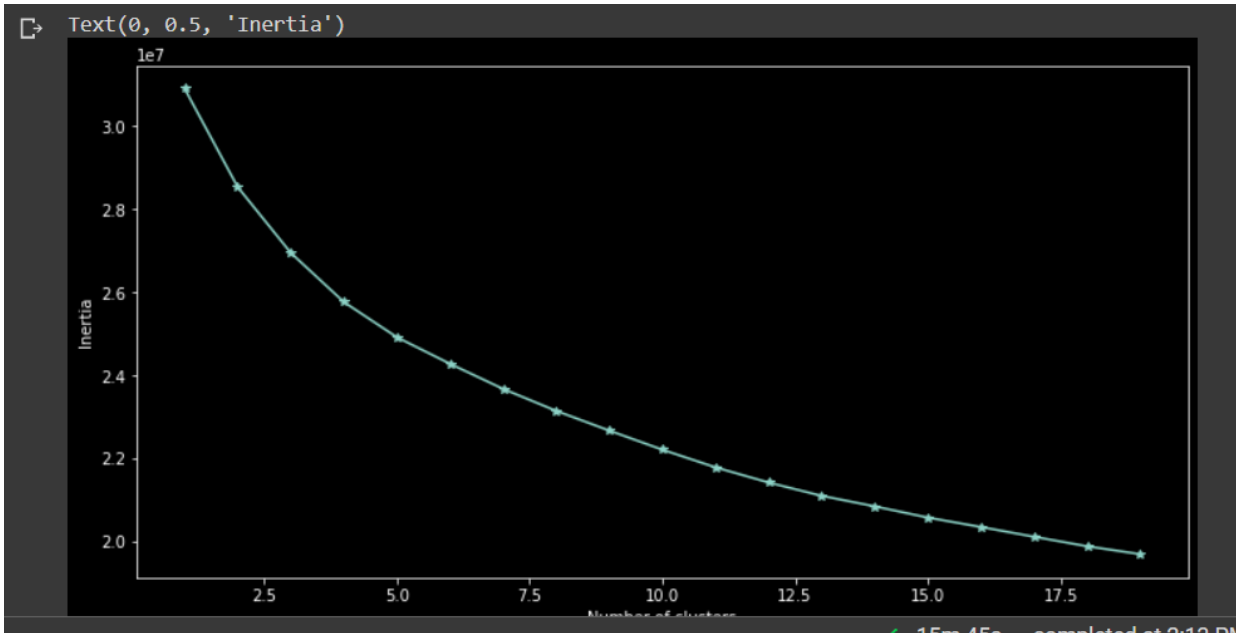
Task 3

Objective

Create a Bag of words based matching/categorization solution on the MNIST-fashion database. Downloaded data from- https://www.kaggle.com/zalando-research/fashionmnist/data - this data had 2 .csv files used as training and testing set images

How did you choose the optimum value of the number of clusters? Mention (<100 words) in assignment report.

