index

January 24, 2022

1 Multicollinearity of Features - Lab

1.1 Introduction

In this lab, you'll identify multicollinearity in the Ames Housing dataset.

1.2 Objectives

You will be able to: * Create a scatter matrix and correlation matrix * Assess Intpret the output of a correlation matrix * Identify if variables are exhibiting collinearity * Decide how to address the collinearity in the data set

1.3 Correlation matrix for the Ames Housing data

1.3.1 Import data

Let's reimport the Ames Housing data assign the numeric variables we want to keep to `numeric_vars`.

[62]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

pd.options.display.max_columns = 999
pd.options.display.max_rows = 999

ames = pd.read_csv('ames.csv')

numeric_vars = ['LotFrontage', 'LotArea', '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']
```

1.3.2 Create processed

Create a new dataframe named `ames_preprocessed` that contains only the features in `numeric_vars`.

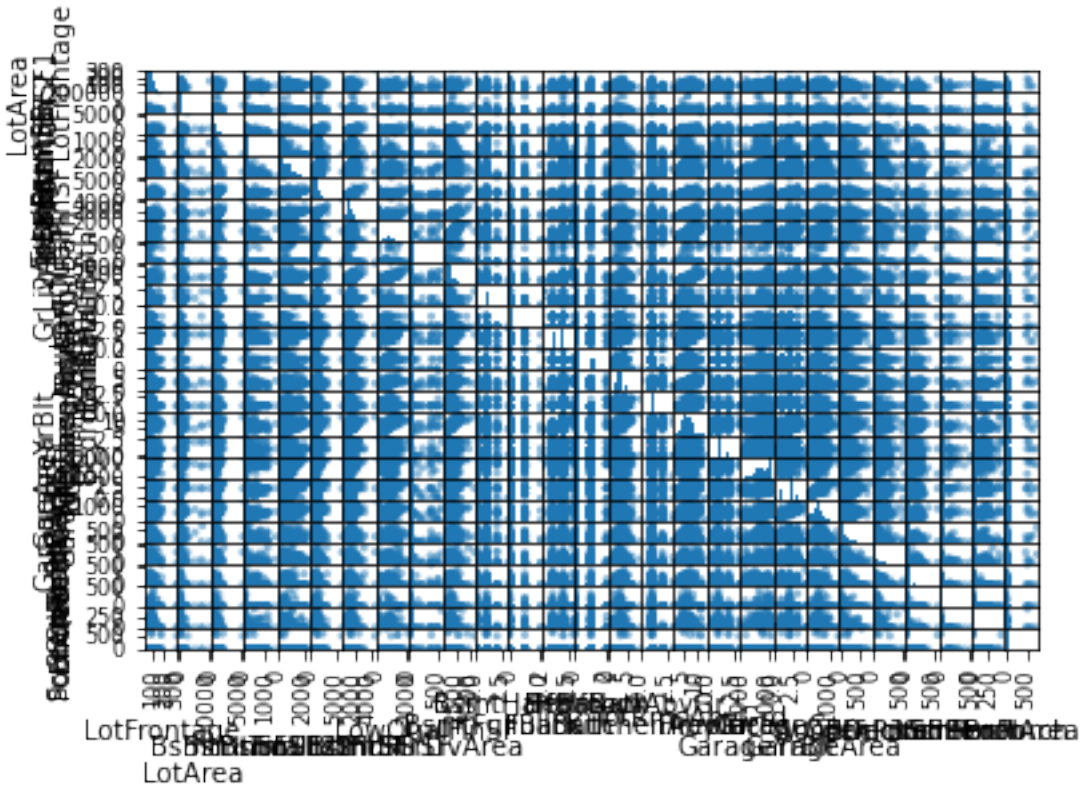
[63]:
```
# create single dataframe called ames_preprocessed

ames_preprocessed = ames[numeric_vars]
```

1.4 Scatter matrix

Create the scatter matrix for the Ames Housing data. This takes a few minutes to load!

[6]:
```
# use pd.plotting.scatter_matrix
pd.plotting.scatter_matrix(ames_preprocessed);
```



The scatter matrix took a while to load and is hard to read. Run the code below to see if adjusting some of the visualization settings helps.

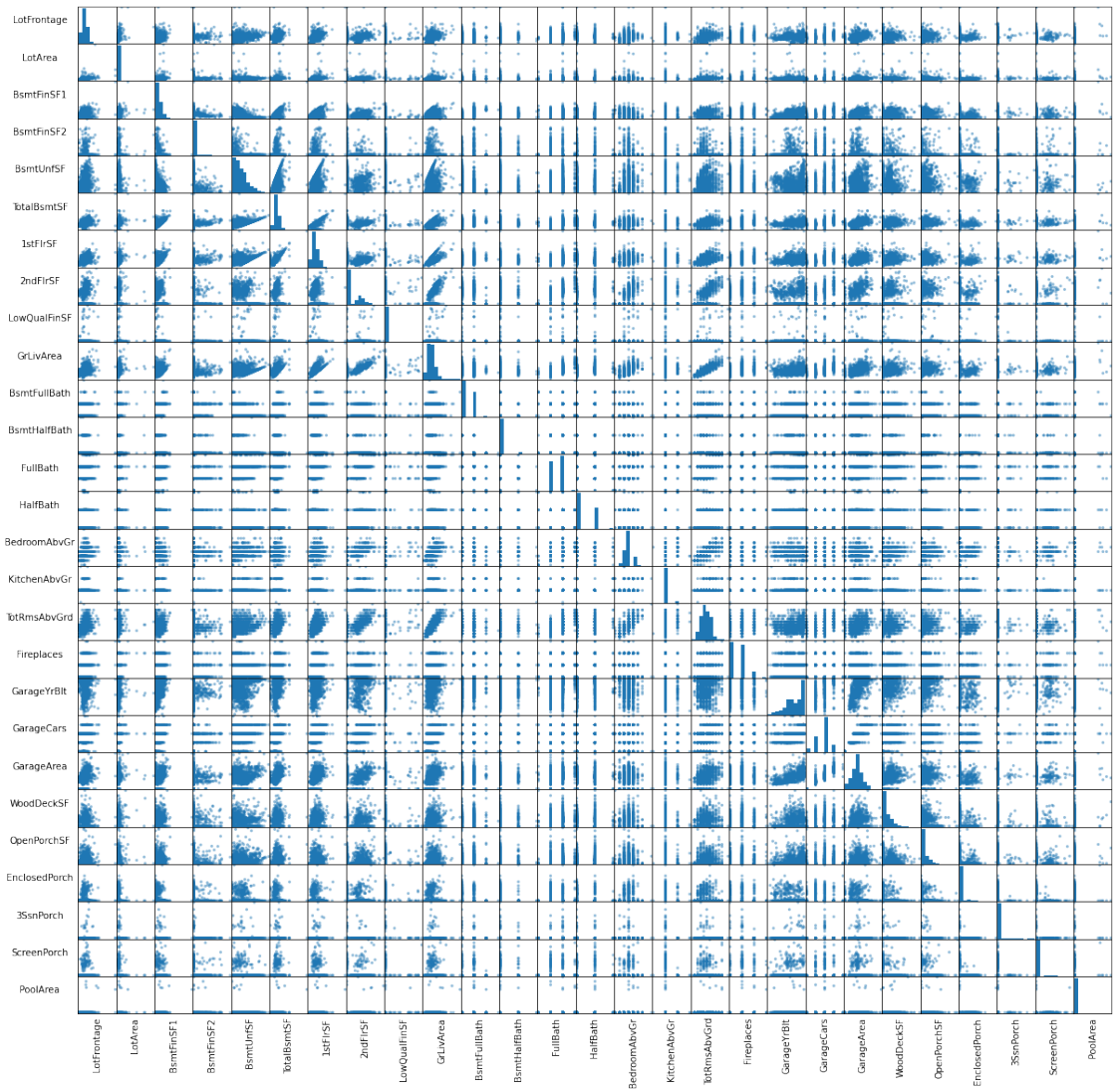
[7]:
```
sm = pd.plotting.scatter_matrix(ames_preprocessed, figsize=[20, 20]);

# Rotates the text
[s.xaxis.label.set_rotation(90) for s in sm.reshape(-1)]
[s.yaxis.label.set_rotation(0) for s in sm.reshape(-1)]

#May need to offset label when rotating to prevent overlap of figure
[s.get_yaxis().set_label_coords(-1,0.5) for s in sm.reshape(-1)]

#Hide all ticks
[s.set_xticks(()) for s in sm.reshape(-1)]
[s.set_yticks(()) for s in sm.reshape(-1)]

