

國立雲林科技大學工業工程與管理系碩士在職專班

碩士論文

Department of Industrial Engineering and Management

National Yunlin University of Science & Technology

Master Thesis

以純 Youbike 資料識別 Covid-19

對臺北市區活動的影響

Using Youbike Data to Identify the
Impact of Covid-19 on Activities in Downtown Taipei

陳詳元

Hsiang-Yuan Chen

指導教授：陳奕中博士

Advisor: Yi-Chung Chen, Ph.D.

中華民國 111 年 2 月

January 2022

摘要

從 2019 年末 Covid-19 疫情發生至今，已對世界各國的生活產生重大的影響，因其傳染途徑主要為透過飛沫傳染，而各國政府之主要防疫措施除要求配戴口罩外，就是減少民眾群聚之機會，以臺灣為例，企業推行居家上班、學校推行遠距教學、民眾減少旅遊、減少店內用餐改以外帶..等防疫措施實行，從上述情形我們認知 Covid-19 已對民眾生活的有所改變，而針對此問題，我們選擇與民眾生活最貼近的公共交通工具 YouBike 歷史租賃資料進行探討，不採用主觀的認知或專業知識理論來進行分析資料，而是以純資料的觀點來做分析，我們將資料進行清洗降維後，運用群集分析得出各筆數據的分群結果，再與 Covid-19 有關之特徵值資料結合後，使用隨機森林找出特徵因子的重要度，並嘗試在不加入特定知識下的情況下，進行 Covid-19 對民眾生活的影響的探討，而其實驗流程證實了我們方法的可行性。

關鍵字：Covid-19、YouBike、集群分析、隨機森林

Abstract

Since its outbreak in late 2019, Covid-19 has had a significant impact worldwide. As the main route of transmission is through breathing in contaminated droplets in the air, anti-epidemic measures include wearing masks and reducing contact, especially among large groups of people. In Taiwan for example, many enterprises have implemented home-based work, schools have implemented distance teaching, tourism has been reduced, and many restaurants are offering take-out over sit-down meals. These measures have profoundly transformed the way people live. In this study, we analyze the historical records of the public transportation system YouBike. This analysis is conducted without subjective or professional knowledge, but rather from an informative perspective. After cleaning and reducing the dimensionality of the data, we applied cluster analysis and combined the eigenvalue data of Covid-19 to find the importance of the eigenfactors using the random forest approach. Our research aim was to explore the impact of Covid-19 on people's lives without the filter of specific knowledge. Our experimental results confirm the feasibility of our method.

Keyword : *Covid-19 、YouBike 、cluster analysis 、random forest*

誌謝

首先感謝指導教授陳奕中老師對於論文上的指導與細心教學，同時每週也都撥出時間與我們一同討論，提供不同的思考方向與執行方法，得以讓完全沒有大數據分析基礎的我，能逐漸一步一步完成論文；在 Python 與 Matlab 的軟體操作及程式撰寫部份，非常感謝佳昇、承安以及冠博三位學弟的協助教導，讓我可以順利完成資料分析結果的取得，使論文有突破性的進展。

在雲科大工管系就讀的時間裡，除感謝工管系提供良好的學習環境外；更開心能遇到工工班同學們，在這段時間大家一同討論課業、相互加油、得以堅持下來完成學業；尤其是博揆、玲瑜、聿楹這三位同學，一同在論文撰寫的道路上，互相幫忙、互相督促，有你們的相挺，我才有辦法堅持到最後。

能有動力再重回校園就讀研究所，非常感謝前公司的好同事與好朋友們，你們優秀的工作能力與學識，讓我發現了自己的不足，產生想追上你們的念頭，進而決定進修研究所，其中非常感謝資穎的鼓勵，讓總是說自己學歷不佳的我完成心中的缺憾。

最後謝謝家人的支持與信任，在這求學的期間給予支持與鼓勵，讓總是三分鐘熱度的我能專心與堅持的完成學業，謝謝你們。

陳詳元謹致

中華民國 111 年 2 月

目錄

摘要	i
Abstract.....	ii
誌謝	iii
目錄	iv
表目錄	vii
圖目錄	ix
第一章 緒論	1
1.1 研究背景與動機	1
1.2 研究目的	2
1.3 研究流程	3
第二章 文獻探討	5
2.1 YouBike 使用型態研究	5
2.2 YouBike 轉乘行為特性分析	6
2.3 公用自行車租賃影響因素	6
2.4 集群分析	6
2.5 隨機森林	8
第三章 資料集介紹	10
3.1 資料範圍	10
3.2 YouBike 開放資料平台資料格式	10
3.3 各 YouBike 租用站原始資料範例	11
3.4 Covid-19 原始資料範例	12
3.5 降雨量資料	12
3.6 Google trend 資料	13
3.7 資料彙整說明	13
第四章 研究方法	16
4.1 研究流程	16

4.2	資料的前處理	16
4.3	特徵值標準化	17
4.4	k-means 分群後進行資料標籤	17
4.4.1	集群分析 K-means	18
4.4.2	集群數量 k 值的選定	20
4.4.3	分群資料筆數平衡與重新分群合併縮減	20
4.5	訓練隨機森林找出特徵因子重要度	22
4.6	特徵因子重要度探討	25
第五章	研究結果	26
5.1	資料集介紹	26
5.2	實驗環境設定	28
5.2.1	特徵因子選擇	28
5.2.2	實驗使用之軟體套件	30
5.3	捷運類 YouBike 租用站影響因子重要度探討	30
5.3.1	捷運類 YouBike 租用站資料分群	30
5.3.2	捷運類 YouBike 租用站資料平衡與分群合併縮減	31
5.3.3	捷運類 YouBike 租用站特徵值資料進行隨機森林訓練與因子重要度計算 31	
5.3.4	捷運類 YouBike 租用站影響因子重要度探討	34
5.4	公園類 YouBike 租用站影響因子重要度探討	37
5.4.1	公園類 YouBike 租用站資料分群	37
5.4.2	公園類 YouBike 租用站資料平衡與分群合併縮減	38
5.4.3	公園類 YouBike 或用站特徵值資料進行隨機森林訓練與因子重要度計算 38	
5.4.4	公園類 YouBike 租用站影響因子重要度探討	41
5.5	學校類 YouBike 租用站影響因子重要度探討	44
5.5.1	學校類 YouBike 租用站資料分群	44
5.5.2	學校類 YouBike 租用站資料平衡與分群合併縮減	45

5.5.3	學校類 Youbike 租用站特徵值資料進行隨機森林訓練與因子重要度計算	45
5.5.4	學校類 Youbike 租用站影響因子重要度探討.....	47
第六章	結論	50
6.1	研究結論	50
6.2	研究建議	51
	參考文獻	52



表目錄

表 3-1 YouBike 資料欄位說明.....	11
表 3-2 YouBike 租用站原始資料範例.....	12
表 3-3 Covid-19 原始資料範例.....	12
表 3-4 降雨量資料範例.....	13
表 3-5 Google trend 資料範例.....	13
表 3-6 各 YouBike 租用站每日各時區租用次數.....	14
表 3-7 各 YouBike 租用站各時區月平均租用次數.....	14
表 3-8 各 YouBike 租用站各時區月平均租用次數.....	14
表 3-9 各 YouBike 租用站各時區租用次數縮減比例.....	14
表 3-10 特徵值組合.....	15
表 4-1 捷運類租用站 1 月份分群結果.....	21
表 4-2 捷運類租用站 1 月分群合併縮減結果.....	22
表 5-1 去除日租用人次與月平均租用人次未達 10 人次之資料.....	27
表 5-2 各分類站總租用次數.....	28
表 5-3 捷運類租用站分群結果_輪廓系數.....	30
表 5-4 捷運類 YouBike 租用站 1 月份分群結果.....	31
表 5-5 捷運類 YouBike 租用站 1-7 月分群結果.....	31
表 5-6 捷運類 YouBike 租用站 1-7 月分群合併縮減結果.....	31
表 5-7 捷運類 YouBike 租用站 1 月份特徵值與分群結果結合.....	32
表 5-8 捷運類 YouBike 租用站 1 月份資料混淆矩陣分類結果.....	32
表 5-9 捷運類 YouBike 租用站 1 月份因子重要度.....	32
表 5-10 捷運類 YouBike 租用站資料 1-7 月混淆矩陣分類結果.....	33
表 5-11 捷運類 YouBike 租用站 1-7 月因子重要度.....	33
表 5-12 捷運類 YouBike 租用站 1-7 月各因子重要度排序.....	34
表 5-13 公園類租用站分群結果_輪廓系數.....	37
表 5-14 公園類 YouBike 租用站 1 月份分群結果.....	38

表 5-15 公園類 YouBike 租用站 1-7 月分群結果	38
表 5-6 公園類 YouBike 租用站 1-7 月分群合併縮減結果.....	38
表 5-17 公園類 YouBike 租用站 1 月份特徵值與分群結果結合	39
表 5-18 公園類 YouBike 租用站 1 月份資料混淆矩陣分類結果	39
表 5-19 公園類 YouBike 租用站 1 月份因子重要度	40
表 5-20 公園類 YouBike 租用站資料 1-7 月混淆矩陣分類結果.....	40
表 5-21 公園類 YouBike 租用站 1-7 月因子重要度	41
表 5-22 公園類 YouBike 租用站 1-7 月各因子重要度排序	42
表 5-23 公園類租用站分群結果_輪廓系數	44
表 5-24 學校類 YouBike 租用站 1 月份分群結果	44
表 5-25 學校類 YouBike 租用站 1-7 月分群結果	45
表 5-26 學校類 YouBike 租用站 1-7 月分群合併縮減結果.....	45
表 5-27 學校類 YouBike 租用站 1 月份特徵值與分群結果結合	46
表 5-28 學校類 YouBike 租用站 1 月份資料混淆矩陣分類結果	46
表 5-29 學校類 YouBike 租用站 1 月份因子重要度	46
表 5-30 學校類 YouBike 租用站資料 1-7 月混淆矩陣分類結果.....	47
表 5-21 學校類 YouBike 租用站 1-7 月因子重要度	47
表 5-32 學校類 YouBike 租用站 1-7 月各因子重要度排序	48

圖目錄

圖 1-1 研究目的步驟.....	3
圖 1-2 本研究流程圖.....	4
圖 4-1 研究流程圖.....	16
圖 4-2 k-means 示意圖-1	19
圖 4-3 k-means 示意圖-2	19
圖 4-4k-means 示意圖-3	19
圖 4-5k-means 示意圖-4	19
圖 4-6 k-means 示意圖-5	20
圖 4-7 捷運類租用站分群預測結果.....	22
圖 5-1 捷運類 YouBike 租用站各月份特徵因子直方圖.....	35
圖 5-2 1-7 月下雨天數與日平均降雨量.....	36
圖 5-3 1-7 月份確診人數變化圖.....	37
圖 5-4 公園類 YouBike 租用站各月份特徵因子直方圖.....	42
圖 5-5 學校類 YouBike 租用站各月份特徵因子直方圖.....	48

第一章 緒論

1.1 研究背景與動機

Covid-19於2019年末在中華人民共和國湖北省武漢市首次被發現，隨後在2020年全面擴散至世界各國，演變成一場全球淪陷的傳染性疾病，其感染人數遠超過2002年發生的SARS，自此它影響絕大部份人類的生活方式，以臺灣為例，部份企業讓員工改以居家上班或辦公室人員分流、學校推行遠距教學、旅遊景點進行人數管制..等措施，皆是為了減少人群接觸以避免疫情擴大，基於上述情況，我們認知Covid-19的出現造成民眾生活習性上的改變，而實際上是否有影響，則需透過真實的數據資料來求證。

食衣住行育樂等六個項目幾乎囊括民眾生活的全部，其中的「行」，交通擔任促進整個社會運作的重要角色，也是民眾生活的核心工具，以臺北市為例，就民眾生活機能來說，公共自行車YouBike是取得最便利以及使用成本最低的大眾交通工具，且不受時間限制24小時皆可借還車。臺北市YouBike租車站截至2021年12月31日止，YouBike1.0共設有399個站點，其設置的區域遍及捷運站、學校、醫院、旅遊景點、公園、住宅區、運動場所、交通運輸站等...，可說是涵蓋整個臺北市的生活圈，就使用面來說，YouBike已成為民眾每日交通運輸所需的第一哩與最後一哩之交通工具，可說是與臺北市民眾的通勤生活畫上等號。

