PROJECT REPORT-2

My github account link-https://github.com/arpmit-a

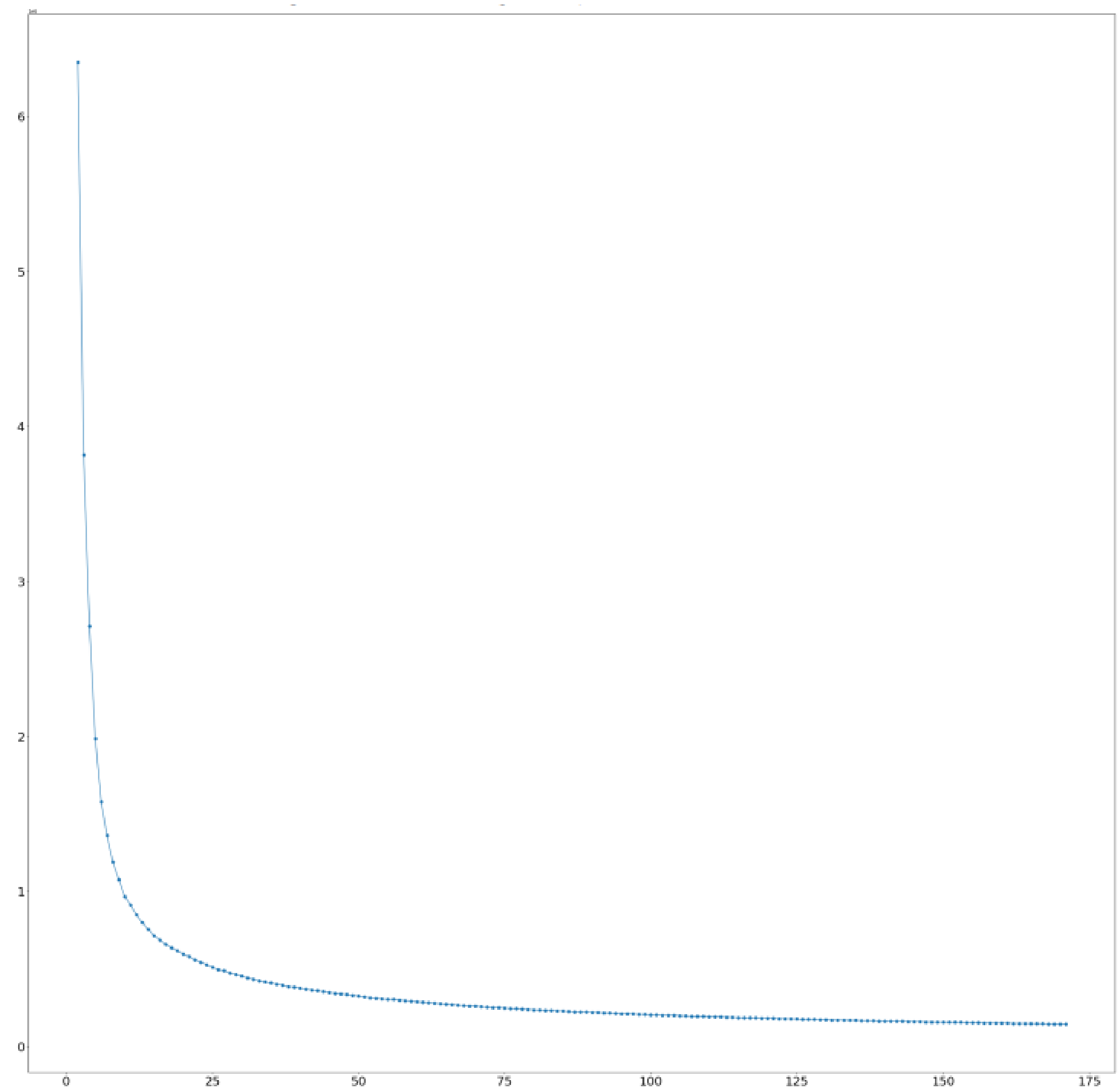
The link of my google colab note book - https://colab.research.google.com/drive/1XOadqIq53ExZCL2gfEUhUe9QRDZeNBoD?usp=sharing

