Heart Disease

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
df =pd.read_csv("heart-disease.csv")
df.info()
df.isnull().sum()
plt.figure(figsize=(20,10))
sns.heatmap(data = df.corr(), annot = True)
plt.show()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       303 non-null    int64
 1   sex       303 non-null    int64
 2   cp        303 non-null    int64
 3   trestbps  303 non-null    int64
 4   chol      303 non-null    int64
 5   fbs       303 non-null    int64
 6   restecg   303 non-null    int64
 7   thalach   303 non-null    int64
```

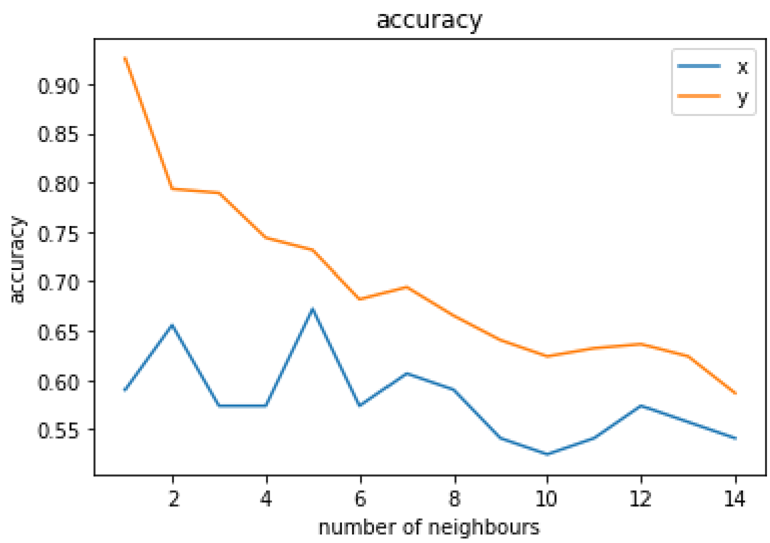
```
df.head()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	th
0	63	1	3	145	233	1	0	150	0	2.3	0	0	
1	37	1	2	130	250	0	1	187	0	3.5	0	0	
2	41	0	1	130	204	0	0	172	0	1.4	2	0	
3	56	1	1	120	236	0	1	178	0	0.8	2	0	
4	57	0	0	120	354	0	1	163	1	0.6	2	0	

```
X = df.iloc[:,[4,5,8]].values
y = df.iloc[:,[13]].values
import numpy as np
n = np.arange(1,15)
train_accuracy = np.empty(len(n))
test_accuracy = np.empty(len(n))
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state=42
knn = KNeighborsClassifier(n_neighbors=15)

knn.fit(X_train, y_train)
for i, k in enumerate(n):
    knn = KNeighborsClassifier(n_neighbors=k)
    knn.fit(X_train, y_train)

    # Compute training and test data accuracy
    train_accuracy[i] = knn.score(X_train, y_train)
    test_accuracy[i] = knn.score(X_test, y_test)

# Generate plot
plt.title("accuracy")
plt.xlabel("number of neighbours")
plt.ylabel("accuracy")

plt.plot(n, test_accuracy, label = 'Testing dataset Accuracy');
plt.plot(n, train_accuracy, label = 'Training dataset Accuracy');
plt.legend(["x","y"],loc = "upper right");
```

```
plt.show();
print("accuracy: ",knn.score(X_test, y_test))
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:11: DataConversionWarnir
  # This is added back by InteractiveShellApp.init_path()
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:14: DataConversionWarnir
```




```
accuracy:  0.5409836065573771
```

```
import numpy as np
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
from sklearn.cluster import KMeans
data = pd.read_csv('heart-disease.csv')
```

```
x = data.iloc[:,[4,5,8,13]] # 1t for rows and second for columns
kmeans = KMeans(n_clusters=2, init='random', n_init=10, max_iter=300,tol=1e-04,random_stat
kmeans.fit(x)
identified_clusters = kmeans.fit_predict(x)
identified_clusters
data_with_clusters = data.copy()
data_with_clusters['Clusters'] = identified_clusters
plt.scatter(data_with_clusters['trestbps'],data_with_clusters['chol'],data_with_clusters['
```

```
<matplotlib.collections.PathCollection at 0x7fe773dbc410>
```