In [110…
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import math
```

In [111…
```
df = pd.read_csv("G://Heart2.csv")
df.head(5)
```

Out[111…

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	
0	63	1	typical	145	233	1	2	150	0	2.3	3	0	
1	67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3	
2	67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2	reve
3	37	1	nonanginal	130	250	0	0	187	0	3.5	3	0	
4	41	0	nontypical	130	204	0	2	172	0	1.4	1	0	

In [112…
```
df.head(14)
```

Out[112…

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	
0	63	1	typical	145	233	1	2	150	0	2.3	3	0	
1	67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3	
2	67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2	re
3	37	1	nonanginal	130	250	0	0	187	0	3.5	3	0	
4	41	0	nontypical	130	204	0	2	172	0	1.4	1	0	
5	56	1	nontypical	120	236	0	0	178	0	0.8	1	0	
6	62	0	asymptomatic	140	268	0	2	160	0	3.6	3	2	
7	57	0	asymptomatic	120	354	0	0	163	1	0.6	1	0	
8	63	1	asymptomatic	130	254	0	2	147	0	1.4	2	1	re
9	53	1	asymptomatic	140	203	1	2	155	1	3.1	3	0	re
10	57	1	asymptomatic	140	192	0	0	148	0	0.4	2	0	
11	56	0	nontypical	140	294	0	2	153	0	1.3	2	0	
12	56	1	nonanginal	130	256	1	2	142	1	0.6	2	1	
13	44	1	nontypical	120	263	0	0	173	0	0.0	1	0	re

In [113…
```
df['Thal'] = pd.factorize(df['Thal'])[0].astype(np.int)
df.head()
```

Out[113…

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Tha
0	63	1	typical	145	233	1	2	150	0	2.3	3	0	
1	67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3	
2	67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2	

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Tha
3	37	1	nonanginal	130	250	0	0	187	0	3.5	3	0	
4	41	0	nontypical	130	204	0	2	172	0	1.4	1	0	

In [114…
```
df['ChestPain'] = pd.factorize(df['ChestPain'])[0].astype(np.int)
df['AHD'] = pd.factorize(df['AHD'])[0].astype(np.int)
```

In [115…
```
df.head()
```

Out[115…

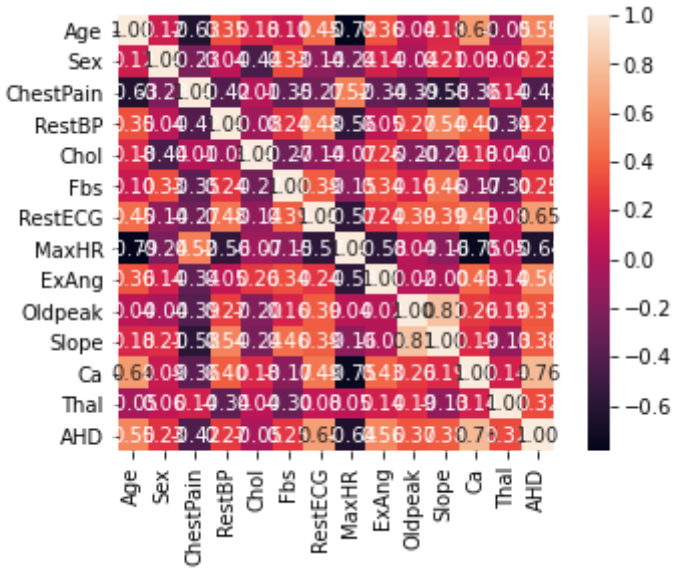
	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal
0	63	1	0	145	233	1	2	150	0	2.3	3	0	0
1	67	1	1	160	286	0	2	108	1	1.5	2	3	1
2	67	1	1	120	229	0	2	129	1	2.6	2	2	2
3	37	1	2	130	250	0	0	187	0	3.5	3	0	1
4	41	0	3	130	204	0	2	172	0	1.4	1	0	1

In [116…
```
cols = ['Age','Sex','ChestPain','RestBP','Chol','Fbs','RestECG','MaxHR','ExAr
```

In [117…
```
import numpy as np

cm = np.corrcoef(df[cols].values.T)
hm = sns.heatmap(cm,
                 cbar=True,
                 annot=True,
                 square=True,
                 fmt='.2f',
                 annot_kws={'size': 10},
                 yticklabels=cols,
                 xticklabels=cols)

plt.show()
```



In [118…
```
df.drop(['ChestPain','MaxHR'],axis=1,inplace=True)
```

In [119…
```
df.head()
```

Out[119…

	Age	Sex	RestBP	Chol	Fbs	RestECG	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	63	1	145	233	1	2	0	2.3	3	0	0	0
1	67	1	160	286	0	2	1	1.5	2	3	1	1
2	67	1	120	229	0	2	1	2.6	2	2	2	1
3	37	1	130	250	0	0	0	3.5	3	0	1	0
4	41	0	130	204	0	2	0	1.4	1	0	1	0

In [152…
```
def inf(p,n):
    diff = -p/14
    diff1 = -n/14
    ans = diff*(math.log(-1*diff)) + diff1*(math.log(diff1*-1))
    return ans
```

In [153…
```
df['AHD'].value_counts()
```

Out[153…
```
0    8
1    6
Name: AHD, dtype: int64
```

In [154…
```
inf(8,6)
```

Out[154…
```
0.6829081047004717
```

In [124…
```
df.head()
```

Out[124…

	Sex	Fbs	ExAng	Slope	Ca	Thal	AHD
0	1	1	0	3	0	0	0
1	1	0	1	2	3	1	1
2	1	0	1	2	2	2	1
3	1	0	0	3	0	1	0
4	0	0	0	1	0	1	0

In [125…
```
df1 = df[(df['Sex'] == 1) & (df['AHD'] == 1)]
df1['AHD'].value_counts()
```

Out[125…
```
1    5
Name: AHD, dtype: int64
```

In [126…
```
df1 = df[(df['Sex'] == 1) & (df['AHD'] == 0)]
df1['AHD'].value_counts()
```

Out[126…
```
0    5
Name: AHD, dtype: int64
```

In [147…
```
df['Slope'] = [0 if x == 1 else 1 if x == 2 else 2 for x in df['Slope']]
df.head()
```

Out[147…

	Sex	Fbs	ExAng	Slope	Ca	Thal	AHD
0	1	1	0	1	0	0	0
1	1	0	1	0	3	1	1

	Sex	Fbs	ExAng	Slope	Ca	Thal	AHD
2	1	0	1	0	2	2	1
3	1	0	0	1	0	1	0
4	0	0	0	2	0	1	0

In [181…
```
def info_col(col_name,n):
    l = []
    for i in range(0,n):
        df1 = df[(df[col_name] == i) & (df['AHD'] == 1)]
        if df1.empty:
            l.append(0)
        else:
            l.append(int(df1['AHD'].value_counts()))
        df1 = df[(df[col_name] == i) & (df['AHD'] == 0)]
        if df1.empty:
            l.append(0)
        else :
            l.append(int(df1['AHD'].value_counts()))
    j = 0
    s = 0
    while j < len(l):
        if l[j]>0 and l[j+1]>0:
            s = s + ((l[j]+l[j+1])/14)*inf(int(l[j]),int(l[j+1]))
        j += 2
    return s
```

In [187…
```
info_col('Sex',2)
```

Out[187…
```
0.673487303713993
```

In [188…
```
info_col('Fbs',2)
```

Out[188…
```
0.6535027367351511
```

In [185…
```
info_col('ExAng',2)
```

Out[185…
```
0.5966591891645707
```

In [182…
```
info_col('Thal',3)
```

Out[182…
```
0.5709878794955071
```

In [184…
```
info_col('Ca',4)
```

Out[184…
```
0.3267543504994274
```

In [186…
```
info_col('Slope',3)
```

Out[186…
```
0.4313867113134571
```

In [ ]: