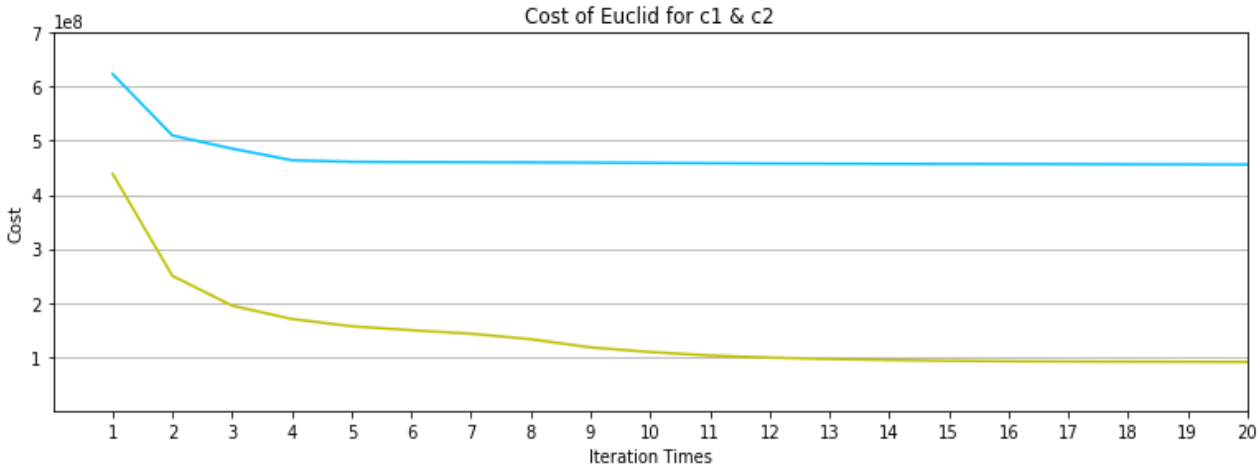
Part2 Report (named as Report.pdf)

For question (a), you should show:

1. A plot of cost vs. iteration for 2 initialization strategies(c1 and c2) for (a)

	C1	C2
Round 1	550117.1	1433739
Round 2	464869.3	1084489
Round 3	470897.4	973431.7
Round 4	483914.4	895934.6
Round 5	489216.1	865128.3
Round 6	487629.7	845846.6
Round 7	483711.9	827219.6
Round 8	475330.8	803590.3
Round 9	474871.2	756039.5
Round 10	457232.9	717332.9
Round 11	447494.4	694587.9
Round 12	450915	684444.5
Round 13	451250.4	674574.7
Round 14	451974.6	667409.5
Round 15	451570.4	663556.6
Round 16	452739	660162.8
Round 17	453082.7	656041.3
Round 18	450583.7	653036.8
Round 19	450368.7	651112.4
Round 20	449011.4	649689



2. Percentage improvement values and your explanation for (a)

Percentage Improvement of c1, c2 for the Euclidean centriod computing

```
In [312]: #Percentage improvement values and my explanation: impovement_of_c1_Eucid

          temp = abs(costList_c1_Euclid[19] - costList_c1_Euclid[0])/ costList_c1_Euclid[0]
          impovement_of_c1_Eucid = str(round(temp*100,3))+'%'
          impovement_of_c1_Eucid

Out[312]: '26.885%'

In [313]: #Percentage improvement values and my explanation: costList_c2_Euclid
          temp = abs(costList_c2_Euclid[19] - costList_c2_Euclid[0])/ costList_c2_Euclid[0]
          impovement_of_c2_Eucid = str(round(temp*100,3))+'%'
          impovement_of_c2_Eucid

Out[313]: '79.438%'
```

解釋:

當我們使用 c1 為起始 centroids(圖-黃色線)，並用 Euclidean distance 聚類，經過 20 輪迭代，我們的誤差 cost 降低了 26.885%

當我們使用 c2 為起始 centroids(圖-藍色線)，並用 Euclidean distance 聚類，經過 20 輪迭代，我們的誤差 cost 降低了 79.438%

這樣的結果表示，當我們採用 Euclidean distance 時，選擇彼此間越遠越好的起始中心點(centroids)是有利的，應當採用 c2.txt，而非隨機選取的 c1.txt

