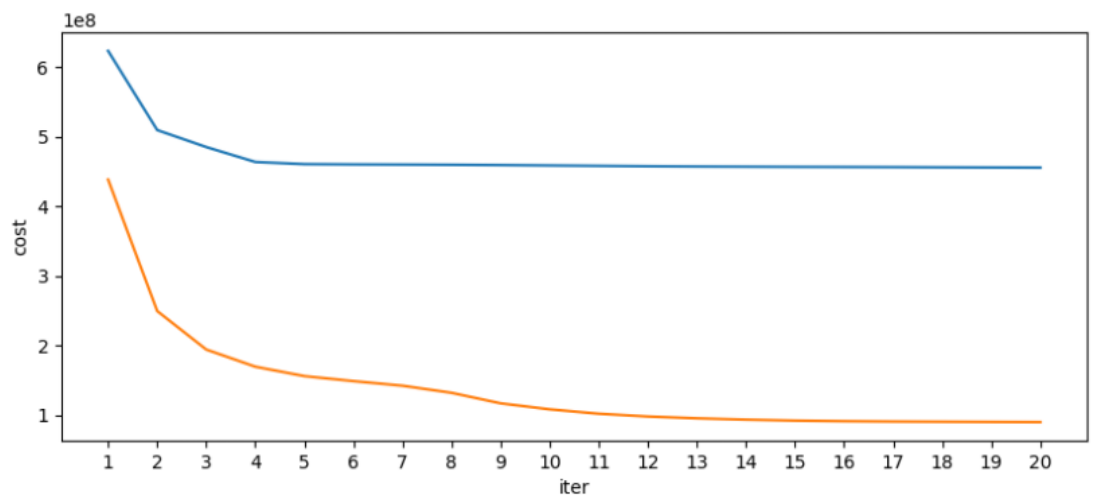


I. question a

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

| | c1 | c2 |
|----|--------------|--------------|
| 0 | 6.236603e+08 | 4.387478e+08 |
| 1 | 5.098629e+08 | 2.498039e+08 |
| 2 | 4.854807e+08 | 1.944948e+08 |
| 3 | 4.639970e+08 | 1.698048e+08 |
| 4 | 4.609693e+08 | 1.562957e+08 |
| 5 | 4.605378e+08 | 1.490942e+08 |
| 6 | 4.603131e+08 | 1.425085e+08 |
| 7 | 4.600035e+08 | 1.323039e+08 |
| 8 | 4.595705e+08 | 1.171710e+08 |
| 9 | 4.590211e+08 | 1.085474e+08 |
| 10 | 4.584907e+08 | 1.022372e+08 |
| 11 | 4.579442e+08 | 9.827802e+07 |
| 12 | 4.575580e+08 | 9.563023e+07 |
| 13 | 4.572901e+08 | 9.379331e+07 |
| 14 | 4.570506e+08 | 9.237713e+07 |
| 15 | 4.568922e+08 | 9.154161e+07 |
| 16 | 4.567036e+08 | 9.104557e+07 |
| 17 | 4.564042e+08 | 9.075224e+07 |
| 18 | 4.561778e+08 | 9.047017e+07 |
| 19 | 4.559869e+08 | 9.021642e+07 |

1.



2.

B. Percentage improvement values and your explanation for (a):

1. C1 這個沒有特別挑過起始點的群集，最後改善率只有 26.8%
而 C2 盡可能挑相鄰比較遠的點則有 79%的改善率
這顯示了起始點對於 K-MEAN 演算法有非常大的影響
若沒有優化過的話，這個演算法在歐基里德距離下可能會收斂在一個不是非常好的區間。

```
In [200]: caculate_percentage(c1_cost)
```

```
Out[200]: 26.885383292517282
```

```
In [201]: caculate_percentage(c2_cost)
```

```
Out[201]: 79.43775029159902
```

C. The Euclidean and Manhattan Distances for all pairs of centroids, with 2 initialization strategies. (共有四個表格)

```
caculate_Euclidean_cent_dest(c1_center,K)
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-----|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 0 | 0.0 | 692.157887 | 3490.258640 | 205.750279 | 346.718823 | 512.612247 | 444.731001 | 566.201992 | 1282.770845 | 307.669128 |
| 1 | NaN | 0.000000 | 2798.801053 | 897.658986 | 1038.826888 | 1204.078199 | 1136.327344 | 1257.449528 | 669.890228 | 412.076077 |
| 2 | NaN | NaN | 0.000000 | 3695.114191 | 3836.906638 | 4002.689083 | 3934.871559 | 4056.135573 | 2294.579642 | 3195.923901 |
| 3 | NaN | NaN | NaN | 0.000000 | 142.438874 | 309.506324 | 241.730115 | 363.262895 | 1474.945421 | 504.634116 |
| 4 | NaN | NaN | NaN | NaN | 0.000000 | 167.149800 | 99.545543 | 220.901784 | 1615.852353 | 646.930564 |
| 5 | NaN | NaN | NaN | NaN | NaN | 0.000000 | 67.911861 | 53.789891 | 1782.203049 | 814.076150 |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 121.633720 | 1715.253200 | 746.335559 |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 1835.639672 | 867.823079 |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 975.320423 |
| 9 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 |

1.

```
caculate_Euclidean_cent_dest(c2_center,K)
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-----|--------------|--------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 0 | 0.0 | 15760.122472 | 14110.834391 | 9045.320235 | 5567.684524 | 1924.624082 | 1100.859050 | 402.890550 | 2105.442576 | 3169.003773 |
| 1 | NaN | 0.000000 | 11524.505650 | 6743.884100 | 10192.525007 | 14455.119372 | 14682.450993 | 15362.417961 | 13674.707531 | 12597.039560 |
| 2 | NaN | NaN | 0.000000 | 9545.879403 | 10883.382188 | 12233.959805 | 13208.002934 | 13786.484183 | 12508.957381 | 11938.376127 |
| 3 | NaN | NaN | NaN | 0.000000 | 3494.222416 | 7718.222010 | 7957.775949 | 8644.807041 | 6947.820636 | 5876.330200 |
| 4 | NaN | NaN | NaN | NaN | 0.000000 | 4404.562591 | 4492.458214 | 5169.937291 | 3488.158519 | 2407.918794 |
| 5 | NaN | NaN | NaN | NaN | NaN | 0.000000 | 1182.864189 | 1615.788236 | 1313.327493 | 2153.771472 |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 698.488136 | 1010.197665 | 2085.460676 |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 1702.792658 | 2768.607719 |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 1080.534944 |
| 9 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 |

2.

```
caculate_Manhattan_cent_dest(c1_center,K)
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-----|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 0 | 0.0 | 725.419181 | 3792.931961 | 201.761134 | 371.686629 | 573.088824 | 495.515034 | 636.531831 | 1727.399277 | 404.984098 |
| 1 | NaN | 0.000000 | 3067.512780 | 927.180315 | 1097.105810 | 1298.508005 | 1220.934215 | 1361.951012 | 1001.980095 | 320.435083 |
| 2 | NaN | NaN | 0.000000 | 3994.693095 | 4164.618590 | 4366.020785 | 4288.446995 | 4429.463792 | 2065.532685 | 3387.947863 |
| 3 | NaN | NaN | NaN | 0.000000 | 169.925494 | 371.327690 | 293.753900 | 434.770697 | 1929.160411 | 606.745232 |
| 4 | NaN | NaN | NaN | NaN | 0.000000 | 201.402195 | 123.828405 | 264.845202 | 2099.085905 | 776.670727 |
| 5 | NaN | NaN | NaN | NaN | NaN | 0.000000 | 77.573790 | 63.443007 | 2300.488100 | 978.072922 |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 141.016797 | 2222.914310 | 900.499132 |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 2363.931107 | 1041.515929 |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 1322.415178 |
| 9 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 |

3.

```
: caculate_Manhattan_cent_dest(c2_center,K)
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 0 | 0.0 | 15757.884176 | 20197.003176 | 9519.040843 | 5585.675176 | 3077.724176 | 1304.403401 | 466.237579 | 2362.883758 | 3341.586372 |
| 1 | NaN | 0.000000 | 4439.119000 | 6238.843333 | 10172.209000 | 12680.160000 | 14453.480775 | 15291.646598 | 13395.000418 | 12416.297804 |
| 2 | NaN | NaN | 0.000000 | 10677.962333 | 14611.328000 | 17119.279000 | 18892.599775 | 19730.765598 | 17834.119418 | 16855.416804 |
| 3 | NaN | NaN | NaN | 0.000000 | 3933.365667 | 6441.316667 | 8214.637442 | 9052.803264 | 7156.157085 | 6177.454471 |
| 4 | NaN | NaN | NaN | NaN | 0.000000 | 2507.951000 | 4281.271775 | 5119.437598 | 3222.791418 | 2244.088804 |
| 5 | NaN | NaN | NaN | NaN | NaN | 0.000000 | 1773.320775 | 2611.486598 | 714.840418 | 263.862196 |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 838.165823 | 1058.480357 | 2037.182971 |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 1896.646180 | 2875.348793 |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 978.702614 |
| 9 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 |

