

```
[12]: import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns

In [13]: df=pd.read_csv('D:\Ananya\heart.csv')
df.head()

In [14]: df.head()

Out[14]:
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

```


In [15]: df.describe()

Out[15]:
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000
mean	54.434146	0.695010	0.942439	131.611707	246.000000	0.149268	0.529756	148.114146	0.336585	1.071512	1.386366	0.754146	2.323902	0.513171
std	9.472290	0.460373	1.028641	17.516178	51.56251	0.356627	0.527878	23.005724	0.472772	1.175093	0.617785	1.030798	0.620660	0.500070
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	48.000000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	132.000000	0.000000	0.000000	1.000000	0.000000	2.000000	0.000000
50%	56.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	152.000000	0.000000	0.000000	0.000000	2.000000	2.000000	1.000000
75%	61.000000	1.000000	2.000000	140.000000	275.000000	0.000000	1.000000	166.000000	1.000000	1.800000	2.000000	1.000000	3.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.000000	1.000000

```


In [16]: df.dtypes

Out[16]:
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	int64													
sex	int64													
cp	int64													
trestbps	int64													
chol	int64													
fbs	int64													
restecg	int64													
thalach	int64													
exang	int64													
oldpeak	float64													
slope	int64													
ca	int64													
thal	int64													
target	int64													
dtype:	object													

```


In [17]: plt.hist(df['age'],bins=[20,30,40,50,60,70,80],edgecolor='black')
plt.title('age')

Out[17]:
```

Text(0.5, 1.0, 'age')



```


In [18]: plt.hist(df['trestbps'],bins=[90,100,110,120,130,140,150,160,170,180,190,200],edgecolor='black')
plt.title('resting blood pressure')

Out[18]:
```

Text(0.5, 1.0, 'resting blood pressure')



```


In [ ]:

In [18]: plt.hist(df['chol'], bins = 7, edgecolor = 'black')
plt.title('Cholesterol')

Out[18]:
```

Text(0.5, 1.0, 'Cholesterol')



```


In [11]: plt.hist(df['thalach'], bins = [70,80,90,100,110,120,130,140,150,160,170,180,190,200], edgecolor = 'black')
plt.title('Max Heart Rate')

Out[11]:
```

Text(0.5, 1.0, 'Max Heart Rate')



```


In [12]: plt.hist(df['oldpeak'], bins = 5, edgecolor = 'black')
plt.title('ST Depression')

Out[12]:
```

Text(0.5, 1.0, 'ST Depression')



```


In [13]: plt.scatter(df['age'],df['trestbps'], s=30, c = '#b6eb7a', edgecolor = 'green', linewidth = 1, alpha = 0.8)
plt.xlabel('Age')
plt.ylabel('resting Blood Pressure')
plt.title('Age vs RBP')

Out[13]:
```

Text(0.5, 1.0, 'Age vs RBP')



```


In [14]: plt.scatter(df['age'],df['chol'], s=30, c = '#9bdeac', edgecolor = 'green', linewidth = 1, alpha = 0.8)
plt.xlabel('Age')
plt.ylabel('cholesterol')
plt.title('Age vs Cholesterol')

Out[14]:
```

Text(0.5, 1.0, 'Age vs Cholesterol')



```


In [15]: plt.scatter(df['age'],df['thalach'], s=30, c = '#b6eb7a', edgecolor = 'green', linewidth = 1, alpha = 0.8)
plt.xlabel('Age')
plt.ylabel('Max Heart Rate')
plt.title('Age vs Max Heart Rate')

Out[15]:
```

Text(0.5, 1.0, 'Age vs Max Heart Rate')



```


In [16]: plt.scatter(df['age'],df['oldpeak'], s=30, c = '#a8df65', edgecolor = 'green', linewidth = 1, alpha = 0.8)
plt.xlabel('Age')
plt.ylabel('ST depression')
plt.title('Age vs ST depression')

Out[16]:
```

Text(0.5, 1.0, 'Age vs ST depression')



```


In [ ]:

In [18]: sns.jointplot(x=df['thalach'], y=df['chol'], data=df)

Out[18]:
```

<seaborn.axisgrid.JointGrid at 0x2336e35210>



```


In [ ]:

