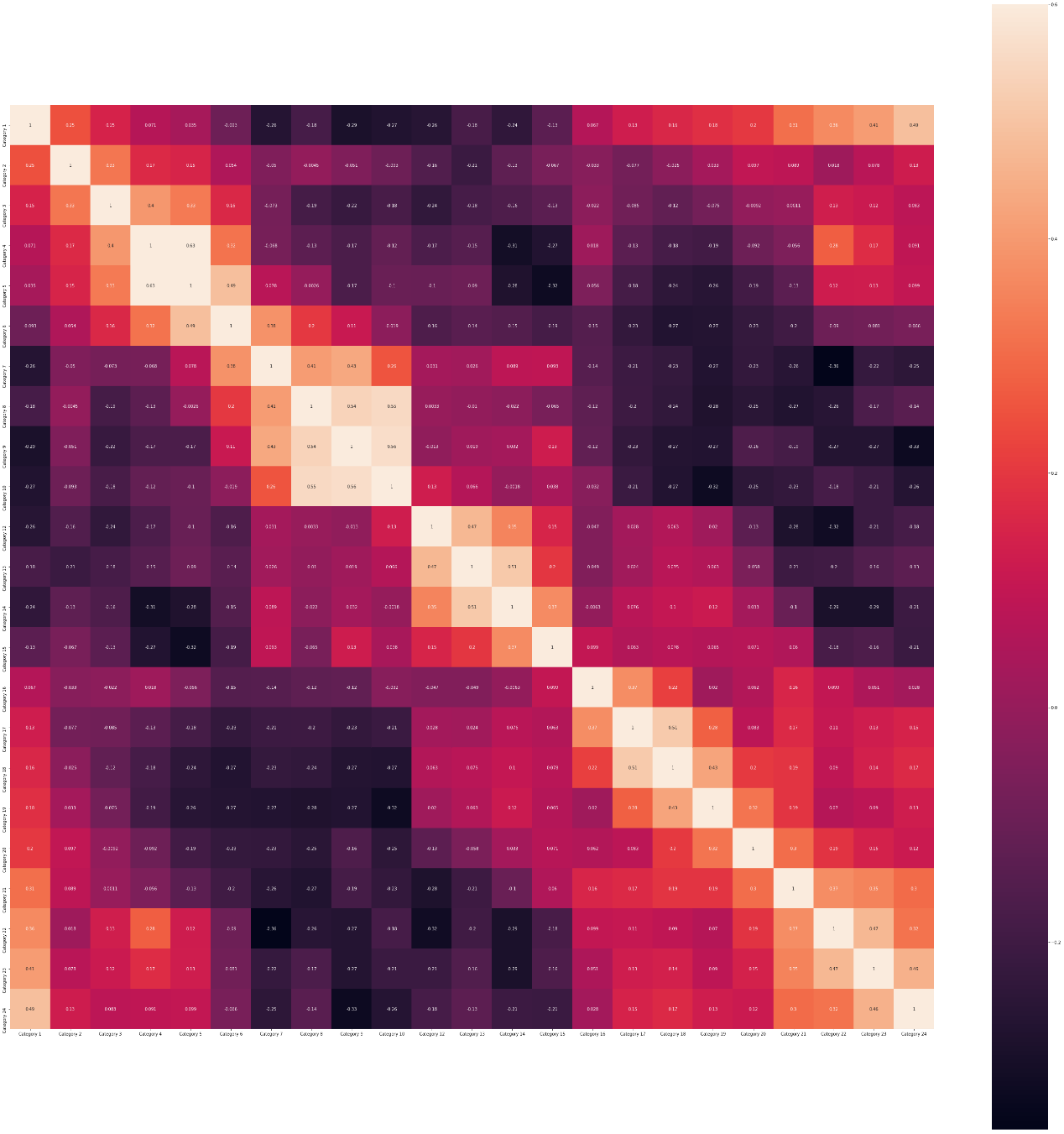
**Data Analytics Homework4**

**張庭維 F74046488**

這次作業我所採用的資料為google\_review\_rates的資料集，裡面描述了試用者在google地標上的評分平均，而剛開始時我用heatmap來觀察有哪幾個變數是比較有關係的得到的heatmap如下

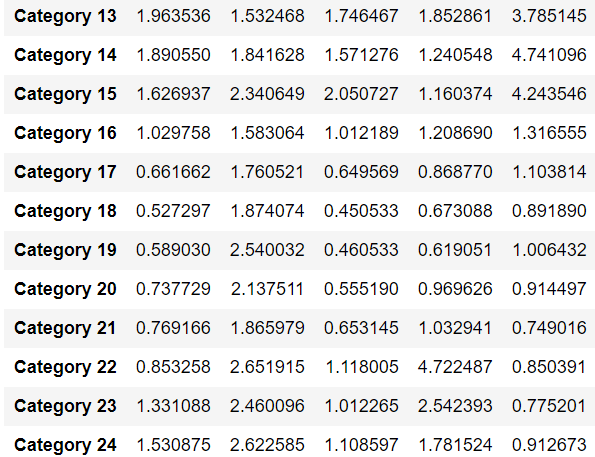
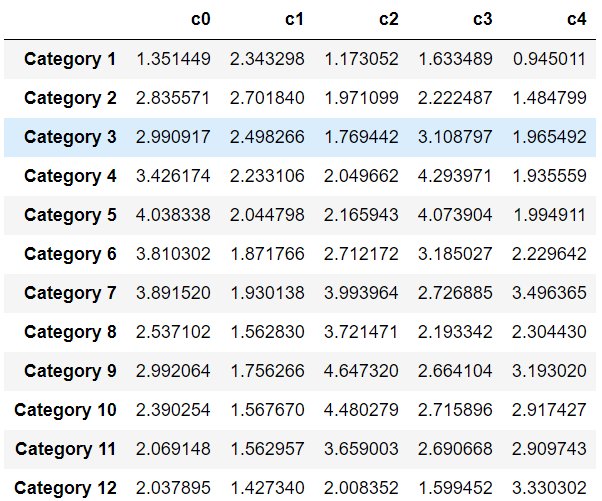
其中比較有關係的分別為:

1. 公園與電影院
2. 動物園與餐廳
3. 動物園與酒吧
4. Pizza店與旅館
5. 泳池與跳舞俱樂部

分別的解釋如下:

1. 公園與電影院的評分有關我認為因為當人們在餐廳吃飽後就會在附近的公園散步，而餐廳會影響散步時的心情，且若在餐廳外等待要進入用餐時也會再外面的公園散步，因此公園與餐廳的評價是有關係的。
2. 動物園與餐廳我認為跟公園與餐廳的關係是類似的，兩個關係是相輔相臣的
3. 動物園與酒吧我認為這是個比較偏向旅遊的選擇，當我們下午去動物園後，晚上便會去酒吧放鬆今天一整天旅遊的疲憊，因此這個評價事有關連的。
4. Pizza店與旅館我認為關係很大，就表面上來看去住旅館的人若不想出門就有很大的機會叫外送pizza，而飯店的距離遠近交通便利都會互相影響到pizza店的評價。
5. 泳池與跳舞俱樂部我認為是他們會開在附近，又或者會是同一家公司，因此品質便會互相影響。

接著我將使用者分群用以區分使用者的類型種類，得到的結果如下



我們可以觀察第一群人對於戲院、博物館以及購物中心的評價偏高，對於游泳池以及運動場的評價偏低，我認為第一群人可能比較像是婦女的類型，對於逛街參觀展覽比較有興趣，但是對於運動提不起勁。

而第二類人對於景點的評價基本上都不高，而其中比較高的項目偏向風景類，像是花園，觀景台，標誌景物等，這些我推測比較像是老年人喜歡的項目，且他們對於很多的東西偏向嚴格，所以給的評價都不太高。

第三類人我認為是中年男子，他們對於餐廳以及酒吧的評價都很高，且他們不喜歡去美容店，從這些特徵很明顯可以推論出這應該是中年男子的族群。

第四類人我認為並沒有特定的年齡層族群，而是一群喜歡風景戶外的使用者，他們喜歡公園喜歡風景，且不喜歡在室內活動。

第五類我認為青少年，尤其是他們對於果汁酒吧的評價極高，且他們也會去參觀藝廊，且他們並不喜歡靜態的行程，像是看風景去公園之類的因此可以推論應該為青少年使用者。

方法:

我先採用了heatmap這些資料有甚麼特性，後來再使用kmeans的clustering方法來切分種類，而種類的話我發現如果切太細我們會沒有辦法去解釋每一個族群，因為切分態系會造成許多相似的資料會增加資料解釋的難度。

改善:

我認為還有很多改善的空間，像是我們可以試著透過每個人的評分來建立關聯式規則(association rule)來看通常人們的行程會是怎麼走的，我會假設如果他沒有評分的話代表他沒有去過，而這樣的話我們可以藉由哪些景點為一個族群來推論走這些景點的族群為那些，且他們的滿意度如何。