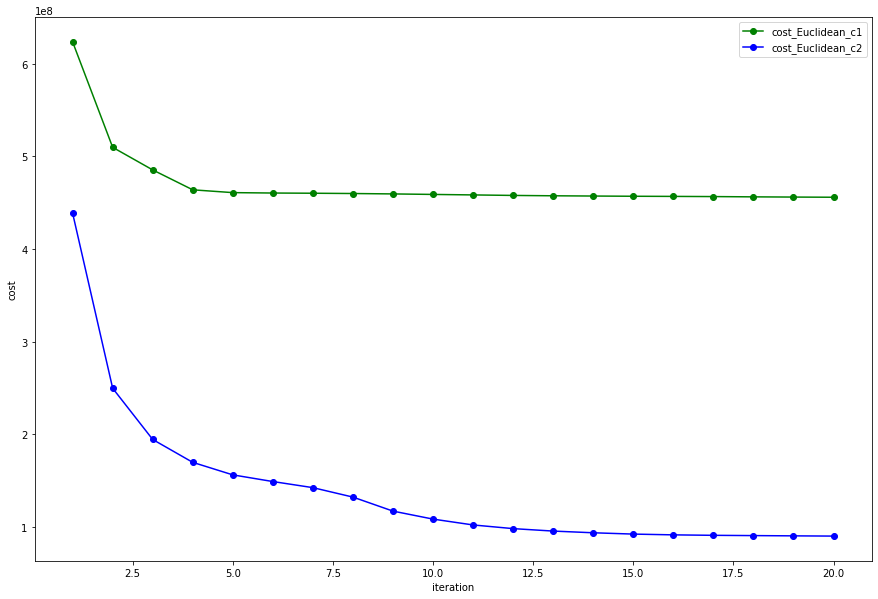
Kmeans Report

110062643 嚴翎愷

**Euclidean Distance**

1. A plot of cost vs. iteration for 2 initialization strategies(c1 and c2) for Euclidean distance
2. Percentage improvement values and your explanation for Euclidean distance

C1: 26.885383292518288 ; C2: 79.437750291599

**Reason**: 可以發現選取 C2 作為 initial centroid 比較適合。因為 C1 是取前 10 個點視為 centroid，這 10 個點可能原本就屬於同個 cluster，所以效果比較不好。但 C2 的取法所形成的 cluster 可以比較分散，故分群的效果比較好。

1. The Euclidean and Manhattan Distances for all pairs of centroids, with 2 initialization strategies:
   1. C1 with Euclidean Distance



* 1. C1 with Manhattan Distance



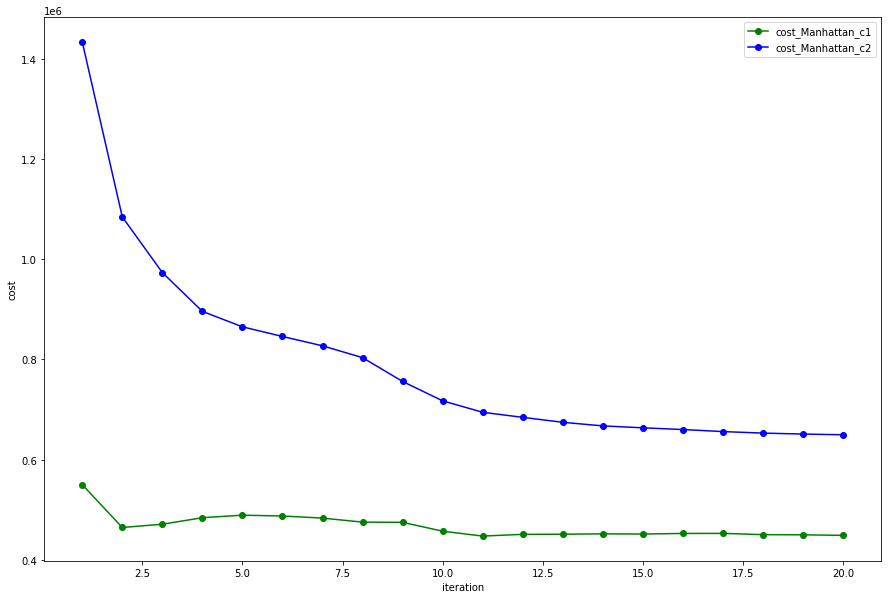
* 1. C2 with Euclidean Distance



* 1. C2 with Manhattan Distance



**Manhattan Distance**

1. A plot of cost vs. iteration for 2 initialization strategies (c1 and c2) for Manhattan distance****
2. Percentage improvement values and your explanation for Euclidean distance

C1: 18.393840107083165; C2: 54.68569434813371

Reason: 由上圖發現，在 Manhattan Distance 下，C2 的 cost 明顯大於 C1，而在 improvement 上面， 相比 Euclidean Distance，在 Manhattan Distance 上選取C2 的改善幅度沒有 Euclidean 大。

1. The Euclidean and Manhattan Distances for all pairs of centroids, with 2 initialization strategies:
2. C1 with Euclidean Distance
3. C1 with Manhattan Distance
4. C2 with Euclidean Distance
5. C2 with Manhattan Distance