

Assignment 7

Author: Rohit Pendse

Roll No: 33358

Problem Statement

Data Visualization on Air quality and Heart

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

%matplotlib inline
```

```
In [2]: df = pd.read_csv('heart.csv')
```

```
In [3]: df
```

Out[3]:

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope
0	1	63	1	typical	145	233	1	2	150	0	2.3	3
1	2	67	1	asymptomatic	160	286	0	2	108	1	1.5	2
2	3	67	1	asymptomatic	120	229	0	2	129	1	2.6	2
3	4	37	1	nonanginal	130	250	0	0	187	0	3.5	3
4	5	41	0	nontypical	130	204	0	2	172	0	1.4	1
...
298	299	45	1	typical	110	264	0	0	132	0	1.2	2
299	300	68	1	asymptomatic	144	193	1	0	141	0	3.4	2
300	301	57	1	asymptomatic	130	131	0	0	115	1	1.2	2
301	302	57	0	nontypical	130	236	0	2	174	0	0.0	2
302	303	38	1	nonanginal	138	175	0	0	173	0	0.0	1

303 rows × 15 columns



```
In [4]: df.isnull().sum()
```

```
Out[4]: Unnamed: 0      0
Age          0
Sex          0
ChestPain    0
RestBP       0
Chol         0
Fbs         0
RestECG      0
MaxHR        0
ExAng        0
Oldpeak      0
Slope        0
Ca           4
Thal         2
AHD          0
dtype: int64
```

```
In [5]: df1 = pd.read_csv('forestfires.csv')
```

```
In [6]: df1
```

```
Out[6]:
```

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.00
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.00
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.00
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.00
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.00
...
512	4	3	aug	sun	81.6	56.7	665.6	1.9	27.8	32	2.7	0.0	6.44
513	2	4	aug	sun	81.6	56.7	665.6	1.9	21.9	71	5.8	0.0	54.29
514	7	4	aug	sun	81.6	56.7	665.6	1.9	21.2	70	6.7	0.0	11.16
515	1	4	aug	sat	94.4	146.0	614.7	11.3	25.6	42	4.0	0.0	0.00
516	6	3	nov	tue	79.5	3.0	106.7	1.1	11.8	31	4.5	0.0	0.00

517 rows × 13 columns

```
In [7]: df.columns
```

```
Out[7]: Index(['Unnamed: 0', 'Age', 'Sex', 'ChestPain', 'RestBP', 'Chol', 'Fbs',
              'RestECG', 'MaxHR', 'ExAng', 'Oldpeak', 'Slope', 'Ca', 'Thal', 'AHD'],
              dtype='object')
```

```
In [8]: df1.columns
```

```
Out[8]: Index(['X', 'Y', 'month', 'day', 'FFMC', 'DMC', 'DC', 'ISI', 'temp', 'RH',
              'wind', 'rain', 'area'],
              dtype='object')
```

```
In [9]: df1.isnull().sum()
```

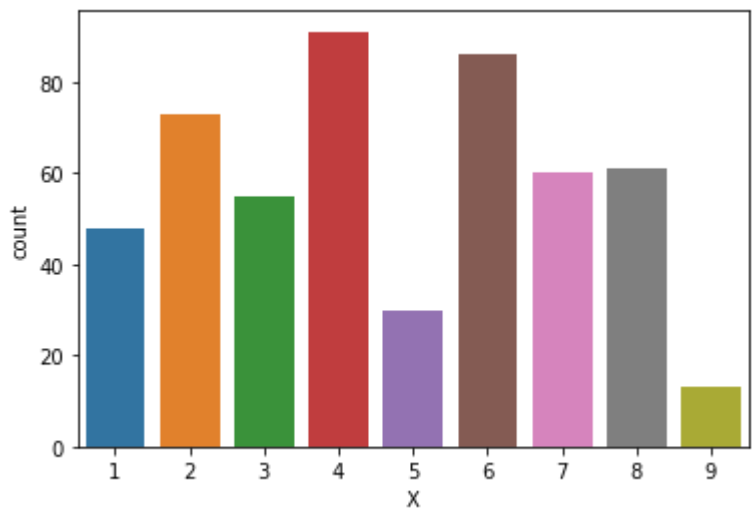
```
Out[9]: X      0
        Y      0
        month  0
        day    0
        FFMC   0
        DMC    0
        DC     0
        ISI    0
        temp   0
        RH     0
        wind   0
        rain   0
        area   0
        dtype: int64
```

```
In [10]: df1['X'].value_counts()
```

```
Out[10]: 4      91
         6      86
         2      73
         8      61
         7      60
         3      55
         1      48
         5      30
         9      13
         Name: X, dtype: int64
```

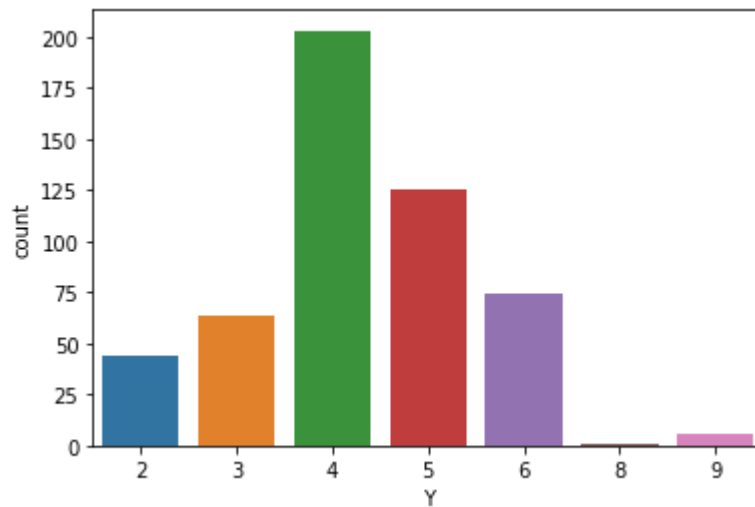
```
In [11]: sns.countplot(df1['X'])
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x16d7554d808>
```



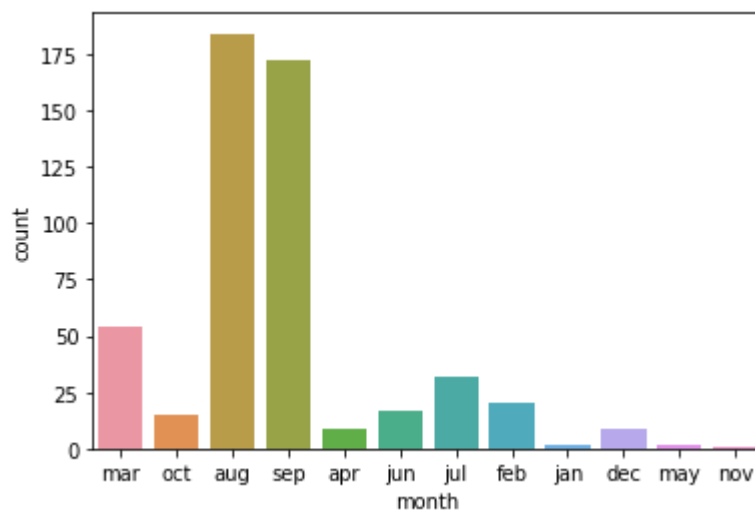
```
In [12]: sns.countplot(df1['Y'])
```

```
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x16d76d13f88>
```



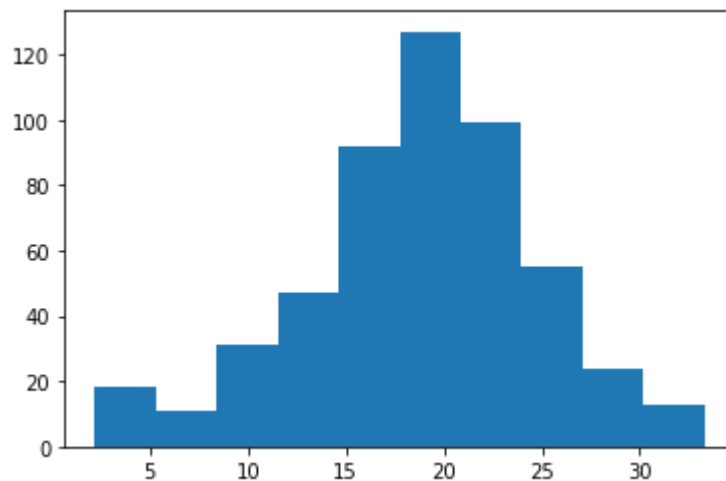
```
In [13]: sns.countplot(df1['month'])
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x16d76da78c8>
```



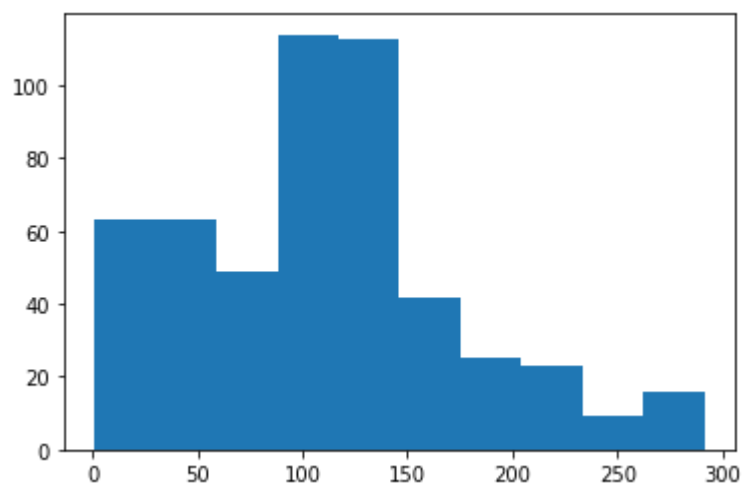
```
In [14]: plt.hist(df1['temp'])
```

