 **krishnaik06** Add files via upload

Latest commit 8c0a8c3 on Sep 8, 2019 [History](#)

👤 1 contributor

852 lines (852 sloc) | 66.3 KB

<>

📄

Raw

Blame

🖥

✎

🗑

Handle Test Data set

```
In [70]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [71]: test_df=pd.read_csv('test.csv')
```

```
In [72]: test_df.shape
```

Out[72]: (1459, 80)

```
In [4]: test_df.head()
```

Out[4]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	ScreenPorch	PoolArea	PoolQC	Fence	MiscFeat
0	1461	20	RH	80.0	11622	Pave	NaN	Reg	Lvl	AllPub	...	120	0	NaN	MnPrv	NaN
1	1462	20	RL	81.0	14267	Pave	NaN	IR1	Lvl	AllPub	...	0	0	NaN	NaN	Gar2
2	1463	60	RL	74.0	13830	Pave	NaN	IR1	Lvl	AllPub	...	0	0	NaN	MnPrv	NaN
3	1464	60	RL	78.0	9978	Pave	NaN	IR1	Lvl	AllPub	...	0	0	NaN	NaN	NaN
4	1465	120	RL	43.0	5005	Pave	NaN	IR1	HLS	AllPub	...	144	0	NaN	NaN	NaN

5 rows × 80 columns

```
In [ ]:
```

```
In [171]: #check null values
test_df.isnull().sum()
```

Out[171]:

Id	0
MSSubClass	0
MSZoning	4
LotFrontage	227
LotArea	0
Street	0
Alley	1352
LotShape	0
LandContour	0
Utilities	2
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      1
Exterior2nd      1
MasVnrType       16
MasVnrArea       15
ExterQual        0
ExterCond        0
Foundation       0
...
HalfBath         0
BedroomAbvGr     0
KitchenAbvGr     0
KitchenQual      1
TotRmsAbvGrd     0
Functional       2
Fireplaces       0
FireplaceQu      730
GarageType       76
GarageYrBlt      78
GarageFinish     78
GarageCars       1
GarageArea       1
GarageQual       78
GarageCond       78
PavedDrive       0
WoodDeckSF       0
OpenPorchSF      0
EnclosedPorch    0
3SsnPorch        0
ScreenPorch      0
PoolArea         0
PoolQC          1456
Fence            1169
MiscFeature      1408
MiscVal          0
MoSold           0
YrSold           0
SaleType         1
SaleCondition    0
Length: 80, dtype: int64
```

```
In [73]: ## Fill Missing Values

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

```
In [74]: test_df['MSZoning']=test_df['MSZoning'].fillna(test_df['MSZoning'].mode()[0])
```

```
In [7]: test_df.shape
```

Out[7]: (1459, 80)

In [8]: test_df.drop(['Alley'],axis=1,inplace=True)

In [9]: test_df.shape

Out[9]: (1459, 79)

In [10]: test_df['BsmtCond']=test_df['BsmtCond'].fillna(test_df['BsmtCond'].mode()[0])
test_df['BsmtQual']=test_df['BsmtQual'].fillna(test_df['BsmtQual'].mode()[0])

In [11]: test_df['FireplaceQu']=test_df['FireplaceQu'].fillna(test_df['FireplaceQu'].mode()[0])
test_df['GarageType']=test_df['GarageType'].fillna(test_df['GarageType'].mode()[0])

In [12]: test_df.drop(['GarageYrBlt'],axis=1,inplace=True)

In [13]: test_df.shape

Out[13]: (1459, 78)

In [14]: test_df['GarageFinish']=test_df['GarageFinish'].fillna(test_df['GarageFinish'].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)

In [15]: test_df.shape

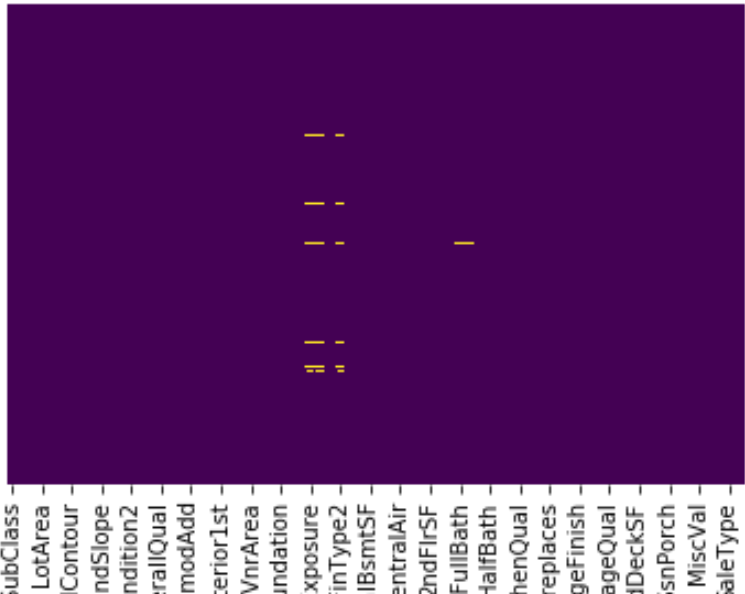
Out[15]: (1459, 75)

In [16]: test_df.drop(['Id'],axis=1,inplace=True)

In [17]: test_df['MasVnrType']=test_df['MasVnrType'].fillna(test_df['MasVnrType'].mode()[0])
test_df['MasVnrArea']=test_df['MasVnrArea'].fillna(test_df['MasVnrArea'].mode()[0])

In [18]: sns.heatmap(test_df.isnull(),yticklabels=False,cbar=False,cmap='viridis')

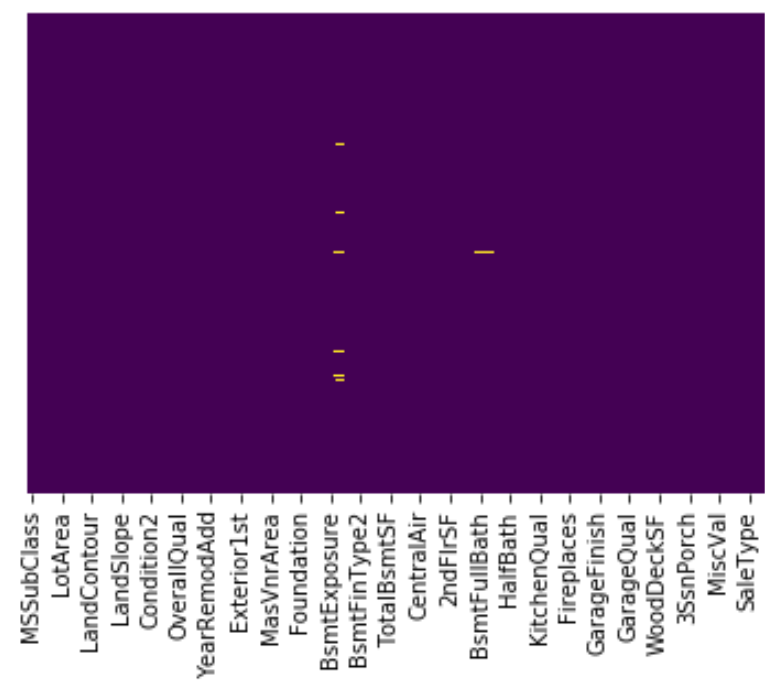
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x21918b69cf8>



```
In [19]: test_df['BsmtExposure']=test_df['BsmtExposure'].fillna(test_df['BsmtExposure'].mode()[0])
```

```
In [26]: sns.heatmap(test_df.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

Out[26]: <matplotlib.axes._subplots.AxesSubplot at 0x2191be249b0>



```
In [21]: test_df['BsmtFinType2']=test_df['BsmtFinType2'].fillna(test_df['BsmtFinType2'].mode()[0])
```

```
In [34]: test_df.loc[:, test_df.isnull().any()].head()
```

Out[34]:

	Utilities	Exterior1st	Exterior2nd	BsmtFinType1	BsmtFinSF1	BsmtFinSF2	BsmtUnfSF	TotalBsmtSF	BsmtFullBath	BsmtHalfBath	KitchenQual	Function
0	AllPub	VinylSd	VinylSd	Rec	468.0	144.0	270.0	882.0	0.0	0.0	TA	Typ
1	AllPub	Wd Sdng	Wd Sdng	ALQ	923.0	0.0	406.0	1329.0	0.0	0.0	Gd	Typ
2	AllPub	VinylSd	VinylSd	GLQ	791.0	0.0	137.0	928.0	0.0	0.0	TA	Typ
3	AllPub	VinylSd	VinylSd	GLQ	602.0	0.0	324.0	926.0	0.0	0.0	Gd	Typ
4	AllPub	HdBoard	HdBoard	ALQ	263.0	0.0	1017.0	1280.0	0.0	0.0	Gd	Typ

```
In [42]: test_df['Utilities']=test_df['Utilities'].fillna(test_df['Utilities'].mode()[0])
test_df['Exterior1st']=test_df['Exterior1st'].fillna(test_df['Exterior1st'].mode()[0])
test_df['Exterior2nd']=test_df['Exterior2nd'].fillna(test_df['Exterior2nd'].mode()[0])
test_df['BsmtFinType1']=test_df['BsmtFinType1'].fillna(test_df['BsmtFinType1'].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['TotalBsmtSF'].mean())
test_df['BsmtFullBath']=test_df['BsmtFullBath'].fillna(test_df['BsmtFullBath'].mode()[0])
test_df['BsmtHalfBath']=test_df['BsmtHalfBath'].fillna(test_df['BsmtHalfBath'].mode()[0])
test_df['KitchenQual']=test_df['KitchenQual'].fillna(test_df['KitchenQual'].mode()[0])
```

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

In [43]: test_df.shape

Out[43]: (1459, 74)

In [69]: test_df.to_csv('formulatedtest.csv',index=False)

In []: