Programming Assignment 2 Deadline: 1/2

**112753207 資碩計一 張詠軒**

此次作業主要目的在讓同學學習運用 Python 由超級市場的顧客資料(marketData) 中，瞭解客戶，例如可以輕鬆融合目標客群，以便賦予營銷團隊意義並相應地規劃戰略。

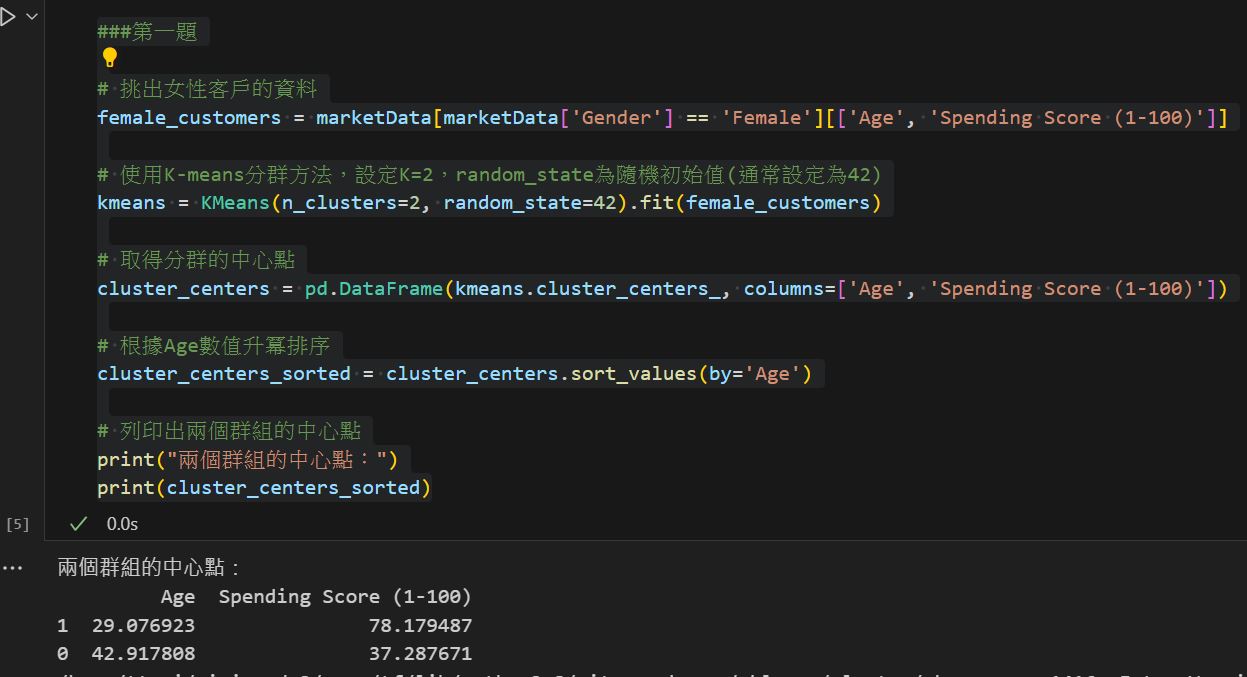
顧客資料包含：CustomerID、Age、Gender、Annual Income、Spending Score，其中 Spending Score 是根據顧客行為和購買物品所設定。

* 作業給定的 marketData 已經附在 Moodle 平台上。
* 作業每人繳交一份報告，檔案類型以 **pdf** 為限。

上傳檔名格式為學號**\_P2**，例如：**112753XXX\_P2.pdf**

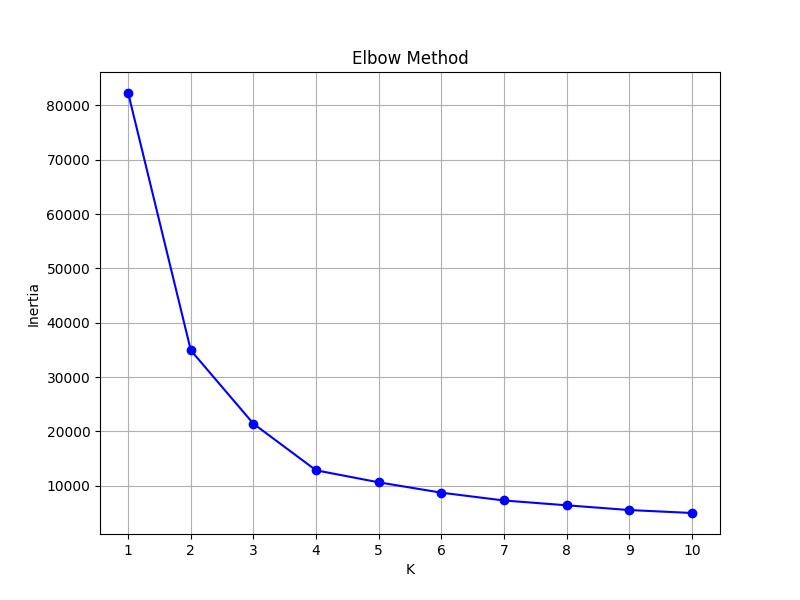
* 此次作業可以使用現有套件執行運算。

1. 從 marketData 的資料中，取女性客戶的兩個欄位: **Age**、**Spending Score** 進行客戶分群，請使用 K-means 分群法，當 K = 2 時，請列出每一群的中心點，例如C1中心點: Age=33.3、Spending Score=87.1。2 個中心點的列表請根據中心點的 Age 數值升冪排序。



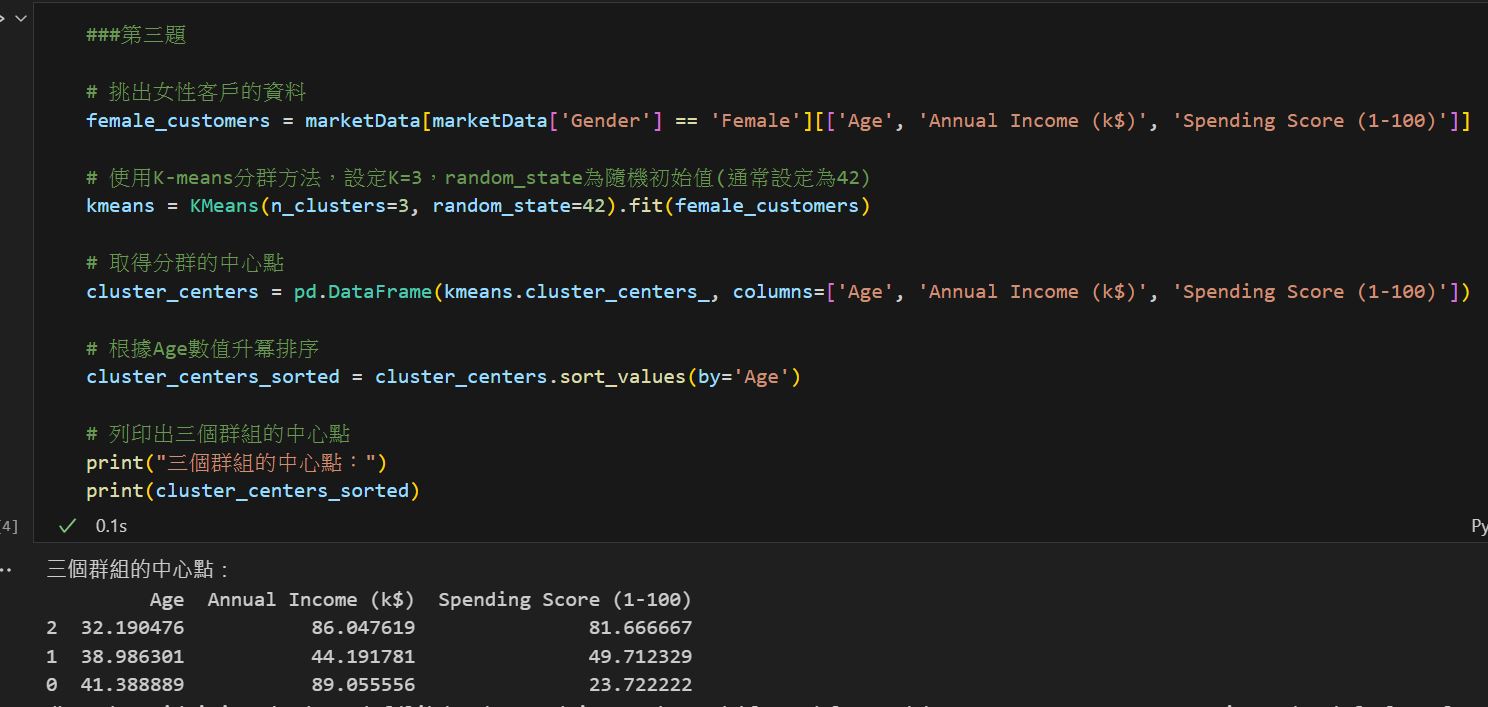
第一群組的年齡中心點約為29歲，平均消費分數為78.18；而第二群組的年齡中心點約為42.92歲，平均消費分數則為37.29。這可能代表著年輕女性顧客更傾向於有較高的消費分數，而年長女性則可能有較低的消費傾向。

1. 承第1題，請利用 Elbow 方法找出 K 應該設置多少？



K值應該設置為4。

1. 從 marketData 的資料中，取所有客戶的三個欄位: **Age**、**Annual Income**、**Spending Score** 進行客戶分群，請使用 K-means 分群法， 當 K = 3 時，請列出每一群的中心點，例如C1中心點: Age=33.3、Annual Income=87.1、Spending Score=88.1。3 個中心點的列表請根據中心點的 Age 數值升冪排序。

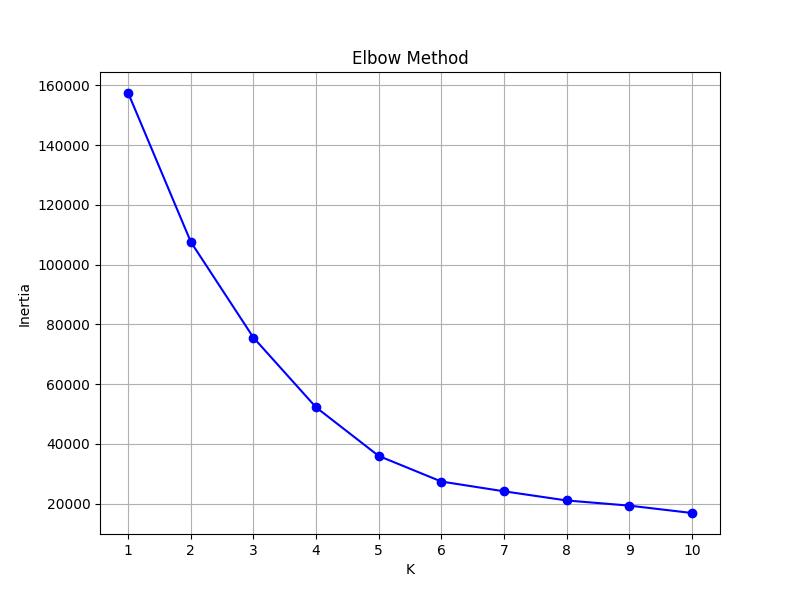


第一群組的年齡中心點約為32.19歲，年收入約為86.05千元，消費分數約為81.67。

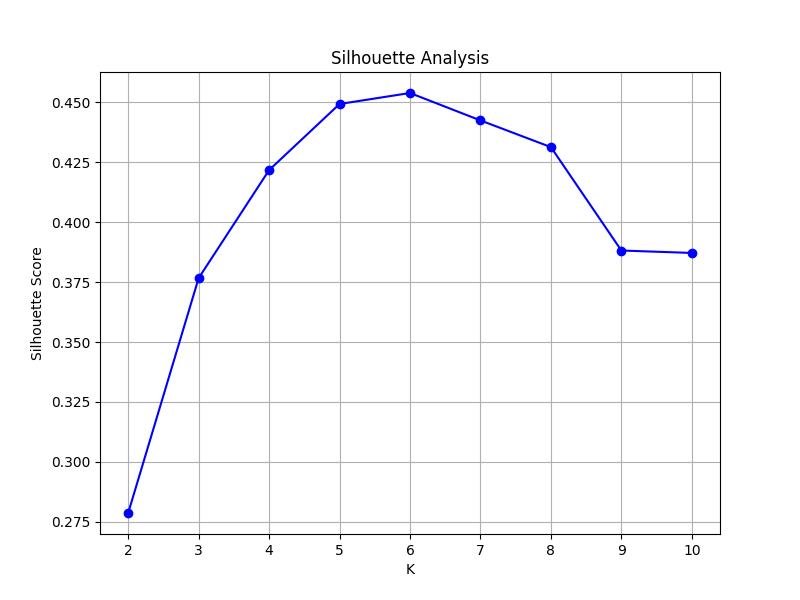
第二群組的年齡中心點約為38.99歲，年收入約為44.19千元，消費分數約為49.71。

第三群組的年齡中心點約為41.39歲，年收入約為89.06千元，消費分數約為23.72。

1. 承第3題，請利用 Elbow 方法找出 K 應該設置多少？



K值應該設置為6。



使用Silhouette分析方法去做驗證，得出K值應該設置為6。

1. 承第4題的K值設置，假設現在有一個行銷活動，請問你要怎麼透過Kmeans分群結果進行篩選，選擇一群目標客群，請列出此群的中心點，並解釋你的理由。

選擇目標客群時，我們可以考慮不同群組的特徵，以及該群組是否符合行銷活動的目標和需求。假設我們想要找出一個具有潛力的目標客群。

在K=3的情況下，我們有三個群組，每個群組有不同的特徵。假設：

第一群組特徵：平均年齡較年輕，年收入和消費分數都比較高。

第二群組特徵：年齡稍大，年收入中等，消費分數在中等偏下。

第三群組特徵：年齡較大，年收入較高，但消費分數較低。

如果希望找到一個潛在的高價值客戶群，可能會考慮第一群組，因為他們年輕且有較高的消費傾向和收入。因此，我們可以選擇第一群組作為潛在的目標客群。

我們可以列出第一群組的中心點，這些中心點代表了該群組在特徵上的平均值：

年齡約為32.19歲，年收入約為86.05千元，消費分數約為81.67。

這個群組的客戶可能是對於高價值產品或高消費活動感興趣的客戶，因為他們有較高的消費傾向和收入，並且相對年輕。所以，針對這個群組進行行銷活動可能會更有潛力帶來較高的銷售和回饋。