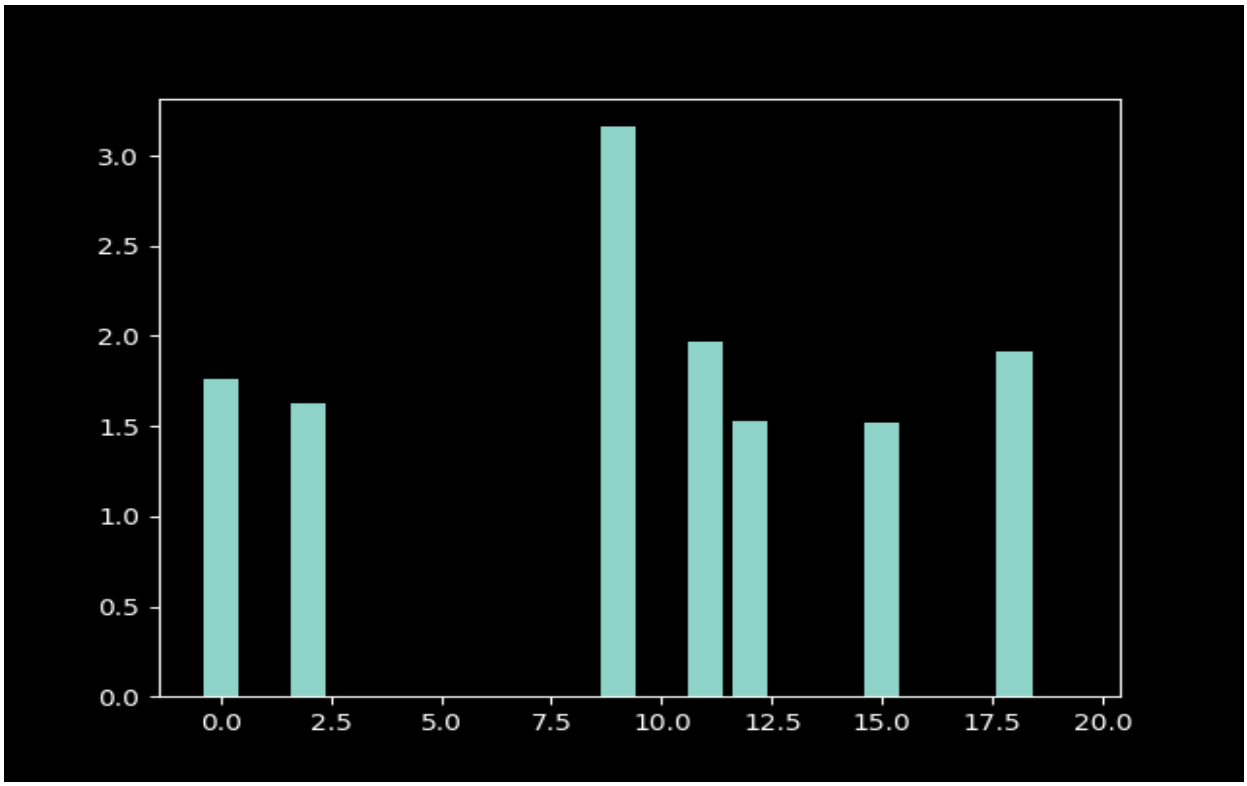


Explanation :
K-means is an unsupervised algorithm as it clusters the data into *k* clusters, even if *k* is not the right number of clusters to use. Therefore, when using k-means clustering, to determine the right value of k i applied elbow method
I chose a value of *k* that still has a low SSE(sum of squared errors), and the elbow usually represents where we start to have diminishing returns by increasing *k* , also we had to check for the accuracy of our entire code simultaneously for different values of k

Therefore, k=20

ComputeHistogram() – takes as input a feature vector and the visual dictionary matrix and generates the histogram using soft assignment (giving weight to the next nearest neighbor)

Output:



MatchHistogram() – the function compares two histograms and returns the distance.

```
from scipy.stats import norm
def MatchHistograms(i: int,j:int):
    plt.imshow(train_imgs[i], cmap='gray')
    plt.show()
    plt.imshow(train_imgs[j], cmap='gray')
    plt.show()
    a=tfidf[i]
    b=tfidf[j]
    print(np.dot(a,b))
    cosine_similarity = np.dot(a,b) / ( (np.dot(a,a) **.5) * (np.dot(b,b) ** .5) )
    return cosine_similarity
```

Display the Accuracy (1 Mark).

```
PROBLEMS    OUTPUT    DEBUG CONSOLE    TERMINAL

) is deprecated due SIFT tranfer to the main r
C:\Users\Tiya\AppData\Local\Programs\Python\Py
  warnings.warn(
Total Hits Are: 4365
accuracy =  0.4766845036584034
[[756  47  10  28   6  34   3  54  39  10]
 [ 22 588  49  56  81  48  16  24  55  34]
```

Output answer.csv:

```
,True Classes,Predicted Classes
0,T-shirt/top,T-shirt/top
1,Pullover,Coat
2,Pullover,Sandal
3,Dress,Coat
4,Bag,Pullover
5,Shirt,Pullover
6,Sandal,Sandal
7,T-shirt/top,T-shirt/top
8,Dress,Trouser
9,Coat,Coat
10,Coat,Coat
11,Shirt,Coat
12,Bag,Pullover
13,Sandal,Sandal
14,Shirt,Pullover
15,Dress,Pullover
16,Shirt,Coat
17,Coat,Pullover
18,Coat,T-shirt/top
19,Coat,Coat
20,Pullover,Pullover
21,Trouser,Trouser
22,Sandal,Sandal
23,Bag,Bag
24,Coat,Trouser
25,Coat,Sneaker
26,Trouser,Trouser
27,Sandal,Sandal
28,Sneaker,Ankle boot
29,Sneaker,Ankle boot
30,Bag,Ankle boot
31,Trouser,Trouser
32,T-shirt/top,Trouser
```