In [19]: plt.scatter(df['age'],df['oldpeak'], s=30, c = '#e2979c', edgecolor = 'red', linewidth = 1, alpha = 0.8)
plt.xlabel('Age')
plt.ylabel('ST depression')
plt.title('Age vs ST depression')

Out[19]:
```

Text(0.5, 1.0, 'Age vs ST depression')



```


In [24]: sns.jointplot(x=df['chol'], y=df['trestbps'], data=df, kind='kde', shade=True, cmap='Purples_d')

C:\Users\Anagha\Anaconda3\Lib\site-packages\seaborn\axisgrid.py:1826: FutureWarning:
'shade' is now deprecated in favor of 'fill'; setting 'fill=True'
This will become an error in seaborn v0.14.0; please update your code.
  func=seaborn.jointplot(x=x, y=y, **kwargs)

Out[24]:
```

<seaborn.axisgrid.JointGrid at 0x2928c81440>



```


In [ ]:

In [26]: plt.scatter(df['trestbps'],df['thalach'], s=30, c = '#e2979c', edgecolor = 'red', linewidth = 1, alpha = 0.8)
plt.xlabel('resting Blood Pressure')
plt.ylabel('Max Heart Rate')
plt.title('RBP vs Max Heart Rate')

Out[26]:
```

Text(0.5, 1.0, 'RBP vs Max Heart Rate')



```


In [32]: X = df[['age','trestbps','chol','thalach','oldpeak']]
y = df['target']

In [33]: data = X
data_std = (data - data.mean()) / data.std()
data = pd.concat([y, data_std.iloc[:,1:]], axis=1)
data = pd.melt(data, id_vars = 'target', var_name = 'features',
              value_name = 'value')
data.head()

Out[33]:
```

	target	features	value
0	0	age	-0.268306
1	0	age	-0.158080
2	0	age	1.715758
3	0	age	0.723726
4	0	age	0.833952

```


In [ ]:

In [27]: sns.catplot(x='target',y='thalach',data=df,kind='box',hue='exang', palette=sns.color_palette(['#a62447', '#74d4c0']))

Out[27]:
```

<seaborn.axisgrid.FacetGrid at 0x2928e39690>

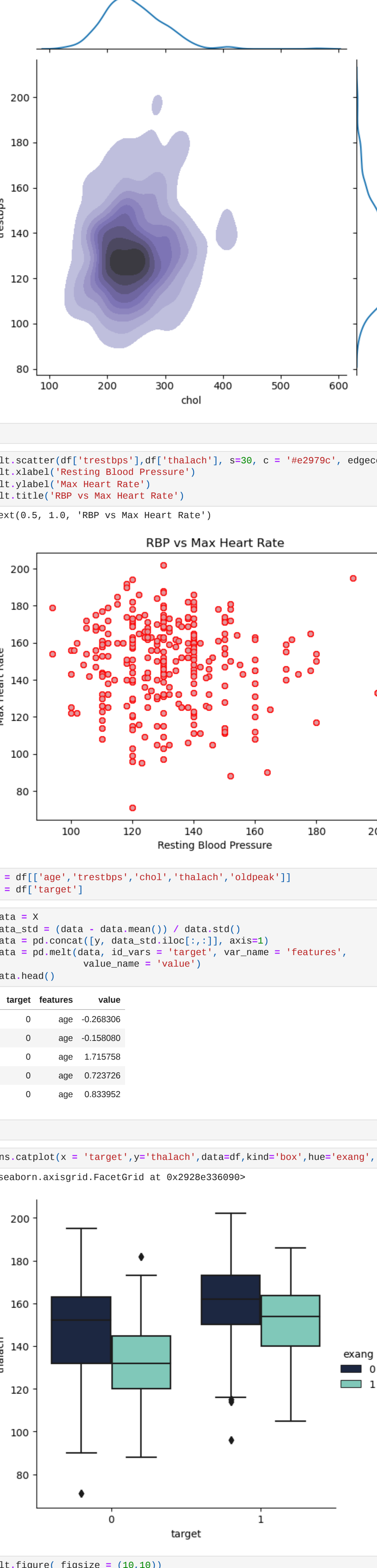


```


In [28]: plt.figure(figsize = (10,10))
sns.heatmap(df.corr(), annot = True)

Out[28]:
```

<Axes: >



```


In [ ]:

In [ ]:
```