plt.show()
```



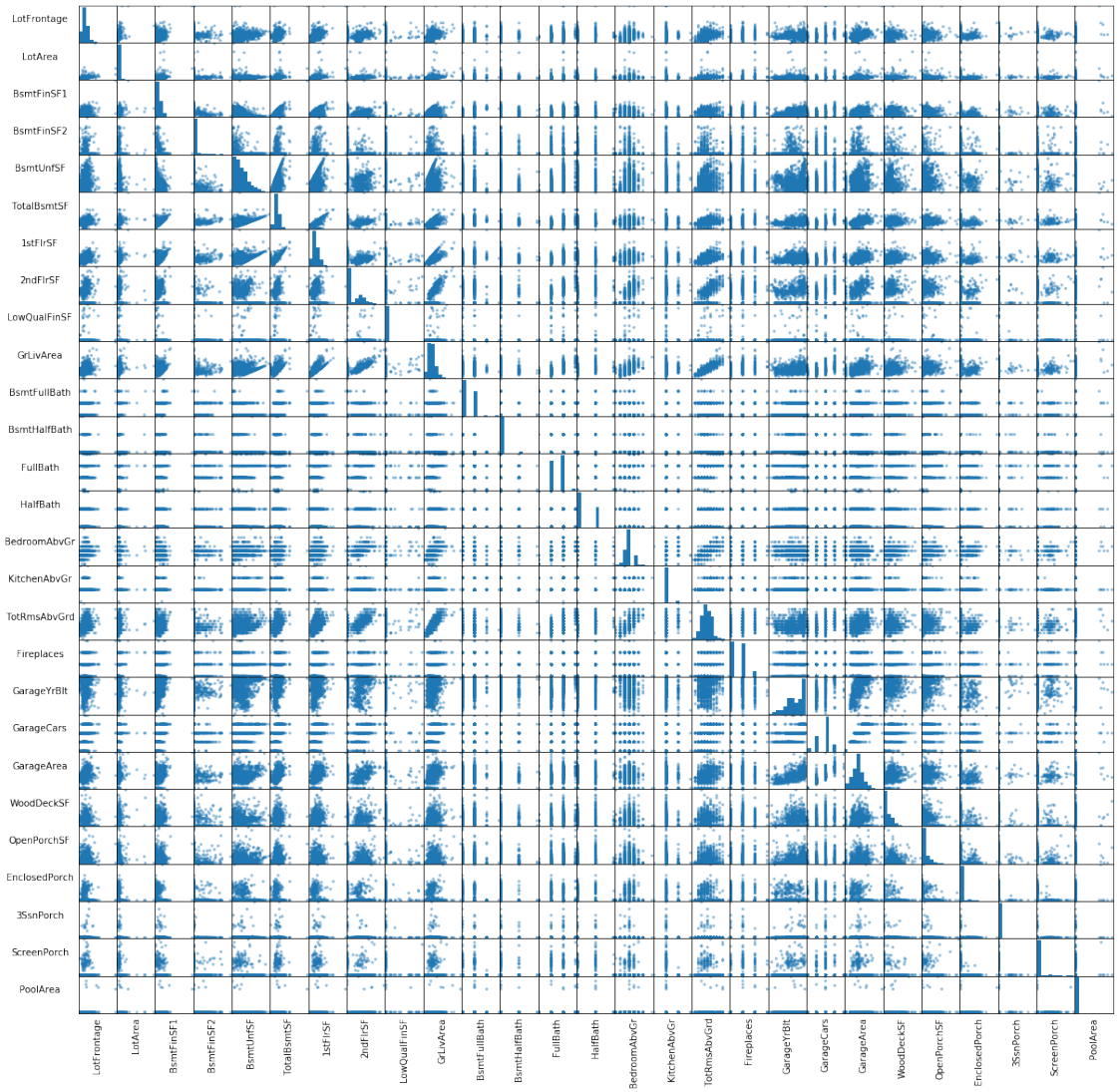
[8]:
```
#__Solution__
sm = pd.plotting.scatter_matrix(ames_preprocessed, figsize=[20, 20]);

# Rotates the text
[s.xaxis.label.set_rotation(90) for s in sm.reshape(-1)]
[s.yaxis.label.set_rotation(0) for s in sm.reshape(-1)]

#May need to offset label when rotating to prevent overlap of figure
[s.get_yaxis().set_label_coords(-1,0.5) for s in sm.reshape(-1)]

#Hide all ticks
[s.set_xticks(()) for s in sm.reshape(-1)]
[s.set_yticks(()) for s in sm.reshape(-1)]

plt.show()
```



The enhanced plot demonstrates that with larger datasets, scatter matricies become less useful. Through careful examination of the matrix it's clear that `TotRmsAbvGrd` seems correlated with `GrLivArea`, but how easy to use would this matrix if a dataset has hundreds or thousands of variables? Also visual approach to finding correlation cannot be automated, so a numeric approach is a good next step.

1.5 Correlation matrix

Next, create and look at the correlation matrix:

[64]:
```
ames_preprocessed.corr()
```

[64]:

	LotFrontage	LotArea	BsmtFinSF1	BsmtFinSF2	BsmtUnfSF	\
LotFrontage	1.000000	0.426095	0.233633	0.049900	0.132644	
LotArea	0.426095	1.000000	0.214103	0.111170	-0.002618	
BsmtFinSF1	0.233633	0.214103	1.000000	-0.050117	-0.495251	
BsmtFinSF2	0.049900	0.111170	-0.050117	1.000000	-0.209294	
BsmtUnfSF	0.132644	-0.002618	-0.495251	-0.209294	1.000000	
TotalBsmtSF	0.392075	0.260833	0.522396	0.104810	0.415360	
1stFlrSF	0.457181	0.299475	0.445863	0.097117	0.317987	
2ndFlrSF	0.080177	0.050986	-0.137079	-0.099260	0.004469	
LowQualFinSF	0.038469	0.004779	-0.064503	0.014807	0.028167	
GrLivArea	0.402797	0.263116	0.208171	-0.009640	0.240257	
BsmtFullBath	0.100949	0.158155	0.649212	0.158678	-0.422900	
BsmtHalfBath	-0.007234	0.048046	0.067418	0.070948	-0.095804	
FullBath	0.198769	0.126031	0.058543	-0.076444	0.288886	
HalfBath	0.053532	0.014259	0.004262	-0.032148	-0.041118	
BedroomAbvGr	0.263170	0.119690	-0.107355	-0.015728	0.166643	
KitchenAbvGr	-0.006069	-0.017784	-0.081007	-0.040751	0.030086	
TotRmsAbvGrd	0.352096	0.190015	0.044316	-0.035227	0.250647	
Fireplaces	0.266639	0.271364	0.260011	0.046921	0.051575	
GarageYrBlt	0.070250	-0.024947	0.153484	-0.088011	0.190708	
GarageCars	0.285691	0.154871	0.224054	-0.038264	0.214175	
GarageArea	0.344997	0.180403	0.296970	-0.018227	0.183303	
WoodDeckSF	0.088521	0.171698	0.204306	0.067898	-0.005316	
OpenPorchSF	0.151972	0.084774	0.111761	0.003093	0.129005	
EnclosedPorch	0.010700	-0.018340	-0.102303	0.036543	-0.002538	
3SsnPorch	0.070029	0.020423	0.026451	-0.029993	0.020764	
ScreenPorch	0.041383	0.043160	0.062021	0.088871	-0.012579	
PoolArea	0.206167	0.077672	0.140491	0.041709	-0.035092	

	TotalBsmtSF	1stFlrSF	2ndFlrSF	LowQualFinSF	GrLivArea	\
LotFrontage	0.392075	0.457181	0.080177	0.038469	0.402797	
LotArea	0.260833	0.299475	0.050986	0.004779	0.263116	
BsmtFinSF1	0.522396	0.445863	-0.137079	-0.064503	0.208171	

BsmtFinSF2	0.104810	0.097117	-0.099260	0.014807	-0.009640
BsmtUnfSF	0.415360	0.317987	0.004469	0.028167	0.240257
TotalBsmtSF	1.000000	0.819530	-0.174512	-0.033245	0.454868
1stFlrSF	0.819530	1.000000	-0.202646	-0.014241	0.566024
2ndFlrSF	-0.174512	-0.202646	1.000000	0.063353	0.687501
LowQualFinSF	-0.033245	-0.014241	0.063353	1.000000	0.134683
GrLivArea	0.454868	0.566024	0.687501	0.134683	1.000000
BsmtFullBath	0.307351	0.244671	-0.169494	-0.047143	0.034836
BsmtHalfBath	-0.000315	0.001956	-0.023855	-0.005842	-0.018918
FullBath	0.323722	0.380637	0.421378	-0.000710	0.630012
HalfBath	-0.048804	-0.119916	0.609707	-0.027080	0.415772
BedroomAbvGr	0.050450	0.127401	0.502901	0.105607	0.521270
KitchenAbvGr	-0.068901	0.068101	0.059306	0.007522	0.100063
TotRmsAbvGrd	0.285573	0.409516	0.616423	0.131185	0.825489
Fireplaces	0.339519	0.410531	0.194561	-0.021272	0.461679
GarageYrBlt	0.322445	0.233449	0.070832	-0.036363	0.231197
GarageCars	0.434585	0.439317	0.183926	-0.094480	0.467247
GarageArea	0.486665	0.489782	0.138347	-0.067601	0.468997
WoodDeckSF	0.232019	0.235459	0.092165	-0.025444	0.247433
OpenPorchSF	0.247264	0.211671	0.208026	0.018251	0.330224
EnclosedPorch	-0.095478	-0.065292	0.061989	0.061081	0.009113
3SsnPorch	0.037384	0.056104	-0.024358	-0.004296	0.020643
ScreenPorch	0.084489	0.088758	0.040606	0.026799	0.101510
PoolArea	0.126053	0.131525	0.081487	0.062157	0.170205

	BsmtFullBath	BsmtHalfBath	FullBath	HalfBath	BedroomAbvGr	\
LotFrontage	0.100949	-0.007234	0.198769	0.053532	0.263170	
LotArea	0.158155	0.048046	0.126031	0.014259	0.119690	
BsmtFinSF1	0.649212	0.067418	0.058543	0.004262	-0.107355	
BsmtFinSF2	0.158678	0.070948	-0.076444	-0.032148	-0.015728	
BsmtUnfSF	-0.422900	-0.095804	0.288886	-0.041118	0.166643	
TotalBsmtSF	0.307351	-0.000315	0.323722	-0.048804	0.050450	
1stFlrSF	0.244671	0.001956	0.380637	-0.119916	0.127401	
2ndFlrSF	-0.169494	-0.023855	0.421378	0.609707	0.502901	
LowQualFinSF	-0.047143	-0.005842	-0.000710	-0.027080	0.105607	
GrLivArea	0.034836	-0.018918	0.630012	0.415772	0.521270	
BsmtFullBath	1.000000	-0.147871	-0.064512	-0.030905	-0.150673	
BsmtHalfBath	-0.147871	1.000000	-0.054536	-0.012340	0.046519	
FullBath	-0.064512	-0.054536	1.000000	0.136381	0.363252	
HalfBath	-0.030905	-0.012340	0.136381	1.000000	0.226651	
BedroomAbvGr	-0.150673	0.046519	0.363252	0.226651	1.000000	
KitchenAbvGr	-0.041503	-0.037944	0.133115	-0.068263	0.198597	
TotRmsAbvGrd	-0.053275	-0.023836	0.554784	0.343415	0.676620	
Fireplaces	0.137928	0.028976	0.243671	0.203649	0.107570	
GarageYrBlt	0.124553	-0.077464	0.484557	0.196785	-0.064518	
GarageCars	0.131881	-0.020891	0.469672	0.219178	0.086106	
GarageArea	0.179189	-0.024536	0.405656	0.163549	0.065253	

WoodDeckSF	0.175315	0.040161	0.187703	0.108080	0.046854
OpenPorchSF	0.067341	-0.025324	0.259977	0.199740	0.093810
EnclosedPorch	-0.049911	-0.008555	-0.115093	-0.095317	0.041570
3SsnPorch	-0.000106	0.035114	0.035353	-0.004972	-0.024478
ScreenPorch	0.023148	0.032121	-0.008106	0.072426	0.044300
PoolArea	0.067616	0.020025	0.049604	0.022381	0.070703

	KitchenAbvGr	TotRmsAbvGrd	Fireplaces	GarageYrBlt	\
LotFrontage	-0.006069	0.352096	0.266639	0.070250	
LotArea	-0.017784	0.190015	0.271364	-0.024947	
BsmtFinSF1	-0.081007	0.044316	0.260011	0.153484	
BsmtFinSF2	-0.040751	-0.035227	0.046921	-0.088011	
BsmtUnfSF	0.030086	0.250647	0.051575	0.190708	
TotalBsmtSF	-0.068901	0.285573	0.339519	0.322445	
1stFlrSF	0.068101	0.409516	0.410531	0.233449	
2ndFlrSF	0.059306	0.616423	0.194561	0.070832	
LowQualFinSF	0.007522	0.131185	-0.021272	-0.036363	
GrLivArea	0.100063	0.825489	0.461679	0.231197	
BsmtFullBath	-0.041503	-0.053275	0.137928	0.124553	
BsmtHalfBath	-0.037944	-0.023836	0.028976	-0.077464	
FullBath	0.133115	0.554784	0.243671	0.484557	
HalfBath	-0.068263	0.343415	0.203649	0.196785	
BedroomAbvGr	0.198597	0.676620	0.107570	-0.064518	
KitchenAbvGr	1.000000	0.256045	-0.123936	-0.124411	
TotRmsAbvGrd	0.256045	1.000000	0.326114	0.148112	
Fireplaces	-0.123936	0.326114	1.000000	0.046822	
GarageYrBlt	-0.124411	0.148112	0.046822	1.000000	
GarageCars	-0.050634	0.362289	0.300789	0.588920	
GarageArea	-0.064433	0.337822	0.269141	0.564567	
WoodDeckSF	-0.090130	0.165984	0.200019	0.224577	
OpenPorchSF	-0.070091	0.234192	0.169405	0.228425	
EnclosedPorch	0.037312	0.004151	-0.024822	-0.297003	
3SsnPorch	-0.024600	-0.006683	0.011257	0.023544	
ScreenPorch	-0.051613	0.059383	0.184530	-0.075418	
PoolArea	-0.014525	0.083757	0.095074	-0.014501	

	GarageCars	GarageArea	WoodDeckSF	OpenPorchSF	EnclosedPorch	\
LotFrontage	0.285691	0.344997	0.088521	0.151972	0.010700	
LotArea	0.154871	0.180403	0.171698	0.084774	-0.018340	
BsmtFinSF1	0.224054	0.296970	0.204306	0.111761	-0.102303	
BsmtFinSF2	-0.038264	-0.018227	0.067898	0.003093	0.036543	
BsmtUnfSF	0.214175	0.183303	-0.005316	0.129005	-0.002538	
TotalBsmtSF	0.434585	0.486665	0.232019	0.247264	-0.095478	
1stFlrSF	0.439317	0.489782	0.235459	0.211671	-0.065292	
2ndFlrSF	0.183926	0.138347	0.092165	0.208026	0.061989	
LowQualFinSF	-0.094480	-0.067601	-0.025444	0.018251	0.061081	
GrLivArea	0.467247	0.468997	0.247433	0.330224	0.009113	

BsmtFullBath	0.131881	0.179189	0.175315	0.067341	-0.049911
BsmtHalfBath	-0.020891	-0.024536	0.040161	-0.025324	-0.008555
FullBath	0.469672	0.405656	0.187703	0.259977	-0.115093
HalfBath	0.219178	0.163549	0.108080	0.199740	-0.095317
BedroomAbvGr	0.086106	0.065253	0.046854	0.093810	0.041570
KitchenAbvGr	-0.050634	-0.064433	-0.090130	-0.070091	0.037312
TotRmsAbvGrd	0.362289	0.337822	0.165984	0.234192	0.004151
Fireplaces	0.300789	0.269141	0.200019	0.169405	-0.024822
GarageYrBlt	0.588920	0.564567	0.224577	0.228425	-0.297003
GarageCars	1.000000	0.882475	0.226342	0.213569	-0.151434
GarageArea	0.882475	1.000000	0.224666	0.241435	-0.121777
WoodDeckSF	0.226342	0.224666	1.000000	0.058661	-0.125989
OpenPorchSF	0.213569	0.241435	0.058661	1.000000	-0.093079
EnclosedPorch	-0.151434	-0.121777	-0.125989	-0.093079	1.000000
3SsnPorch	0.035765	0.035087	-0.032771	-0.005842	-0.037305
ScreenPorch	0.050494	0.051412	-0.074181	0.074304	-0.082864
PoolArea	0.020934	0.061047	0.073378	0.060762	0.054203

	3SsnPorch	ScreenPorch	PoolArea
LotFrontage	0.070029	0.041383	0.206167
LotArea	0.020423	0.043160	0.077672
BsmtFinSF1	0.026451	0.062021	0.140491
BsmtFinSF2	-0.029993	0.088871	0.041709
BsmtUnfSF	0.020764	-0.012579	-0.035092
TotalBsmtSF	0.037384	0.084489	0.126053
1stFlrSF	0.056104	0.088758	0.131525
2ndFlrSF	-0.024358	0.040606	0.081487
LowQualFinSF	-0.004296	0.026799	0.062157
GrLivArea	0.020643	0.101510	0.170205
BsmtFullBath	-0.000106	0.023148	0.067616
BsmtHalfBath	0.035114	0.032121	0.020025
FullBath	0.035353	-0.008106	0.049604
HalfBath	-0.004972	0.072426	0.022381
BedroomAbvGr	-0.024478	0.044300	0.070703
KitchenAbvGr	-0.024600	-0.051613	-0.014525
TotRmsAbvGrd	-0.006683	0.059383	0.083757
Fireplaces	0.011257	0.184530	0.095074
GarageYrBlt	0.023544	-0.075418	-0.014501
GarageCars	0.035765	0.050494	0.020934
GarageArea	0.035087	0.051412	0.061047
WoodDeckSF	-0.032771	-0.074181	0.073378
OpenPorchSF	-0.005842	0.074304	0.060762
EnclosedPorch	-0.037305	-0.082864	0.054203
3SsnPorch	1.000000	-0.031436	-0.007992
ScreenPorch	-0.031436	1.000000	0.051307
PoolArea	-0.007992	0.051307	1.000000

Return `True` for positive or negative correlations that are bigger than 0.75 in the correlation matrix:

[65]:
```
abs(ames_preprocessed.corr())>0.75
```

[65]:

	LotFrontage	LotArea	BsmtFinSF1	BsmtFinSF2	BsmtUnfSF	\
LotFrontage	True	False	False	False	False	
LotArea	False	True	False	False	False	
BsmtFinSF1	False	False	True	False	False	
BsmtFinSF2	False	False	False	True	False	
BsmtUnfSF	False	False	False	False	True	
TotalBsmtSF	False	False	False	False	False	
1stFlrSF	False	False	False	False	False	
2ndFlrSF	False	False	False	False	False	
LowQualFinSF	False	False	False	False	False	
GrLivArea	False	False	False	False	False	
BsmtFullBath	False	False	False	False	False	
BsmtHalfBath	False	False	False	False	False	
FullBath	False	False	False	False	False	
HalfBath	False	False	False	False	False	
BedroomAbvGr	False	False	False	False	False	
KitchenAbvGr	False	False	False	False	False	
TotRmsAbvGrd	False	False	False	False	False	
Fireplaces	False	False	False	False	False	
GarageYrBlt	False	False	False	False	False	
GarageCars	False	False	False	False	False	
GarageArea	False	False	False	False	False	
WoodDeckSF	False	False	False	False	False	
OpenPorchSF	False	False	False	False	False	
EnclosedPorch	False	False	False	False	False	
3SsnPorch	False	False	False	False	False	
ScreenPorch	False	False	False	False	False	
PoolArea	False	False	False	False	False	

	TotalBsmtSF	1stFlrSF	2ndFlrSF	LowQualFinSF	GrLivArea	\
LotFrontage	False	False	False	False	False	
LotArea	False	False	False	False	False	
BsmtFinSF1	False	False	False	False	False	
BsmtFinSF2	False	False	False	False	False	
BsmtUnfSF	False	False	False	False	False	
TotalBsmtSF	True	True	False	False	False	
1stFlrSF	True	True	False	False	False	
2ndFlrSF	False	False	True	False	False	
LowQualFinSF	False	False	False	True	False	
GrLivArea	False	False	False	False	True	
BsmtFullBath	False	False	False	False	False	
BsmtHalfBath	False	False	False	False	False	
FullBath	False	False	False	False	False	

```
HalfBath            False     False     False         False      False
BedroomAbvGr        False     False     False         False      False
KitchenAbvGr        False     False     False         False      False
TotRmsAbvGrd        False     False     False         False       True
Fireplaces          False     False     False         False      False
GarageYrBlt         False     False     False         False      False
GarageCars          False     False     False         False      False
GarageArea          False     False     False         False      False
WoodDeckSF          False     False     False         False      False
OpenPorchSF         False     False     False         False      False
EnclosedPorch       False     False     False         False      False
3SsnPorch           False     False     False         False      False
ScreenPorch         False     False     False         False      False
PoolArea            False     False     False         False      False

              BsmtFullBath  BsmtHalfBath  FullBath  HalfBath  BedroomAbvGr  \
LotFrontage          False         False     False     False         False
LotArea              False         False     False     False         False
BsmtFinSF1           False         False     False     False         False
BsmtFinSF2           False         False     False     False         False
BsmtUnfSF            False         False     False     False         False
TotalBsmtSF          False         False     False     False         False
1stFlrSF             False         False     False     False         False
2ndFlrSF             False         False     False     False         False
LowQualFinSF         False         False     False     False         False
GrLivArea            False         False     False     False         False
BsmtFullBath          True         False     False     False         False
BsmtHalfBath         False          True     False     False         False
FullBath             False         False      True     False         False
HalfBath             False         False     False      True         False
BedroomAbvGr         False         False     False     False          True
KitchenAbvGr         False         False     False     False         False
TotRmsAbvGrd         False         False     False     False         False
Fireplaces           False         False     False     False         False
GarageYrBlt          False         False     False     False         False
GarageCars           False         False     False     False         False
GarageArea           False         False     False     False         False
WoodDeckSF           False         False     False     False         False
OpenPorchSF          False         False     False     False         False
EnclosedPorch        False         False     False     False         False
3SsnPorch            False         False     False     False         False
ScreenPorch          False         False     False     False         False
PoolArea             False         False     False     False         False

              KitchenAbvGr  TotRmsAbvGrd  Fireplaces  GarageYrBlt  \
LotFrontage          False         False       False        False
LotArea              False         False       False        False
```

```
BsmtFinSF1            False        False       False        False
BsmtFinSF2            False        False       False        False
BsmtUnfSF             False        False       False        False
TotalBsmtSF           False        False       False        False
1stFlrSF              False        False       False        False
2ndFlrSF              False        False       False        False
LowQualFinSF          False        False       False        False
GrLivArea             False         True       False        False
BsmtFullBath          False        False       False        False
BsmtHalfBath          False        False       False        False
FullBath              False        False       False        False
HalfBath              False        False       False        False
BedroomAbvGr          False        False       False        False
KitchenAbvGr           True        False       False        False
TotRmsAbvGrd          False         True       False        False
Fireplaces            False        False        True        False
GarageYrBlt           False        False       False         True
GarageCars            False        False       False        False
GarageArea            False        False       False        False
WoodDeckSF            False        False       False        False
OpenPorchSF           False        False       False        False
EnclosedPorch         False        False       False        False
3SsnPorch             False        False       False        False
ScreenPorch           False        False       False        False
PoolArea              False        False       False        False

              GarageCars  GarageArea  WoodDeckSF  OpenPorchSF  EnclosedPorch  \
LotFrontage        False       False       False        False          False
LotArea            False       False       False        False          False
BsmtFinSF1         False       False       False        False          False
BsmtFinSF2         False       False       False        False          False
BsmtUnfSF          False       False       False        False          False
TotalBsmtSF        False       False       False        False          False
1stFlrSF           False       False       False        False          False
2ndFlrSF           False       False       False        False          False
LowQualFinSF       False       False       False        False          False
GrLivArea          False       False       False        False          False
BsmtFullBath       False       False       False        False          False
BsmtHalfBath       False       False       False        False          False
FullBath           False       False       False        False          False
HalfBath           False       False       False        False          False
BedroomAbvGr       False       False       False        False          False
KitchenAbvGr       False       False       False        False          False
TotRmsAbvGrd       False       False       False        False          False
Fireplaces         False       False       False        False          False
GarageYrBlt        False       False       False        False          False
GarageCars          True        True       False        False          False
```

```
GarageArea           True        True       False        False          False
WoodDeckSF          False       False        True        False          False
OpenPorchSF         False       False       False         True          False
EnclosedPorch       False       False       False        False           True
3SsnPorch           False       False       False        False          False
ScreenPorch         False       False       False        False          False
PoolArea            False       False       False        False          False

               3SsnPorch  ScreenPorch  PoolArea
LotFrontage        False        False     False
LotArea            False        False     False
BsmtFinSF1         False        False     False
BsmtFinSF2         False        False     False
BsmtUnfSF          False        False     False
TotalBsmtSF        False        False     False
1stFlrSF           False        False     False
2ndFlrSF           False        False     False
LowQualFinSF       False        False     False
GrLivArea          False        False     False
BsmtFullBath       False        False     False
BsmtHalfBath       False        False     False
FullBath           False        False     False
HalfBath           False        False     False
BedroomAbvGr       False        False     False
KitchenAbvGr       False        False     False
TotRmsAbvGrd       False        False     False
Fireplaces         False        False     False
GarageYrBlt        False        False     False
GarageCars         False        False     False
GarageArea         False        False     False
WoodDeckSF         False        False     False
OpenPorchSF        False        False     False
EnclosedPorch      False        False     False
3SsnPorch           True        False     False
ScreenPorch        False         True     False
PoolArea           False        False      True
```

Now, include `stack` and `zip` to create a more robust solution that will return the variable pairs from the correlation matrix that have correlations over .75, but less than 1.

[70]:
```
df = ames_preprocessed.corr().abs().stack().reset_index().sort_values(0,
                                                        ascending=False)
df["pairs"] = list(zip(df.level_0, df.level_1))
df.head(15)
df.set_index("pairs")
df.drop(columns = ["level_1", "level_0"], inplace = True)
ind = df.loc[df[0] == float(1)].index
```

```
df.drop(index = ind, inplace = True)
```

[79]:
```
df1 = df.loc[df[0]>0.75]
df1.drop_duplicates(subset = [0], inplace = True)
```

```
<ipython-input-79-5ed87f2c7c19>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df1.drop_duplicates(subset = [0], inplace = True)
```

[80]:
```
df1.head()
```

[80]:

	0	pairs
559	0.882475	(GarageArea, GarageCars)
441	0.825489	(TotRmsAbvGrd, GrLivArea)
167	0.819530	(1stFlrSF, TotalBsmtSF)

Which varibles are highly correlated in the Ames Housing data set?

[ ]:
```
# write answer here

"""
There are three sets of variales that are highly correlated.

Garage Area with Garage Cars, Total Rooms Above Ground with Total Square

Feet of Living Space Above Ground, and First Floor Square Feet with Total

Basement Square Feet.

"""
```

1.6 Make a data decision

Now that you know which variables are correlated with eachother, which would you drop from the dataset?

[ ]:
```
# write answer here
```

[ ]:
```
#__SOLUTIOM__
"""
Since three different pairs of variables are highly correlated, the correct
approach would be to drop one variable from each pair.

One approach would be to drop Garage Area, Total Rooms, and Total Basement
Square Feet.
```

```
Garage Area: The size of the garage is dependent on how many cars are in it.

If you wanted to still keep the information captured by Garage Area,
you could create a new variable "Average space per car" before dropping
Garage Area.

Total Rooms: There are other variables that count the number of kitchens,
bathrooms, bedrooms, etc.

Total Basement Square Feet: The first floor of a building is usually built
upon the foundation, which contains the basement.

To keep the information that there *is* a basement, you could create
a variable "HasBasement", when TotalBsmtSF >0, before deleting the original
variable.
"""
```

1.7 Address the colinearity

Remove the chosen variables from `ames_preprocessed`.

[81]:
```
# write code here

ames_preprocessed.drop(columns = ["GarageArea", "TotRmsAbvGrd","TotalBsmtSF"],
                       inplace = True, axis = 1)
```

```
/opt/anaconda3/envs/learn-env/lib/python3.8/site-
packages/pandas/core/frame.py:4163: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  return super().drop(
```

1.8 Summary

Good job! You got some hands-on practice creating and interpreting a scatter matrix and correlation matrix to identify if variables are collinear in the Ames Housing data set. You also edited the Ames Housing data set so highly correlated variables are removed.