Tarunkumar Pasumarthi
903085537

HW5 Bonus

Part1:

Code:

```
import pandas as pd
df=[0,0,0,0,0]
df[0]= pd.read_csv("sample_data/data0.csv")
df[1]= pd.read_csv("sample_data/data1.csv")
df[2]= pd.read_csv("sample_data/data2.csv")
df[3]= pd.read_csv("sample_data/data3.csv")
df[4]= pd.read_csv("sample_data/data4.csv")


from sklearn.cluster import KMeans as sklearnKMeans
from time import time

t=[]
for d in df:
  kmeans_sk = sklearnKMeans(init="k-means++", n_clusters=5, n_jobs=-1)
  start=time()
  kmeans_sk.fit(d)
  end=time()
  t1= end-start
  t.append(t1)
  print(t1)
```

Results:

```
1.6562762260437012
1.7306795120239258
2.503502607345581
3.9505224227905273
14.963164329528809
```

Part 2:

```
import cudf as gd

cdf=[0,0,0,0,0]
cdf[0]= gd.read_csv("sample_data/data0.csv")
cdf[1]= gd.read_csv("sample_data/data1.csv")
cdf[2]= gd.read_csv("sample_data/data2.csv")
cdf[3]= gd.read_csv("sample_data/data3.csv")
cdf[4]= gd.read_csv("sample_data/data4.csv")

from cuml.cluster import KMeans as cumlKMeans
```

Tarunkumar Pasumarthi
903085537

```
from time import time

t2=[]
for d in cdf:
  kmeans_cuml = cumlKMeans(init="k-
means||", n_clusters=5, oversampling_factor=40)
  start=time()
  kmeans_cuml.fit(d)
  end=time()
  t1= end-start
  t2.append(t1)
  print(t1)
```

results:

```
0.065627622604370
0.073067951202392
0.150350260734558
0.295052242279052
1.496316432952880
```

Part3: