

第 M 組 MDS 期末報告

國企四 許捷翔、國企四 古森允

國企三 朱柏諺、國企三 吳政諒、資管三 蔡可亮

一、研究動機

共享單車不僅提供了便利的交通選擇，還有助於減少城市的交通擁擠和環境污染。然而，共享單車運營商需要有效地管理單車站點，以確保車輛和停車位的有效使用，同時提高營業利潤。為了實現這一目標，本專案旨在最佳化共享單車站點的運營，以提高借用量、會員數和降低成本。

二、研究背景

共享單車已經成為現代城市中的一種受歡迎的運輸方式。在臺灣規模最大的共享自行車為 YouBike，但是詳細的資料難以蒐集。因此本組改以紐約市共享單車系統，同時也是美國最大的共享單車系統的 Citi Bike 借用資料做分析，了解影響 Citi Bike 的靜態與動態變因並建立預測未來借用量，以規劃人力與資源的調度。

三、問題定義

目標展開議題樹如圖一所示：



(圖一)

第一層將利潤展開為提升營收與成本。首先討論營收的部分，我們希望透過平衡站點的供給與需求，讓每個區域的站點都有車可借。為此我們提出的解決方案是預測下一季的常駐停車柱數量與預測各時段所需的調度人員數量。

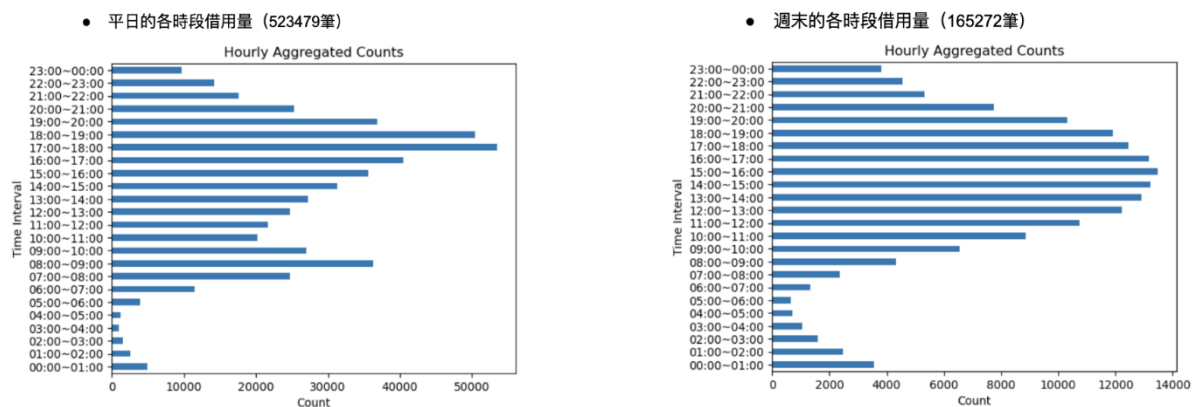
降低成本面，我們又再細分為降低人力調度成本與降低會員制度的回饋金兩部分：

在降低人力調度成本部分，我們分析了時間相關的動態資料對需求量的影響，以此與預測下一季各時段的常駐與臨時人員的配置與建立小區間的需求預測模型。在降低會員調度回饋金部分，Citi Bike 會提供幫忙調度車輛到熱門區域的會員回饋金，因此我們希望透過這樣的區間預測模型能控制會員回饋金發放。

四、研究方法

(一) 資料探索

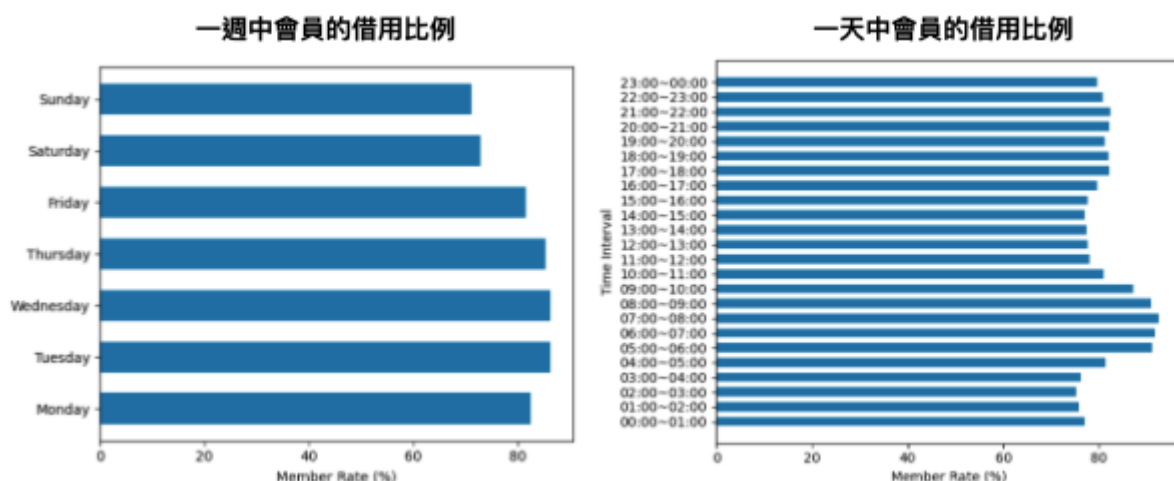
本研究最主要的資料集為 kaggle 上的 2023 年五月 Citi Bike 的借用資料。為了能夠更準確的掌握 Citi Bike 用戶借用單車的狀況，我們分別分析平日與假日各時段的單車借用狀況，分析結果如下圖（圖二）所示：



(圖二)

(圖二) 結果判讀：可以發現平日的 8:00~9:00 以及 16:00~19:00 為借用的尖峰時段；假日的 14:00~17:00 為借用的尖峰時段。推測原因為：平日的借用需求源自於上班族有通勤的需求，因此會有較高的借用量需求；假日下午為大眾出門放風的時段，加上並沒有急迫的時間壓力，因此多選擇以騎乘 Citi Bike 作為其代步工具及休閒娛樂。

接著分析 Citi Bike 週一到週日和一天中各時段的會員借用比例，以了解會員和非會員各自的借用狀況，分析結果如（圖三）所示：



(圖三)

(圖三) 結果判讀：可以發現平日會員的借用比例（超過 80%）多於假日（70%），而一天中早上的會員比例最高（超過 90%）、下午和凌晨的會員比例最低（75%），推測上班族為會員的比例高，而週末將共享單車做為休閒娛樂工具的使用者是會員的比例低，因此可以選擇在週末下午的時段安排員工在熱門站點駐點，先了解借用者對於站點增設和車量調度的需求後，再推廣其加入會員，以提高使用者的顧客終生價值和增加公司的營收。

1. 2023 年 5 月的 Citi Bike 借用資料以及 2021 年 11 月到 2023 年 10 月的 Citi Bike 借用資料

本研究 2021 年 11 月到 2023 年 10 月的 Citi Bike 借用資料的資料來源為 Citi Bike NYC 的官網。前者主要用於後續的各站點的借用量預測以及停車柱、常駐人員調整；後者則用於與週期性變數 (如天氣) 較為 LSTM 模型及時預測模型。（由於資料集過大，因此未上傳至繳交區）

2. 天氣資料

本研究天氣資料的來源為 timeanddate.com，使用其歷史觀測資料，經過預處理成虛擬變數之後包含以下欄位：

欄位名稱	Temp	Wind	Humidity	Barometer	Visibility	Weather Description (Dummy)
單位	°C	km/h	%	mbar	km	不適用
範例值	20	9	72	1010	16	Heavy snow, Sunny, Rainy...計 30 種

3. 車流量資料

本研究車流量資料來源為 NYC OpenData，透過網頁中的資料庫篩選功能，篩選出 2023 年 5 月的曼哈頓市中心三個 Neighborhood Areas 的平均車流量資料，並星期以及時段欄位讓該資料集方便與主要的資料集（2023 年五月 Citi Bike 的借用資）做整合。我們也將 WktGeom 欄位轉換為易於視覺化的經緯度格式，並創建 Longitude、Latitude 欄位。

欄位名稱	weekday	hour	Longitude	Latitude	average_vol	neighbourhood
單位	不適用	不適用	不適用	不適用	輛 / hr	不適用
範例值	Friday	00:00 ~ 01:00	-73.98975655	40.77256852	22.25	Upper West Side

4. 地區人口特徵、電信以及公共建設普及率資料

本研究地區人口特徵、電信以及公共建設普及率資料來源為 NYC OpenData，透過網頁中的資料庫篩選功能篩選曼哈頓的資料，其中包含近 30 個曼哈頓的 Neighborhood Areas 的各項資料，其中跟借用量較為相關的變數為人口密度、公共建設數量和公共建設普及率資料。

為了方便定位 Neighborhood Areas 以減少後續資料整合的複雜性，因此取該 Neighborhood Areas 的中心點為其經緯度座標，並新增兩欄位至資料表。另外，也透過資料表中的 NTA（Neighborhood Areas 的編號）欄位對照出相應的 Neighborhood Areas 並新增該欄位，以利後續與其他資料表相互整合。下表整理幾個比較具有代表性的欄位為例：

欄位名稱	Total Population	Population Density (per Sq. Mi.)	Low-Income Housing	Number of Public Schools	Street Furniture: Bike Shelters	Empire City Subway Coverage
單位	人	人/m^2	棟	所	個	不適用
範例值	52,529	82,076.56	2,270	20	3	Empire City Subway Coverage

(二) 資料預處理

天氣資料由於 timeanddate.com 並非固定觀測，雖資料多集中在 51 分觀測，但仍不免出現一個小時觀測多次，或連續多小時未觀測的狀況，因此為了能夠得到連續的每小時資料，缺失值的補齊方法如下：

- 一小時多筆資料：

連續變數：取平均值

Dummy Variables：取最大值（即該小時出現過的天氣都算入）

- 該小時沒有資料：

連續變數：取前一筆資料與後一筆資料的線性平均，如 01:00 時溫度 10 度，下一筆資料為 04:00 溫度 16 度，則 02:00 為 12 度、03:00 為 14 度。

Dummy Variables：取前一筆資料與後一筆資料的中間作為分割點

(三) 預測曼哈頓各地 (Neighborhood Areas) 下一季的使用量

資料範疇

此部分採用的資料集為「2023 年 5 月的 Citi Bike 借用量資料」和「地區人口特徵、電信以及公共建設普及率資料」所合併而成的資料集。另外，我們也再從「2023 年 5 月的 Citi Bike 借用量資料」中計算出各地的會員組成比例，並新增該欄位。原始資料集的變數總數為 125 個。

資料前處理

1. 變數初步挑選

因為部分欄位為電信及施工成本相關資料，對於使用量的預測幫助不大，因此從 125 個地區資料中篩選出 44 項影響使用量的關鍵因素。

2. 數據轉換

因為每個地區的大部分的人口特徵資料在不同筆資料中值都是相同的，因此將這些欄位資料集用 NTA 欄位組成群組，並用取平均的方式來整合這些值；各地建築類型的資料由於資料維度相較於 Neighborhood Areas 更為細緻，所以將這部分的欄位資料集用 NTA 欄位組成群，並用加總的方式來整合這些值。

3. 刪除高相關變數

由於許多的變數之間都存有高度相關性問題，若刪除所有具有高度相關性問題的變數，將大幅降低資料集的變數量。因此，我再調整過刪除變數門檻多次後，最終刪除同時與其他 3 個變數相關性達 0.7 以上之變數，讓資料集可以呈現大部分的資訊量，而不至於因共線性問題而衍生過度配適或多重共線性風險。

4. 重要變數挑選

由於在進行 VIF 多重共線性檢定後，我們發現部分變數隱含多重共線性問題。因此，我們選用 Lasso Regression 來挑選重要變數，同時幫助降低資料集中的多重共線性問題。最終，納入模型的變數量只剩 37 個。

線性回歸分析結果 (Fordward Selection / Pc = 0.05)

