covid-19-task-4-1

June 12, 2025

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
```

```python
cov_data=pd.read_csv("/content/covid-dataset.csv")
```

```python
cov_data.head()
```

```
   fever  bodyPain  age  runningNose  diffBreath  infectionProb
0    102         0    9            0          -1              0
1    102         0   10            0           0              1
2    104         0   33            1          -1              0
3    101         1   59            0           1              0
4     99         0   98            0           0              0
```

```python
cov_data.shape
```

```
(2575, 6)
```

```python
cov_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2575 entries, 0 to 2574
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   fever          2575 non-null   int64
 1   bodyPain       2575 non-null   int64
 2   age            2575 non-null   int64
 3   runningNose    2575 non-null   int64
 4   diffBreath     2575 non-null   int64
 5   infectionProb  2575 non-null   int64
dtypes: int64(6)
memory usage: 120.8 KB
```

```python
cov_data.isnull().sum()
```

```
[ ]: fever          0
     bodyPain       0
     age            0
     runningNose    0
     diffBreath     0
     infectionProb  0
     dtype: int64
```

```
[ ]: cov_data.isnull().sum()
```

```
[ ]: fever          0
     bodyPain       0
     age            0
     runningNose    0
     diffBreath     0
     infectionProb  0
     dtype: int64
```

```
[ ]: x=cov_data
```

```
[ ]: print(x)
```

```
           fever  bodyPain  age  runningNose  diffBreath  infectionProb
     0        102         0    9            0          -1              0
     1        102         0   10            0           0              1
     2        104         0   33            1          -1              0
     3        101         1   59            0           1              0
     4         99         0   98            0           0              0
     ...      ...       ...  ...          ...         ...            ...
     2570      99         0   90            0           0              1
     2571     100         0   53            0          -1              1
     2572     101         0   44            1           0              0
     2573     102         0   97            0          -1              1
     2574     104         1   62            1          -1              1

     [2575 rows x 6 columns]
```
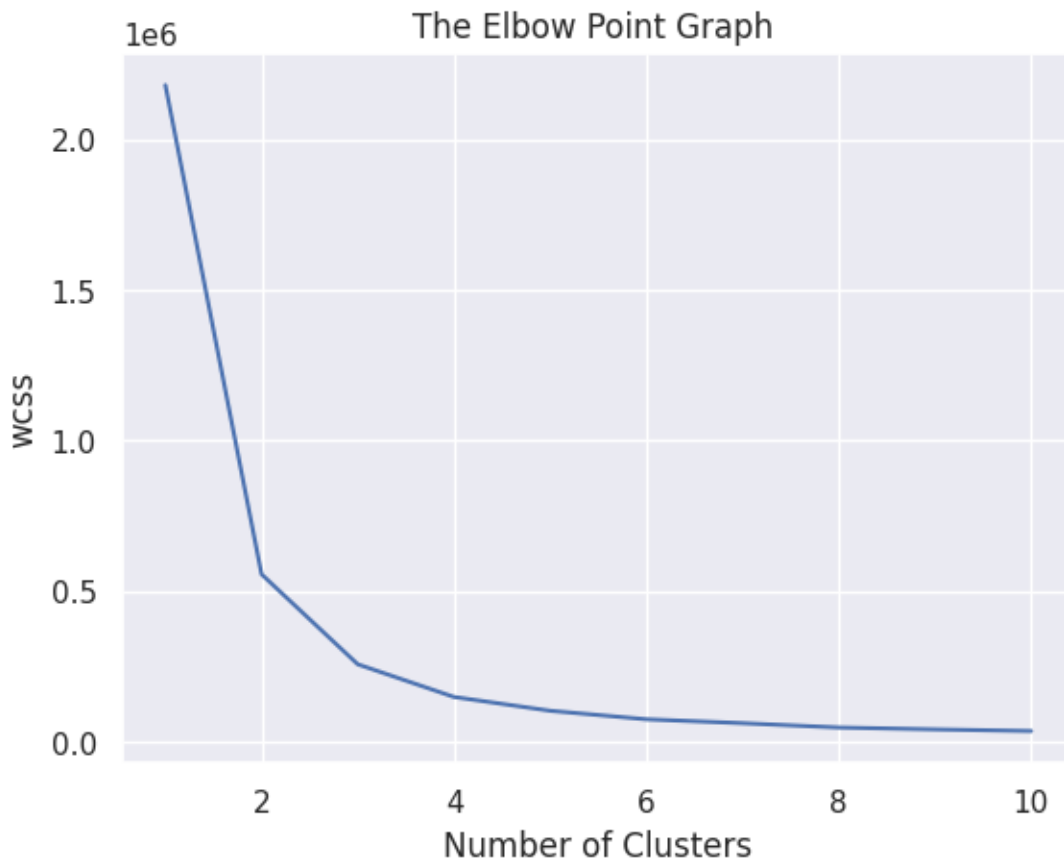
```
[44]: wcss = []

      for i in range(1,11):
        kmeans=KMeans(n_clusters=i,init='k-means++',random_state=42)
        kmeans.fit(x)
        wcss.append(kmeans.inertia_)
```

```
[45]: kmean=KMeans(n_clusters=5,init='k-means++',random_state=42)
```

```
[51]: covid=kmean.fit_predict(x)
```

```
[46]: sns.set()
      plt.plot(range(1,11),wcss)
      plt.title('The Elbow Point Graph ')
      plt.xlabel('Number of Clusters')
      plt.ylabel('wcss')
      plt.show()
```

The Elbow Point Graph


```
[47]: plt.figure(figsize=(7,7))
      plt.scatter(x.iloc[covid==0,0],x.iloc[covid==0,1],s=50,c='brown',label='cluster␣
       ↪1')
      plt.scatter(x.iloc[covid==1,2],x.iloc[covid==1,2],s=50,c='blue',label='cluster␣
       ↪2')
      plt.scatter(x.iloc[covid==2,3],x.
       ↪iloc[covid==2,4],s=50,c='yellow',label='cluster 3')

      plt.scatter(kmean.cluster_centers_[:,0],kmean.cluster_centers_[:
       ↪,1],s=300,c='red',label='centroids')
```

```
[47]: <matplotlib.collections.PathCollection at 0x7db038c28290>
```