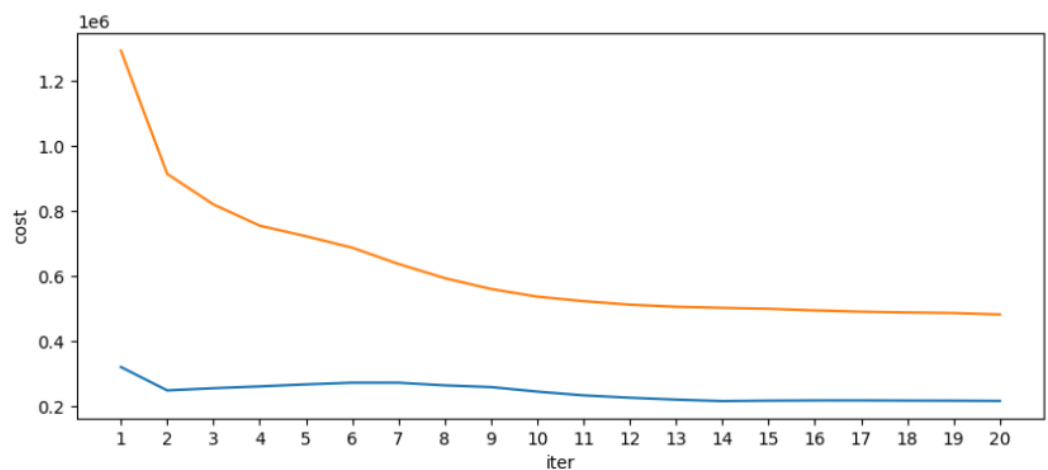
4.

II. question b

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

| | c1 | c2 |
|----|---------------|--------------|
| 0 | 320999.635000 | 1.293148e+06 |
| 1 | 249189.482721 | 9.144685e+05 |
| 2 | 255924.308876 | 8.206620e+05 |
| 3 | 261497.385403 | 7.554983e+05 |
| 4 | 267695.131222 | 7.230270e+05 |
| 5 | 272871.102442 | 6.878002e+05 |
| 6 | 272777.598086 | 6.377756e+05 |
| 7 | 264658.690328 | 5.942001e+05 |
| 8 | 259378.790623 | 5.607968e+05 |
| 9 | 245542.047041 | 5.373011e+05 |
| 10 | 233965.438696 | 5.234947e+05 |
| 11 | 226775.883719 | 5.126663e+05 |
| 12 | 220952.277462 | 5.062781e+05 |
| 13 | 216455.242024 | 5.028378e+05 |
| 14 | 217765.068398 | 5.000986e+05 |
| 15 | 218438.903704 | 4.950856e+05 |
| 16 | 218509.334248 | 4.911071e+05 |
| 17 | 217971.525654 | 4.888196e+05 |
| 18 | 217665.483348 | 4.870105e+05 |
| 19 | 216910.600134 | 4.826190e+05 |

1.



2.

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

1. 在曼哈頓距離下，其結果與歐基里德距離類似，

C1 這個沒有特別挑過起始點的群集，最後改善率只有 32.4%

而 C2 盡可能挑相鄰比較遠的點則有 62.6%的改善率
 這顯示了起始點對於 K-MEAN 演算法有非常大的影響
 若沒有優化過的話，這個演算法在歐基里德距離下可能會收斂在一個不是非常好的區間。

```
In [210]: caculate_percentage(c1_cost)
```

```
Out[210]: 32.42652748365107
```

```
In [211]: caculate_percentage(c2_cost)
```

```
Out[211]: 62.678745370752495
```

2.

C. The Euclidean and Manhattan Distances for all pairs of centroids, with 2 initialization strategies. (共有四個表格)

```
caculate_Euclidean_cent_dest(c1_center,K)
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-----|------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|-------------|
| 0 | 0.0 | 1063.57039 | 10261.505612 | 225.806397 | 392.119358 | 564.798196 | 496.025855 | 614.565839 | 2307.351485 | 408.616816 |
| 1 | NaN | 0.00000 | 9200.740779 | 1287.201104 | 1451.352090 | 1624.522097 | 1555.834936 | 1674.280532 | 1257.666717 | 658.516328 |
| 2 | NaN | NaN | 0.000000 | 10486.525230 | 10651.527955 | 10824.765427 | 10755.997278 | 10874.560374 | 7958.818624 | 9853.154102 |
| 3 | NaN | NaN | NaN | 0.000000 | 166.437220 | 339.008172 | 270.224831 | 388.790204 | 2532.955899 | 634.246175 |
| 4 | NaN | NaN | NaN | NaN | 0.000000 | 173.244351 | 104.488542 | 223.060368 | 2698.776367 | 800.241002 |
| 5 | NaN | NaN | NaN | NaN | NaN | 0.000000 | 68.820685 | 49.864818 | 2871.802772 | 973.218447 |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 118.649672 | 2803.008572 | 904.411441 |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 2921.601619 | 1023.013097 |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 1898.948384 |
| 9 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 |

1.

```
caculate_Euclidean_cent_dest(c2_center,K)
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-----|--------------|--------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 0 | 0.0 | 15752.182979 | 14104.908463 | 9037.368673 | 3452.221883 | 3979.561973 | 1291.949724 | 501.593106 | 2127.506082 | 2983.079053 |
| 1 | NaN | 0.000000 | 11524.505650 | 6743.884100 | 12429.412206 | 12013.553364 | 14499.915727 | 15255.688911 | 13675.574843 | 12788.368193 |
| 2 | NaN | NaN | 0.000000 | 9545.879403 | 11288.836228 | 10841.503110 | 13018.859563 | 13707.399513 | 12385.179985 | 11941.357947 |
| 3 | NaN | NaN | NaN | 0.000000 | 5685.612386 | 5286.011596 | 7771.311576 | 8537.833796 | 6941.936439 | 6059.969422 |
| 4 | NaN | NaN | NaN | NaN | 0.000000 | 716.222585 | 2160.300322 | 2953.744702 | 1331.347777 | 653.167392 |
| 5 | NaN | NaN | NaN | NaN | NaN | 0.000000 | 2704.044600 | 3487.424934 | 1886.984803 | 1229.458820 |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 794.283570 | 838.781261 | 1712.242433 |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 1627.317856 | 2481.938861 |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 893.055394 |
| 9 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 |

2.

caculate_Manhattan_cent_dest(c1_center,K)

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-----|-------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|--------------|
| 0 | 0.0 | 1328.002439 | 12274.290614 | 256.783868 | 437.525799 | 643.629193 | 564.365568 | 704.643698 | 2745.083772 | 480.191085 |
| 1 | NaN | 0.000000 | 10946.288175 | 1584.786307 | 1765.528238 | 1971.631632 | 1892.368007 | 2032.646138 | 1417.081333 | 847.811354 |
| 2 | NaN | NaN | 0.000000 | 12531.074482 | 12711.816413 | 12917.919807 | 12838.656182 | 12978.934313 | 9529.206842 | 11794.099529 |
| 3 | NaN | NaN | NaN | 0.000000 | 180.741931 | 386.845325 | 307.581700 | 447.859831 | 3001.867640 | 736.974953 |
| 4 | NaN | NaN | NaN | NaN | 0.000000 | 206.103393 | 126.839769 | 267.117899 | 3182.609571 | 917.716884 |
| 5 | NaN | NaN | NaN | NaN | NaN | 0.000000 | 79.263624 | 61.014506 | 3388.712965 | 1123.820278 |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 140.278130 | 3309.449340 | 1044.556653 |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 3449.727470 | 1184.834783 |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 2264.892687 |
| 9 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 |

3.

caculate_Manhattan_cent_dest(c2_center,K)

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 0 | 0.0 | 15749.521886 | 20188.640886 | 9510.678552 | 4203.232136 | 5386.699386 | 1571.295025 | 572.111257 | 2578.389226 | 3310.606227 |
| 1 | NaN | 0.000000 | 4439.119000 | 6238.843333 | 11546.289750 | 10362.822500 | 14178.226861 | 15177.410628 | 13171.132660 | 12438.915659 |
| 2 | NaN | NaN | 0.000000 | 10677.962333 | 15985.408750 | 14801.941500 | 18617.345861 | 19616.529628 | 17610.251660 | 16878.034659 |
| 3 | NaN | NaN | NaN | 0.000000 | 5307.446417 | 4123.979167 | 7939.383528 | 8938.567295 | 6932.289326 | 6200.072326 |
| 4 | NaN | NaN | NaN | NaN | 0.000000 | 1183.467250 | 2631.937111 | 3631.120878 | 1624.842910 | 892.625909 |
| 5 | NaN | NaN | NaN | NaN | NaN | 0.000000 | 3815.404361 | 4814.588128 | 2808.310160 | 2076.093159 |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 999.183767 | 1007.094201 | 1739.311202 |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 2006.277969 | 2738.494969 |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 | 732.217000 |
| 9 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.000000 |

4.