# FEATURE ENGINEERING

 import pandas as pd import numpy as np import matplotlib.pyplot as PI t %matplotlib inline import seaborn as sns import warnings warnings . filterwarnings( ignore ' )

## init •y: 146: UserWarning: A NumPy versio

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | n >=1.16.5 and <1.23.0 is required for this version warnings. warn(f"A NumPy version >={np\_minversion} | of SciPy (detected version and <{np\_maxversion}" | 1 . 26.0 |  |  |  |  | | pep\_data=pd . read\_csv ( " PEPI. csv" ) |  |  |  |  |  |  | |  |  |  |  |  |  |  | | pep\_data . head ( ) |  |  |  |  |  |  | | Id MSSubClass MSZoning LotFrontage LotArea Street | LotShape LandContour | Utilities | PoolArea | PoolQC | Fence | MiscFeaturc | |
|  |

Out [3] :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 60 | 65.0 | 8450 | Pave | NaN | Reg | Lvl | AllPub | o | NaN | NaN | NaB |
|  | 2 | 20 | 80.0 | 9600 | Pave | NaN | Reg | Lvl | AllPub | o | NaN | NaN | NaB |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. 3 60 2. 4 70 3. 5 60 4. rows x 81 columns | 68.0  60.0  84.0 | 11250  9550  14260 | Pave  Pave  Pave | NaN  NaN  NaN | IRI  IRI  IRI | Lvl  Lvl  Lvl | AllPub o  AllPub o  AllPub o | NaN  NaN  NaN | NaN  NaN  NaN | NaB |
| pep\_data . shape |  |  |  |  |  |  |  |  |  |  |

In

(1460, 81)

Out [4] :

* dtypes=pd . DataFrame(pep\_data . dtypes . value\_counts ( ) . reset\_index(name= I col\_types ' ) ) dtypes

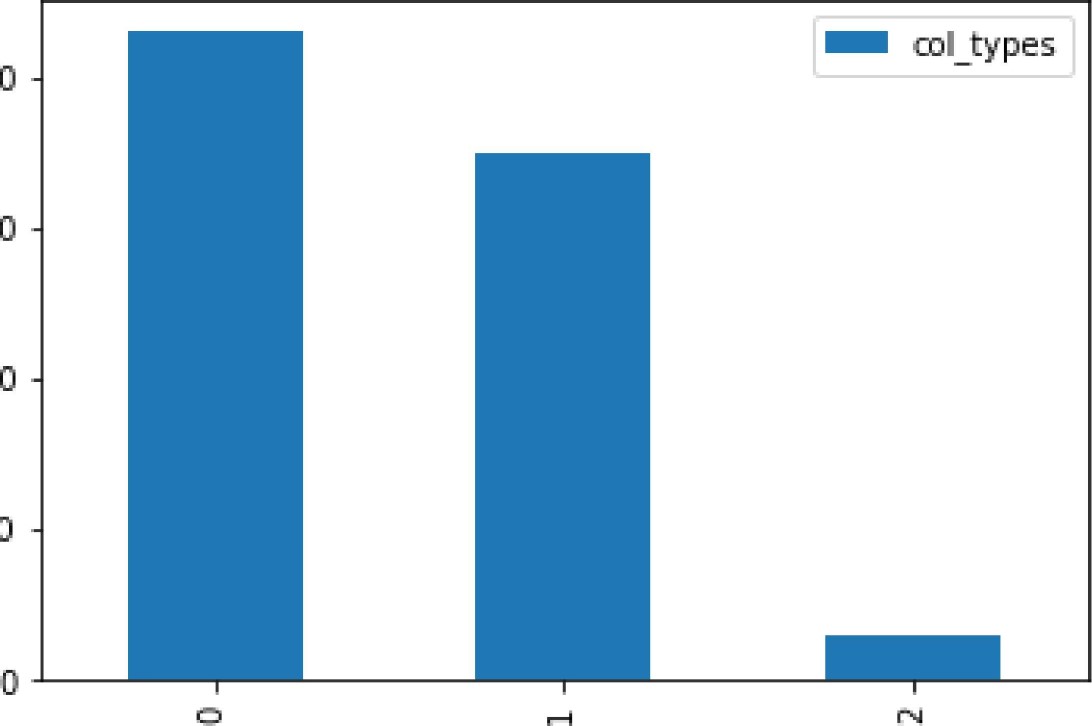
Out [5] : index col\_types



|  |  |
| --- | --- |
| O object | 43 |
| int64 | 35 |
| 2 float64 | 3 |

* dtypes . plot (kind= ' bar' )

|  |  |
| --- | --- |
| Out[6] : | <AxesSubp10t : > |

40

30

20

10

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | for col in pep\_data.columns:  if print(f"{col} {pep\_data [col] . isna( ) . sum()} | {col}-null\_perc . | {pep\_data col] . isna( ) . sum( )/pep\_data . shape [e] } " | |

LotFrontage 259 , LotFrontage-nu11\_perc o. 1773972602739726

Alley 1369 , Alley-null\_perc . e. 9376712328767123

MasVnrType8  MasVnrType-nu11\_perc e. 005479452054794521

MasVnrArea8 MasVnrArea-nu11\_perc e. 005479452054794521 BsmtQua1BsmtQua1-nu11\_perc . e. 025342465753424658

BsmtCondBsmtCond-nu11\_perc . o .025342465753424658

BsmtExposure 38 BsmtExposure-nu11\_perc . o. 026027397260273973 BsmtFinType1 37 BsmtFinType1-nu11\_perc e. 025342465753424658

BsmtFinType2 38 BsmtFinType2-null perc e. 026027397260273973

|  |  |
| --- | --- |
| Electrical | 1 Electrical-null\_perc e. 0006849315068493151 |
| FireplaceQu | 690 FireplaceQu-nu11\_perc 0.4726027397260274 |
| Garage Type | 81 , GarageType-nu11\_perc . o .05547945205479452 |
| GarageYrB1t | 81 GarageYrB1t-nu11\_perc . o .05547945205479452 |

GarageFinish 81 GarageFinish-nu11\_perc . o .05547945205479452

GarageQua1 81 , GarageQua1-nu11\_perc . o .05547945205479452

GarageCond 81 , GarageCond-nu11\_perc . o .05547945205479452

PoolQC 1453 PoolQC-nu11\_perc e. 9952054794520548

Fence 1179 Fence-null\_perc . e. 8075342465753425

MiscFeature . 1406 , MiscFeature-nu11\_perc e. 963013698630137

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | pep\_data . info( ) | |

<class 'pandas. core. frame.DataFrame' > Rangelndex: 1460 entries, 0 to 1459 Data columns (total 81 columns) :

|  |  |  |  |
| --- | --- | --- | --- |
|  | Column | Non-Null Count | Dtype |



|  |  |  |  |
| --- | --- | --- | --- |
| 0 | Id | 1460 non-null | int64 |
| 1 | MSSubC1ass | 1460 non-null | int64 |
| 2 | MSZoning | 1460 non-null | object |
| 3 | LotFrontage | 1201 non-null | float64 |
| 4 | LotArea | 1460 non-null | int64 |
| 5 | Street | 1460 non-null | object |
| 6 | Alley | 91 non-null | object |
| 7 | LotShape | 1460 non-null | object |
| 8 | LandContour | 1460 non-null | object |
| 9 | Utilities | 1460 non-null | object |
| 10 | LotConfig | 1460 non-null | object |
| 11 | LandS10pe | 1460 non-null | object |
| 12 | Neighborhood | 1460 non-null | object |
| 13 | Conditionl | 1460 non-null | object |
| 14 | Condition2 | 1460 non-null | object |
| 15 | BldgType | 1460 non-null | object |
| 16 | HouseSty1e | 1460 non-null | object |
| 17 | OverallQua1 | 1460 non-null | int64 |
| 18 | OverallCond | 1460 non-null | int64 |
| 19 | YearBui1t | 1460 non-null | int64 |
| 20 | YearRemodAdd | 1460 non-null | int64 |
| 21 | RoofSty1e | 1460 non-null | object |
| 22 | RoofMat1 | 1460 non-null | object |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 23 | Exteriorlst | 1460 | non-null | object | |
| 24 | Exterior2nd | 1460 | non-null | object | |
| 25 | MasVnrType | 1452 | non-null | object | |
| 26 | MasVnrArea | 1452 | non-null | float64 | |
| 27 | ExterQua1 | 1460 | non-null | object | |
| 28 | ExterCond | 1460 | non-null | object | |
| 29 | Foundation | 1460 | non-null | object | |
| 30 | BsmtQua1 | 1423 | non-null | object | |
| 31 | BsmtCond | 1423 | non-null | object | |
| 32 | BsmtExposure | 1422 | non-null | object | |
| 33 | BsmtFinType1 | 1423 | non-null | object | |
| 34 | BsmtFinSF1 | 1460 | non-null | int64 | |
| 35 | BsmtFinType2 | 1422 | non-null | object | |
| 36 | BsmtFinSF2 | 1460 | non-null | int64 | |
| 37 | BsmtUnfSF | 1460 | non-null | int64 | |
| 38 | TotalBsmtSF | 1460 | non-null | int64 | |
| 39 | Heating | 1460 | non-null | object | |
| 40 | HeatingQC | 1460 | non-null | object | |
| 41 | CentralAir | 1460 | non-null | object | |
| 42 | Electrical | 1459 | non-null | object | |
| 43 | 1stF1rSF | 1460 | non-null | int64 | |
| 44 | 2ndF1rSF | 1460 | non-null | int64 | |
| 45 | LowQua1FinSF | 1460 | non-null | int64 | |
| 46 | GrLivArea | 1460 | non-null | int64 | |
| 47 | BsmtFu11Bath | 1460 | non-null | int64 | |
| 48 | BsmtHa1fBath | 1460 | non-null | int64 | |
| 49 | FullBath | 1460 | non-null | int64 | |
| 50 | HalfBath | 1460 | non-null | int64 | |
| 51 | BedroomAbvGr | 1460 | non-null | int64 | |
| 52 | KitchebvGr | 1460 | non-null | int64 | |
| 53 | KitchenQua1 | 1460 | non-null | object | |
| 54 | TotRmsAbvGrd | 1460 | non-null | int64 | |
| 55 | Functiol | 1460 | non-null | object | |
| 56 | Fireplaces | 1460 | non-null | int64 | |
| 57 | FireplaceQu | 770 non-null | | object | |
| 58 | GarageType | 1379 non-null | | object | |
| 59 | GarageYrB1t | 1379 non-null | | float64 | |
| 60 | GarageFinish | 1379 non-null | | object | |
| 61 | GarageCars | 1460 non-null | | int64 | |
| 62 | GarageArea | 1460 non-null | | int64 | |
| 63 | GarageQua1 | 1379 non-null | | object | |
| 64 | GarageCond | 1379 non-null | | object | |
| 65 | PavedDrive | 1460 non-null | | object | |
| 66 | WoodDeckSF | 1460 non-null | | int64 | |
| 67 | OpenPorchSF | 1460 non-null | | int64 | |
| 68 | EnclosedPorch | 1460 non-null | | int64 |
| 69 | 3SsnPorch | 1460 non-null | | int64 |
| 70 | ScreenPorch | 1460 non-null | | int64 |
| 71 | PoolArea | 1460 non-null | | int64 |
| 72 | PoolQC | 7 non-null | | object |
| 73 | Fence | 281 non-null | | object |
| 74 | MiscFeature | 54 non-null | | object |
| 75 | MiscVa1 | 1460 non-null | | int64 |
| 76 | MoS01d | 1460 non-null | | int64 |
| 77 | YrS01d | 1460 non-null | | int64 |
| 78 | SaleType | 1460 non-null | | object |
| 79 | SaleCondition | 1460 non-null | | object |
| 80 | SalePrice | 1460 non-null | | int64 |

## dtypes: float64(3), int64(35), object(43) memory usage: 924.0+ KB

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | pep\_data . shape | |

(1460, 81) Out[9] :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | |  | | --- | | for col in pep\_data.columns:  print("{} { } " . format(col,  . unique())) | | | | | |
|  | |  | | --- | | pep\_data [ " PoolQC" ] . isna() ) ] | | | | | |
| cat\_cols=[] for col in pep\_data.columns: if pep\_data[col] .dtype = num\_cols . append (col) else:  cat \_cols. append (col) | ' int64 | or  .dtype | 'float64' : |
|  | |  | | --- | | print(num\_cols, cat \_cols) print(len(num\_cols), len(cat\_cols)) | | | | | |
|  | |  | | --- | | num\_cols\_data=pep\_data [num\_ cols ] num\_cols\_data . head ( ) | | | | | |



Out [11] : Id MSSubClass LotFrontage LotArea OverallQual OverallCond YearBuiIt YearRemodAdd MasVnrArea BsmtFinSF1 WoodDeckSF Ope

  1 60 65.0 8450 7 5 2003 2003 196.0 706

 2 20 80.0 9600 6 8 1976 1976 0.0 978 298

1. 3 60 68.0 11250 7 5 2001 2002 162.0 486 
2. 4 70 60.0 9550 7 5 1915 1970 0.0 216 
3. 5 60 84.0 14260 8 5 2000 2000 350.0 655 192
4. rows x 38 columns



|  |
| --- |
| [cat\_ cols] cat\_cols\_data . head ( ) |

In

Out [12] : MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition 1GarageType GarageFinish

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Pave | NaN | Reg | Lvl | AllPub | Inside | Gtl | CollgCr | Norm | Attchd | RFn |
|  |  | Pave | NaN | Reg | Lvl | AllPub |  | Gtl |  | Feedr | Attchd | RFn |
| 2 |  | Pave | NaN | IRI | Lvl | AllPub | Inside | Gtl | CollgCr | Norm | Attchd | RFn |
| 3 |  | Pave | NaN | IRI | Lvl | AllPub | Corner | Gtl | Crawfor | Norm | Detchd | Unf |
| 4 |  | Pave | NaN | IRI | Lvl | AllPub |  | Gtl | NoRidge | Norm | Attchd | RFn |
| 5 rows x 43 columns | | |



 num\_cols\_data . isna() . sum()

|  |  |  |
| --- | --- | --- |
| " LotFrontage", "GarageYrB1t" ] import plotly.express as PX fig=px.histogram(num\_cols\_data, | # | color= 'YrSold ' |

fig . update\_ layout (bargap=e. 15) fig . show()

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  | | --- | --- | | fig=px.histogram(pep\_data, x=miss[l], fig . update\_ layout (bargap=e. 1) fig . show() | color-- ' SaleCondition ' ) | |

|  |
| --- |
| num\_cols\_data [miss] . describe( ) |

In [15] :

Out[15] : LotFrontage GarageYrBIt

|  |  |  |
| --- | --- | --- |
| count | 1 201 .oooooo | 1379.000000 |
| mean | 70.049958 | 1978.506164 |
| std | 24.284752 | 24.689725 |
| min | 21 .oooooo | 1900.000000 |
| 25% | 59.000000 | 1961 .oooooo |
| 50% | 69.000000 | 1980.000000 |

LotFrontage GarageYrBIt

75% 80.000000 2002.000000

max 313.000000 2010.000000

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  | | --- | --- | | num\_cols\_data[ ' GarageYrB1t ' ] .fillna(num\_cols\_data[ ' GarageYrB1t ' ] . median() , | inp1ace=True) | |

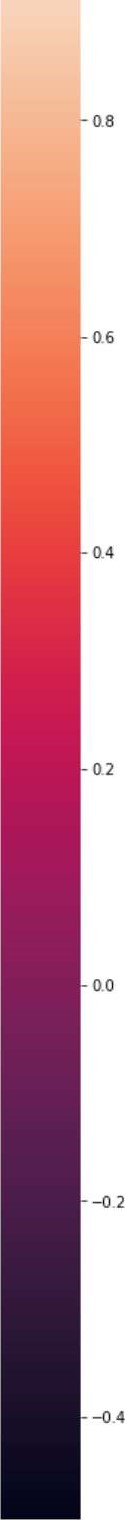
|  |
| --- |
| fig=px.histogram(num\_cols\_data, fig . update\_ layout (bargap=e. 15) fig . show() |

In [17] :

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  | | --- | --- | | fig=px.histogram(pep\_data, x=miss[l], fig . update\_ layout (bargap=e. 1) fig . show() | color= ' SaleCondition ' ) | |

|  |
| --- |
| corr=num\_cols\_data . corr( ) plt.figure(figsize=(35,20)) sns.heatmap(corr, annot=True) |

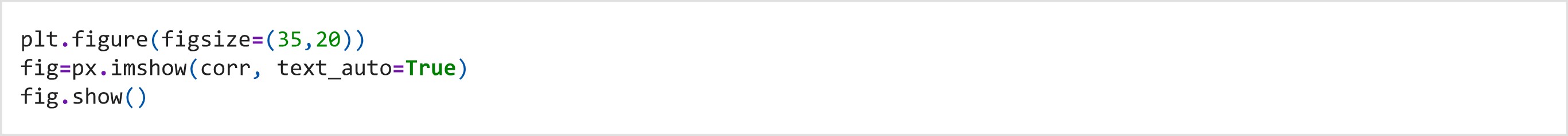
In [19] :

Out[19] :

Ιη [24] :