以往的研究多為使用公共自行車取得方便且貼近生活區域的特性來探討民眾的生活習性，例如翁雅君(2015)透過時間序列方法分析臺北市信義區 YouBike 使

用狀況，曾怡之(2019)以租用時間探討公共自行車使用者移動的時空特徵，鄭群彥(2014)則是進行臺北公共自行車租賃系統使用型態之分析等...。上述研究的結論皆是宣稱他們的方法確實對於區域或是全局的自行車使用特性分析有效，然而就我們的觀點來看，上述研究仍有改進空間。公共自行車的租賃紀錄屬於巨量資料範疇，應以資料導向進行數據的分析與解讀，而這些研究者多是基於某個特定的知識或是依據自身的專業先行解讀數據後再進行分析，研究本身雖然不會產生太大的偏差，但是以大數據分析的角度來看，可能會存在遺漏一些重要資訊的情形。

1.2 研究目的

針對上述問題的探討，本研究提出以純資料的觀點來看，在不使用任何知識下透過分析 YouBike 租賃資料，再結合 Covid-19 資料來探討疫情期間對民眾的生活習性是否有受影響，我們研究目的之做法步驟如圖 1-1 所示，步驟 1 為資料前處理，步驟 2 為 k-means 分群，步驟 3 為使用隨機森林找出特徵因子重要度，步驟 4 為特徵因子重要度探討；首先我們的方法是將租賃站一天 24 小時的租賃資料，劃分為八個時段，以此當作一筆資料，舉例，如我們有 399 個租賃站及 7 個月的租賃歷史資料，就會產生 8 萬 3790 筆資料的資料集，並依據此資料集進行資料清洗、降維，再投入集群分析，接續將分群結果結合 Covid-19 相關的特徵因子後再投入隨機森林之中，接著依據隨機森林得出之特徵因子數據與搭配質化資料針對不同生活圈之租賃站進行討論，例如每個特徵因子在每個月份對於相同類別租賃站的租用量影響程度排名是否相同，綜合以上，我們得以藉由 YouBike 的租賃資

料探討 Covid-19 疫情對於民眾生活習性的影響，爾後可做為當傳染類疫情再度爆發時的商業策略參考依據。

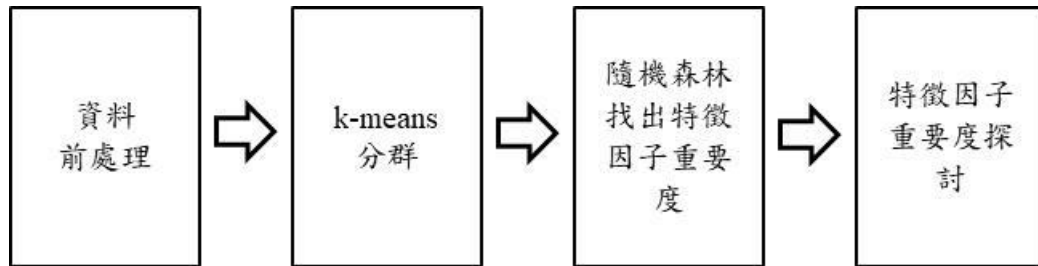


圖 1-1 研究目的步驟

1.3 研究流程

本研究架構如圖 1-2 所示共有八個環節。(1)確認研究主題。(2)文獻探討：對相關文獻進行研討。(3)資料蒐集：資料取得及清洗。(4)特徵值標準化：針對特徵值數據進行標準化處理。(5)集群分析與資料標籤：將各筆數據分群結果作為資料標籤。(6)隨機森林訓練與特徵因子重要度：經由隨機森林訓練找出特徵因子重要度。(7)研究結果：依特徵因子重要度探討對各類別 YouBike 租用站的影響。(8)結論與建議：總結本研究成果及後續研究建議。

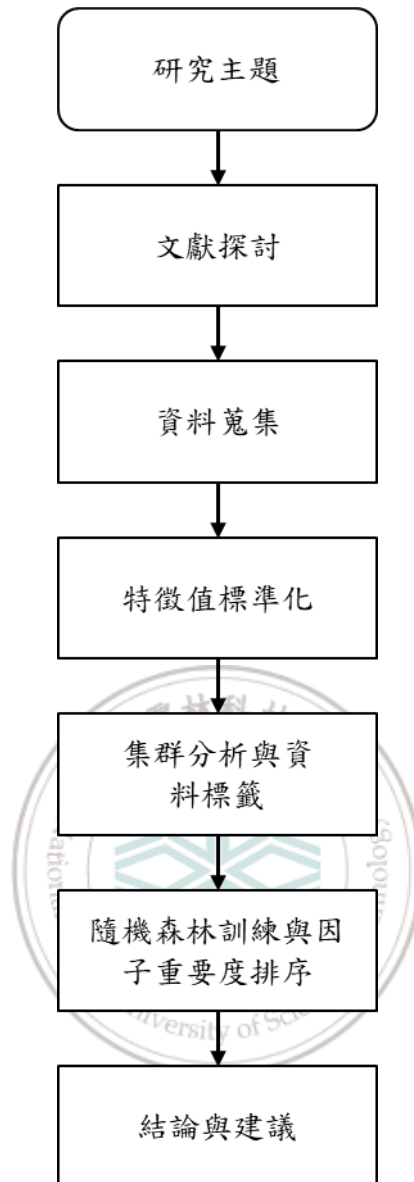


圖 1-2 本研究流程圖

第二章 文獻探討

本章節介紹與本研究相關之文獻，包含公共自行車轉乘分析與特性、集群分析及隨機森林方法。

2.1 YouBike 使用型態研究

臺北市 YouBike 系統於 2009/3/11 開始試營運，市政府也將 YouBike 定義成公共運輸系統中第一哩以及最後一哩之綠色交通工具，余書玫(2009)使用問卷調查民眾使用公共自行車之目的發現，以休閒、遊憩目的最多，其次則是通勤使用，白詩榮(2012)的問卷調查中還發現，少部分則是作為購物與運動健身。此外，黃仁皇(2009)經由問卷調查得知，公共自行車使用者主要使用時段為上午 06:00-08:59、下午 17:00-19:59 以及晚上 20:00-22:59，多為 30 分鐘內的短程騎乘，且每日使用次數約 1-2 次，劉仁龍、鄭麗淑(2017)報告統計 30 分鐘內短程騎乘者比率為 77.5%，平日租借高峰為上午 7-9 時、下午 17-19 時，平日次高峰為中午 12-13 時、夜晚 21 時，假日高峰則落在 12-18 時，李俊毅(2019)提出 Youbike 租借行為受平假日、站點天候、地理條件、租賃站規模等因素影響，上述研究指出站點間使用特性不同，以及使用目的不同，由此可知 YouBike 使用族群多為短時間、距離的使用，並作為最後一哩路的通勤工具，可藉此分析都市內各區生活習性的變化。

2.2 YouBike 轉乘行為特性分析

葉耀墩、林昭賢、劉心荷(2015)針對 YouBike 騎乘用途之特性進行問卷調查，統計結果顯示民眾對於 YouBike 的使用上有幾點特性：於平日使用 YouBike 者佔 49%、假日使用 YouBike 者佔 39%；在平日的使用上，主要以 YouBike 做為通勤用途佔 37%，假日的使用則做為休閒娛樂為主佔 37%；使用 YouBike 後轉乘的運具，捷運佔 55%為最多，其次為步行佔 26%。上述調查結果顯示大部分 YouBike 的使用者皆是把 YouBike 當做來往捷運站最後一哩路的交通工具。

2.3 公用自行車租賃影響因素

李俊毅(2019)利用迴歸分析及皮爾森相關係數探討臺北市公共自行車 2016 年 11 月至 2017 年 10 月之間的租借量與天候之間的關係，發現雨量與租借量顯著負相關。黃晏珊(2015)使用 2014 年 10 月至 2015 年 3 月份之歷史資料分析發現，位處捷運站周邊之租賃站平日因通勤族需求較常發生缺車狀況；假日則是遊憩需求導致鄰近休閒觀光租賃站較容易出現缺車狀況，車位及車輛使用不均的情況就需要透過資源調度來解決，這方面就衍生出許多需求預測的研究。

2.4 集群分析

集群分析法是資料探勘領域中非常重要的使用技術，此技術的目的主要是將大量的資料依據其資料的相似度或相異度區分為若干個集群，而所有集群皆具備有「相同集群內之資料相似性高，而不同集群間之資料差異性高」的特性，意指性質越接近的，越會相聚在一起，也就是「集群」的概念。

集群分析的方法主要可分為階層式集群方法(Hierarchical cluster methods)和非階層式集群方法(Non-Hierarchical cluster methods)兩大類。

Ward, J. H. (1963)提出之華德法為階層式集群法，其方法先將每一樣本資料視為一集群，再依序將集群合併，合併順序的安排是基於方差分析概念，以離差平方和為相似性測度。同集群樣本之間的離差平方和最小，集群與集群間的離差平方和最大。J. B. MacQueen (1967)提出的K-means為非階層式集群方法，事先設定數目為 k ，將各事物點分割成 k 個原始集群後，計算某一事物點到各集群重心的歐幾里得距離，接著分派此事物點至距離最近的集群，並重新計算各集群的重心，進行重複計算及分派，直到各事物點都不必重新分派到其他集群為止。Punj、Stewart(1983)歸納階層式集群法與非階層式集群法兩種方法之優劣，得出對 K-平均數法而言，若以隨機方法選取起始點，其集群效果較差，但若以集群之重心或中位數為起始點，則其集群效果要比華德法與平均連鎖法好。

集群分析被廣泛使用在不同領域，吳伯書(2019)使用集群分析結合新北市警局提供之資料，利用機器學習的方式，從中找出一些影響少年犯罪的關鍵成因。蔡儀君(2005)對樣本基本資料結構與特性進行描述，接著將智慧型手機使用的相關題項找出，並進行因素分析找出因素構面作為分群變數，藉由兩階段分群法進行分群，探討其各群間相關之特性與智慧型手機使用之目的。林佳薇(2015)使用台灣地區北、中、南、東四區域 12 個雨量站，1947 至 2012 共 66 年的月雨量資料，以分位迴歸探討不同位數 SPI、乾旱延時、乾旱量、乾旱頻率系列的時間變化趨勢，並以集群分析探討各測站乾旱趨勢的空間關聯性。凌子平(2015)應用空間限制集群分析法，分別劃設 1996 年與 2011 年兩個年度之生活圈，並進一步將

分群成果與現今行政轄區及國土計畫中所劃設之生活圈相比較，以檢視臺灣現今所產生之空間發展議題。羅婉瑞(2010)使用集群分析探討航空公司營運績效，選出兩組分析對象，分別是全球客運營收排名前十大航空公司與亞太航空協會(Association of Asia Pacific Airlines, AAPA)成員航空公司，使用上述各航空公司西元 2004 年至 2008 年年度報表資料，將我國兩大航空公司與上述兩研究對象以集群分析做分組與排序比較。在集群分析分組後，以各小組的航空公司獲利能力指標做為被解釋變數。

2.5 隨機森林

隨機森林是由決策樹發展出來的演算法，屬於集成學習的方法，在其他領域也被廣泛使用證明其可用性，決策樹是一種非線性的預測方法，透過一套二元的選擇規則，基於預測的變數做出預測。隨機森林屬於整體學習中袋裝法的一種，是透過多顆決策樹共同做出決策，可以減少過度擬合(overfitting)，對共線性以及缺失的資料也比較不敏感。徐翊菁(2020)，以隨機森林預測心房顫動發作，實驗結果顯示性能指標中的準確性、靈敏性、特異性、精確度皆能在 90% 以上。呂昀(2019)，應用隨機森林預測線上飯店之點閱與預訂，實驗結果證明單純找出特徵因子再進行飯店是否被預訂的準確度不佳，但如經過 k-means 分群後，可有效提升隨森林的預測準確率。鍾佩茶(2021)，隨機森林之預測準確度較時間序列模型佳，且利用轉折點法發現，隨機森林模型在預測銷售額之預測趨勢，能高度重合真實值，亦即具有高度的預測精確度。葉子維(2018)，本研究應用決策樹、隨機森林及線性判別分析預測潛在行動銀行使用者，並進行預測結果比較分析，研究

結果顯示，使用隨機森林法的整體準確率最高，線性判別分析的靈敏度較好。郭家奴(2014)，本研究以隨機森林水位預報模式搭配隨機森林雨量預報模式進行前置時間 1~3 小時之水位預報，其結果顯示前置時間 1~2 小時之預報有良好的預報能力。



第三章 資料集介紹

本研究利用臺北市政府提供之 YouBike 站點的租借資料、中央氣象局網頁提供的降雨量資料、Google trend 的 Covid-19 主題搜尋筆數資料以及疾管署網頁提供的確診人數資料等，進行 Covid-19 疫情對民眾生活習性改變的探討，我們將上述資料集經資料清洗、降維、篩選、分群標籤等處理方式後，最後將資料投入隨機森林找出特徵因子重要度數據進行探討。

3.1 資料範圍

本研究的臺北市 YouBike 租賃明細資料為來至於臺北市政府交通局提供，資料範圍為 2020/1/1 至 2020/7/31 止，其租用站點設置點遍及捷運站、學校、醫院、旅遊景點、公園、住宅區、運動場等場所，合計共有 399 個 YouBike 租用站，資料筆數共計有 920 萬 3624 筆。

3.2 YouBike 開放資料平台資料格式

該資料集取自 YouBike 借還資料，資料欄位包含交易序號、卡號、扣款時間、扣款金額、借車時間、借車站代號等共有 23 個欄位，其格式如表 3-1。

表 3-1 YouBike 資料欄位說明

欄位名稱	資料型式	欄位說明
card_physical_id	nvarchar(20)	交易序號
card_txn_subtype_id	nvarchar(20)	卡號
debit_time	datetime	扣款時間
debit_money	int	扣款金額
rent_time	datetime	借車時間
rent_station_code	nvarchar(20)	借車站代號
rent_station	nvarchar(40)	借車場站
rent_column	nvarchar(40)	借車車柱
bike_id	nvarchar(20)	自行車編號
return_time	datetime	還車時間
return_station_code	nvarchar(20)	還車站代號
return_station	nvarchar(40)	還車場站
return_column	nvarchar(40)	還車車柱
rent	nvarchar(20)	租用
mobile_phone_id	nvarchar(20)	手機序號
card_type	nvarchar(20)	卡片種類
fee_type	nvarchar(20)	費率別
sp_id	nvarchar(20)	業者代碼
data_type	nvarchar(20)	根據檔案是 IPASS 或
SrcUpdateTime	datetime	整批資料更新時間
UpdateTime	datetime	存入 DB 時間
InfoTime	datetime	該筆資料更新時間
InfoDate	date	該筆資料更新時間

3.3 各 YouBike 租用站原始資料範例

各站租借原始資料範例如表 3-2，其中「借車時間」欄位為各租用站點的車輛實際借出時間，為後續研究運用，我們將資料格式分拆為「日期」及「時間」兩欄，「借車場站」為 YouBike 租用站之名稱，原始資料共有 920 萬 3624 筆。

表 3-2 YouBike 租用站原始資料範例

借車時間	借車場站	日期	時間
2020/1/4 19:04	捷運市政府站(3 號出口)	2020/1/4	19
2020/1/4 11:35	富台公園	2020/1/4	11
2020/1/4 21:29	捷運台電大樓站(2 號出口)	2020/1/4	21
2020/1/4 22:56	捷運公館站(2 號出口)	2020/1/4	22
2020/1/4 10:07	羅斯福寧波東街口	2020/1/4	10
2020/1/4 14:53	國興青年路口	2020/1/4	14
...
2020/1/4 18:20	臺北南山廣場	2020/1/4	18
2020/1/4 08:45	中山行政中心	2020/1/4	8

3.4 Covid-19 原始資料範例

Covid-19 原始資料範例如表 3-3，資料欄位包含個案研判日、個案公佈日、縣市別、新增確診人數等 4 個欄位，Covid-19 原始資料共 124 筆。

表 3-3 Covid-19 原始資料範例

個案研判日	個案公佈日	縣市別	新增確診人數
2020-01-22	2020/1/22	境外移入	1
2020-01-24	2020/1/24	境外移入	2
2020-01-26	2020/1/26	境外移入	1
2020-01-27	2020/1/27	境外移入	1
...
2020-07-24	2020/7/24	境外移入	3
2020-07-27	2020/7/27	境外移入	4

3.5 降雨量資料

降雨量資料範例如表 3-4，資料欄位包含日期、降雨量等 2 個欄位，降雨量資料共 213 筆。

表 3-4 降雨量資料範例

日期	降雨量
2020/1/1	0
...	...
2020/7/26	38.5
2020/7/27	3.5
2020/7/28	0.5
2020/7/29	0
2020/7/30	0

3.6 Google trend 資料

Google trend 資料範例如表 3-5，資料欄位包含日期、Google trend 等 2 個欄位，Google trend 資料共 213 筆。

表 3-5 Google trend 資料範例

日期	Google trend
2020/1/1	0
2020/1/2	0
2020/1/3	1
2020/1/4	1
2020/1/5	2
...	...
2020/7/28	8
2020/7/29	7
2020/7/30	6
2020/7/31	6

3.7 資料彙整說明

YouBike 原始資料經清洗降維後，我們將每個 YouBike 租用站 0~24 時切分為 8 個時區，再分別統計出每日每個時區的總借用次數，最後再進行資料統計，產出三個新的資料集，分別為各 YouBike 租用站每日各時區租用次數(表 3-6)、各 YouBike 租用站各時區月平均租用次數(表 3-7)、各 YouBike 租用站各時區年平均

租用次數(3-8)，最後將上述三個資料集進行混合計算，產出一個各 YouBike 租用站各時區租用次數縮減比例資料集(表 3-9)，此資料集用來進行群集分析使用，資料筆數共 91 萬 9006 筆。

表 3-6 各 YouBike 租用站每日各時區租用次數

日期	借車場站	時區	租用次數
2020/1/17	捷運中山國小站(4 號出口)	11-13	23
2020/1/17	捷運中山國小站(4 號出口)	11-13	23
2020/1/17	捷運中山國小站(4 號出口)	14-16	20
2020/1/17	捷運中山國小站(4 號出口)	17-19	47
2020/1/17	捷運中山國小站(4 號出口)	20-22	27

表 3-7 各 YouBike 租用站各時區月平均租用次數

借車場站	時區	月-平日平均 租用次數	月-假日平均 租用次數
捷運中山國小站(4 號出口)	11-13	26.5	21.7
捷運中山國小站(4 號出口)	14-16	27.8	30.3
捷運中山國小站(4 號出口)	17-19	40.1	22.3
捷運中山國小站(4 號出口)	20-22	28.8	18.1

表 3-8 各 YouBike 租用站各時區月平均租用次數

借車場站	時區	年-平日平均 租用次數	年-假日平均 租用次數
捷運中山國小站(4 號出口)	11-13	26.3	31.5
捷運中山國小站(4 號出口)	14-16	25	31.7
捷運中山國小站(4 號出口)	17-19	46.2	31
捷運中山國小站(4 號出口)	20-22	28.9	23.2

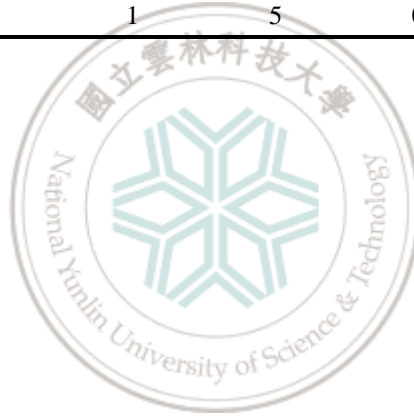
表 3-9 各 YouBike 租用站各時區租用次數縮減比例

日期	借車場站	時區	與今年同 月平日或 假日平均 差異	與上月平日 或假日平均 差異	與今年平 日或假日 平均差異
2020/1/17	捷運中山國小站(4 號出口)	11-13	-0.094	0.013	0.017
2020/1/17	捷運中山國小站(4 號出口)	14-16	-0.245	-0.195	-0.16
2020/1/17	捷運中山國小站(4 號出口)	17-19	0.197	0.29	0.039
2020/1/17	捷運中山國小站(4 號出口)	20-22	-0.039	0.015	-0.066

完成 YouBike 租用站各時區租用次數縮減比例資料集後，我們接續將 Covid-19 原始資料、降雨量原始資料、Google trend 原始資料等三個資料集，再加上寒暑假時間起迄日期，整合成一個新的特徵值組合資料集，此資料機用來進行隨機森林訓練使用，資料筆數共 91 萬 9006 筆。

表 3-10 特徵值組合

大台北 (北北基 桃)確診 人數	非大台北 (北北基 桃)確診 人數	境外確診 人數	前 1 日確 診人數	前 3 日確 診人數	降雨量	Google trend	寒暑假 期間
1	0	0	2	4	0	52	0
1	0	0	2	4	0	52	0
1	0	0	1	5	0	46	0
...
1	0	0	1	5	0	46	0



第四章 研究方法

4.1 研究流程

本研究為依照圖 4-1 之研究流程進行影響各類別 YouBike 租用站之車輛租用量的特徵因子重要度探討，執行步驟：(1) 資料的前處理與選擇：將 YouBike 原始資料進行清洗與分類，並依類別進行資料的擷取。(2)特徵值數據正規化：將特徵值資料之絕對數值修改為相對數值。(3)k-means 分群後進行資料標籤：依經過 k-means 演算法分群後為每筆資料進行標記。(4)集群數量 k 值的選定：以輪廓係數法來決定 k 值。(5)隨機森林找出重要因子：使用隨機森林找出各因子之重要度。

透過上述的流程步驟，依照隨機森林找出之各因子重要度，針對不同生活圈之 YouBike 租賃站租用量影響程度進行探討。

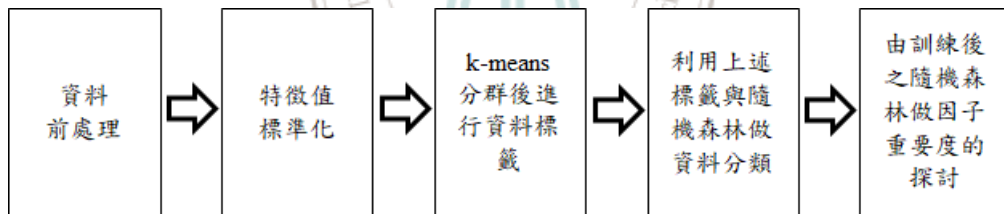


圖 4-1 研究流程圖

4.2 資料的前處理

因本研究為採用非監督式學習的集群分析，故資料不經人為定義之階段，由演算法自行依據資料的數值進行分群，然而資料的前處理仍是進行集群分析所必經的步驟，需要對資料進行(1)缺失、無效值的清洗；(2)資料收斂降維。

本研究首先對於資料內的 YouBike 自行車租借站點進行降維，簡述如下：

1、YouBike 租用站點收斂：

本研究將 2020/1/1-2020/7/31 之租用站點資料，先行整理出每個月所有的租用站點，再將 7 個月份的全部資料統一彙整，後續擷取出每個月份皆有出現之租用站點，此區間如借用站點並非每月份皆有出現，則視為此站點是在這區間內新增或移除之站點，為避免因數據完整性不足之站點影響資料分析運用，故將此類站點於本研究中進行剔除，經收斂後之站點，由原 399 個站縮減為 292 個站。

4.3 特徵值標準化

資料處理的方式繁多，資料之間也常有不同的尺度跟數量差異，當資料間的水平差異較大，或是運用不同尺度計算時，如果我們直接使用原始數據進行資料分析，數值較大的資料很可能會影響了整體分析的結果，進而影響弱勢尺度及數據的作用，為了保證資料不致因尺度及數量影響了綜合分析的結果，因此我們使用資料標準化這個方法，將標準化(Standardization)後的特徵值數據投入隨機森林以找出特徵因子重要度；本研究使用的資料標準化方法為 MaxAbs(最大值絕對值標準化(MaxAbs))，根據最大值的絕對值進行標準化。

4.4 k-means 分群後進行資料標籤

因本研究之資料為無經過專家或專業知識定義後，為每筆資料進行分類處理，僅以純資料面來進行群集分析，故需為每筆資料做標籤，以提供後續隨機森林訓練時使用；資料標籤方式是採用 K-means 分群之結果為每筆資料做標記，而對於 k 值之選定，則先以輪廓係數來決定初次分群之 k 值，再將特徵值與分群資料投入 Matlab R2021a 版，同時將各群資料筆數平衡後，最後使用 Classification

Learner 套件觀看訓練後每筆資料分群的結果，再將依據結果將相近的資料重新合併為同一群，以完成最終每筆資料的標籤作業。

4.4.1 集群分析 K-means

K-means 分群就是對所有數據進行分組，將相似的數據歸類為一起，每一筆數據的能有一個分組，每一組稱作為群集 (Cluster)。那分類根據什麼來定義，常用距離來做運算。K-means 分群其實就有點像是以前學數學時，找重心的概念。下列步驟總結了決定群集成員的過程，而這個過程適用於任何群集數量：

1、步驟 0：預設集群的數量(K)

$$\mu_c^{(0)} \in R^d, c = 1, 2, \dots, k \quad (2-1)$$

2、步驟 1：從猜測各群集的中心位置開始。因為我們尚不知道它們是否真的就是各自群集的中心點，先姑且稱為「偽中心點」。

3、步驟 2：計算每個點到集群「偽中心點」的歐式距離，再將各個資料點分派到距離最近的「偽中心點」。

$$S_c^{(t)} = \{X_i \mid \|X_i - \mu_c^t\| \leq \|X_i - \mu_{c^*}^{(t)}\|, \forall i = 1, \dots, n\} \quad (2-2)$$

4、步驟 3：根據與群集內各成員的距離，重新調整「偽中心點」的位置。

反覆上述步驟 2、3 動作，直到群集不變，群集中心不動為止，以下如圖 4-2 至圖

4-6 簡易說明如下(橘色是群集的初始偽中心點，綠色為新的群集偽中心點)：

步驟 1：開始猜測各群集的中心位置

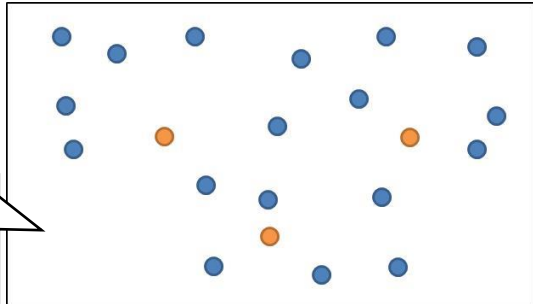


圖 4-2 k-means 示意圖-1

步驟 2：將各個資料點分派到距離最近的「偽中心點」

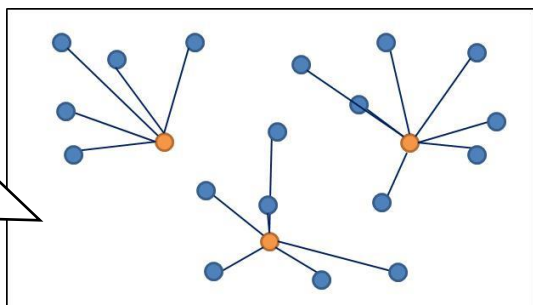


圖 4-3 k-means 示意圖-2

步驟 3：根據與群集內各成員的距離，重新調整「偽中心點」的位置

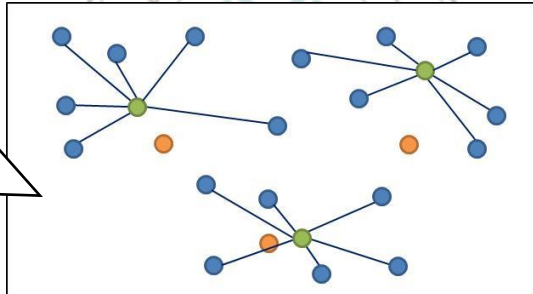


圖 4-4 k-means 示意圖-3

持續重複步驟 2 與步驟 3

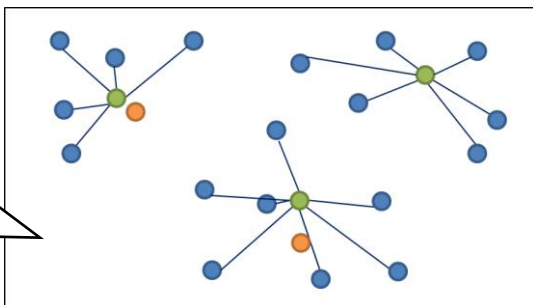


圖 4-5 k-means 示意圖-4

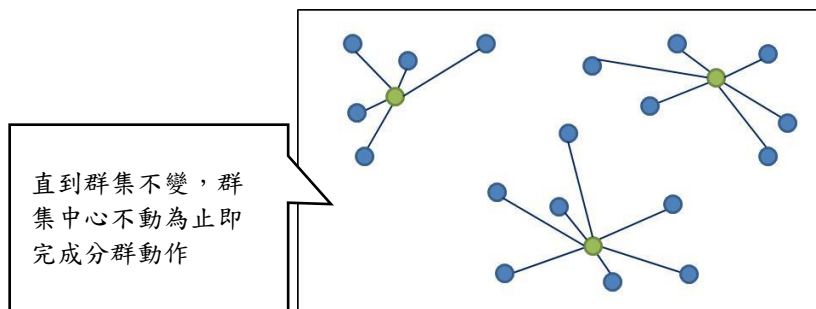


圖 4-6 k-means 示意圖-5

4.4.2 集群數量 k 值的選定

在我們採用為資料標籤之做法後，要確認資料應分為幾群較適當是首先要考量的問題，所以本研究選擇使用輪廓係數 Silhouette coefficient 法來決定合理的 k 值，輪廓係數法介紹如下：輪廓係數是聚類效果好壞的一種評價方式。它最早由 Peter J. Rousseeuw 於 1986 年提出。它結合內聚度和分離度兩種因素。可以用來在相同原始數據的基礎上用來評價不同算法、或者算法不同運行方式對聚類結果所產生的影響。效果好壞的其中一種方法。其概念滿足集群的概念，是找出「組間異質；組內均質」的值，同集群之間距離之距離以 a 表示，異群之間的點平均距離以 b 表示，其得出的 S 值越大，表示集群區隔越明顯，計算完所有樣本的 S_i 均值之後，可用以評估集群是否合理有效。

$$S = \frac{b - a}{\max(a, b)} \quad (2-3)$$

4.4.3 分群資料筆數平衡與重新分群合併縮減

當決定 k 值再以 k-means 分群後，我們觀察資料集發現各群間之有資料不平衡的問題，為避免因資料不平衡而導致隨機森林分類結果不佳，後續我們再透過資料平衡與合併縮減等方式，取得每群資料筆數平衡與最終每筆資料的分群結果實際舉例說明如下：

表 4-1 為捷運類租用站 1 月的分群結果，由此表可看出各群資料筆數不平衡問題，分類在第 1 群的有 840 筆，而第 5 群僅有 303 筆資料，為避免因資料不平衡而導致隨機森林分類結果不佳，故我們後續再透過 MATLAB 進行資料平衡的處理。

本研究於 MATLAB R2021a 版之程式環境下以隨機挑選的方式將每群資料統一皆為 300 筆(作法為取分群資料筆數最低之整數)，後續再使用 Classification Learner 套件觀看訓練後每筆資料分群的結果，分群結果如圖 4-7 所示，簡單說明圖 4-7 之觀看方式，以第 1 群 330 筆資料來說，正確被預測為第 1 群的有 92 筆，被預測為第 2 群的有 74 筆，被預測為第 3 群的有 10 筆，被預測為第 4 群的有 33 筆，被預測為第 5 群的有 91 筆，其餘 2-5 群之觀看方式皆相同；從分群結果可看出，第 3 群有 215 筆資料被預測為第 2 群，第 4 群資料有 153 筆資料被預測為第 2 群，表示第 3、4 群之資料特性與第 2 群相近，故我們將第 3 群與第 4 群之資料重新分類在第 2 群，使得分群數由 5 群縮減為 3 群，經由上述 k-means 分群與 Matlab 資料平衡與分群合併縮減做法，我們取得最終分群合併縮減結果如表 4-2 所示。

表 4-1 捷運類租用站 1 月份分群結果

分群	資料筆數	資料佔比
1	840	27%
2	460	15%
3	730	23%
4	777	25%
5	303	10%

True Class	1	92	74	10	33	91
	2	27	220	19	17	17
	3	23	215	16	27	19
	4	57	153	21	27	42
	5	67	6	2	22	293
		1	2	3	4	5
		Predicted Class				

圖 4-7 捷運類租用站分群預測結果

表 4-2 捷運類租用站 1 月分群合併縮減結果

月份	縮減前分群數	縮減後分群數
1 月	5	3

4.5 訓練隨機森林找出特徵因子重要度

對於特徵值數據計算的選擇，常用之方式有皮爾森相關係數與隨機森林，因皮爾森相關係數之方式是為衡量的是變數之間的線性相關性，而本研究所採用之特徵值並非一定有線性關係，故採用隨機森林較為合適。

隨機森林[23]是由 Breiman 於 2001 年提出。隨機森林也是一種監督式的學習演算法，是以決策樹為基礎學習器的集成學習演算法，它是由 K 顆決策樹所組成，其中 $h(X, \theta_k)$ 是一顆決策樹， X 是隨機森林輸入的樣本集合， θ_k 是獨立的隨機特徵，最後再依投票法選出最佳的樹。

$$\{h(X, \theta_k); k = 1, 2, \dots, k\} \quad (2-4)$$

決策樹的選擇是使用 CART 演算法進行分類，其分類的原則是採用基尼係數 (Gini Index) 來衡量資料集對分類的不純度(impurity)，假設資料集合 X 包含 n 個類別，基尼係數 $Gini(X)$ 定義為：

$$Gini(X) = 1 - \sum_{a=1}^n P_a^2 \quad (2-5)$$

其中 P_a 為在集合 X 中的類別 a 的機率

CART 採用二元分割方式，假設使用條件 C 將資料集 X 分割為 X_1 及 X_2 兩個子集，根據分割條件計算出兩個子集之基尼係數 $Gini_c(X)$ 公式為：

$$Gini_c(X) = \frac{x_1}{x} Gini(X_1) + \frac{x_2}{x} Gini(X_2) \quad (2-6)$$

然而在進行分割條件之選擇，即是選擇具有最小 $Gini_c(X)$ 的特徵條件來進行分割。

假設一個大小為 N 的訓練資料集樣本，隨機森林[24]會採用 bootstrap 的抽樣方法隨機且有放回地抽取 N 個訓練子集，以抽出的子集從 M 個特徵，隨機選取 m 個(條件 $m < M$)，從中計算出一個最具分類能力的特徵以分割該節點，訓練出 N 棵決策樹，其中森林裡的每棵樹之間是沒有關聯的，且都會完整生長，不作任何剪枝，而對於未被抽到的樣本組成了 N 個袋外資料(out-of-bag, OOB)，最後彙總所有決策樹的預測，以多數決的方式來決定最佳的分類結果。

隨機森林是集成式(Ensemble)分類器由多個決策樹 $h_1(X), h_2(X), \dots, h_k(X)$ 組合而成，然而隨機選擇訓練樣本，設其中的 X 為樣本向量， Y 為正確分類的類別向量，則定義邊際函數(Margin Function)為：

$$mg(X,Y) = av_k I\{h_k(X) = Y\} - \max_{l \neq y} av_k I\{h_k(X) = l\} \quad (2-7)$$

其中 $I\{\}$ 為函數指標， av 表示平均值，邊際函數主要是用來衡量平均得票數的程度差異，表示了正確分類 Y 之下 X 的得票數目超過其他錯誤分類的最大得票數目的程度，該值愈大表示分類的可信度愈高。[隨機森林的構建過程如下\(引用自壹讀網頁\[25\]\)](#)：

- 1、從原始訓練集中使用 Bootstrapping 方法隨機有放回採樣取出 m 個樣本，共進行 n_tree 次採樣。生成 n_tree 個訓練集
- 2、對 n_tree 個訓練集，我們分別訓練 n_tree 個決策樹模型
- 3、對於單個決策樹模型，假設訓練樣本特徵的個數為 n ，那麼每次分裂時根據信息增益/信息增益比/基尼指數 選擇最好的特徵進行分裂
- 4、每棵樹都已知這樣分裂下去，知道該節點的所有訓練樣例都屬於同一類。在決策樹的分裂過程中不需要剪枝
- 5、將生成的多顆決策樹組成隨機森林。對於分類問題，按照多棵樹分類器投票決定最終分類結果；對於迴歸問題，由多顆樹預測值的均值決定最終預測結果

本研究為將標準化後特徵因子資料結合標籤資料(分群結果)丟入隨機森林訓練，經由隨機森林訓練產出之結果，我們可得到各特徵因子的重要度數據，然後對各因子重要度進行探討。

4.6 特徵因子重要度探討

本研究為依據隨機森林找出特徵因子重要度數據後，將 1-7 月份的特徵因子重要度數據進行排序與以直方圖方式呈現，再搭配 1-7 月的 Covid-19 確診人數資料與降雨量資料等資料，針對 1-7 月的特徵因子重要度變化逐一進行探討。



第五章 研究結果

本章將討論研究方法之流程，套用在本研究的集群分析及特徵因子要度探討，實驗將臺北市 YouBike 站的租用資料進行資料處理投用於集群分析後，將分群結果與特徵因子結合投入再隨機森林中，依結果探討特徵因子重要度。

5.1 資料集介紹

本研究將 YouBike 原始資料經過清洗、降維、變數處理後，再將租賃資料之全天 24 小時劃分為 2H-4H、5H-7H、...、23H-01H 共八個時區，未滿 60 分鐘以 1 小時計算，例如上午 07:59 分之資料納入 5H-7H 資料集內；下午 20:01 分歸納入 20H-22H 資料集內，整理後之資料共有 91 萬 9006 筆。

經資料觀察發現，如有時區日租用人次與月平均租用人次未達 10 人次之借用站點，經與其它項目計算增減比例後，有數值過大之情況(表 5-1)，為避免影響分群結果，將日借用人次與月平均借用人次未達 10 人次之資料去除，去除後資料筆數縮減為 35 萬 5122 筆。

表 5-1 去除日租用人次與月平均租用人次未達 10 人次之資料

日期	時區	停車場	日租用人次	月平均租用人次	與今年同月平日或假日平均差異	與上月平日或假日平均差異	與今年平日或假日平均差異
2020/2/8	23-01	新明路 321 巷口	3	0.4	650%	329%	1400%
2020/2/11	20-22	新明路 321 巷口	10	2.2	355%	203%	1329%
2020/2/16	2-4	新明路 321 巷口	2	0.2	900%	567%	1900%
2020/2/20	23-01	新明路 321 巷口	3	0.6	400%	233%	1400%
2020/2/21	11-13	向陽南港路口	6	2.4	150%	233%	11%
2020/2/22	8-10	向陽南港路口	4	2.2	82%	233%	-7%
2020/2/23	8-10	向陽南港路口	4	2.2	82%	233%	-7%
2020/2/27	11-13	向陽南港路口	6	2.4	150%	233%	11%
2020/2/15	2-4	新湖國小	2	0.3	567%	1900%	1900%

資料集經上述方清洗後，再將 YouBike 租用站依其名稱或性質做為分類(例：捷運市政府站(3 號出口)歸類為「捷運」類；一江公園歸類為「公園」類；台灣科技大學歸類為「學校」類；台北市政府歸類為「行政機關」類)，共區分 17 類，因資料筆數過於龐大，故將每類別 2020/1/1-2020/7/31 之租用量進行加總，再依總借用量排序，最後採總借用量排名前 3 名之類別做後續的資料研究探討(表 5-2)。

表 5-2 各分類站總租用次數

類別	總租用次數	排序
捷運	168387	1
公園	136061	2
學校	105132	3
景點	41181	4
住宅區	17198	5
...
停車場	3942	15
其他	1729	16
社區	1410	17

5.2 實驗環境設定

5.2.1 特徵因子選擇

(1)大台北(北北基桃)確認診人數：我們想得知整個大台北地區之確診人數多寡是否影響台北市民眾租用 YouBike 之行為。(資料來源：本研究由 COVID-19 全球疫情地圖提供之資料整理而來，網址:<https://covid-19.nchc.org.tw/index.php>)

(2)非大台北(北北基桃)確認診人數：我們想得知非大台北地區之確診人數多寡是否影響台北市民眾租用 YouBike 之行為。(資料來源：本研究由 COVID-19 全球疫情地圖提供之資料整理而來，網址:<https://covid-19.nchc.org.tw/index.php>)

(3)境外確診人數：我們想得知境外確診人數多寡是否影響台北市民眾租用 YouBike 之行為。(資料來源：本研究由 COVID-19 全球疫情地圖提供之資料整理而來，網址:<https://covid-19.nchc.org.tw/index.php>)

(4)前一日確診人數：我們想得知今日台北市民眾租用 YouBike 之行為是否受前一日確診人數之影響。(資料來源：本研究由 COVID-19 全球疫情地圖提供之資料整理而來，網址:<https://covid-19.nchc.org.tw/index.php>)

(5)前三日確診人數：我們想得知今日台北市民眾租用 YouBike 之行為是否受前三日累計確診人數之影響。(資料來源：本研究由 COVID-19 全球疫情地圖提供之資料整理而來，網址:<https://covid-19.nchc.org.tw/index.php>)

(6)降雨量：Youbike 為戶外行交通工具，而雨天會直接影響台北市民眾租用 YouBike 之行為，故將此因子納入特徵值內。(資料來源：本研究由交通部中央氣象局提供之資料整理而來，網址:<https://www.cwb.gov.tw/V8/C/D/DailyPrecipitation.html>)

(7) Google trend 搜尋量：我們想透過民眾對於 Covid-19 的搜尋熱度來觀察是否影響台北市民眾租用 YouBike 之行為，故我們以 Covid-19 此主題作為於 Google trend 搜尋量多寡之標的。(資料來源：本研究由 Google trend 提供之資料整理而來，網址:<https://trends.google.com.tw/trends/?geo=TW>)

(8)寒暑假期間：寒暑假直接影響學生上下學人數多寡，從而影響學校附近 YouBike 租用站之借用量，故將此因子納入特徵值內。(資料來源：本研究由教育部即時新聞提供之資料整理而來，網址 :https://www.edu.tw/News_Content.aspx?n=9E7AC85F1954DDA8&s=F430DC0E892811A6)

5.2.2 實驗使用之軟體套件

本研究使用 Python 之 sklearn.cluster import KMeans 套件進行每筆資料分群，再使用 MATLAB R2021a 版本進行每群資料筆數平衡處理，後續再將分群結果與每筆特徵值資料結合後，以 Python 之 from sklearn.ensemble import RandomForestClassifier 套件觀看隨機森林訓練後之分類結果與特徵因子重要度。

5.3 捷運類 YouBike 租用站影響因子重要度探討

5.3.1 捷運類 YouBike 租用站資料分群

以 2020 年 1 月份捷運類之 YouBike 租用站資料為例，本研究先以輪廓係數決定 k 值之方式為先將 k 值=2~10 之輪廓係數計算出，再將各 k 值間之輪廓係數進行縮減比例計算，固定取縮減比例>-20%之最大 k 值(表 5-3)。

表 5-3 捷運類租用站分群結果_輪廓係數

分群數(k)	輪廓係數	縮減比例
2	365.3782882	
3	258.9716493	-0.29%
4	189.1376571	-0.27%
5	148.4165508	-0.22%
6	123.691048	-0.17%
7	111.0785655	-0.10%
8	102.526787	-0.08%
9	94.5215232	-0.08%
10	86.74305701	-0.08%

在確認 k 值等於 5 之後，我們將使用 k-means 方法來對 3110 筆的樣本進行分群，1 月份資料分群結果如表 5-4：

表 5-4 捷運類 YouBike 租用站 1 月份分群結果

分群	資料筆數	資料佔比
1	840	27%
2	460	15%
3	730	23%
4	777	25%
5	303	10%

經由上述 k-means 分群做法，我們依序完成各捷運類 1~7 月之資料分群，分群結果如表 5-5 所示。

表 5-5 捷運類 YouBike 租用站 1-7 月分群結果

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
分群數	5	4	5	5	5	5	4

5.3.2 捷運類 YouBike 租用站資料平衡與分群合併縮減

以 2020 年 1 月份捷運類之 YouBike 租用站資料為例，經由第 4 章「4.4.3 分群資料筆數平衡與重新分群合併縮減」舉例之做法，我們依序完成捷運類 1~7 月之分群資料，最終 1-7 月分群合併縮減結果如表 5-6 所示。

表 5-6 捷運類 YouBike 租用站 1-7 月分群合併縮減結果

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
分群數	3	3	3	3	3	5	4

5.3.3 捷運類 YouBike 租用站特徵值資料進行隨機森林訓練與因子重要度計算

以 2020 年 1 月份捷運類 YouBike 租用站為例，我們將每群資料筆數平衡處理後，再將每筆特徵值資料與分群結果結合後之數據(表 5-7)丟入隨機森林觀看分類結果與特徵因子重要度，我們使用 80%的資料進行訓練與 20%的資料進行測試，產出之混淆矩陣結果如表 5-8，觀看表 5-8 之數據，第 1 群 58 筆資料裡有 14 筆預測在第 1 群，第 2 群 62 筆資料裡有 36 筆預測在第 2 群，第 3 群 60 筆資料裡

有 52 筆第預測在第 3 群，，全部 180 筆資料裡共有 102 筆分類正確，整體分類之正確率達 57%，表示對於隨機森林的分類結果是尚堪使用的。

接下來我們觀看各特徵因子重要度(表 5-9)，可看出捷運類租用站 2020 年 1 月份以「Google trend 搜尋量」之因子重要度影響最高，而大台北(北北基桃)確認診人數影響最低。

表 5-7 捷運類 YouBike 租用站 1 月份特徵值與分群結果結合

大台北 (北北基 桃)確診 人數	非大台 北(北北 基桃)確 診人數	境外確 診人數	前 1 日 確診人 數	前 3 日 確診人 數	降雨量	Google trend	寒暑假 期間	分群標 籤
0	0	1	0	0	0	0.44	1	1
0	0	0	0	0	0	0.45	1	1
...
0	0	0	0	0	0.02	0.02	0	3

表 5-8 捷運類 YouBike 租用站 1 月份資料混淆矩陣分類結果

Predicted lable	1	14	13	7
	2	19	36	1
	3	25	13	52
		1	2	3
	True label			

表 5-9 捷運類 YouBike 租用站 1 月份因子重要度

特徵因子	因子重要度
大台北(北北基桃)確認診人數	0.322
非大台北(北北基桃)確認診人數	0.394
境外確診人數	0.388
前 1 日確診人數	0.388
前 3 日確診人數	0.438
降雨量	0.455
Google trend 搜尋量	0.466
寒暑假期間	0.422

經由上述做法，我們依序完成各 1-7 月之隨機森林訓練，得出各月份混淆矩陣分類結果(表 5-10)與各月份特徵因子之重要度(表 5-11)。以 1 月份由表 5-10 可看出 6、7 月份之分類準確度低於 50%，表示隨著疫情趨緩，與 Covid-19 有相關之六項特徵因子對於此 2 個月的捷運通勤族租用 YouBike 的行為已無影響，故對於此兩個月的特徵因子重要度探討將不予討論。

表 5-10 捷運類 YouBike 租用站資料 1-7 月混淆矩陣分類結果

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
捷運	57%	61%	77%	66%	64%	27%	32%