```
Out[14]: (array([ 18., 11., 31., 47., 92., 127., 99., 55., 24., 13.]),  
array([ 2.2 ,  5.31,  8.42, 11.53, 14.64, 17.75, 20.86, 23.97, 27.08,  
        30.19, 33.3 ]),  
<a list of 10 Patch objects>)
```



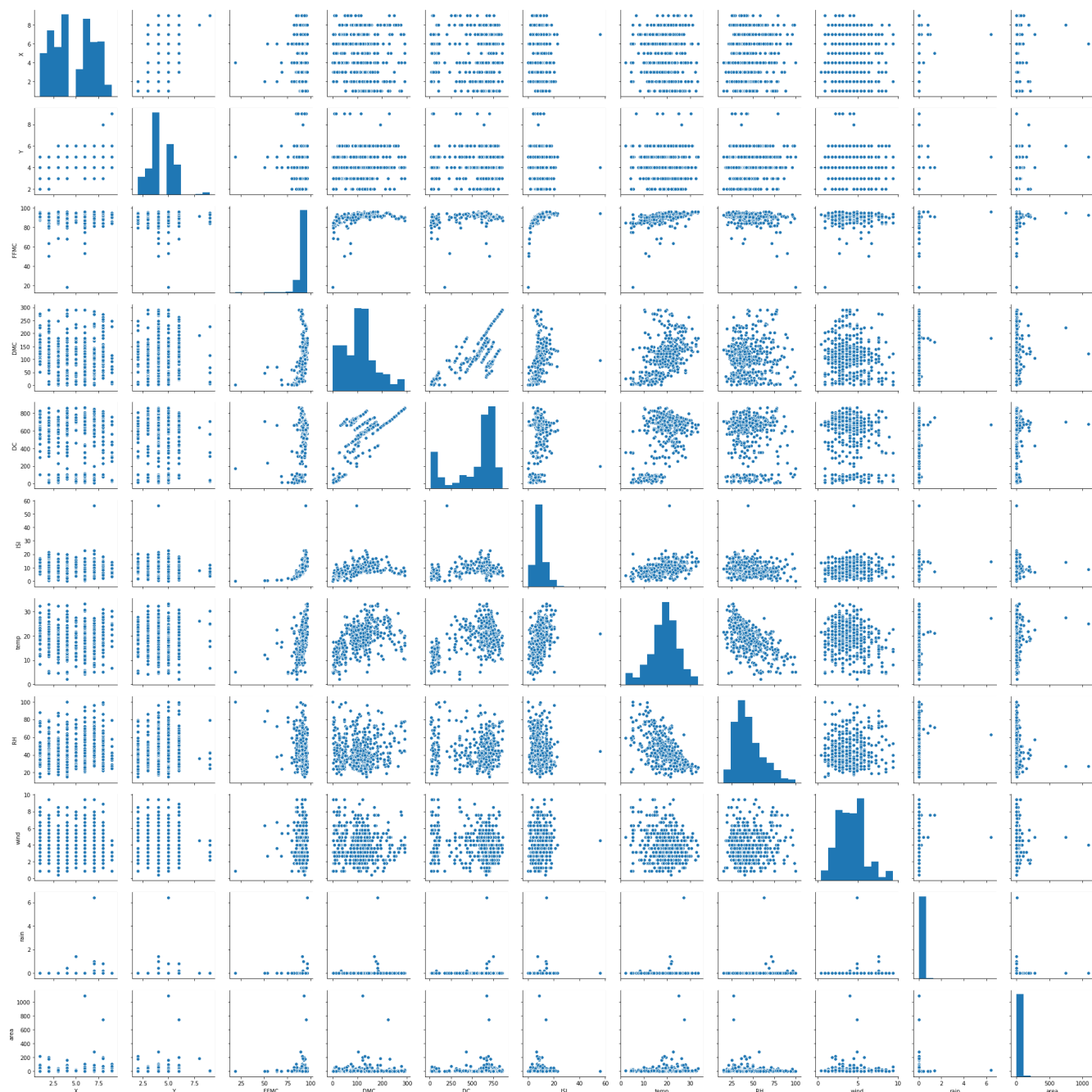
```
In [15]: plt.hist(df1['DMC'])
```

```
Out[15]: (array([ 63.,  63.,  49., 114., 113.,  42.,  25.,  23.,   9.,  16.]),  
array([  1.1 , 30.12, 59.14, 88.16, 117.18, 146.2 , 175.22, 204.24,  
        233.26, 262.28, 291.3 ]),  
<a list of 10 Patch objects>)
```



```
In [16]: sns.pairplot(data=df1)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x16d76f5f108>
```



```
In [17]: sns.pairplot(data=df)
```

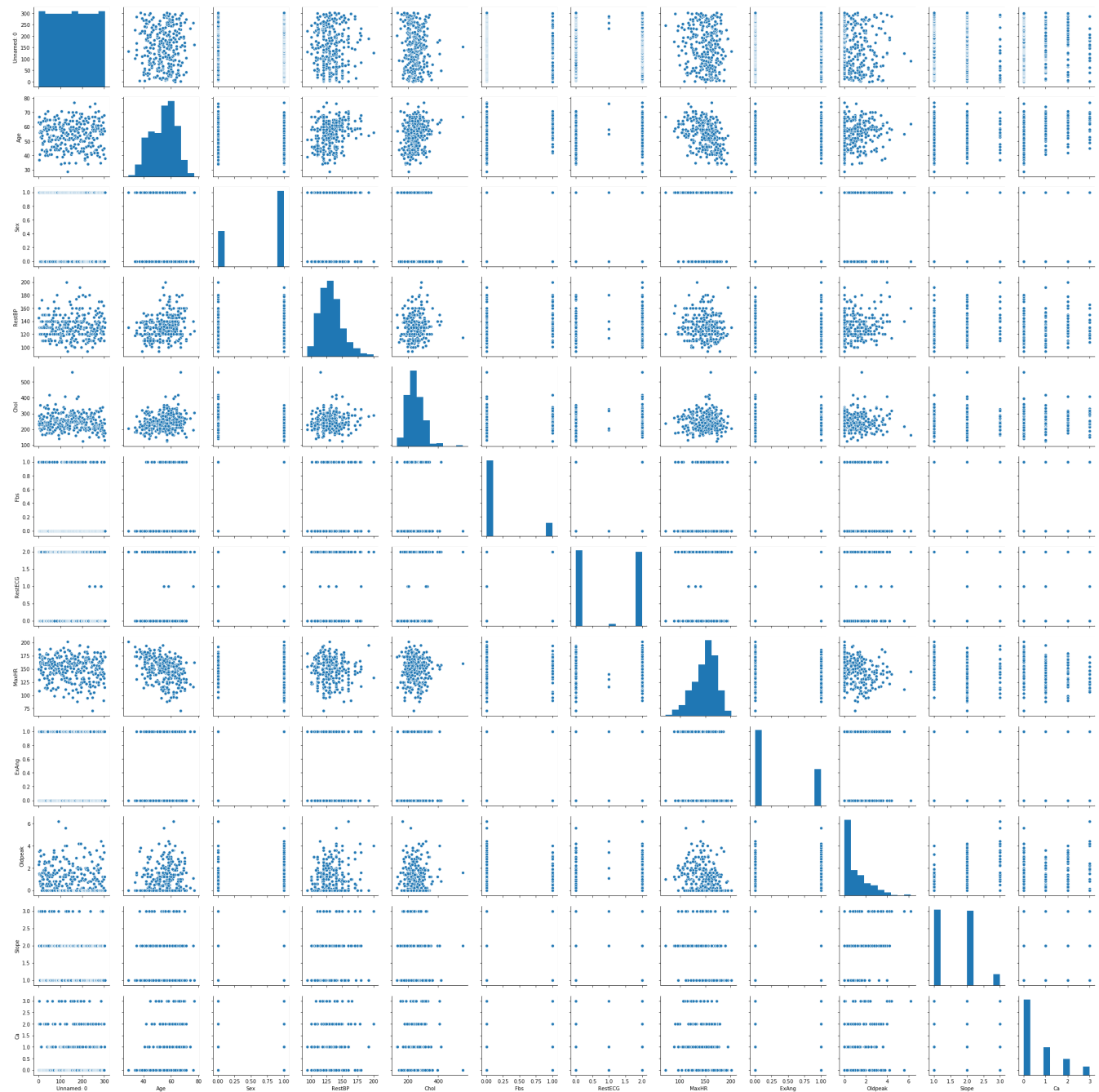
```
C:\Users\omkar\anacondanew\lib\site-packages\numpy\lib\histograms.py:824: RuntimeWarni
ng: invalid value encountered in greater_equal
```

```
keep = (tmp_a >= first_edge)
```

```
C:\Users\omkar\anacondanew\lib\site-packages\numpy\lib\histograms.py:825: RuntimeWarni
ng: invalid value encountered in less_equal
```

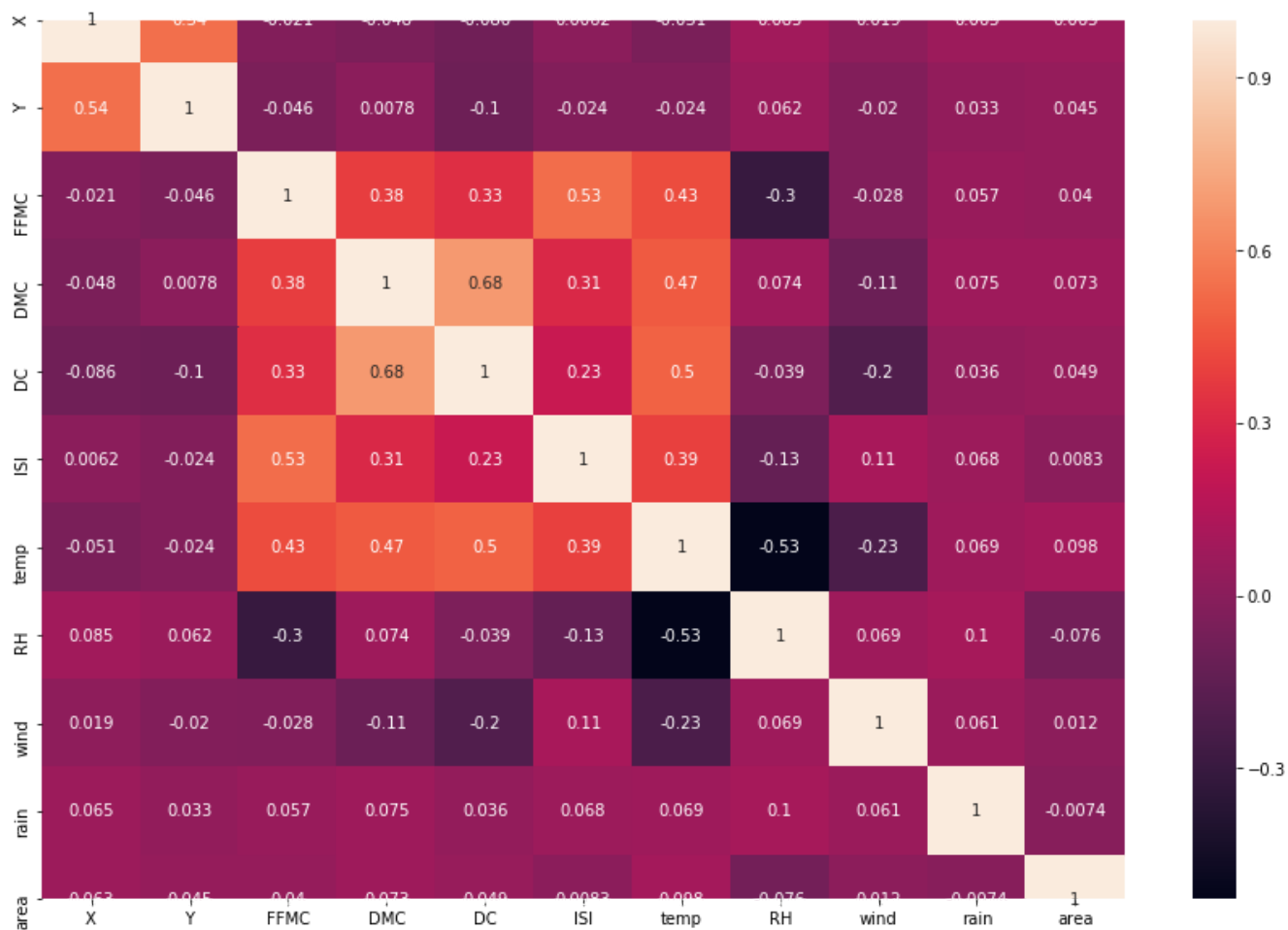
```
keep &= (tmp_a <= last_edge)
```

```
Out[17]: <seaborn.axisgrid.PairGrid at 0x16d7b927a88>
```



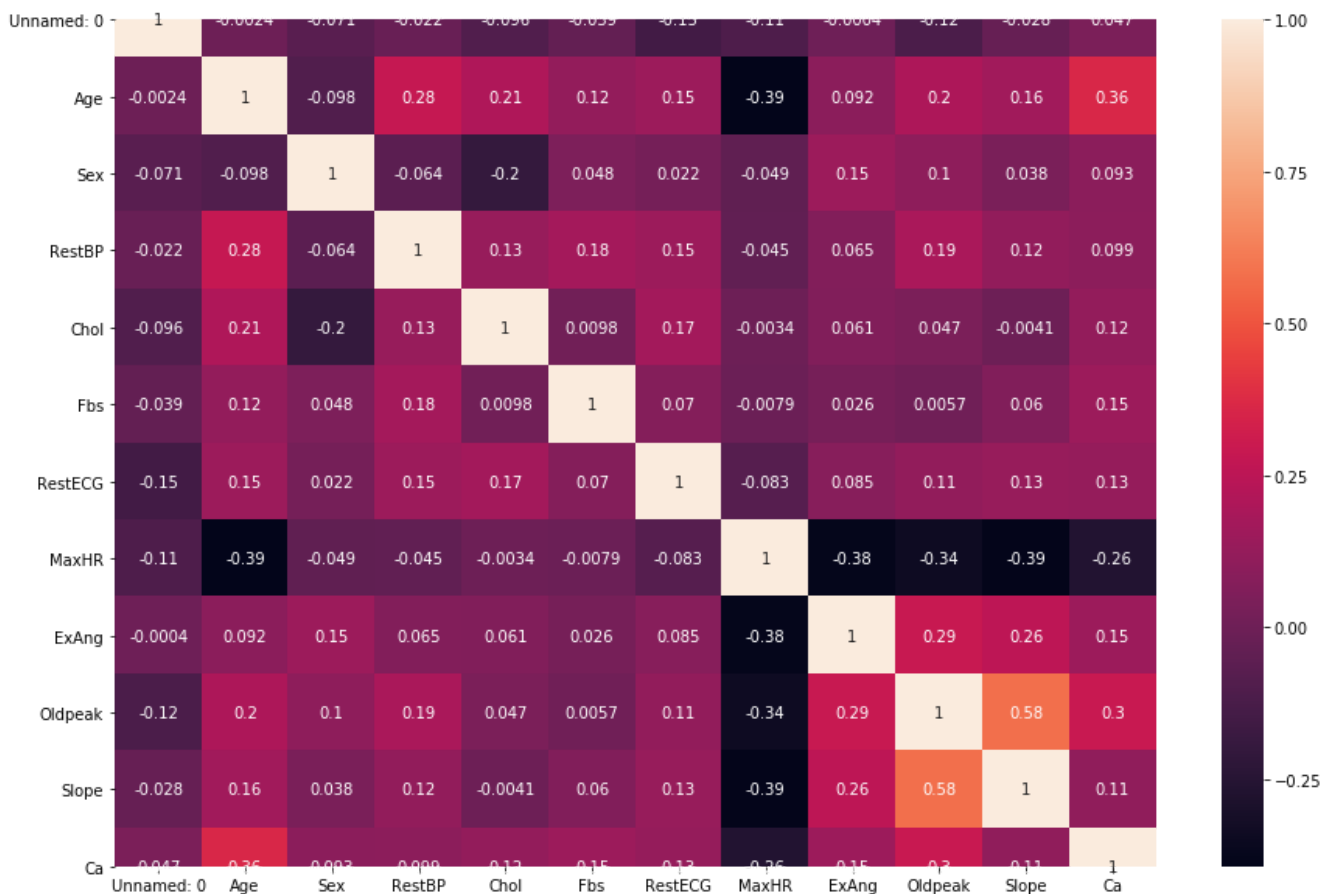
```
In [18]: plt.figure(figsize=(15,10))
sns.heatmap(df1.corr(), annot=True)
```

```
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x16d02dffc88>
```




```
In [19]: plt.figure(figsize=(15,10))
sns.heatmap(df.corr(), annot=True)
```

```
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x16d05879d08>
```



```
In [20]: df.columns
```

```
Out[20]: Index(['Unnamed: 0', 'Age', 'Sex', 'ChestPain', 'RestBP', 'Chol', 'Fbs',
                'RestECG', 'MaxHR', 'ExAng', 'Oldpeak', 'Slope', 'Ca', 'Thal', 'AHD'],
              dtype='object')
```

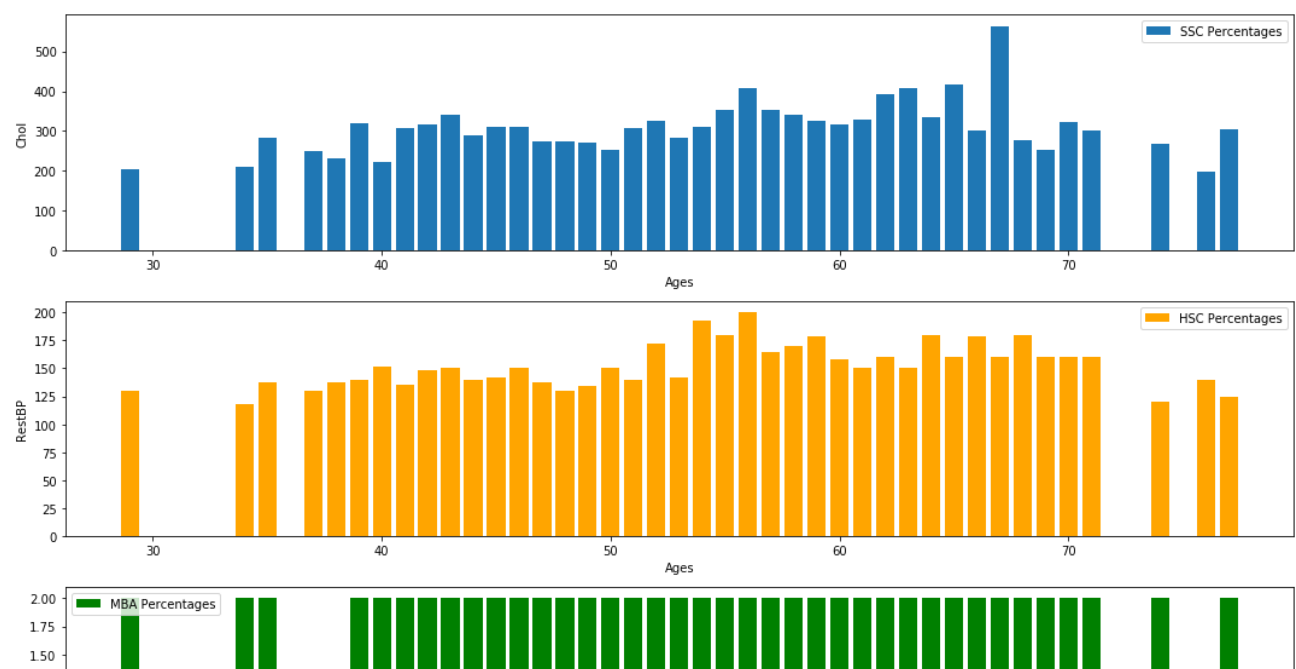
```
In [21]: ages = df['Age']
Chol = df['Chol']
RestBP = df['RestBP']
RestECG = df['RestECG']
fig = plt.figure(figsize=(15,10))

plt.subplot(3,1,1)
plt.bar(ages, Chol, label='SSC Percentages')
plt.legend()
plt.xlabel('Ages')
plt.ylabel('Chol')

plt.subplot(3,1,2)
plt.bar(ages, RestBP, color='orange', label='HSC Percentages')
plt.legend()
plt.xlabel('Ages')
plt.ylabel('RestBP')

plt.subplot(3,1,3)
plt.bar(ages, RestECG,color='g', label='MBA Percentages')
plt.legend()
plt.xlabel('Ages')
plt.ylabel('RestECG')

plt.tight_layout(w_pad=20)
plt.show()
```



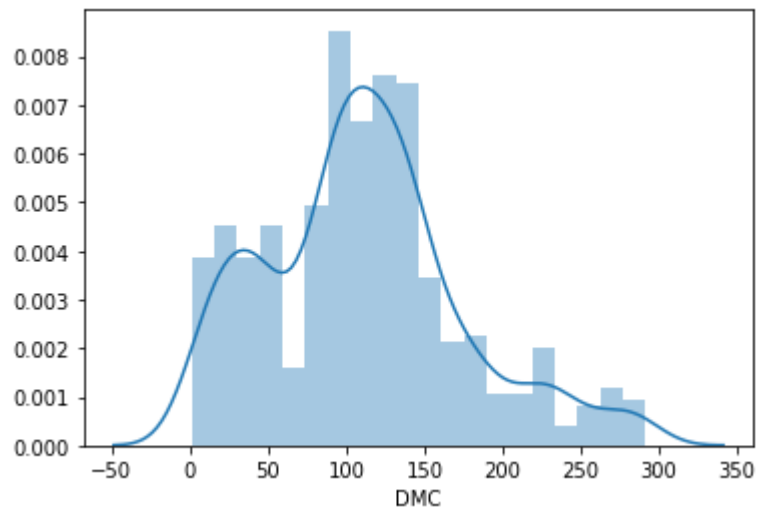
```
In [22]: df1.head()
```

Out[22]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0

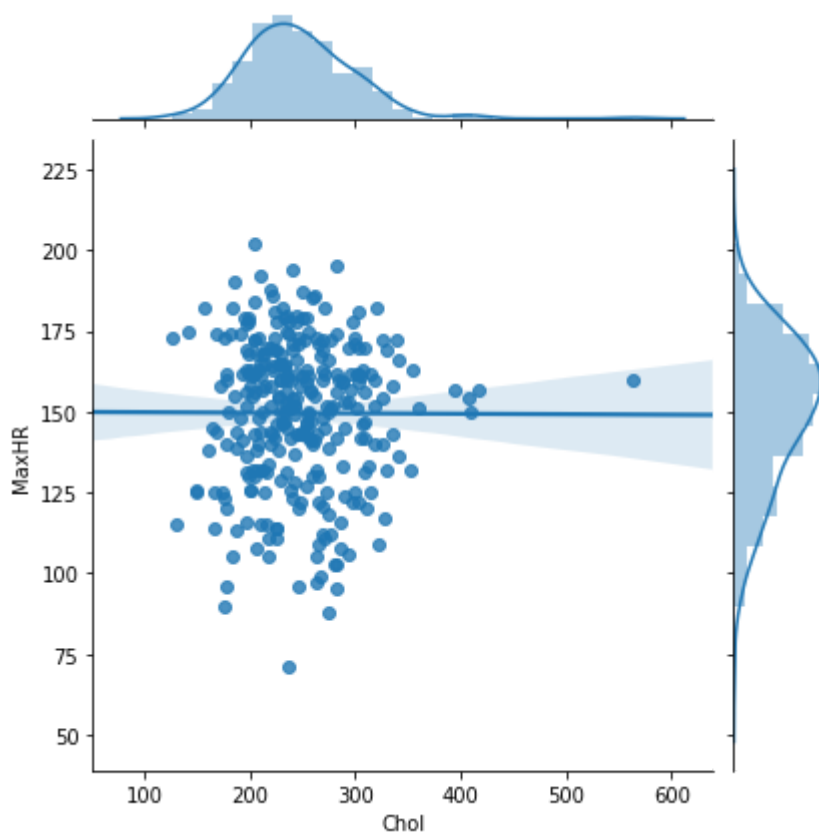
```
In [23]: sns.distplot(df1['DMC'], bins=20)
```

```
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x16d0561e408>
```



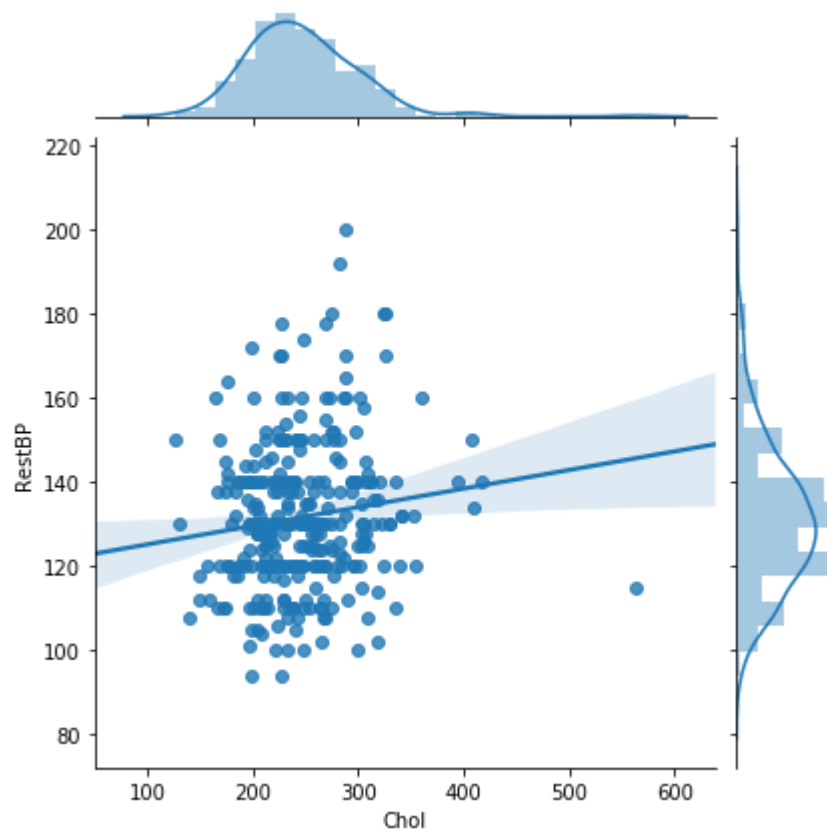
```
In [24]: sns.jointplot(x='Chol', y='MaxHR', data = df, kind='reg')
```

```
Out[24]: <seaborn.axisgrid.JointGrid at 0x16d069b3dc8>
```



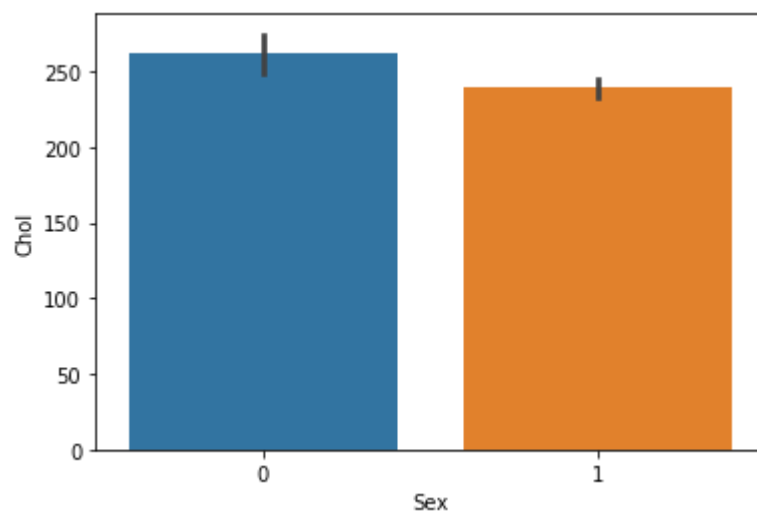
```
In [25]: sns.jointplot(x='Chol', y='RestBP', data = df, kind='reg')
```

```
Out[25]: <seaborn.axisgrid.JointGrid at 0x16d06691188>
```



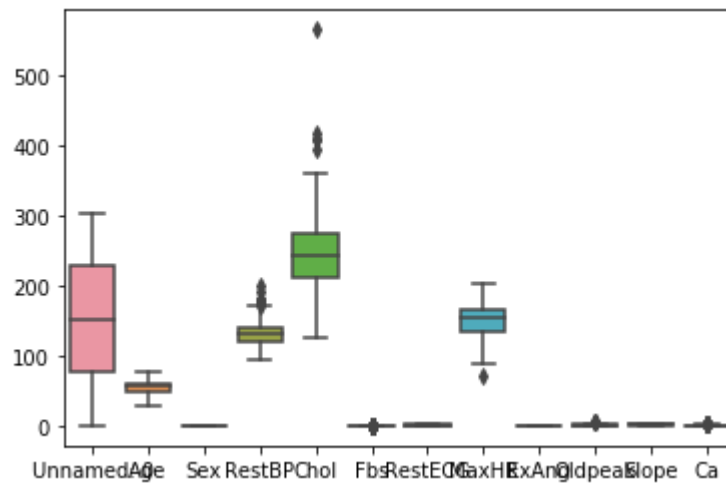
```
In [26]: sns.barplot(x='Sex', y='Chol', data=df, estimator=np.mean)
```

```
Out[26]: <matplotlib.axes._subplots.AxesSubplot at 0x16d06800d08>
```



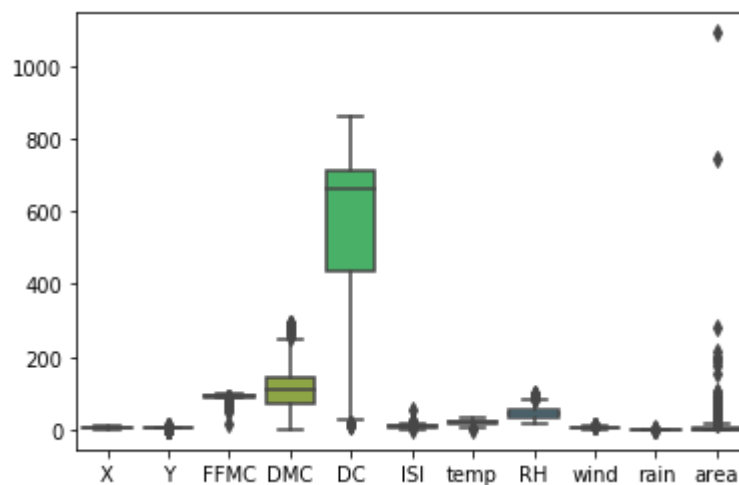
```
In [27]: sns.boxplot(data=df, orient='v')
```

```
Out[27]: <matplotlib.axes._subplots.AxesSubplot at 0x16d067d2e88>
```



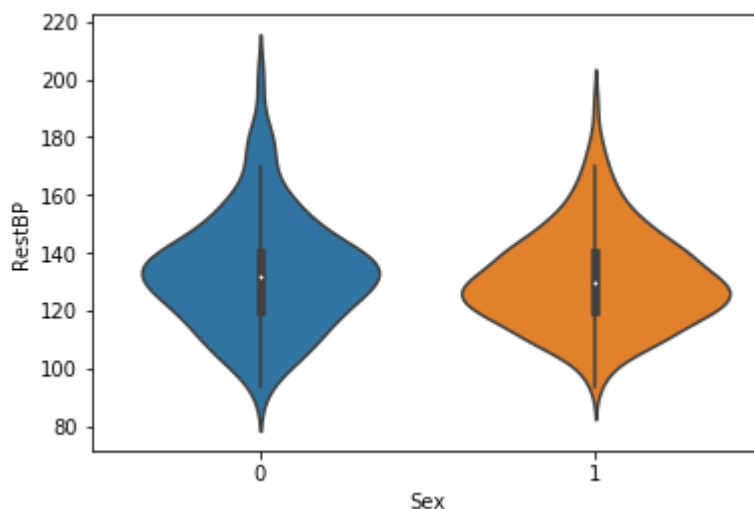
```
In [28]: sns.boxplot(data=df1, orient='v')
```

```
Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x16d06e3f508>
```



```
In [30]: sns.violinplot(x='Sex',y='RestBP', data=df)
```

```
Out[30]: <matplotlib.axes._subplots.AxesSubplot at 0x16d070f0208>
```



In []: