

```
In [1]: pip install openpyxl

Requirement already satisfied: openpyxl in /srv/conda/envs/notebook/lib/python3.7/site-packages (3.0.7)
Requirement already satisfied: et-xmlfile in /srv/conda/envs/notebook/lib/python3.7/site-packages (from openpyxl) (1.1.0)
Note: you may need to restart the kernel to use updated packages.

In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

In [3]: data=pd.read_excel("Ames_housing.xlsx")

In [4]: data

Out[4]:
```

	Order	PID	MS SubClass	MS Zoning	Lot Frontage	Lot Area	Street	Alley	Lot Shape	Land Contour	...	Pool Area	Pool QC	Fence	Misc Feature	Misc Val	Mo Sold	Yr Sold	Sale Type	Sale Condition	SalePrice
0	1	526301100	20	RL	141.0	31770	Pave	NaN	IR1	Lvl	...	0	NaN	NaN	NaN	0	5	2010	WD	Normal	215000
1	2	526350040	20	RH	80.0	11622	Pave	NaN	Reg	Lvl	...	0	NaN	MnPrv	NaN	0	6	2010	WD	Normal	105000
2	3	526351010	20	RL	81.0	14267	Pave	NaN	IR1	Lvl	...	0	NaN	NaN	Gar2	12500	6	2010	WD	Normal	172000
3	4	526353030	20	RL	93.0	11160	Pave	NaN	Reg	Lvl	...	0	NaN	NaN	NaN	0	4	2010	WD	Normal	244000
4	5	527105010	60	RL	74.0	13830	Pave	NaN	IR1	Lvl	...	0	NaN	MnPrv	NaN	0	3	2010	WD	Normal	189900
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2925	2926	923275080	80	RL	37.0	7937	Pave	NaN	IR1	Lvl	...	0	NaN	GdPrv	NaN	0	3	2006	WD	Normal	142500
2926	2927	923276100	20	RL	NaN	8885	Pave	NaN	IR1	Low	...	0	NaN	MnPrv	NaN	0	6	2006	WD	Normal	131000
2927	2928	923400125	85	RL	62.0	10441	Pave	NaN	Reg	Lvl	...	0	NaN	MnPrv	Shed	700	7	2006	WD	Normal	132000
2928	2929	924100070	20	RL	77.0	10010	Pave	NaN	Reg	Lvl	...	0	NaN	NaN	NaN	0	4	2006	WD	Normal	170000
2929	2930	924151050	60	RL	74.0	9627	Pave	NaN	Reg	Lvl	...	0	NaN	NaN	NaN	0	11	2006	WD	Normal	188000

2930 rows x 82 columns

```
In [5]: df=pd.DataFrame(data)
```

```
In [6]: print(df)
```

```
In [7]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2930 entries, 0 to 2929
Data columns (total 82 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   Order               2930 non-null   int64
1   PID                 2930 non-null   int64
2   MS SubClass         2930 non-null   int64
3   MS Zoning           2930 non-null   object
4   Lot Frontage        2440 non-null   float64
5   Lot Area            2930 non-null   int64
6   Street              2930 non-null   object
7   Alley              198 non-null    object
8   Lot Shape           2930 non-null   object
9   Land Contour        2930 non-null   object
10  Utilities           2930 non-null   object
11  Lot Config          2930 non-null   object
12  Land Slope          2930 non-null   object
13  Neighborhood        2930 non-null   object
14  Condition 1         2930 non-null   object
15  Condition 2         2930 non-null   object
16  Bldg Type           2930 non-null   object
17  House Style         2930 non-null   object
18  Overall Qual         2930 non-null   int64
19  Overall Cond        2930 non-null   int64
20  Year Built          2930 non-null   int64
21  Year Remod/Add      2930 non-null   int64
22  Roof Style          2930 non-null   object
23  Roof Matl           2930 non-null   object
24  Exterior 1st        2930 non-null   object
25  Exterior 2nd        2930 non-null   object
26  Mas Vnr Type        2930 non-null   object
27  Mas Vnr Area        2907 non-null   float64
28  Exter Qual          2930 non-null   object
29  Exter Cond          2930 non-null   object
30  Foundation          2930 non-null   object
31  Bsmt Qual           2850 non-null   object
32  Bsmt Cond           2850 non-null   object
33  Bsmt Exposure       2847 non-null   object
34  BsmtFin Type 1      2850 non-null   object
35  BsmtFin SF 1        2929 non-null   float64
36  BsmtFin Type 2      2849 non-null   object
37  BsmtFin SF 2        2929 non-null   float64
38  Bsmt Unf SF         2929 non-null   float64
39  Total Bsmt SF       2929 non-null   float64
40  Heating             2930 non-null   object
41  Heating QC          2930 non-null   object
42  Central Air         2930 non-null   object
43  Electrical          2929 non-null   object
44  1st Flr SF          2930 non-null   int64
45  2nd Flr SF          2930 non-null   int64
46  Low Qual Fin SF     2930 non-null   int64
47  Gr Liv Area         2930 non-null   int64
48  Bsmt Full Bath      2928 non-null   float64
49  Bsmt Half Bath      2928 non-null   float64
50  Full Bath           2930 non-null   int64
51  Half Bath           2930 non-null   int64
52  Bedroom AbvGr       2930 non-null   int64
53  Kitchen AbvGr       2930 non-null   int64
54  Kitchen Qual        2930 non-null   object
55  TotRms AbvGrd       2930 non-null   int64
56  Functional          2930 non-null   object
57  Fireplaces          2930 non-null   int64
58  Fireplace Qu        1508 non-null   object
59  Garage Type         2773 non-null   object
60  Garage Yr Blt       2771 non-null   float64
61  Garage Finish       2771 non-null   object
62  Garage Cars         2929 non-null   float64
63  Garage Area         2929 non-null   float64
64  Garage Qual         2771 non-null   object
65  Garage Cond         2771 non-null   object
66  Paved Drive         2930 non-null   object
67  Wood Deck SF        2930 non-null   int64
68  Open Porch SF       2930 non-null   int64
69  Enclosed Porch      2930 non-null   int64
70  3Ssn Porch          2930 non-null   int64
71  Screen Porch        2930 non-null   int64
72  Pool Area           2930 non-null   int64
73  Pool QC             13 non-null     object
74  Fence               572 non-null     object
75  Misc Feature        106 non-null     object
76  Misc Val            2930 non-null   int64
77  Mo Sold             2930 non-null   int64
78  Yr Sold             2930 non-null   int64
79  Sale Type           2930 non-null   object
80  Sale Condition      2930 non-null   object
81  SalePrice           2930 non-null   int64
dtypes: float64(11), int64(28), object(43)
memory usage: 1.8+ MB
```

```
In [8]: df.describe()
```

```
Out[8]:
```

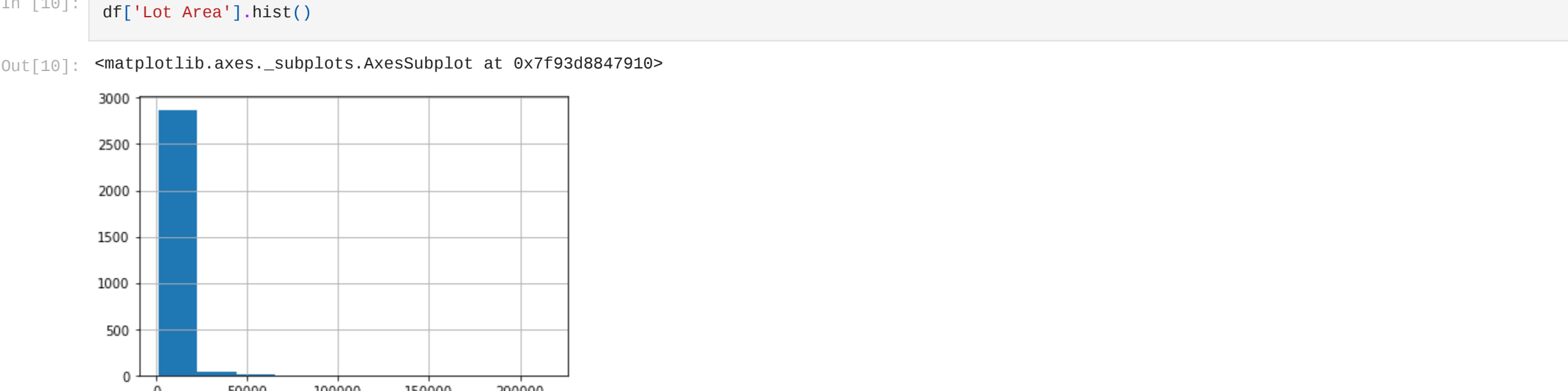
	Order	PID	MS SubClass	Lot Frontage	Lot Area	Overall Qual	Overall Cond	Year Built	Year Remod/Add	Mas Vnr Area	...	Wood Deck SF	Open Porch SF	Enclosed Porch
count	2930.00000	2.9300000e+03	2930.000000	2440.000000	2930.000000	2930.000000	2930.000000	2930.000000	2930.000000	2907.000000	...	2930.000000	2930.000000	2930.000000
mean	1465.50000	7.144645e+08	57.387372	69.224590	10147.921843	6.094881	5.563140	1971.356314	1984.266553	101.896801	...	93.751877	47.533447	23.011604
std	845.96247	1.887308e+08	42.638025	23.365335	7880.017759	1.411026	1.111537	30.245361	20.860286	179.112611	...	126.361562	67.483400	64.139059
min	1.00000	5.263011e+08	20.000000	21.000000	1300.000000	1.000000	1.000000	1872.000000	1950.000000	0.000000	...	0.000000	0.000000	0.000000
25%	733.25000	5.284770e+08	20.000000	58.000000	7440.250000	5.000000	5.000000	1954.000000	1965.000000	0.000000	...	0.000000	0.000000	0.000000
50%	1465.50000	5.354536e+08	50.000000	68.000000	9436.500000	6.000000	5.000000	1973.000000	1993.000000	0.000000	...	0.000000	27.000000	0.000000
75%	2197.75000	9.071811e+08	70.000000	80.000000	11555.250000	7.000000	6.000000	2001.000000	2004.000000	164.000000	...	168.000000	70.000000	0.000000
max	2930.00000	1.007100e+09	190.000000	313.000000	215245.000000	10.000000	9.000000	2010.000000	2010.000000	1600.000000	...	1424.000000	742.000000	1012.000000

8 rows x 39 columns

```
In [9]: plt.plot(df.PID,df.SalePrice,ls='',marker='o',color='red')
```



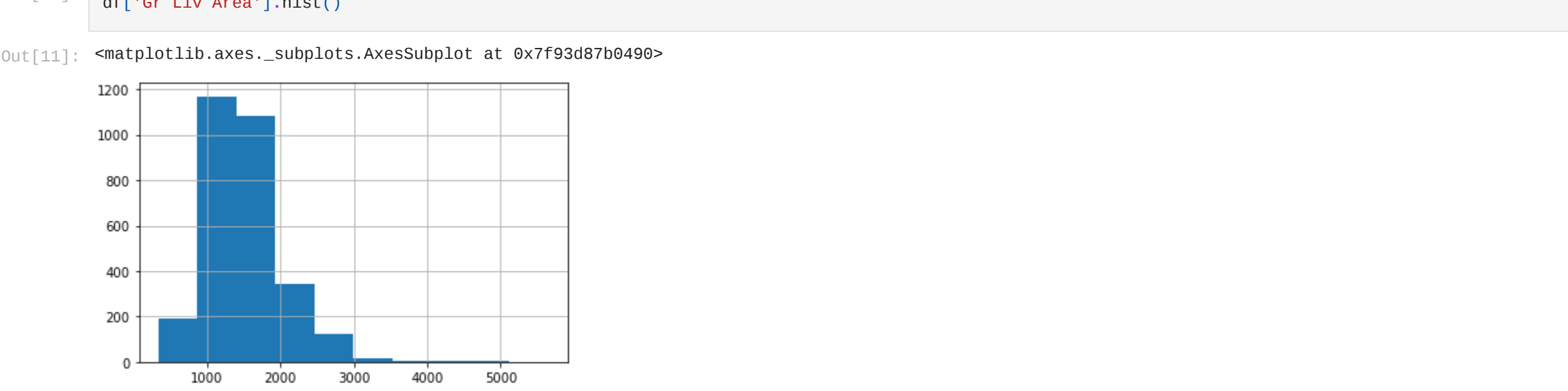
```
In [10]: df['Lot Area'].hist()
```



```
In [11]: df['SalePrice'].hist()
```



```
In [11]: df['Gr Liv Area'].hist()
```



```
In [ ]:
```