Major Project-2 *Applying K-means clustering in spotify music data*

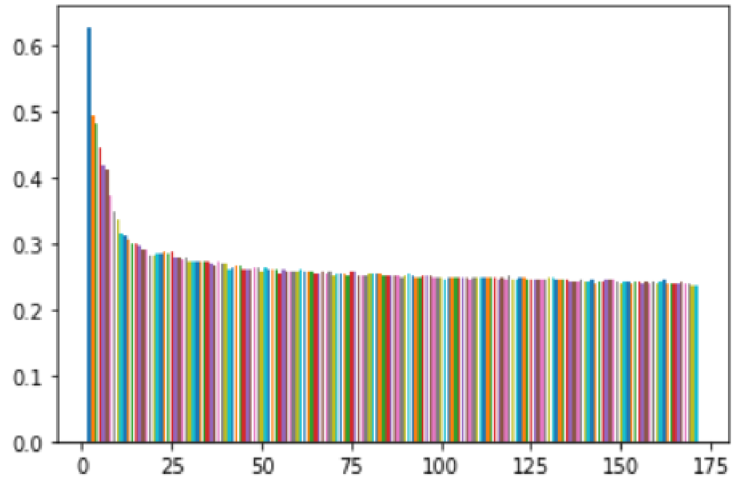
```
import numpy as np import pandas as pd

df= pd.read_csv('/content/genres_v2.csv') df

df.fillna('0')

df.info()

df=df.fillna('0')



ninth_col=df.pop('mode') df.insert(8,
'mode', ninth_col)

eighth_col=df.pop('key') df.insert(7,
'key', eighth_col)

ninteenth_col=df.pop('title') df.insert(19,
'title', ninteenth_col)

df.info()

x=df.iloc[:,0:11].values x

#We can do the visualization part using 2 of the columns but it won't be of much use as th
#So we'll skip the visualization

np.sqrt(29434)# As there are total 29434 points. So number of clusters should range from 2

    171.56339936011994

#Let's apply K-Means clustering by 2 methods #1.
Elbow Method



import matplotlib.pyplot as plt
```

```
 from sklearn.cluster import KMeans

k = range(2,172) # The range is in between 2 and 171

sse = [] # It's a blank list to be used later. The full form of sse = sum of squared errors
```

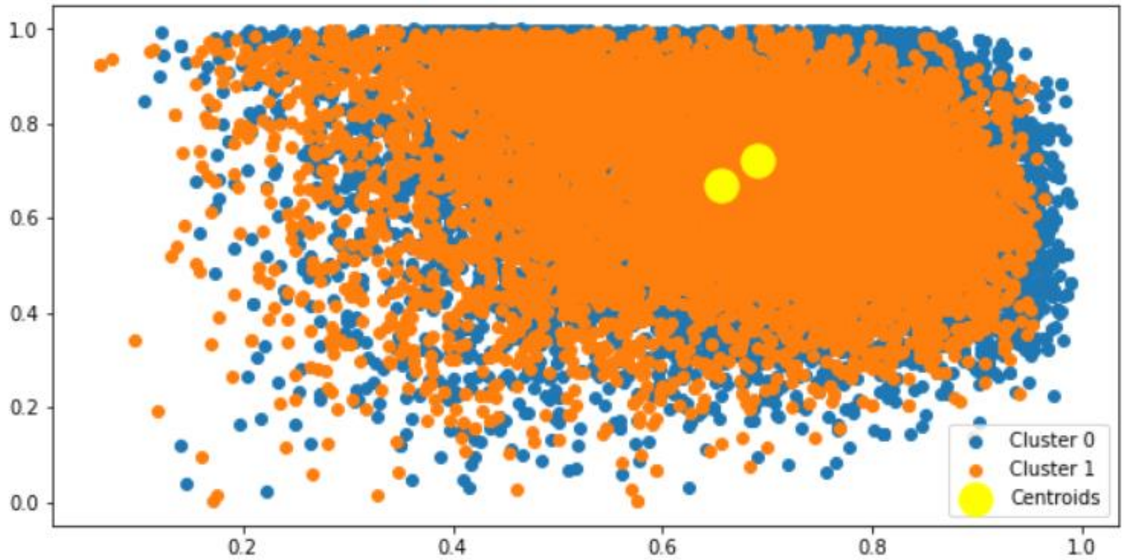
```
 for i in k :
   model_demo = KMeans(n_clusters = i,random_state = 0)   model_demo.fit(x)
 sse.append(model_demo.inertia_) #.inertia_ - calculates the sum of squared error
  f=plt.figure()
 f.set_figwidth(40)
 f.set_figheight(40)
 plt.scatter(k,sse) plt.plot(k,sse)
 matplotlib.pyplot.xticks(fontsize=25)
 matplotlib.pyplot.yticks(fontsize=25)
```



```
from sklearn.metrics import silhouette_score
k = range(2,172) for i in k:
  model_demo = KMeans(n_clusters = i,random_state = 0)
model_demo.fit(x)

 y_pred = model_demo.predict(x)

 print(f"{i} Clusters ,Score = {silhouette_score(x,y_pred)}")

plt.bar(i,silhouette_score(x,y_pred))
```



```
# At k=2 we get max silhouette score...
k = 2
from sklearn.cluster import KMeans

model = KMeans(n_clusters = k,random_state = 0) model.fit(x)

KMeans(n_clusters=2, random_state=0)

y = model.predict(x) # predicted output y array([1, 0,

1, ..., 0, 0, 0], dtype=int32)


plt.figure(figsize = (10,5)) for i in
range(k):
  plt.scatter(x[y == i,0],x[y == i,1],label = f'Cluster {i}')
plt.scatter(model.cluster_centers_[:,0],model.cluster_centers_[:,1],s = 300,c = 'yellow',
```



The problem with this project is that it takes a lot of time to plot the graphs. First 2 plots nearly took 40mins to be shown.