<AxesSubp10t : >

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 0011 0011 0033 0028 | | | | 0013 | 0013 | 41.022 | 005 | 0005 | 0006 | η.0079 0015 | | 001 | 00056 | η.044 00083 | | 0.0023 402 00056 00068 | | | | 0038 | 0003 | 0027 | 002 000012 0017 | | | 0018 | 003 000048 00029 | | | η.047 | 00013 | 0057 η.0062 0021 000071 0022 | | | |
| MSSubClass | 0011 | 039 -0.14 0033 | | | | 0059 | 0028 | 0041 | 0023 | 007 | 0066 | 0.14 | 24 | 4.25 | 031 | 0046 | 0075 | 00035 00023 | | 013 018 | | 0023 | 028 | 004 | 0046 | 0081 | η.04 | 0099 | 0013 00061 0012 | | | 0044 | 0026 | 00083 00077 0014 0021 0084 | | | |
| LotFrontage | 0.011 | 039 0.43 025 | | | | 0059 | 0.12 | 0.089 | 019 | 023 | 005 | 0.13 | 039 | 046 | 0.08 | 0038 | 04 | 01 | 00072 | 0.2 0054 | | 026 | 00061 | 035 | 0.27 | 0067 | 029 | 0.34 | ο. 089 0.15 | | 0011 | 0.07 | 0041 | 0.21 | 00034 0011 00074' 0.35 | | |
|  | 0033 | 014 043 011 | | | | 00056 | 0014 | 0014 |  | 021 | 011 | 00026 | 026 | 03 | 0051 | 00048 | 0.26 | ο 16 | 0048 | 013 0014 | | ο 12 | 0018 | 019 | 0.27 | 0026 | 015 | ο 18 | 017 0085 | | 0018 | 002 | 0043 | 0078 | 0038 00012 0014 | | 026 |
| OverallQual | 0028 | 0033 025 011 | | | | 0092 | 057 | 055 | 041 | 024 | 0059 | 031 | 054 | 048 | 0-3 | 003 | 059 | 011 | η.04 | 055, 027 | | 0.1 | 418 | 043 | 04 | 051 | 06 | 056 | 024 031 | | 011 | 0.03 | 0065 | 0065 | 0031 0-071 0027 | | 079 |
| OverallCond | 0013 | 0059 0059 00056 0092 | | | |  | -0.38 | 0074 | 013 | 4046 | 004 | -0.14 | 017 | η.14 | 0029 | 0025 | 008 | 0055 | ο 12 | 019 0061 | | 0013 | 0087 | 0058 | 0024 | 031 | η.19 | 0.15 | 00033 0033 | | 007 | 0026 | 0055 | 4002 | 0069 00035 0044 | | 0078 |
| \*arBuilt | 0013 | 0028 | 012 | 0014 | 57 | -0.38 |  | 059 | 032 | 025 | 0049 | 015 | 639 | 028 | 001 | 018 | 02 | 019 | 0038 | 047 024 | | 4.071 | 017 | 0096 | 015 | 078 | ι054 | 048 | 022 | 0.19 | -0.39 | 0031 | 005 | 00049 | 0034 0012 0014 | | 052 |
| \*arRemodAdd | 0022 | 0041 | 0.089 | 0014 | Ι055 | 0074 | 059 |  | 018 | 013 | 0068 | 018 | 029 | 024 | 0.14 | 0062 | 029 |  | -0.0Ι2 | ο." • 018 | | -0.041 | 015 | 019 | 011 | 062 | 042 | 0.37 | 021 | 023 | 019 | 0045 | 0039 | 0.0058 | 001 0021 0036 | | 051 |
| MasVnrArea | 005 | 0023 | 0.19 |  | 041 | 013 | 0.32 | ο 18 |  | 026 | 0072 | 011 | 036 | 0.34 | 017 | 0069 | 039 | 0085 | 0027 | 0.28 02 | | οι | 0038 | 028 | 025 | 025 | 0.36 | 037 | 016 | 013 | 011 | 0019 | 0061 | 0012 | 003 0006 00082 | | 048 |
| BsmtFinSFl | 0.005 | 007 | 023 | 021 | 0.24 | 0046 | 025 | 013 | 026 |  | 005 |  | 52 | 045 | 014 | 0065 | 021 | 06 | 0067 | 0059 00043 | |  | η.08Ι | 0.04-4 | 026 | 015 | 0.22 | 03 | 02 | 011 |  | 0026 | 0062 | 0.14 | ο 0036 0016 | 0014 | 039 |
| BsmtFinSF2 | 0.006 | 0066 | 005 | 011 | 0059 | 004 | 0049 | η 068 | -0.072 | 005 |  | -0.21 | 01 | 0097 | 0099 | 0015 00096 | | 016 | 0071 | 0076 0032 | | 0016 | 0041 | η. 035 | 0047 | 0088 | 0038 | 0018 | 0068 | 00031 | 0037 | 003 | 0089 | 0042 | 00049 0015 | 0032 | 0011 |
| BsmtlJnfSF | 00079 4).Ι4 | | 0.13 | η.0026 | 031 | η.Ι4 | 015 | ο 18 | 0.11 |  | 021 |  | 042 | 0.32 | 00045 | 0028 024 | | 042 | 0096 | 029 4.041 | | 017 | 003 | 025 | 0052 | ο 19 | 0.21 | ο. 18 | 00053 | 013 | 00025 0021 | | 0013 | -0.035 | 0024 0035 | οο4Ι | 021 |
| btalBsmtSF | 0015 024 | | 039 | 0.26 |  | η.17 | ο.39 | 029 | 036 | 0.52 | 01 | 0.42 |  | 082 | 417 | 0033 046 | | 0.31 -Ο.ΟΟΟ3Ι 0.32 4049 | | | | 005 | 0069 | 029 | 034 | 0.31 | 0.43 | 0.49 | 023 | 025 | 4095 0037 | | ο 084 | 0.13 | 41.018 0013 | η.0Ι5 | 061 |
| 1stFlrSF | 001 025 | | 046 | 03 | 48 |  | 028 | 024 | 034 | 045 | 0097 | 032 | 082 |  | 02 | .0.014 057 | | 024 0002 038 012 | | | | 013 | 0068 | 041 | 0.41 | 022 | 044 | 049 | 024 | 021 | 0065 0056 | | 0089 | 013 | 0021 0031 | 0014 | 61 |
| 2ndFlrSF | 00056 031 | | 008 | 0051 | 03 | 0029 | 0.01 | 014 | 017 | 0.14 | 0099 00045 | | 017 | 02 |  | 0063 | | 017 0024 0.42 061 | | | | 05 | 0059 | 062 | 019 | 0068 | 018 | 0.14 | 0092 | 021 | 0062 0024 | | 0041 | 0081 | 0016 0035 | 0029 | 032 |
| LowQualFinSF | 0.044 0046 | | 0038 | 00048 | 003 | 0.025 | 018 | 0062 | 0069 | 0065 | 0015 0028 | | 0033 | 0014 | 0063 | 0.13 | | 0047 00058000071 4.027 | | | | 0.11 | ο. 0075 | 013 | 4.021 | 0027 | .0.094 | 0068 | 0025 | 0018 | 0061 00043 0027 | | | 0062 | 00038 0022 | .0.029 | 0026 |
| &LivArea | 10083 0075 | | 04 | 026 | 5 | 408 | 0.2 | 029 | 039 | 021 | 00096 024 | | 045 | 057 | 069 | α 13 | | 0035 | 0019 |  | 042 | 952 |  |  | 046 | 022 | 047 | 047 | 025 | 0.33 | 00091 0021 01 | | | αι 7 | η.0024 005 | 0037 | 071 |
| BsmtFullBath | 00023 00035 01 | | | 016 | 011 | 0055 | 019 | ο 12 | 0085 | 6 | 016 042 | | 031 | 024 | 017 | 0047 ο.035 | |  | 015 | 0065 | 0031 | 015 | 0042 | 0053 | 014 | ο 12 | 0\_13 | 018 | 018 | 0067 | 005 000011 0023 | | | 0068 | 0023 0025 | 0067 | 023 |
| BsmtHalfBath | -0.02 00023 00072 0048 | | | | 004 | 012 | 0038 | 0012 | 0027 | 0067 | 0071 0096 000031 0002 | | | | 0024 00058 .0.019 | | | 015 |  | 0055 | η 012 | 0047 | 0038 | 0024 | 0.029 | 0075 | 0021 | 0025 | 004 | 0025 00086 0035 0032 | | | | 0.02 | 00074 0033 | 0047 | 0017 |
| FullBath | 60056 013 | | 02 013 | | 055 | 019 | 047 | 044 | 0.28 | 0059 | 0076 029 032 | | | 038 | 042 | | 063 | 0065 | 0055 |  | 014 | 036 | 013 | ΐΦ55 | 024 | 047 | 047 | 041 | 019 | 026 012 0035 00081 | | | | 005 | 0014 0056 | 002 | 056 |
| HalfBath | 00068 ο 18 | | 0.054 0014 | | 027 | 0061 | 0.24 | ο 18 | 02 | ο. 0043 0032 0041 0049 | | | | η.12 | η.027 | | 042 | -0.03Ι | η.0Ι2 | ο. 14 |  | 023 | 0068 | 034 | 02 | 019 | 022 | 0.16 | 011 | 02 -0.095 0005 0072 | | | | 0022 | 00013 0009 | 41.01 | 028 |
| BedroomAbvGr | 0038 0023 | | 026 0.12 | |  | 0013 | 0071 | 0041 |  | 0016 017 005 | | | | 0.13 | 0.5 011 | | 052 | 015 | 0047 | 0.36 | 023 |  | 02 |  | 011 |  | 0086 | 0065 | 0047 | 0094 0042 η.024 0044 | | | | 0071 | 00078 0047 | 0036 | οι 7 |
| KitchebvGr | 0003 028 | | 00061 0018 | | 018 | 0087 | 017 | 015 | 0038 | 0081 0041 003 0069 | | | | 0068 | 0059 00075 | | οι | 0042 | 0038 | 013 | 0068 | 02 |  | 026 | 012 |  | 0051 | 0064 | 009 | 007 0037 0025 0052 | | | | 0015 | 0062 0027 | 0032 | 014 |
| %tRmsAbvGrd | 0.027 004 | | 035 | 019 | 043 | 0058 | 0096 | 019 | 0.28 | 0044 0035 025 029 | | | | 041 | 013 | | 083 | 0053 | 0024 | 055 | 034 | 068 | 026 |  | 033 | 014 | 036 | 034 | 017 | 023 00042 00067 0059 | | | | 0084 | 0025 0037 | 0035 | 053 |
| fjreplaces | 002 0046 | | 0.27 | 0.27 | 04 | 0024 | 015 | 011 | 025 | 026 0047 0052 034 | | | | 0.41 | 0.19 0021 | | 046 | 014 | 0029 | 024 | 0.2 | 011 | 13.12 | 033 |  | 0043 | 03 | 0.27 | 02 | 017 0025 0011 018 | | | | 0095 | 00014 0046 | 0024 | 047 |
| &rageYrBlt | .000Ι2 0081 | | 0067 | 0026 | 051 | 4.31 | ο. 78 | 062 | 025 | 0.15 -0.088 0.19 031 | | | | 022 | 0068 η.027 | | 022 | 0.12 | -0.075 | o.4j | 019 | 4.06 |  | ο 14 | 0043 |  | 0.47 | 0.47 | 022 | 022 028 0023 0076 | | | | 4015 | -0.032 00049 00008 | | 047 |
| &rageCars | 0017 004 | | 029 | 015 | 06 | 019 | 054 | 042 | 036 | 022 0038 021 043 | | | | 044 | 018 0094 | | 047 | 013 | 0021 | 047 | 022 | 0086 | 0051 | 036 | 03 | 047 |  | 088 | 023 | 021 015 0036 005 | | | | 0021 | 0043 0041 0039 | | 064 |
| &rageArea | 0018 0099 | | 0.34 | 018 | 056 | 015 | 0.48 | 0.37 | 037 | 03 0018 0.18 049 | | | | 049 | 014 η. 068 | | 047 | 018 | 0025 | 0.41 | 016 | 0065 | 0064 | 034 | 027 | 047 | 0.88 |  | 022 | 024 012 0035 0051 | | | | 0061 | 0027 0028 0027 | | 062 |
| Vb0dDeckSF | 003 0013 | | 0089 | 0.17 | 0.24 | 00033 | 0.22 | 021 | 0.16 | 02 0068 00053 023 | | | | 024 | 0092 0025 | | 0.25 | 0.18 | 004 | 0.19 | 011 | 0047 | 009 | 017 | 02 | 022 | 023 | 0.22 |  | 0059 0.13 0033 .0.074 | | | | 0073 00096 0021 0022 | | | 0.32 |
|  | .0004800061 | | 015 | 0085 | 031 | 0033 | 019 | 023 | 013 | 011 00031 013 025 | | | | 021 | 021 0018 | | 033 | 0067 | 0025 | 026 | 02 | 0094 | -0.07 | 0.23 | 0.17 | 022 | 021 | 024 | 0059 | 0093 00058 0074 | | | | 0061 0019 0071 0058 | | | 032 |
| EnclosedPorch | 00029 0012 | | 0011 | 0018 |  | 007 | 039 | 019 |  | 43.1 0037 00025 0095 | | | | 0065 | 0062 0061 | | 00091 | 005 | 00086 | 012 | 0095 | 0042 | 0037 | 00042 | 0025 | 028 | 015 | 012 | 013 | 0093 0037 4.083 | | | | 0054 0018 0029 00099 013 | | | |
| 3SsnPorch | 0,047 0044 | | 007 | 002 | 003 | 0.026 | 0031 | 0.045 | 0.019 | 0026 003 0021 0037 | | | | 0.056 | 0024 00043 0021 000011 0035 | | | | | 0035 | 4.005 | -0,024 | 0025 00067 | | 0.011 | 0023 | 0036 | 0035 | 0033 00058 0037 0031 | | | | | 0008 000035 0029 0019 0.045 | | | |
| \*reenPorch | 00013 0026 | | 0041 | 0043 | 0065 | 0.055 | -0.05 | 0039 | 0061 | 0062 0089 0013 0084 | | | | 0089 | 0041 0027 01 0023 0032 00081 0072 | | | | | | | 0044 | 0052 0059 | | 0.18 | 0076 | 005 | 0051 | 0074 0074 0083 0031 | | | | | 0051 0032 0023 0011 0.11 | | | |
| PoolArea | 0.057 ο. 0083 | | 0.21 | 0078 | 0065 | 0002 0.0049 00058 | | | 0012 | ο 14 0042 0035 013 | | | | 013 | 0081 0.062 017 0068 002 0.05 0022 | | | | | | | 0071 | 0015 ο. 084 | | 0095 | 0015 | 0021 | 0061 | 0073 0061 0054 0051 | | | | | 003 0034 406 0092 | | | |
| MiscVal 00062 00077 00034 | | | | 0038 | 0031 | 0069 0034 001 | | | 003 | 00036 00049 0024 0018 | | | |  | 0016 00038 00024 0023 00074 0014 00013 00078 | | | | | | | | 0062 0025 | | 00014 0032 | | 0043 | 0027 00096 0019 0018 000035 0032 | | | | | | 003 00065 00049 0021 | | | |
| MoSold 0021 0.014 0011 | | | | 00012 | 0071 00035 0012 0021 | | | | 0006 | 0016 0015 0035 0013 | | | | 0031 | 0035 0022 005 0025 0033 0056 0009 0047 | | | | | | | | 0027 0037 | | 0046 00049 | | 0041 | 0028 0021 0071 0029 0029 0023 | | | | | | 0034 00065 4115 0046 | | | |
| 000071 4.021 00074 | | | | | 0027 0044 0.036 00082 0014 0032 0041 0015 | | | | | | | | | 0014 | 0029 η.029 0037 0067 4047 -0.02 4.01 0036 | | | | | | | | 0032 0035 | | 0024 000083 0039 | | | 0027 0022 0058 00099 0019 | | | | | | 006 00049 η.Ι5 4.029 | | | |
| f ¯  SlePrice 0022 0084 0.35 0.26 | | | | | 079 \_0.078 \_0 21 | | | | | | | | |  | 032 n026 071 023 0017 5 028 017 | | | | | | | | 014 153 | | 047 047 | | | 032 032 013 0045 | | | | | | 0092 0021 0-046 n029 | | | |

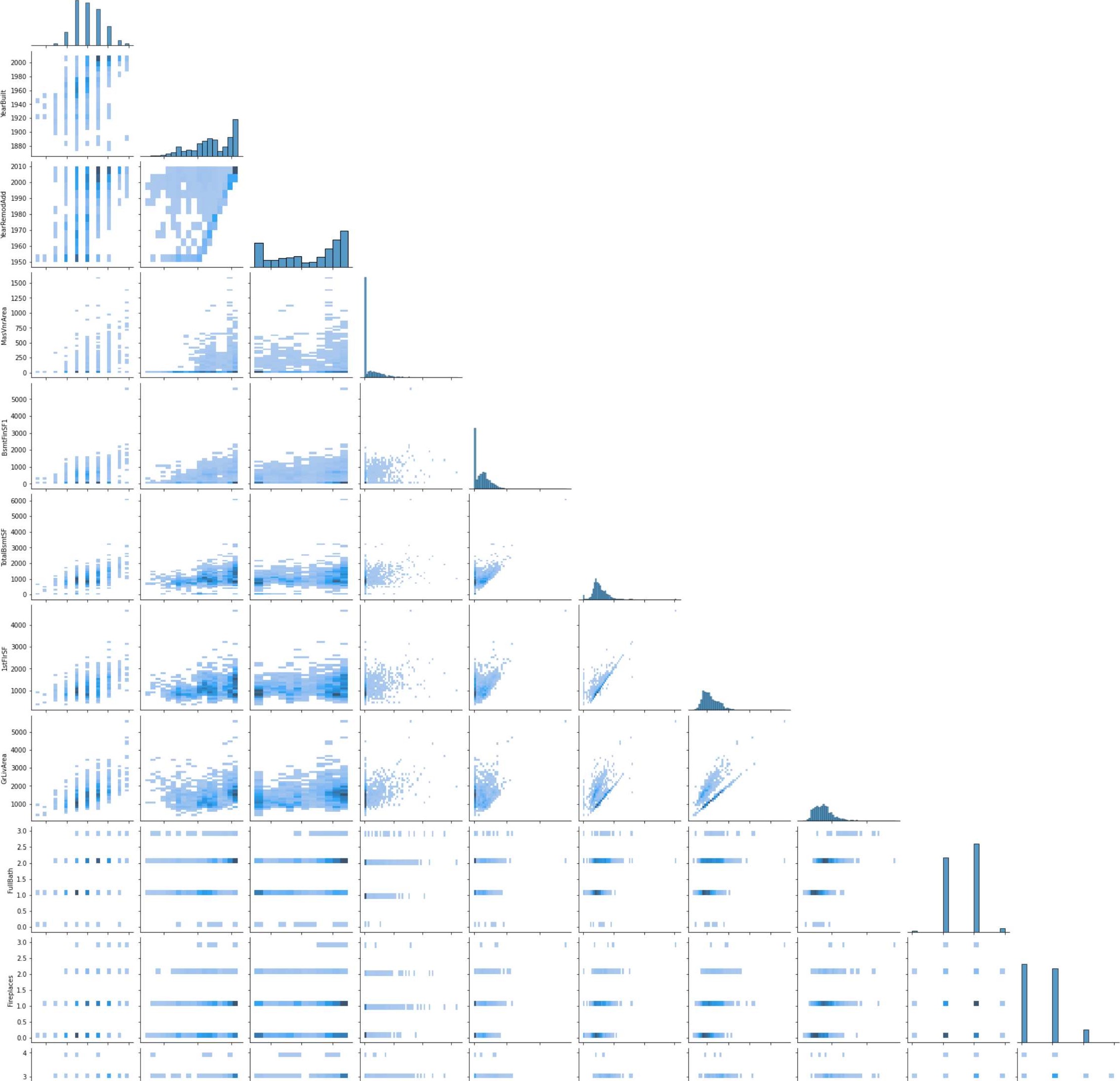


< Figure size  with 0 Axes >

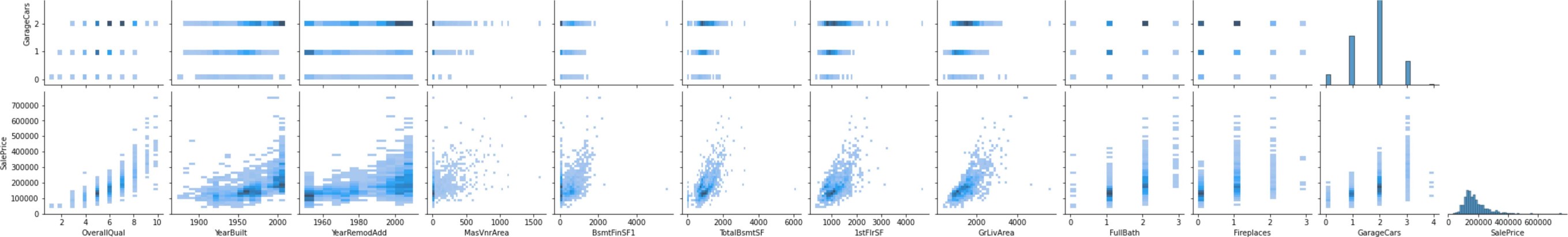
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | | --- | --- | | significant\_features=[ ' OverallQua1 ' ,  'YearBui1t ' , 'YearRemodAdd' , 'MasVnrArea  'TotalBsmtSF ' , '1stF1rSF ' ,  GrLivArea ' ' FullBath ' , ' Fireplaces'  'GarageCars ' , ' SalePrice | ' BsmtFinSF1 ' , | | | | | |
| [ significant\_features ] , | kind= ' hist ' , | diag\_kind= hist ' , | corner-True) |

<seaborn.axisgrid.PairGrid at ox2268d709130> Out[23] :

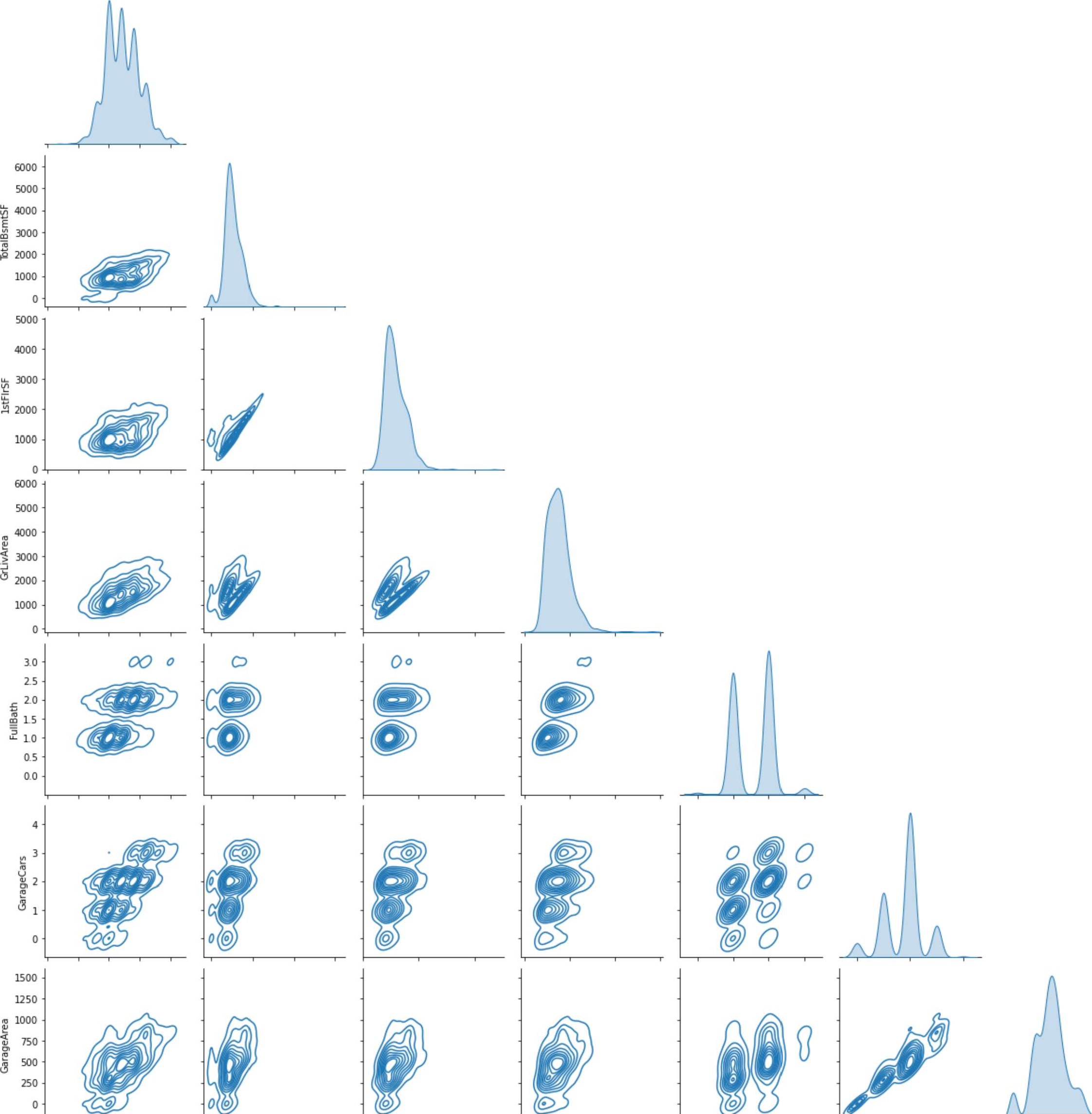
< Figure size  with e Axes >

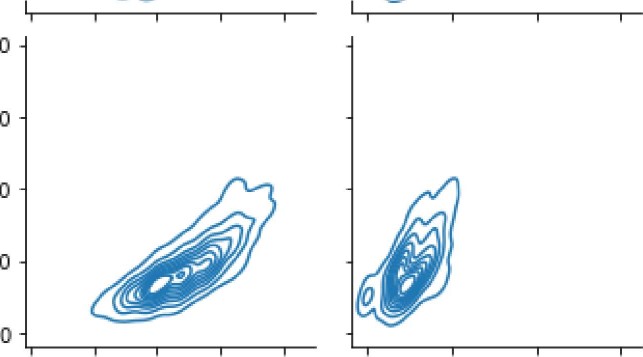
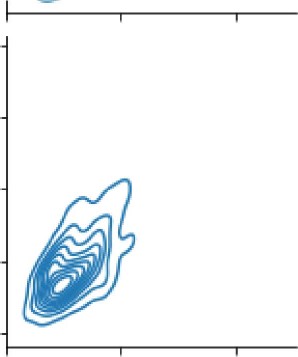


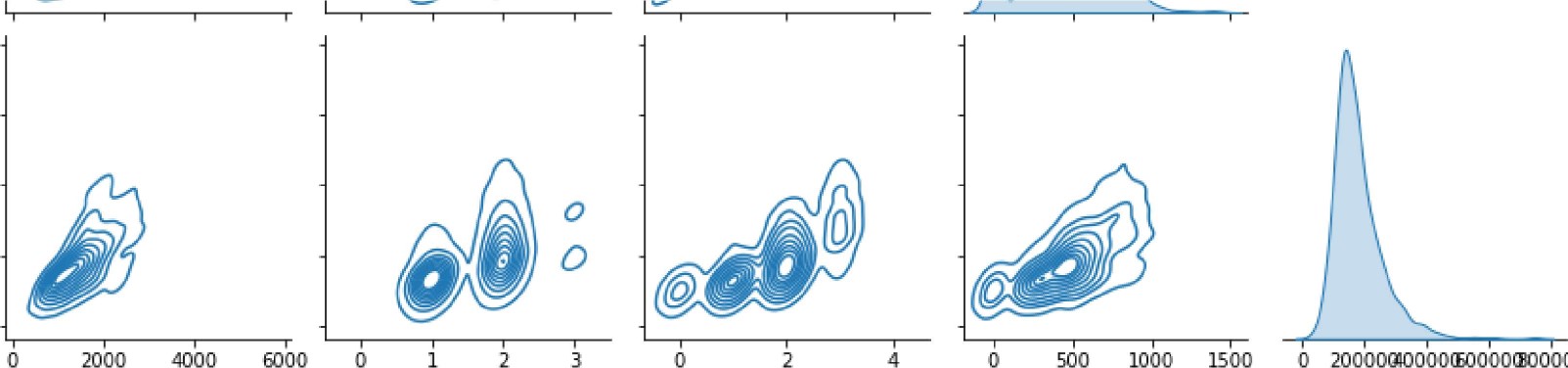
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| sns . pairplot ( | [ significant\_features ] , | kind= kde' , | ' kde' , | corner=True) |

<seaborn .axisgrid.PairGrid at Ox289e0474f40> Out [27] :

Feature\_Engineering



800000



0

20000ßoooocooooæooooo

600000 a. 400000

200000

0.0 25 5.0 7.5 10.0 o 2000 4000 6000  2000 4000

OverallQuaI Zta18smtSF 1stFlrSF GrLivArea Fu113ath GarageCars GarageArea rice

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | cat cols data.isna() •sum() | |

Out [25] : MSZoning

## Street

Alley 1369

LotShape

LandContour

Utilities

LotConfig

LandS10pe

Neighborhood

Conditionl

Condition2 Bldg Type

HouseSty1e RoofSty1e

RoofMat1

Exteriorlst

Exterior2nd

|  |  |
| --- | --- |
| MasVnrType | 8 |

ExterQua1

ExterCond

Foundation

|  |  |
| --- | --- |
| BsmtQua1  BsmtCond | 37 |
| BsmtExposure BsmtFinType1 | 38 |
| BsmtFinType2 | 38 |

Heating

## HeatingQC CentralAir

|  |  |
| --- | --- |
| Electrical  KitchenQua1 Functiol | 1 |
| FireplaceQu | 690 |
| Garage Type | 81 |
| GarageFinish | 81 |
| GarageQua1 | 81 |
| GarageCond  PavedDrive | 81 |
| PoolQC | 1453 |
| Fence | 1179 |
| MiscFeature  Sale Type SaleCondition dtype: int64 | 1406 |

|  |  |  |
| --- | --- | --- |
| missl=[ 'MasVnrType' , 'BsmtQua1 ' , 'BsmtCond' , ' BsmtExposure ' ' BsmtFinType1 1 , ' BsmtFinType2' , GarageCond ' moremiss=[ 'Alley' , ' FireplaceQu I , ' PoolQC' , ' Fence ' , 'MiscFeature' ] | ' Electrical ' , | 'Garage Type' , |

n 24 ] :

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  |  | | --- | --- | --- | --- | --- | | for | col in missl:  print( { } " . format | [col] . value\_counts (normalize=True, | dropna=Fa1se))) # , | dropna=False | | | | | | | |
|  | |  |  |  |  | | --- | --- | --- | --- | | for | col in missl:  print("mean of missing value perc of { } is | { : . 2f}" . format (col, | cat\_cols\_data[col] .isnull() . mean())) | | | | | | | |
| mean | of missing value | perc | of | MasVnrType is 0.01 |
| mean | of missing value | perc | of | BsmtQua1 is 0.03 |
| mean | of missing value | perc | of | BsmtCond is 0.03 |
| mean | of missing value | perc | of | BsmtExposure is  0.03 |
| mean | of missing value | perc | of | BsmtFinType1 is  0.03 |
| mean | of missing value | perc | of | BsmtFinType2 is  0.03 |
| mean | of missing value | perc | of | Electrical is 0.00 |
| mean | of missing value | perc | of | Garage Type is 0.06 |
| mean | of missing value | perc | of | GarageFinish is 0.06 |
| mean | of missing value | perc | of | GarageQua1 is  0.06 |
| mean | of missing value | perc | of | GarageCond is  0.06 |
|  | |  | | --- | | ## create copy of original variable to new one for i, j in zip(missl, imputed) :  cat\_cols\_data [j ] [i] . copy( ) | | | | | | | |
|  | |  |  |  | | --- | --- | --- | | random\_ sample random\_ sample | = cat\_cols\_data[ 'GarageQua1 ' ] .dropna() . sample(cat\_cols\_data[ 'GarageQua1 ' ] .isnull() . sum(), | random\_state=43) | | | | | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | | --- | --- | | random\_ sample = cat\_cols\_data[ 'GarageQua1 ' ] .dropna() .  'GarageQua1 ' ] .isnull() . sum(), random\_ sample | random\_state=43) | |  |  | | random\_ sample . index=cat\_cols\_data [ [ GarageQua1 ' ] . isnull ( ) ] . index |  | |

In [74] :

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | | --- | --- | | cat\_cols\_data . loc [  [ ' GarageQua1 ' ] . isnull ( ) , new\_GarageQua1 ' ]=random\_sample |  | |  |  | | cat\_cols\_data[ ' new\_GarageQua1 ' ] |  | |  |  | | cat\_cols\_data[ ' new\_GarageQua1 ' ] . value\_ counts( ) |  | |
|  |
|  |
|  | |  | | --- | | # In above piece of code we had done that we created sample which is random values in Alley variable which of Lenght equa | |
|  | |  | | --- | | ' 'for col in cat cols data. columns: if col in moremiss:  random\_sample\_train=x\_train[col] .dropna() . .isnull() . sum(), random\_state=e, replace=True) random\_sample\_test=x\_train[col] .dropna() . .isnull() . sum(), random\_state=0, replace=True) else:  random\_sample\_train=x\_train[col] .dropna() .  . isnull() . sum(), random\_state=e) random\_sample\_test=x\_train[col] . dropna() . sample(x\_test[col] .isnull( ) . sum(), random\_state=0) ' | |

Splitting dataset to impute missing value treatment

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  | | --- | --- | --- | --- | | import sklearn from sklearn.model\_selection import train\_test\_split x\_train, x\_test, y\_train, y\_test=train\_test\_split(cat\_cols\_data, | 'SalePrice' ] , | test size-0.25, | random\_state=43) | |

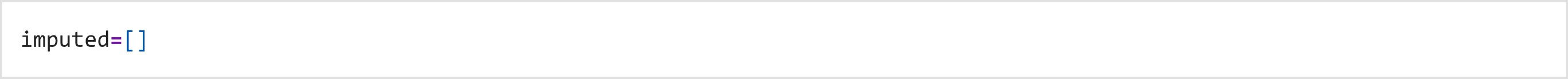
|  |
| --- |
| ## create copy of original variable to new one for i, j in zip(missl, imputed) : x\_train [j ] [i] . copy ( ) x\_test [j . copy() |

In [31] :

|  |
| --- |
| def  . dropna() . .isnull() .sum(), random\_state=0) random\_sample\_test=x\_train[col] .dropna() . . isnull() . sum(), random\_ sample\_train . index=x\_train [x\_train [col] . isnull ( ) ] . index random\_sample\_test . index=x\_test [x\_test [ col] . isnull ( ) ] . index x\_train.loc[x\_train[col] . isnull ( ) , d[col]]=random\_sample\_train x\_test.loc[x\_test[col] . isnull ( ) , d[col] ]=random\_sample\_test |

In [32] :

In [33] : we had done that we created sample which is random values in selected variable which of Length e # In above piece of code

In [29] : 

In [30] : i in missl: for i: ' new + i imputed . append (i)

In [34] :

|  |  |
| --- | --- |
| Out [34] : | [ ' new\_MasVnrType' , |

imputed

new BsmtQua1 , newBsmtCond' , new BsmtExposure new BsmtFinType1 ' , new BsmtFinType2' ,

' new Electrical ' , new\_GarageType' , new GarageFinish ' , newGarageQua1 , new GarageCond ' ]

|  |  |
| --- | --- |
| d=dict() for i, j in zip(missl, | imputed) : |

In [35] :



|  |
| --- |
| for col in missl:  random\_sample\_new(col) |

In [36] :

|  |
| --- |
| x train. columns |

In [37] :

'MSZoning' , 'Street' , 'Alley ' , ' LotShape' , ' LandContour' , 'Utilities ' ,

Out[37] : ' ' LandS10pe' , 'Neighborhood' , 'Conditionl' , 'Condition? ,

## ' LotConfig

Bldg Type' , HouseSty1e' , ' RoofSty1e' , ' RoofMat1' , ' Exteriorlst ' ,

## Exterior2nd' , MasVnrType ' , ' ExterQua1' , ' ExterCond' , ' Foundation

BsmtQua1% 'BsmtCond% ' BsmtExposure' BsmtFinType1' , BsmtFinType2' ,

Heating' ' HeatingQC , 'CentralAir% ' Electrical ' , 'KitchenQua1 ,

## Functiol% ' FireplaceQu' GarageType ' 'GarageFinish' , 'GarageQua1 ' ,

'GarageCond' , 'PavedDrive' PoolQC' , ' Fence' ' MiscFeature

SaleType' , ' SaleCondition' , 'new\_MasVnrType' , ' new\_BsmtQua1 ' ,

' new BsmtCond ' , 'new\_BsmtExposure' new\_BsmtFinType1 ' ,

'new\_BsmtFinType2' , 'new Electrical ' , 'new\_GarageType' ,

' new\_GarageFinish' , 'new\_GarageQua1' , I new\_GarageCond' ] , dtype= ' object' )

|  |  |
| --- | --- |
| x\_train.drop(moremiss, axis-I, x\_test.drop(moremiss, axis-I, | inp1ace=True) inp1ace=True) |

In [38] :

In [39] :

x\_train.drop(missl, axis-I, inp1ace=True) x\_test.drop(missl, axis-I, inp1ace=True)

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  | | --- | --- | | for col in x train. columns:  . format(x\_train[col] . value\_counts(norma1ize=Fa1se, dropna=Fa1se))) # | dropna=FaLse | |
|  | |  |  |  |  | | --- | --- | --- | --- | | fig=px.bar(x\_train, x: ' new\_BsmtExposure' , fig . show() | Y= ' new\_BsmtCond ' , | color= 'CentralAir I , | title: ' Bar chart ' ) | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | fig=px.box(x\_train, y='HeatingQC' , fig .  ' Box plot ) fig . show() | 'CentralAir' , | color= ' new\_GarageType ' ) | |



In [43] :



In [46] :





|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| #new cat col data=pd. concat([x\_train,  #new cat col data. shape | | x\_test]) | | |
| #new cat COL data=pd. concat([new\_cat\_col\_data,  #new cat COL data. shape | | | pep\_data[ 'SaLePrice ' ] ] , | axis=l) |
| #new cat col data. drop( 'SaLePrice ' | axis=l, inpLace=True) | | | |

#new cat COL data. drop(missl, axis=l, inpLace=True)

x\_train [x\_train . isna() . any (axis-I) ]

x\_train.isna() .sum()

# MODEL BUILDING

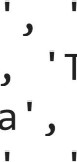
[49] :

num cols data. columns

Index( [ ' ld ' , 'MSSubC1ass ' , ' LotFrontage' , ' LotArea ' , 'OverallQua1 '

## Out [49] : ,

'OverallCond j , 'YearBui1t'YearRemodAdd , MasVnrArea BsmtFinSF1 ,

 'BsmtFinSF2% ' BsmtUnfSF j TotalBsmtSF 1stF1rSF 2ndF1rSF

 LowQua1FinSF GrLivArea 'BsmtFu11Bath% BsmtHa1fBath' , ' FullBath%

'HalfBath j , BedroomAbvGr i , KitchebvGr' , 'TotRmsAbvGrd% ' Fireplaces 

### 'GarageYrB1t' , 'GarageCars ' ' GarageArea' , ' WoodDeckSF ' , 'OpenPorchSF' ,



|  |  |  |
| --- | --- | --- |
| ' EnclosedPorch' , ' 3SsnPorch' , ' ScreenPorch ' ,  'MoS01d' , 'YrS01d ' , 'SalePrice' ] , dtype= ' object ' ) | PoolArea' , | 'MiscVa1' , |

|  |
| --- |
| selected\_cols=[ ' LotFrontage' , ' LotArea' , 'OverallQua1' ,  'YearBui1t ' , 'YearRemodAdd ' , ' MasVnrArea ' ' BsmtFinSF1' ,  ' BsmtUnfSF ' , 'TotalBsmtSF ' , '1stF1rSF' , '2ndF1rSF ' ,  ' GrLivArea ' ' FullBath' , 'TotRmsAbvGrd' , ' Fireplaces '  'GarageYrB1t' , 'GarageCars ' , 'GarageArea ' , 'SalePrice' ] |

[50] :

|  |  |
| --- | --- |
| corr=num\_cols\_data [ selected \_cols ] . corr ( ) plt.figure(figsize=(35,20)) sns.heatmap(corr, annot=True, cmap=sns. color\_palette( "coolwarm" , | 12) ) |

[51] :

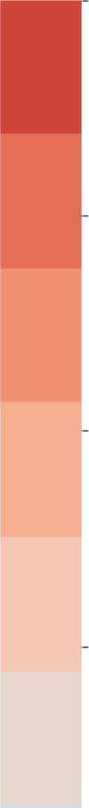
<AxesSubp10t

### Out[51] : : >

10

LotFrontage 0.43 025 012 0089 0.19 023 0.39 046 0.08 04 02 035 0.27 0067 0.29 034 0.35

LotArea 043 011 0014 0014 0.1 021 0.0026 0.26 03 0051 0.26 013 019 027 0.026 0.15 018 026

 OverallQual - 0.25 011 057 055 041 024 031 0.54 048 03 0.59 0.55 043 04 051 06 056 079-0.8

&arBuilt 0.12 0.014 059 032 0.25 0\_15 0.39 028 001 047 0.096 0.15 0.78 0.54 048 052

&arRemodAdd 0089 0014 0.55 059 018 013 018 029 024 014 029 O. 44 019 0.11 0.62 042 037

0.6

MasVnrArea - 0.19 041 0.32 018 0.26 0.11 0.36 034 0.17 0.39 0.28 028 025 025 0.36 037 0.48

BsmtfinSF1 - 023 021 024 025 013 026 0.52 045 014 0.21 0-059 0.044 026 0.15 0.22 03

BsmtUnfSF 013 00026 031 015 0.18 011 0.5 042 032 00045 0.24 0.29 025 0052 019 0.21 018 0.21

-0.4

btalBsmtSF 0.39 0.26 054 0.39 029 0.36 052 0.42 082 0.17 0.45 0.32 0.29 034 0.31 043 0.49 061

1stFlrSF 046 0.3 048 028 024 034 045 032 0.82 0.57 038 041 0.41 044 049 0.61

-02 hdF1rSF 008 0.051 0.3 001 0.14 017 014 00045 0.17 042 062 019 0068 0.18 014 0.32

 GLivArea 04 026 059 0.29 039 021 0-24 0.45 057 069 083 0.46 022 0.47 047 0.71

FullBath 02 013 055 047 044 028 0059 029 0.32 038 042 0.63 055 0.24 047 047 041 0.56

0.35 019 0.43 0096 0.19 028 0044 0.25 0.29 041 062 0.83 0.55 033 0.14 0.36 034 053

Fireplaces 0.27 027 0.4 015 011 026 0052 034 041 019 0.46 0.24 033 0043 03 027 0.47

GarageYrBlt 0067 0.026 051 078 062 025 015 019 0.31 022 0068 022 047 014 0043 047 047 047

GarageCars - 0.29 015 0.6 054 0.42 036 022 0.21 0.43 044 018 0.47 0.47 036 0.47 088

GarageArea 034 018 056 048 037 037 0.3 0.18 0.49 049 014 047 041 034 0.47 088 062

0.35 026 052 0.51 048 039 021 0.61 061 032 on 0.56 053 047 047 062

LotFrontage LotArea OverallQual &arBuilt &arRemodAdd MasVnrArea BsmtfinSF1 BsmtUnfSF %talBsmtSF 1stFlrSF 2ndFlrSF &LjvArea FullBath %tRmsAbvGrd Fireplaces GarageYrBlt GarageCars GarageArea

# Removed multicollinearity from above heatmap remaining variables

|  |  |
| --- | --- |
| new\_selected\_cols=[ ' OverallQua1 ' ,  'YearBui1t ' , 'YearRemodAdd' , 'MasVnrArea '  'TotalBsmtSF ' , '1stF1rSF ' ,  'GrLivArea ' FullBath ' , ' Fireplaces'  'GarageCars ' , ' SalePrice | ' BsmtFinSF1' , |

In [ 52] :

|  |
| --- |
| from scipy. stats import chi 2\_contingency |

[53] :

Checking Relation exists between Categorical variables using chi square analysis

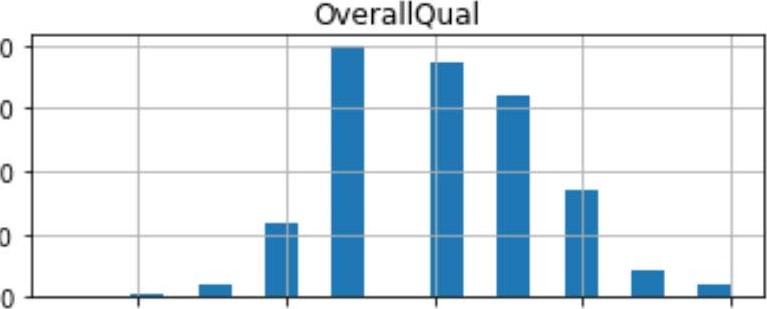
|  |
| --- |
| for coll in x train. columns:  for c012 in x train. columns: if coll!=c012:  contingency\_table - pd. crosstab(x\_train[coll], x\_train[c012] ) chi 2, p, = chi 2\_contingency(contingency\_table) if p«e.05 •.  print(f" {c} Chi-squared statistic for {coll, c012}: {chi 2} p-value is {p}" ) |

|  |
| --- |
| new\_num\_data=num\_cols\_data[new\_selected\_cols] |

In [56] :

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | PIt . figure 10)) new\_num\_data.hist(bins=20, 10)) PIt . show() | |

< Figure size  with e Axes >

400

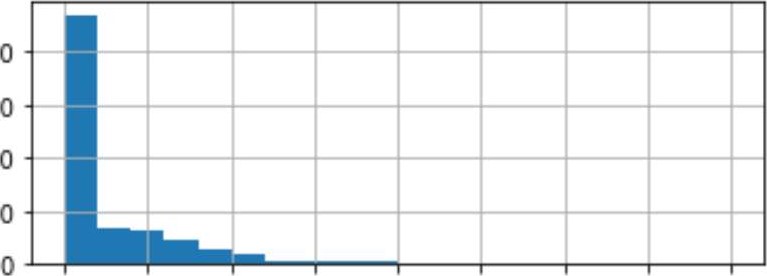
300

200

100

2 4 6 8 10

MasVnrArea

800

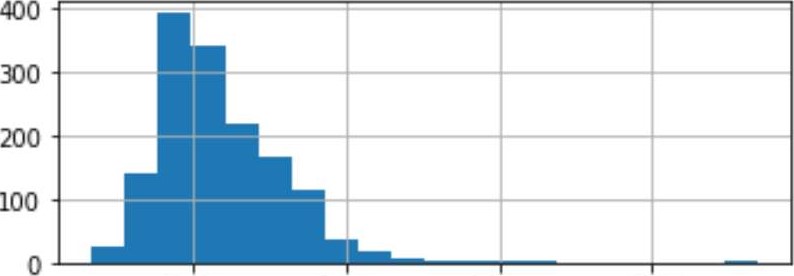
600

400

200

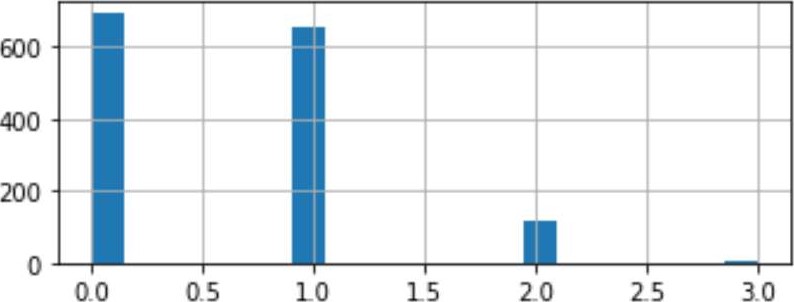
0 200 400 600 800 1000 1200 1400 1600

1stFlrSF

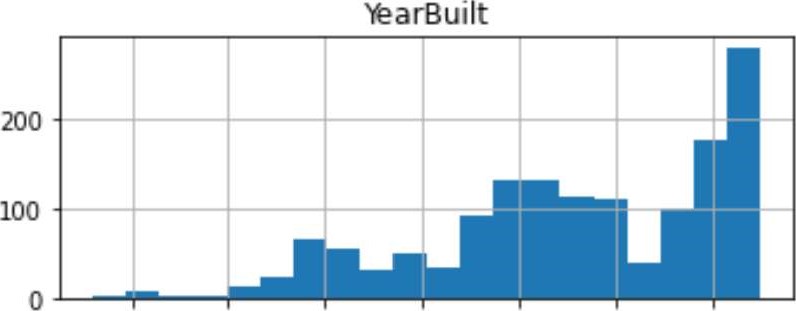


1000 2000 3000 4000

fireplaces

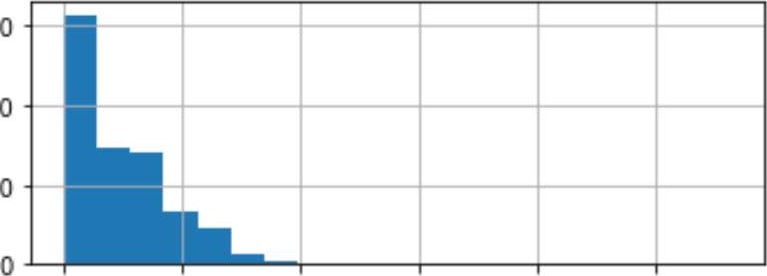






1880 1900 1920 1940 1960 1980 2000

BsmtFinSF1

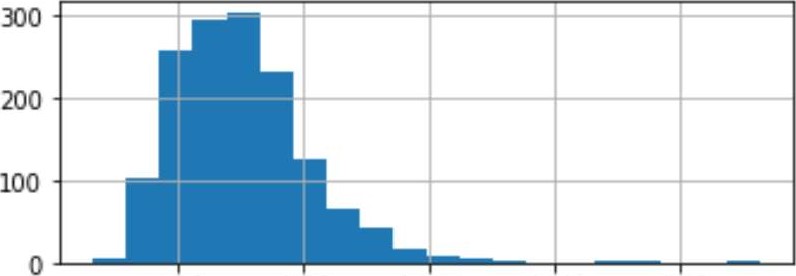
600

400

200

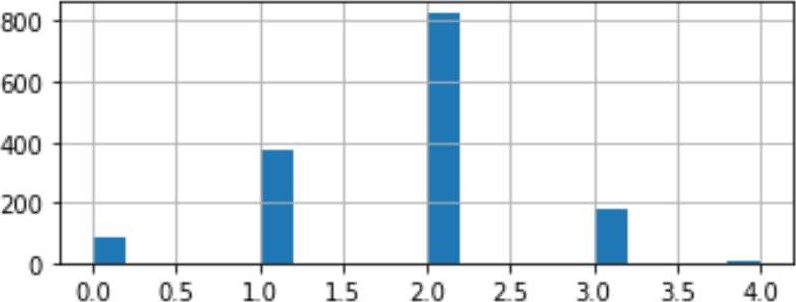
o 1000 2000 3000 4000 5000

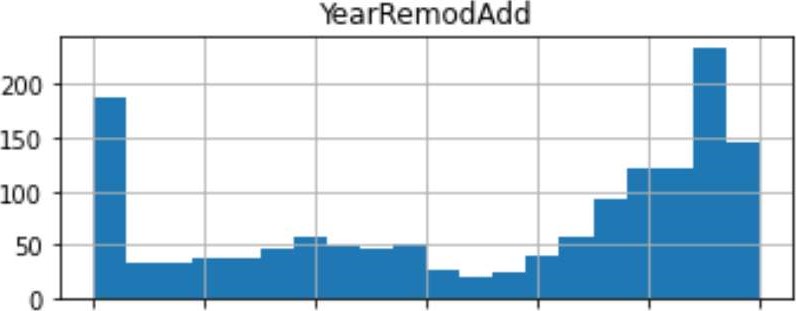
GrLivArea



1000 2000 3000 4000 5000

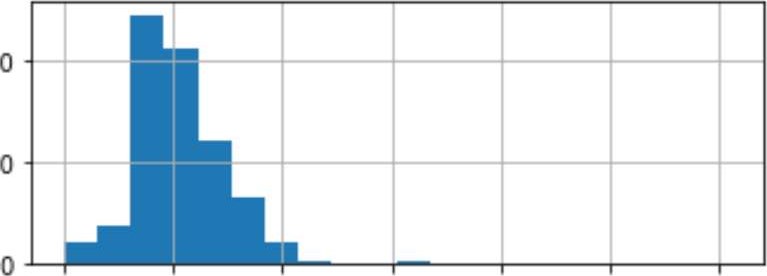
GarageCars





1950 1960 1970 1980 1990 2000 2010

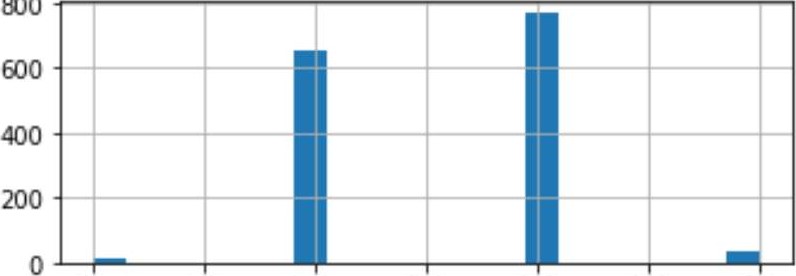
TotalBsmtSF

400

200

o 1000 2000 3000 4000 5000 6000

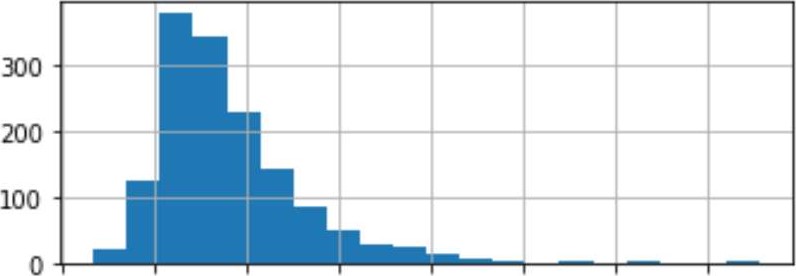
FullBath



800

00 0.5 1.0 15 20 25 3.0

SalePrice



0 100000 200000 300000 400000 500000 600000 700000

|  |  |
| --- | --- |
| new\_num\_data [ "MasVnrArea " ] . fillna ( new\_num\_data [ "MasVnrArea " ] . mode( ) [0] , | inp1ace=True) |

In [59] :

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  | | --- | --- | | X=new\_num\_data . drop( ' SalePrice ' Y=new\_num\_data[ ' SalePrice ] | axis-1) | |
|  | |  |  | | --- | --- | | from sklearn.model\_selection import train\_test\_split from sklearn. linear\_model import LinearRegression, Ridge, from sklearn.metrics import r2\_score, mean\_squared\_error, | Lasso mean absolute error | |
|  | |  | | --- | | x\_train, x\_test, y\_train, y\_test=train\_test\_sp1it(X,Y,test\_size=0.25, random\_state=43) Ir=LinearRegression() | |

Feature Engineering

\_

|  |
| --- |
| def  :  mod. fit (x\_train, y\_train) y\_pred=mod . predict (x\_test)  (y\_test, y\_pred) mse=mean\_squared\_error(y\_test, y\_pred) mae=mean\_absolute\_error(y\_test, y\_pred) print( 'Model' , mod) print( ' R-2 ' , rsquare) print( 'MSE' , mse) print( 'MAE' , mae) |

[67] :

In [68] :

model\_res(lr)

Model LinearRegression()

R-2 0.8074675718780645

MSE 1270444006.2814314 MAE 22945.611469390115

* + from sklearn.tree import DecisionTreeRegressor
  + dtr=DecisionTreeRegressor() lasso-Lasso() ridge-Ridge()
  + model\_res(dtr)

Model DecisionTreeRegressor() R-2 0.6919156662625392

MSE 2032924526.2417808 MAE 29605.606849315067

* + model\_res(lasso)

Model Lasso()

R-2 0.8074685895545486 MSE 1270437291.0438886

MAE 22945.20766545359

Feature\_Engineering

In [73] : model\_res ( ridge)

Model Ridge()

R-2 0.8074953307860377

MSE 1270260836.4196086 MAE 22938.73651512525

Our model will predict with 80% variability

