執行平台：Colab
執行工具：Pyspark
執行資料：2019年yellow taxi的數據、2016年6月yellow taxi數據
Q1: What are the most pickups and drop offs region?

2019年yellow taxi數據
利用平行處理方式個別計算每個地區的上下車次數

上車地區：237
下車地區：236

上車地區	上車次數(由高至低)
237	3641682
161	3450649
236	3291351
162	3046788
186	3027440

下車地區	下車次數(由高至低)
236	3429838
161	3261232
237	3256021
170	2643630
230	2596508

2016年6月yellow taxi數據
利用KMeans作為clustering之演算法 (取15個中心點，並取點數最高前五個)，計算時扣除掉在海上的點或是不在紐約中的點，並且因為儲存的資料內容是經緯度，所以計算距離時再換算成經緯度的真實距離，以判斷最接近的中心點位置，並將中心點對應回網站提供的區域圖，換成中心點對應的地區代號。

上車地區：162
下車地區：79

上車地區	上車次數(由高至低)
162	1294220
140	1139854
13	1103497
236	1082312

97	972711

下車地區	下車次數(由高至低)
79	1558172
141	1462727
237	1446984
68	1296210
263	1213418

Q2: When are the peak hours and off-peak hours for taking taxi? hint: You can count the number of pickups in different hours of day.

2019年yellow taxi數據
利用平行處理方式個別計算每個時間點的上下車次數

尖峰時段：18點
離峰時段：18點

上車尖峰時段(小時)	上車次數(由高至低)
18	5539365
19	5233239
17	5025338
20	4776659
15	4733331

下車尖峰時段(小時)	下車次數(由高至低)
18	5602221
19	5468455
20	4837712
17	4832199

21	4739179

Q3: What are the differences between short and long distance trips of taking a taxi?
hint: First, you should define what short and long distance trips are. You may observe the results of Q1 and Q2

短距離與長距離的差別在於上下車的地點不同，長距離指的是上下車在不同地區，而短距離指的是上下車在相同地區。也就是說，當長距離旅乘次數增加時，容易造成上車次數最多的地區與下車次數最多的地區並不一定相符，而且搭車尖峰時段可能會略微不同，反之，若搭乘計程車者幾乎都是行駛短程距離，則上下車最高次數的地區及交通的尖峰時段都會趨近相同。