3. The Euclidean and Manhattan Distances for all pairs of centroids, with 2 initialization strategies. (4 tables in total)

過程中以 Euclidean 計算中點 (進行 classify)

(a) 以 C1 為起始點， 使用 Euclidean 計算任兩點之中間點距離-

1	2	3	4	5	6	7	8	9	10
0	692.1579	3490.259	205.7503	346.7188	512.6122	444.731	566.202	1282.771	307.6691
0	0	2798.801	897.659	1038.827	1204.078	1136.327	1257.45	669.8902	412.0761
0	0	0	3695.114	3836.907	4002.689	3934.872	4056.136	2294.58	3195.924
0	0	0	0	142.4389	309.5063	241.7301	363.2629	1474.945	504.6341
0	0	0	0	0	167.1498	99.54554	220.9018	1615.852	646.9306
0	0	0	0	0	0	67.91186	53.78989	1782.203	814.0762
0	0	0	0	0	0	0	121.6337	1715.253	746.3356
0	0	0	0	0	0	0	0	1835.64	867.8231
0	0	0	0	0	0	0	0	0	975.3204
0	0	0	0	0	0	0	0	0	0

(b) 以 C2 為起始點， 使用 Euclidean 計算任兩點之中間點距離-

1	2	3	4	5	6	7	8	9	10
0	15760.12	14110.83	9045.32	5567.685	1924.624	1100.859	402.8905	2105.443	3169.004
0	0	11524.51	6743.884	10192.53	14455.12	14682.45	15362.42	13674.71	12597.04
0	0	0	9545.879	10883.38	12233.96	13208	13786.48	12508.96	11938.38
0	0	0	0	3494.222	7718.222	7957.776	8644.807	6947.821	5876.33
0	0	0	0	0	4404.563	4492.458	5169.937	3488.159	2407.919
0	0	0	0	0	0	1182.864	1615.788	1313.327	2153.771
0	0	0	0	0	0	0	698.4881	1010.198	2085.461
0	0	0	0	0	0	0	0	1702.793	2768.608
0	0	0	0	0	0	0	0	0	1080.535
0	0	0	0	0	0	0	0	0	0

(c) 以 **C1** 為起始點，使用 **Manhattan** 計算任兩點之中間點距離-

1	2	3	4	5	6	7	8	9	10
0	728.9243	3797.899	212.1811	374.8904	577.4021	499.1579	645.7698	1731.064	406.7012
0	0	3072.889	935.8853	1100.833	1303.896	1225.352	1372.092	1005.293	490.9281
0	0	0	4001.038	4170.305	4372.789	4294.953	4440.72	2513.423	3396.42
0	0	0	0	171.3652	375.2479	296.2547	443.4984	1934.087	609.7493
0	0	0	0	0	204.5229	125.5968	272.9349	2102.865	779.3972
0	0	0	0	0	0	79.40168	69.58988	2306.38	983.0197
0	0	0	0	0	0	0	147.8657	2227.556	904.3703
0	0	0	0	0	0	0	0	2374.545	1050.916
0	0	0	0	0	0	0	0	0	1327.584
0	0	0	0	0	0	0	0	0	0

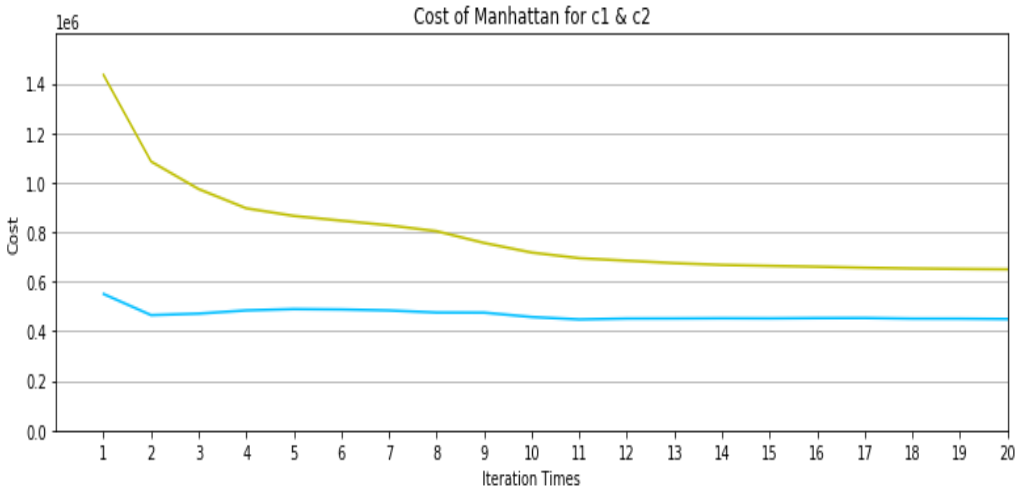
(d) 以 **C2** 為起始點，使用 **Manhattan** 計算任兩點之中間點距離-

1	2	3	4	5	6	7	8	9	10
0	15772.61	20215.65	9533.171	5604.2	3088.054	1311.039	471.2657	2369.412	3349.657
0	0	16003.5	7219.197	10221.03	16105.35	14909.17	15434.46	13950.58	12776.88
0	0	0	10690.48	14613.55	17509.9	18912.61	19748.94	17851.81	16873.24
0	0	0	0	3935.293	8896.389	8228.355	9065.404	7168.733	6190.679
0	0	0	0	0	5893.07	4696.975	5221.253	3737.707	2564.171
0	0	0	0	0	0	1781.823	2619.811	2162.802	3337.746
0	0	0	0	0	0	0	840.7225	1068.94	2137.788
0	0	0	0	0	0	0	0	1901.209	2883.735
0	0	0	0	0	0	0	0	0	1176.45
0	0	0	0	0	0	0	0	0	0

For question (b), you should show:

1. A plot of cost vs. iteration for 2 initialization strategies(c1 and c2) for (b

	C1	C2
Round 1	550117.1	1433739
Round 2	464869.3	1084489
Round 3	470897.4	973431.7
Round 4	483914.4	895934.6
Round 5	489216.1	865128.3
Round 6	487629.7	845846.6
Round 7	483711.9	827219.6
Round 8	475330.8	803590.3
Round 9	474871.2	756039.5
Round 10	457232.9	717332.9
Round 11	447494.4	694587.9
Round 12	450915	684444.5
Round 13	451250.4	674574.7
Round 14	451974.6	667409.5
Round 15	451570.4	663556.6
Round 16	452739	660162.8
Round 17	453082.7	656041.3
Round 18	450583.7	653036.8
Round 19	450368.7	651112.4
Round 20	449011.4	649689



2. Percentage improvement values and your explanation for (b)

```
In [20]: #Percentage improvement values and my explanation: impovement_of_c1_Maha

         temp = abs(costList_c1_Maha[19] - costList_c1_Maha[0])/ costList_c1_Maha[0]
         impovement_of_c1_Maha = str(round(temp*100,3))+'%'
         impovement_of_c1_Maha

Out[20]: '18.379%'
```

```
In [49]: #Percentage improvement values and my explanation: impovement_of_c2_Maha
         temp = abs(costList_c2_Maha[19] - costList_c2_Maha[0])/ costList_c2_Maha[0]
         impovement_of_c2_Maha = str(round(temp*100,3))+'%'
         impovement_of_c2_Maha

Out[49]: '54.686%'
```

解釋:

當我們使用 c1 為起始 centroids(圖-藍色線)，並用 Manhattan distance 聚類，經過 20 輪迭代，我們的誤差 cost 降低了 18.379%，最終誤差小於 c2 之誤差

當我們使用 c2 為起始 centroids(圖-黃色線)，並用 Manhattan distance 聚類，經過 20 輪迭代，我們的誤差 cost 降低了 54.686%，最終誤差高於 c1 之誤差

這樣的結果表示，當我們採用 Manhattan distance 計算 centroids 時，
在迭代次數少時，c2 成本遠高於 c1 成本，而在迭代過程中，c2 有大量的修正，儘管最終結果，c2 成本仍然略高於 c1。
因此，若使用 Manhattan distance 計算 centroids 時，隨機選擇起始中心點(centroids)在迭代次數小的狀況是有利的， 但是整體來說，選擇盡量遠的點作為中心點(c2.txt)會隨著迭代數量增加而大幅修正。

3. The Euclidean and Manhattan Distances for all pairs of centroids, with 2 initialization strategies. (4 tables in total)

以 Manhattan 計算中點 (進行 classify)

(e) 以 C1 為起始點， 使用 Euclidean 計算任兩點之中間點距離-

1	2	3	4	5	6	7	8	9	10
0	2219.177	9948.044	528.6998	413.3651	827.7189	681.035	917.1274	832.1474	729.0563
0	0	7767.946	2734.05	2628.491	3044.478	2898.713	3133.46	1812.455	1491.357
0	0	0	10433.06	10361.37	10773.53	10626.49	10862.97	9340.275	9236.84
0	0	0	0	221.3728	375.1562	249.3792	457.2597	1156.583	1251.158
0	0	0	0	0	415.99	270.7488	505.0711	1171.964	1137.135
0	0	0	0	0	0	147.047	89.49092	1529.464	1553.124
0	0	0	0	0	0	0	236.5146	1391.55	1407.404
0	0	0	0	0	0	0	0	1613.556	1642.129
0	0	0	0	0	0	0	0	0	709.4078
0	0	0	0	0	0	0	0	0	0

(f) 以 C2 為起始點， 使用 Euclidean 計算任兩點之中間點距離-

1	2	3	4	5	6	7	8	9	10
0	514.627	1571.243	1338.161	3022.661	2006.703	9032.333	15747.23	14100.14	5554.787
0	0	1081.379	827.8407	2511.459	1637.729	8521.198	15239.88	13684.61	5047.516
0	0	0	566.551	1649.389	910.9944	7588.405	14328.23	12643.99	4167.637
0	0	0	0	1684.516	1405.109	7694.277	14412.06	13125.35	4219.761
0	0	0	0	0	2124.263	6009.82	12731.4	12006.39	2542.569
0	0	0	0	0	0	7742.628	14474.55	12167.79	4452.972
0	0	0	0	0	0	0	6743.884	9545.879	3494.222
0	0	0	0	0	0	0	0	11524.51	10192.53
0	0	0	0	0	0	0	0	0	10883.38
0	0	0	0	0	0	0	0	0	0

(g) 以 **C1** 為起始點，使用 **Manhattan** 計算任兩點之中間點距離-

1	2	3	4	5	6	7	8	9	10
0	2341.017	11929.3	651.1875	496.3315	947.7432	770.7374	1056.8	1260.511	737.7136
0	0	9597.441	2778.946	2830.145	3280.359	3104.286	3388.983	2380.461	1605.27
0	0	0	12323.29	12421.26	12871.48	12695.55	12979.13	10775.94	11196.79
0	0	0	0	335.9512	558.4693	382.4633	667.5332	1653.826	1379.165
0	0	0	0	0	452.8613	276.3265	561.8492	1755.106	1226.66
0	0	0	0	0	0	177.5932	110.2176	2205.307	1677.667
0	0	0	0	0	0	0	287.4297	2028.902	1500.993
0	0	0	0	0	0	0	0	2314.667	1786.811
0	0	0	0	0	0	0	0	0	1006.368
0	0	0	0	0	0	0	0	0	0

(h) 以 **C2** 為起始點，使用 **Manhattan** 計算任兩點之中間點距離-

1	2	3	4	5	6	7	8	9	10
0	602.9548	2102.554	1430.209	3211.456	3281.488	9517.668	15757.69	20200.26	5588.854
0	0	1500.825	833.4303	2613.997	2682.569	8918.813	15335.96	19602.26	5123.067
0	0	0	674.8276	2062.251	1358.796	7771.222	14980.06	18111.89	4768.923
0	0	0	0	1784.512	1855.58	8090.51	14506.49	18775.12	4293.502
0	0	0	0	0	3413.036	6312.53	12922.93	16995.13	2710.056
0	0	0	0	0	0	9116.025	16325.27	17521.52	6110.833
0	0	0	0	0	0	0	7219.197	10690.48	3935.293
0	0	0	0	0	0	0	0	16003.5	10221.03
0	0	0	0	0	0	0	0	0	14613.55
0	0	0	0	0	0	0	0	0	0