表 5-11 捷運類 YouBike 租用站 1-7 月因子重要度

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
大台北(北北基桃)							
確認診人數	0.32	0.41	0.45	0.40	0.32	0.16	0.22
非大台北(北北基桃)							
確認診人數	0.39	0.39	0.34	0.42	0.32	0.16	0.22
境外確診人數	0.39	0.33	0.67	0.62	0.32	0.18	0.23
前 1 日確診人數	0.39	0.43	0.66	0.63	0.38	0.19	0.27
前 3 日確診人數	0.44	0.52	0.63	0.64	0.45	0.20	0.25
降雨量	0.46	0.40	0.71	0.66	0.64	0.25	0.27
Google trend 搜尋量	0.47	0.54	0.62	0.65	0.56	0.21	0.23
寒暑假期間	0.42	0.40	0.30	0.31	0.32	0.16	0.22

5.3.4 捷運類 YouBike 租用站影響因子重要度探討

表 5-12 為捷運類 YouBike 租用站 1-7 月各因子重要度排序，為使數據方便觀看與探討，我們將因子重要度以排名表(表 5-12) 與直方圖(圖 5-1)之方式呈現，以進行因子重要度探討。

表 5-12 捷運類 YouBike 租用站 1-7 月各因子重要度排序

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
大台北(北北基桃) 確認診人數	8	4	6	7	5	6	6
非大台北(北北基桃) 確認診人數	5	7	7	6	6	7	7
境外確診人數	6	8	2	5	7	5	4
前 1 日確診人數	7	3	3	4	4	4	2
前 3 日確診人數	3	2	4	3	3	3	3
降雨量	2	6	1	1	1	1	1
Google trend 搜尋 量	1	1	5	2	2	2	5
寒暑假期間	4	5	8	8	8	8	8

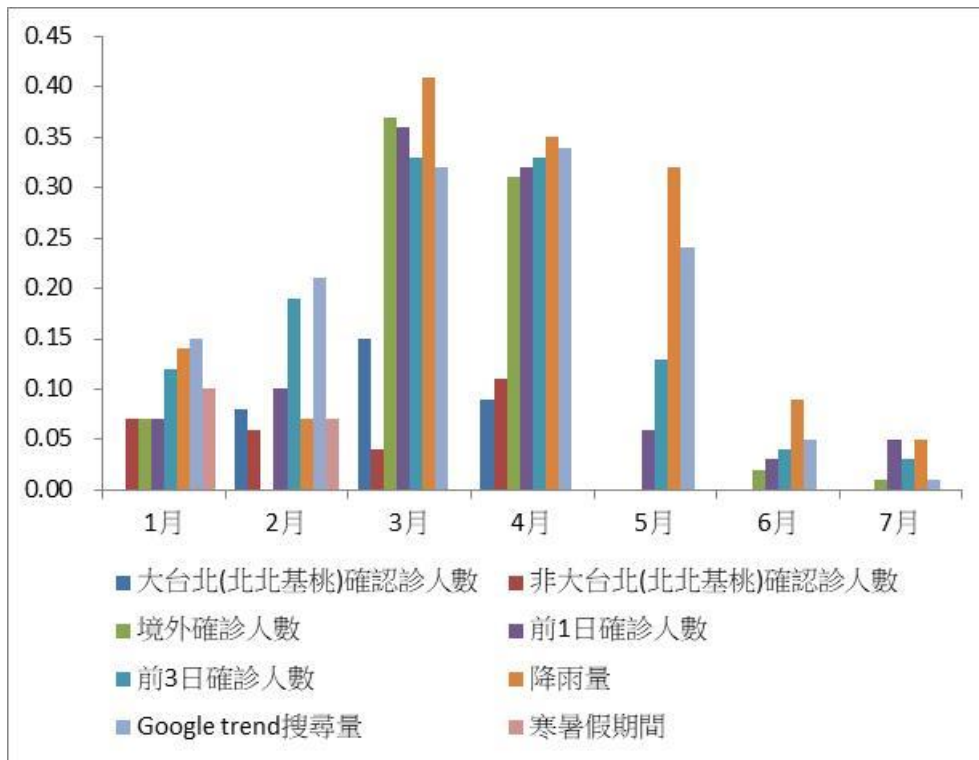


圖 5-1 捷運類 YouBike 租用站各月份特徵因子直方圖

從圖 5-1 觀看每月份各因子重要度變化，先行觀看與 Covid-19 無相關之「降雨量」與「寒暑假期間」兩因子，從各因子重要度排名(表 5-12)來看，其中降雨量在 1-5 月之中，除 2 月分外其餘月份重要度數據皆排名前二名，表示有無下雨對於捷運通勤族租用 YouBike 的行為有高度的影響，且從圖 5-1 搭配圖 5-2 可看出，日平均降雨量愈高，降雨量與他因子重要度數據差異愈大，而「寒暑假期間」1 月與 2 月之重要度排名在第 4 與第 5 名，其餘月份皆排名最末，其中因 1 月與 2 月份裡的寒假期間學生不用上課，表示寒假期間不用上課對學生租用 YouBike 的行為是有影響的；1 月份與 2 月份以「Google trend 搜尋量」重要度數據最高，1-2 月份為疫情剛出現之初期，民眾當時對 Covid-19 的資訊多為透過網路資訊瞭解其傳染特徵，其傳染途徑為通過口沫傳染，且在密閉空間危險度更高，當民眾從網

路上提升對 Covid-19 的瞭解程度後，減少了搭乘捷運的意願，表示民眾對 Covid-19 的瞭解程度增加與租用 YouBike 的行為是有影響的；3 月份以「境外確診人數」重要度數據最高，由圖 5-3 可知，3 月份境外確診人數整體升高至 258 例，遠超過 1 月與 2 月合計之確診人數，面對突然升高之疫情導致民眾心理恐慌，從而影響民眾搭乘捷運的意願，表示面臨境外確診人數攀升對民眾租用 YouBike 的行為是有影響的；4 月與 5 月以「Google trend 搜尋量」重要度數據最高，隨著 3 月與 4 月整體確診人數升溫，使得 4 與 5 月份 Google trend 搜尋量仍維持高檔，顯示民眾仍處於對疫情擔心的狀態與周遭安全的不確定性，減少了搭乘捷運的意願，表示民眾對 Covid-19 的擔心程度與租用 YouBike 的行為是有影響的。

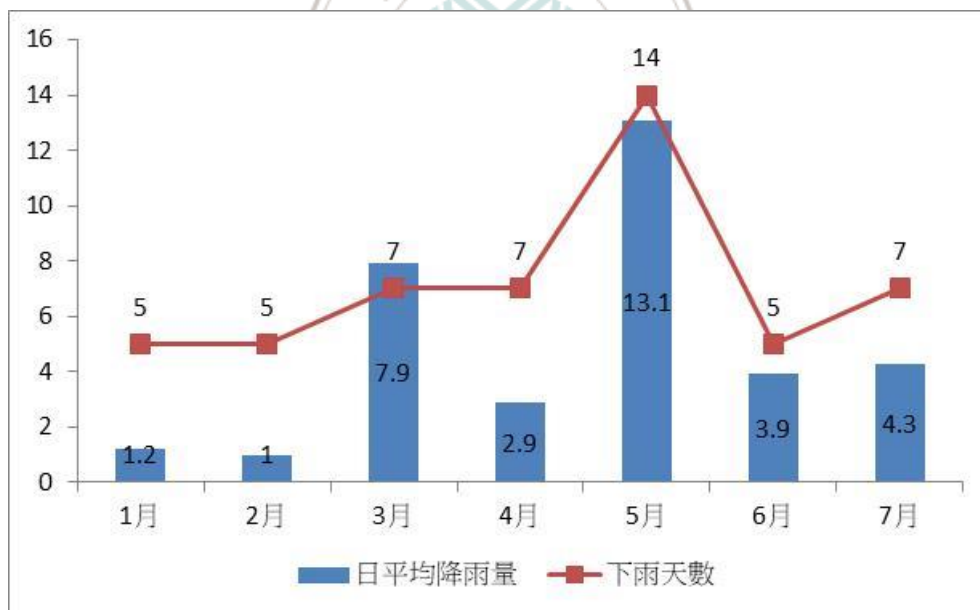


圖 5-2 1-7 月下雨天數與日平均降雨量

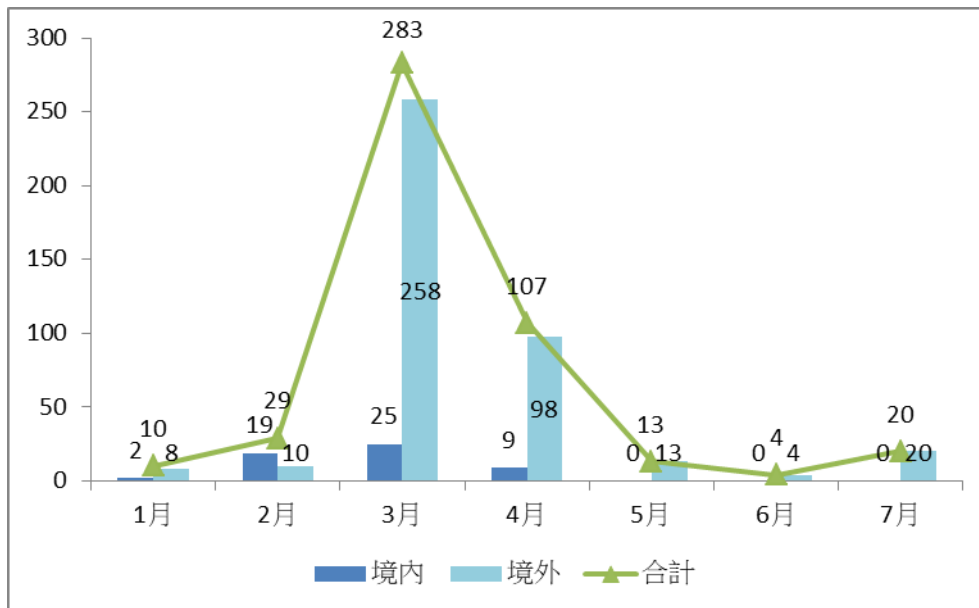


圖 5-3 1-7 月份確診人數變化圖

5.4 公園類 YouBike 租用站影響因子重要度探討

5.4.1 公園類 YouBike 租用站資料分群

以 2020 年 1 月份公園類之 YouBike 租用站資料為例，本研究先以輪廓係數決定 k 值之方式為先將 k 值=2~10 之輪廓係數計算出，再將各 k 值間之輪廓係數進行縮減比例計算，固定取縮減比例>-20%之最大 k 值(表 5-13)。

表 5-13 公園類租用站分群結果_輪廓係數

分群數(k)	輪廓係數	縮減比例
2	1135.140792	
3	755.2961494	-33%
4	585.4169879	-22%
5	504.5564273	-14%
6	433.2308974	-14%
7	381.041378	-12%
8	335.1108725	-12%
9	304.6163456	-9%
10	276.7448815	-9%

在確認 k 值等於 4 之後，我們將使用 k-means 方法來對 6141 筆的樣本進行運算，結果如表 5-14：

表 5-14 公園類 YouBike 租用站 1 月份分群結果

分群	資料筆數	資料佔比
1	2277	37%
2	556	9%
3	2092	34%
4	1216	20%

經由上述 k-means 分群做法，我們依序完成各公園類 1~7 月之資料分群，分群結果如表 5-15 所示。

表 5-15 公園類 YouBike 租用站 1-7 月分群結果

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
分群數	4	4	3	5	4	4	4

5.4.2 公園類 YouBike 租用站資料平衡與分群合併縮減

以 2020 年 1 月份公園類之 YouBike 租用站資料為例，經由第 4 章「4.4.3 分群資料筆數平衡與重新分群合併縮減」舉例之做法，我們依序完成公園類 1~7 月之分群資料，最終 1-7 月分群合併縮減結果如表 5-6 所示。

表 5-6 公園類 YouBike 租用站 1-7 月分群合併縮減結果

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
分群數	2	3	3	3	3	4	4

5.4.3 公園類 YouBike 或用站特徵值資料進行隨機森林訓練與因子重要度計算

以 2020 年 1 月份公園類 YouBike 租用站為例，我們將每群資料筆數平衡處理後，再將每筆特徵值資料與分群結果結合後之數據(表 5-17)丟入隨機森林觀看分類結果與特徵因子重要度，我們使用 80% 的資料進行訓練與 20% 的資料進行測

試，產出之混淆矩陣結果如表 5-18，觀看表 5-18 之數據，第 1 群 102 筆資料裡有 84 筆第預測在第 1 群，第 2 群 98 筆資料裡有 57 筆第預測在第 2 群，全部 200 筆資料裡共有 141 筆分類正確，整體分類之正確率達 71%，表示對於隨機森林的分類結果是可使用的。

接下來我們觀看各因子重要度(表 5-19)，可得出公園類租用站 2020 年 1 月份以「前 3 日確診人數」之因子重要度影響最高，而大台北(北北基桃)確認診人數影響最低。

表 5-17 公園類 YouBike 租用站 1 月份特徵值與分群結果結合

大台北 (北北基 桃)確診 人數	非大台 北(北北 基桃)確 診人數	境外確 診人數	前 1 日 確診人 數	前 3 日 確診人 數	降雨量	Google trend	寒暑假 期間	分群標 籤
0	0	0.5	0.33	0.8	0	0.71	1	1
0	0	0.5	0.33	0.8	0	0.71	1	1
...
0	0	0	0.67	0.6	0.03	0.85	1	2
0	0	0	0.67	0.6	0.03	0.85	1	2

表 5-18 公園類 YouBike 租用站 1 月份資料混淆矩陣分類結果

Predicted label	1	84	41
	2	18	57
		1	2
True label			

表 5-19 公園類 YouBike 租用站 1 月份因子重要度

特徵因子	因子重要度
大台北(北北基桃)確認診人數	0.49
非大台北(北北基桃)確認診人數	0.56
境外確診人數	0.66
前 1 日確診人數	0.67
前 3 日確診人數	0.70
降雨量	0.68
Google trend 搜尋量	0.67
寒暑假期間	0.69

經由上述做法，我們依序完成各 1-7 月之隨機森林訓練，得出各月份混淆矩陣分類結果(表 5-20)與各月份特徵因子之重要度(表 5-21)。由表 5-20 可看出 6、7 月份之分類準確度低於 50%，表示隨著疫情趨緩，與 Covid-19 有相關之六項特徵因子對於此 2 個月的公園類戶外活動族租用 YouBike 的行為已無影響，故對於此 2 個月的特徵因子重要度探討將不予討論。

表 5-20 公園類 YouBike 租用站資料 1-7 月混淆矩陣分類結果

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
捷運	71%	76%	62%	68%	73%	39%	48%

表 5-21 公園類 YouBike 租用站 1-7 月因子重要度

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
大台北(北北基桃) 確認診人數	0.49	0.56	0.4	0.53	0.46	0.3	0.33
非大台北(北北基桃) 確認診人數	0.56	0.5	0.29	0.5	0.46	0.3	0.33
境外確診人數	0.66	0.52	0.48	0.61	0.5	0.33	0.38
前 1 日確診人數	0.67	0.6	0.48	0.57	0.55	0.32	0.36
前 3 日確診人數	0.7	0.6	0.53	0.65	0.62	0.33	0.43
降雨量	0.68	0.63	0.5	0.63	0.73	0.38	0.41
Google trend 搜尋 量	0.67	0.71	0.56	0.62	0.61	0.26	0.38
寒暑假期間	0.69	0.55	0.29	0.45	0.46	0.3	0.33

5.4.4 公園類 YouBike 租用站影響因子重要度探討

表 5-22 為公園類 YouBike 租用站 1-7 月各因子重要度排序，為使數據方便觀看與探討，我們將因子重要度以排名表(表 5-22)與直方圖(圖 5-4)之方式呈現，以進行因子重要度探討。

表 5-22 公園類 YouBike 租用站 1-7 月各因子重要度排序

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
大台北(北北基桃)							
確認診人數	8	5	6	6	6	5	6
非大台北(北北基桃)							
確認診人數	7	8	7	7	7	6	7
境外確診人數	6	7	4	4	5	2	4
前 1 日確診人數	5	4	5	5	4	4	5
前 3 日確診人數	1	3	2	1	2	3	1
降雨量	3	2	3	2	1	1	2
Google trend 搜尋量	4	1	1	3	3	7	3
寒暑假期間	2	6	8	8	8	8	8

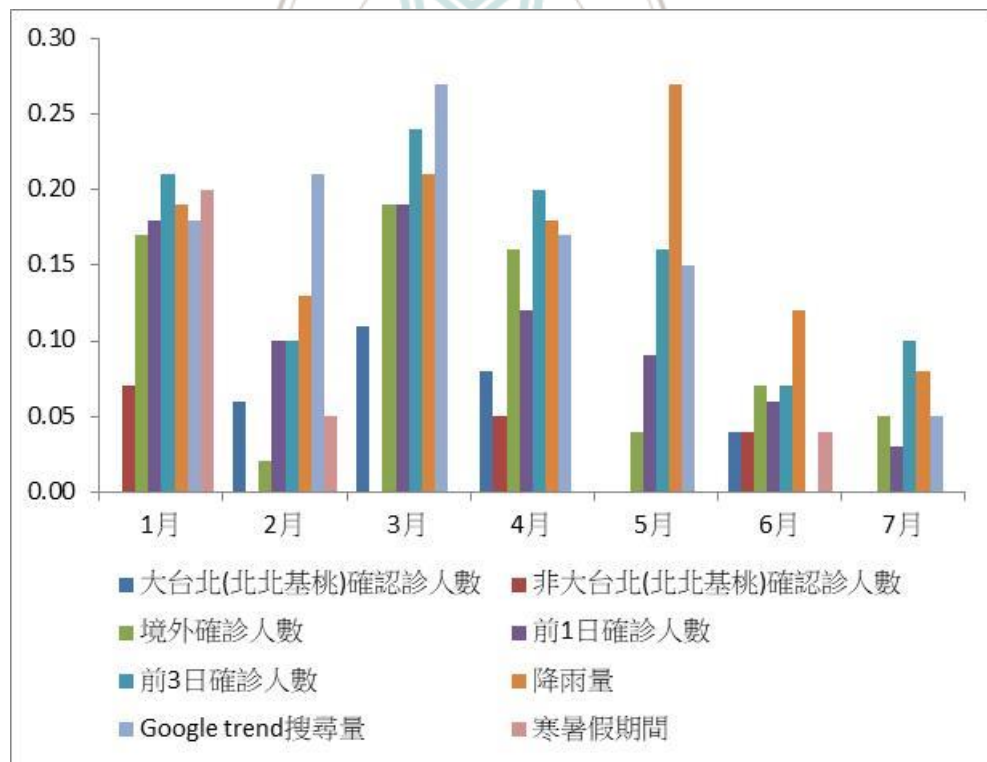


圖 5-4 公園類 YouBike 租用站各月份特徵因子直方圖

從圖 5-4 觀看每月份各因子重要度變化，先行觀看與 Covid-19 無相關之「降雨量」與「寒暑假期間」兩因子，從各因子重要度排名(表 5-22)來看，其中降雨量在 1-5 月之中，重要度數據皆排名前三名，表示有無下雨對於公園戶外活動族租用 YouBike 的行為有高度的影響，且從圖 5-4 搭配圖 5-2 可看出，5 月份之日平均降雨量與下雨天數升高，降雨量與他因子重要度數據差異愈大，而「寒暑假期間」1 月與 2 月之重要度排名在第 2 與第 6 名，其餘月份皆排名最末，其中因 1 月與 2 月份裡的寒假期間學生不用上課，表示寒假期間不用上課對學生提升了外出活動租用 YouBike 的行為是有影響的；當中 1、4、5 月份以「前 3 日確診人數」重要度數據最高，顯示當確診人數累計增加上升，使民眾擔心被傳染的風險提高，進而減少至公園活動的意願，表示確診人數的累計增加對於民眾租用 YouBike 的行為是有影響的；2 月份與 3 月份以「Google trend 搜尋量」重要度數據最高，2 月份為疫情剛出現之初期，民眾從網路上提升對 Covid-19 的瞭解程度後，減少了出門的意願，雖然公園為空曠地區，較密閉空間不易感染，但仍會擔心有感染風險而減少外出，表示民眾對 Covid-19 的瞭解程度與租用 YouBike 的行為是有影響的；3 月為疫情之高峰，使得 Google trend 搜尋量仍維持高檔，顯示民眾仍處於對疫情擔心的狀態與周遭安全的不確定性，減少了外出的意願，表示民眾對 Covid-19 的擔心程度與租用 YouBike 的行為是有影響的。

5.5 學校類 YouBike 租用站影響因子重要度探討

5.5.1 學校類 YouBike 租用站資料分群

以 2020 年 1 月份學校類之 YouBike 租用站資料為例，本研究先以輪廓係數決定 k 值之方式為先將 k 值=2~10 之輪廓係數計算出，再將各 k 值間之輪廓係數進行縮減比例計算，固定取縮減比例>-20%之最大 k 值(表 5-23)。

表 5-23 公園類租用站分群結果_輪廓係數

分群數(k)	輪廓係數	縮減比例
2	1078.339313	
3	741.354609	-31%
4	588.8027857	-21%
5	505.2010451	-14%
6	452.3815852	-10%
7	403.1878251	-11%
8	364.5020512	-10%
9	334.5540088	-8%
10	305.4553261	-9%

在確認 k 值等於 4 之後，我們將使用 k-means 方法來對 4192 筆的樣本進行運算，結果如表 5-24：

表 5-24 學校類 YouBike 租用站 1 月份分群結果

分群	資料筆數	資料佔比
1	330	8%
2	1630	39%
3	1330	32%
4	902	22%

經由上述 k-means 分群做法，我們依序完成各學校類 1~7 月之資料分群，分群結果如表 5-25 所示。

表 5-25 學校類 YouBike 租用站 1-7 月分群結果

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
分群數	4	4	4	4	4	4	4

5.5.2 學校類 YouBike 租用站資料平衡與分群合併縮減

以 2020 年 1 月份學校類之 YouBike 租用站資料為例，經由第 4 章「4.4.3 分群資料筆數平衡與重新分群合併縮減」舉例之做法，我們依序完成學校類 1~7 月之分群資料，最終 1-7 月分群合併縮減結果如表 5-6 所示。

表 5-26 學校類 YouBike 租用站 1-7 月分群合併縮減結果

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
分群數	2	3	2	3	3	3	4

5.5.3 學校類 Youbike 租用站特徵值資料進行隨機森林訓練與因子重要度計算

以 2020 年 1 月份學校類 YouBike 租用站為例，我們將每群資料筆數平衡處理後，再將每筆特徵值資料與分群結果結合後之數據(表 5-27)丟入隨機森林觀看分類結果與特徵因子重要度，我們使用 80%的資料進行訓練與 20%的資料進行測試，產出之混淆矩陣結果如表 5-28，觀看表 5-28 之數據，第 1 群 57 筆資料裡有 52 筆第預測在第 1 群，第 2 群 63 筆資料裡有 42 筆第預測在第 2 群，全部 120 筆資料裡共有 94 筆分類正確，整體分類之正確率達 78%，表示對於隨機森林的分類結果是可使用的。

接下來我們觀看各因子重要度(表 5-29)，可得出學校類租用站 2020 年 1 月份以「前 3 日確診人數」之因子重要度影響最高，而大台北(北北基桃)確認診人數影響最低。

表 5-27 學校類 YouBike 租用站 1 月份特徵值與分群結果結合

大台北 (北北基 桃)確診 人數	非大台 北(北北 基桃)確 診人數	境外確 診人數	前 1 日 確診人 數	前 3 日 確診人 數	降雨量	Google trend	寒暑假 期間	分群標 籤
0	0	0.5	0.33	0.8	0	0.71	1	1
0	0	0.5	0.33	0.6	0.09	0.84	1	1
...
0	0	0	0.67	0.6	0.03	0.85	1	2
0	0	0	0.33	0.2		0.41	1	2

表 5-28 學校類 YouBike 租用站 1 月份資料混淆矩陣分類結果

Predicted label	1	52	21
	2	5	42
		1	2
True label			

表 5-29 學校類 YouBike 租用站 1 月份因子重要度

特徵因子	因子重要度
大台北(北北基桃)確認診人數	0.48
非大台北(北北基桃)確認診人數	0.52
境外確診人數	0.65
前 1 日確診人數	0.70
前 3 日確診人數	0.75
降雨量	0.68
Google trend 搜尋量	0.73
寒暑假期間	0.74

經由上述做法，我們依序完成各 1-7 月之隨機森林訓練，得出各月份混淆矩陣分類結果(表 5-30)與各月份特徵因子之重要度(表 5-31)，由表 5-30 可看出 2 月份之分類準確度低於 50%，2 月為寒假期間，學生人數大幅減少，故不對 2 月的特徵因子重要度做探討；6、7 月份之分類準確度低於 50%，表示隨著疫情趨緩，

與 Covid-19 有相關之六項特徵因子對於此 2 個月的學校類學生族租用 YouBike 的行為已無影響，故對於此 2 個月的特徵因子重要度探討將不予討論。

表 5-30 學校類 YouBike 租用站資料 1-7 月混淆矩陣分類結果

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
捷運	78%	47%	64%	55%	53%	47%	34%

表 5-21 學校類 YouBike 租用站 1-7 月因子重要度

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
大台北(北北基桃) 確認診人數	0.48	0.35	0.46	0.37	0.33	0.28	0.21
非大台北(北北基 桃)確認診人數	0.52	0.31	0.44	0.40	0.33	0.28	0.21
境外確診人數	0.65	0.31	0.60	0.49	0.33	0.35	0.21
前 1 日確診人數	0.70	0.33	0.50	0.43	0.40	0.29	0.23
前 3 日確診人數	0.75	0.40	0.53	0.53	0.41	0.31	0.24
降雨量	0.68	0.39	0.58	0.53	0.51	0.45	0.34
Google trend 搜尋 量	0.73	0.41	0.58	0.46	0.39	0.39	0.25
寒暑假期間	0.74	0.31	0.44	0.33	0.33	0.28	0.21

5.5.4 學校類 Youbike 租用站影響因子重要度探討

表 5-32 為學校類 Youbike 租用站 1-7 月各因子重要度排序，為使數據方便觀看與探討，我們將因子重要度以排名表(表 5-32)與直方圖(圖 5-5)之方式呈現，以進行因子重要度探討。

表 5-32 學校類 YouBike 租用站 1-7 月各因子重要度排序

	1 月	2 月	3 月	4 月	5 月	6 月	7 月
大台北(北北基桃)							
確認診人數	8	4	6	7	5	6	5
非大台北(北北基桃)							
確認診人數	7	6	7	6	6	7	6
境外確診人數	6	7	1	3	7	3	7
前 1 日確診人數	4	5	5	5	3	5	4
前 3 日確診人數	1	2	4	1	2	4	3
降雨量	5	3	2	2	1	1	1
Google trend 搜尋量	3	1	3	4	4	2	2
寒暑假期間	2	8	8	8	8	8	8

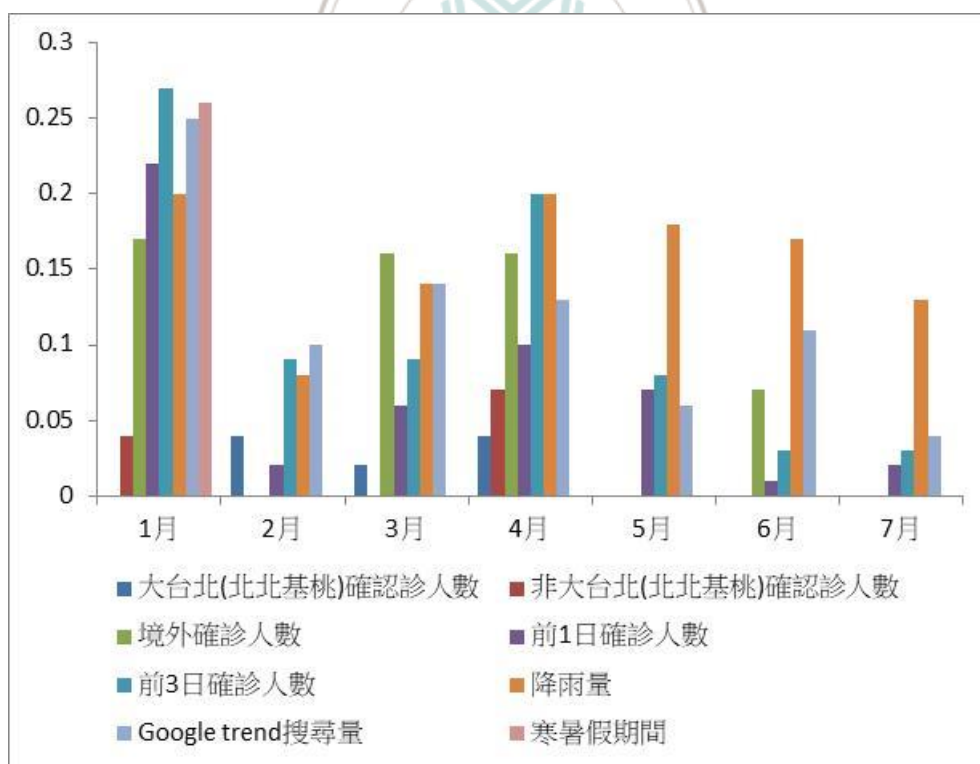


圖 5-5 學校類 YouBike 租用站各月份特徵因子直方圖

從圖 5-5 觀看每月份各因子重要度變化，先行觀看與 Covid-19 無相關之「降雨量」與「寒暑假期間」兩因子，從各因子重要度排名(表 5-32)來看，在 1-5 月之中，除 1 月分外其餘月份重要度數據皆排名前 3 名，表示有無下雨對於學校學生租用 YouBike 的行為有高度的影響；「寒暑假期間」1 月之重要度排名在第 2 名，其餘月份皆排名最末，表示有無上課對學生租用 YouBike 的行為是有影響的；1、4、5 月份以「前 3 日確診人數」重要度數據最高，顯示當確診人數持續增加，使民眾擔心被傳染的風險提高，部分家長可能傾向親自接送上下課，表示確診人數的累計增加對於學生租用 YouBike 的行為是有影響的；3 月份以「境外確診人數」重要度數據最高，由圖 5-3 可看出 3 月份境外確診人數整體升高至 258 例，面對突然升高之疫情導致民眾心理恐慌，從而影響學生去學校上課的方式(家長接送頻度提高)，表示面臨疫情的突然攀升對於學生租用 YouBike 的行為是有影響的。

第六章 結論

6.1 研究結論

本研究運用集群分析法以及隨機森林以大數據分析的分式進行資料分析，經過此分析方式，我們發現 Covid-19 疫情的起伏對於民眾租用 YouBike 的行為是有影響的，而影響的層面包含民眾對疫情資訊的搜尋熱度、當日發布之確診人數、累計的確診人數等，其中捷運周圍的 YouBike 站點，當疫情發生時，以「Google trend」搜尋量對民眾租用 YouBike 的行為影響最大，而在 3 月份疫情高峰時，以「境外確診人數」民眾對租用 YouBike 的行為影響最大；公園周圍的 YouBike 站點，當疫情發生時，以「前 3 日確認人數」對民眾租用 YouBike 的行為影響最大，而在 3 月份疫情高峰時，以「Google trend」搜尋量民眾對租用 YouBike 的行為影響最大，學校周圍的 YouBike 站點，當疫情發生時，以「境外確診人數」民眾對租用 YouBike 的行為影響最大，而在 3 月份疫情高峰時，以「境外確診人數」民眾對租用 YouBike 的行為影響最大，最後 6 月與 7 月隨著疫情降溫，隨機森林分類預測結果皆未達 50%，表示與 Covid-19 有關之因子，對於民眾租用 YouBike 之行為已無影響，故本研究認為以資料集的角度探討 Covid-19 對於臺北市民眾生活習性的影響是可行的，此結果也能提供政府防疫政策的推行或 YouBike 公司的商業策略運用。

6.2 研究建議

因考量實驗的時間與成本之因素，本研究是將各 YouBike 租用站依其名稱或性質做分類後，僅取總租用量加總前 3 名之 YouBike 類別租用站做資料探討，無法將全部 YouBike 類別進行探討實為研究上的缺憾，後續建議可從不同方式為 YouBike 租用站進行分類，例如可依臺北市行政區(信義區、萬華區、北投區...等)做為 YouBike 租用站之分類方式，而各行政區皆涵蓋不同類別之 YouBike 租用站，相信對於單一類別之探討或許更有參考價值；2021 年 Covid-19 疫情相較於 2020 年更為嚴峻，本研究 YouBike 資料僅至 2020 年 7 月，實為可惜，倘若可取得 2021 年整年份之資料，再搭配不同的 YouBike 租用站分類進行探討，相信可以讓此份研究更為豐富。



參考文獻

1. 余書玫(2009)，公共自行車租借系統選擇行為之研究，國立交通大學交通運輸研究所。
2. 黃仁皇(2009)，公共自行車騎乘特性、服務便利性、騎乘滿意度之相關研究—以台北市微笑單車為例。
3. 劉仁龍、鄭麗淑(2017)，大數據分析臺北市公共自行車使用特性，臺北市政府交通局統計室編號：106—18。
4. 李俊毅(2019)，臺北市空氣品質與天氣概況影響公共自行車租借行為之探討國立臺灣師範大學地理學系。
5. 葉耀墩、林昭賢、劉心荷(2015)，公共自行車 YouBike 營運資料分析及精進作為，新北市政府年度自行研究報告。
6. 鍾智林、簡佑勳(2014)，公共自行車時空分析法之構建與營運策略改善 --以臺北微笑自行車為例，都市交通 29:1 2014.06[民 103.06] 頁 1—10。
7. 黃晏珊(2015)，臺北市公共自行車系統營運特性分析，淡江大學運輸管理學系碩士班。
8. 徐翊菁(2020)，基於隨機森林分類之心房顫動發作預測，國立臺灣海洋大學電機工程學系。
9. 呂昀(2019)，應用隨機森林預測線上飯店之點閱與預訂，國立臺北科技大學經營管理系。
10. 翁雅君(2015)，透過時間序列方法分析 YouBike 使用狀況：以信義區為例，國立雲林科技大學資訊管理系。

11. 曾怡之(2019)，以租用時間探討公共自行車使用者移動的時空特徵，中國文化大學地學研究所地理組。
12. 鍾佩綦(2021)，隨機森林法建構之銷售預測模型之研究—以電商平台「一抹丘」為例，中原大學國際經營與貿易研究所
13. 葉子維(2018)，顧客消費行為分析及行動銀行使用預測-決策樹、隨機森林與判別分析之比較，國立台北大學統計學系
14. 郭家姩(2014)，隨機森林在河川水位即時預報之應用，國立成功大學水利及海洋工程學系
15. Ward, J.H.(1963), "Hierarchical grouping to optimize an objective function" Journal of the American Statistical Association, Vol. 58, No. 301, pp. 236-244.
16. MacQueen, J. B. (1967), "Some methods for classification and analysis of multivariate observations", in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol.1, pp. 281-297.
17. Punj, G. and Stewart, D. W. (1983) Cluster analysis in marketing research: review and suggestions for application, Journal of Marketing Research, pp. 134-148.
18. 吳伯書(2019)，少年犯罪高風險集群分析研究-以新北為例，國防大學資訊管理系
19. 蔡儀君(2016)，應用集群分析對於智慧型手機用目的之探討，國立政治大學統計學系
20. 林佳薇(2015)，以分位數迴歸及集群分析探討台灣地區時間與空間乾旱趨勢變化，國立成功大學水利及海洋工程學系

21. 凌子平(2015)，應用空間限制集群分析法劃設臺灣生活圈，國立成功大學都市計劃學系
22. 羅婉瑞(2010)，以集群分析探討航空公司營運績效，國立臺灣海洋大學航運管理系
23. 黃琮智(2019)，隨機森林結合基因演算法於鐵達尼分類問題，臺北市立大學資訊科學系碩士在職專班
24. 許宸綺(2021)，用心跳間隔時間之自身相關性質診斷充血性心力衰竭-以區辨分析、決策樹與隨機森林及比較，東海大學統計學系
25. Python 機器學習實踐：隨機森林算法訓練及調參(無日期)
<https://read01.com/ez6Aam6.html>

