillna(test_df['BsmtHa
lfBath'].mode()[0])
test_df['KitchenQual']=test_df['KitchenQual'].fillna(test_df['KitchenQ
ual'].mode()[0])
test_df['Functional']=test_df['Functional'].fillna(test_df['Functional
'].mode()[0])
test_df['GarageCars']=test_df['GarageCars'].fillna(test_df['GarageCars
'].mean())
test_df['GarageArea']=test_df['GarageArea'].fillna(test_df['GarageArea
'].mean())
test_df['SaleType']=test_df['SaleType'].fillna(test_df['SaleType'].mod
e()[0])
```

```
test_df.shape
```

```
(1459, 74)
```

```
test_df.to_csv('ModifiedTest.csv',index=False)
```