OLS Regression Results						
Dep. Variable:	group_count	R-squared:	0.973			
Model:	OLS	Adj. R-squared:	0.965			
Method:	Least Squares	F-statistic:	124.3			
Date:	Sat, 09 Dec 2023	Prob (F-statistic):	2.69e-15			
Time:	21:34:20	Log-Likelihood:	-301.95			
No. Observations:	28	AIC:	617.9			
Df Residuals:	21	BIC:	627.2			
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-8516.5871	7836.662	-1.087	0.289	-2.48e+04	7780.643
bldgclass_sum	0.0271	0.002	12.173	0.000	0.022	0.032
Street Furniture: Parking Pay Stations	137.5436	26.332	5.223	0.000	82.783	192.304
bldgclass_W	-643.7581	182.306	-3.531	0.002	-1022.883	-264.633
Empire City Subway Coverage (Percentage)	4.065e+04	1.22e+04	3.328	0.003	1.52e+04	6.61e+04
Low-Income Housing (NYCHA)	1.3834	0.480	2.881	0.009	0.385	2.382
bldgclass_Q	-313.8403	124.751	-2.516	0.020	-573.274	-54.407
Omnibus:	1.793	Durbin-Watson:	2.253			
Prob(Omnibus):	0.408	Jarque-Bera (JB):	0.829			
Skew:	0.389	Prob(JB):	0.661			
Kurtosis:	3.322	Cond. No.	1.33e+07			

(圖四)

結果判讀：透過以上線性回歸分析所呈現之結果，我們可以簡單依各變數的回歸係數分成兩類 — 建築物總數、付費停車場數、地鐵覆蓋率及低收入戶數係數為正；教育機構數、戶外休閒設施數係數為負。以下將針對這些變數的係數各自做判讀：

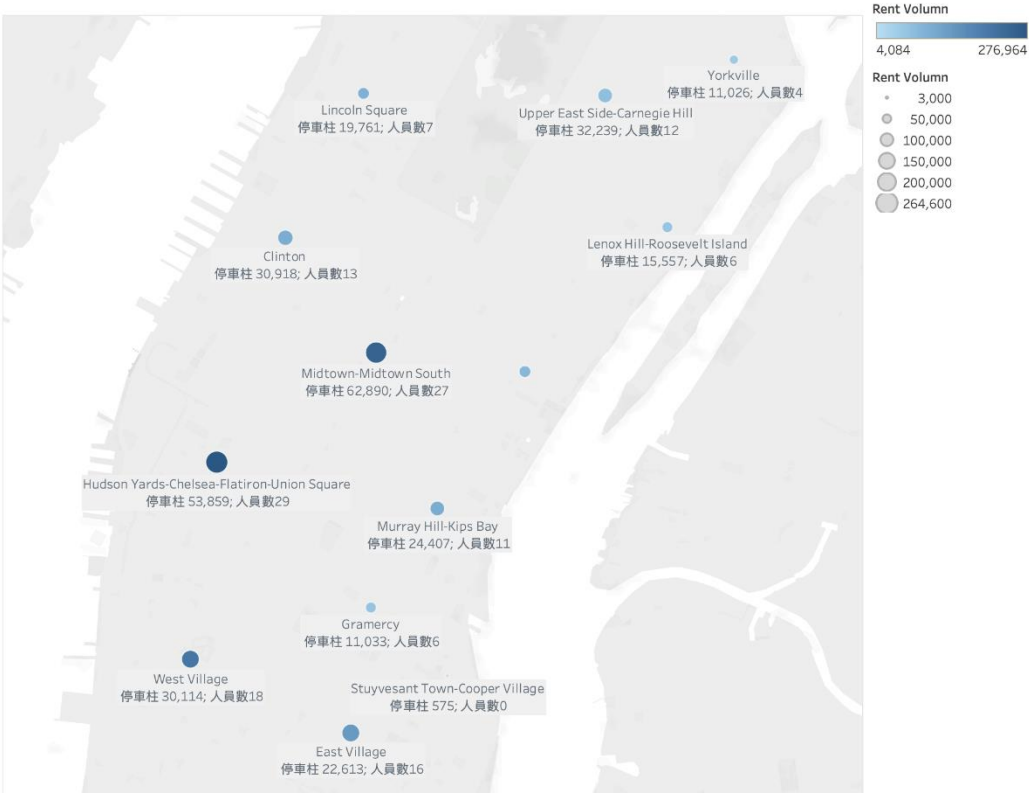
- 建築物總數的係數為正：推測原因為建築物總數越多的地區，人口總數也會越多，因此騎乘 Citi Bike 的需求也大幅增加。
- 付費停車場數的係數為正：推測原因為付費停車場數多的地區商業行為較為蓬勃發展，因此將吸引大量人潮前往該地區，而讓 Citi Bike 的借用需求大幅增加。
- 地鐵覆蓋率的係數為正：推測原因為通勤族常常會先從住家騎乘 Citi Bike 至地鐵站，接著再從目的地的地鐵站騎乘 Citi Bike 至公司辦公，所以讓 Citi Bike 的借用需求大幅增加。
- 低收入戶數係數的係數為正：推測原因為 Citi Bike 相對於其他的大眾運輸工具有較低的成本，因此低收入戶的家庭成員更傾向以 Citi Bike 作為其代步工具。
- 教育機構數的係數為負：推測原因為學校的腹地範圍中鮮少有適合 Citi Bike 騎乘的路段，學生或是訪客大多以步行的方式在校園內移動。加上固定使用自行車在校園間穿梭的同學也更傾向於購買自行車而非租用，因此讓 Citi Bike 的借用需求大幅降低。
- 戶外休閒設施數的係數為負：推測原因為騎乘 Citi Bike 也是一種休閒娛樂的方式，因此若是戶外休閒設施數增加，相當於讓 CikeBike 的替代品變數，因而讓 Citi Bike 的借用需求大幅降低。

模型之延伸應用

將下一季的建築物總數、付費停車場數、地鐵覆蓋率、低收入戶數、教育機構數、戶外休閒設施數等變數的預測值納入線性回歸模型，便能求出下一季各地的借用需求量。同時，我們也加入一些簡單的運算邏輯（圖五），得出下一季各地區的停車柱需求以常駐調度人員數量（圖六）。

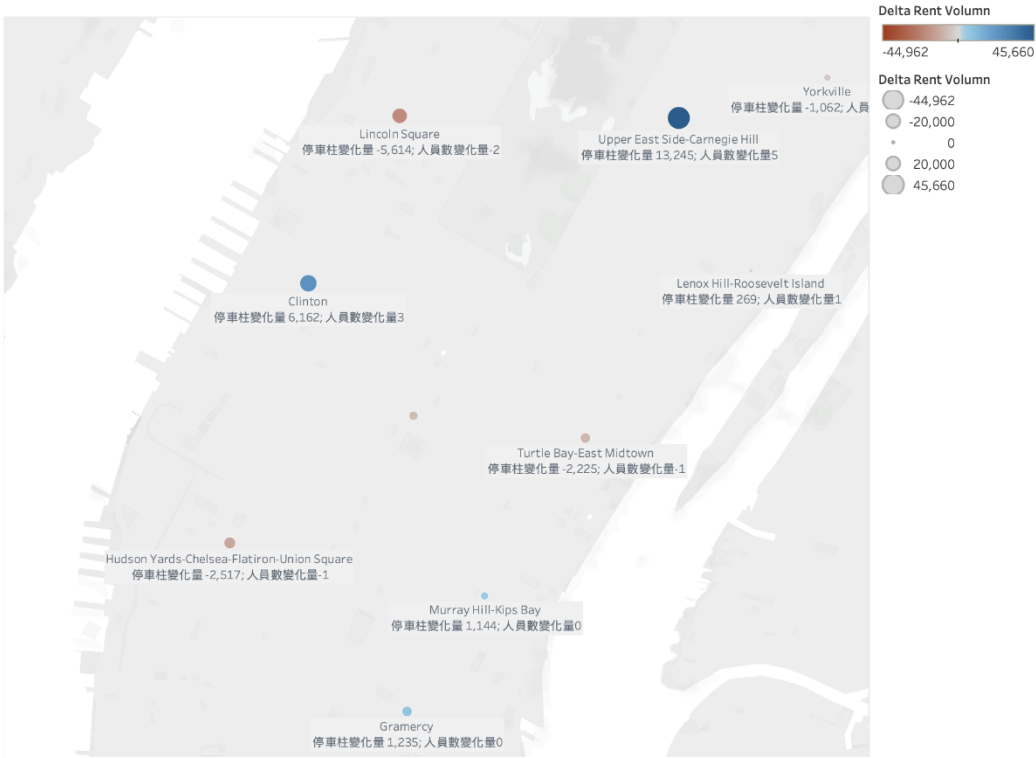
借用量跟停車柱數量之間的轉換方法		
1	假設	各地當前的停車柱數量已經達到最配置： $(\text{各地當前借用量}) / (\text{當前停車柱數量}) = \text{各地最適比例}$
	說明	預測出各地的借用量以後，再將借用量除以各地的最適比例，得出該季各地預計配置的停車柱數量
借用量跟常駐調度人員數量之間的轉換方法		
2	假設	Citibike約5%的營收花費在調度人員上，因此 $(\text{借用量} * \text{單次借用金額} * 5\%)$ 得各站點人力成本
	說明	$(\text{調度人員平均薪資} / \text{各站點人力成本}) = \text{各站點的常駐調度人員數量}$
	實務	由於上述變數除了「借用量」其他都是固定的，所以直接在計算完一個站點後，依循規律把各地區的 $(\text{借用量} / 9000)$ 得每個地區的常駐調度人員數量。

（圖五）



（圖六）

除此之外，為了方便企業端進行決策，因此我們將下一季所預測出的借用需求量、停車柱需求以常駐調度人員數量與上一季的數據相減，得出下一季借用需求量、停車柱需求以常駐調度人員數量的增（減）量（圖七）。



（圖七）

資料前處理

1. 整理天氣變數

針對二元的天氣變數，將 2023/5 的 15 種進行獨熱編碼，另外由於每小時最多只會有兩種天氣變數，因此進行特徵融合，以避免因絕大部分的天氣變數為 0 所導致之共線性問題。

最後整理成 4 大天氣變數，且若任一個子分類為 1 則新變數亦為 1，分類方式如下：

- Sunny：Clear、Sunny、Partly sunny、Passing clouds、Scattered clouds
- Cloudy：Partly cloudy、More clouds than sun、Mostly cloudy、Broken clouds、Cloudy、Overcast、Low clouds
- Rainy：Light rain、Rain
- Foggy：Fog

2. 進行分群和數據轉換

將 2023/5 於曼哈頓之 219 萬筆借用樣本資料依照每天每小時進行分群，計算出每天每小時的借用量、借用之會員比例，並將濕度轉換成百分比、車流量則取時段內樣本的平均值。

3. 遺漏值填補

數據集中僅有風力有遺漏值，因此針對缺漏值的部分取前一小時和後一小時的平均值，而連續缺漏的時段由於前後風力皆為 0，因此將其直接補 0。

4. 共線性檢定

針對連續變數的共線性問題，將其進行 VIF 檢定，並刪除結果大於 10 的變數，包含氣壓、能見度。

線性回歸分析結果

- 每小時借用量與天氣變數和車流量之間的關係（圖八）

	coef	std err	t	P> t	[0.025	0.975]
const	-2387.4228	989.353	-2.413	0.016	-4329.717	-445.129
average_vol	14.2424	0.397	35.895	0.000	13.463	15.021
Sunny	9.4882	868.171	0.011	0.991	-1694.903	1713.879
Cloudy	244.8884	859.900	0.285	0.776	-1443.264	1933.041
Rainy	-1317.0857	269.146	-4.894	0.000	-1845.471	-788.700
Foggy	-316.1605	862.813	-0.366	0.714	-2010.031	1377.710
Temp	132.7179	12.669	10.476	0.000	107.847	157.589
Wind	-5.8137	9.207	-0.631	0.528	-23.890	12.262
Humidity	-3.1878	3.695	-0.863	0.389	-10.442	4.067

（圖八）

結果判讀：透過 OLS 回歸模型，並將顯著水準設為 0.05，可以推論當車流量越多時，借用量也會越高、雨天時借用量每小時會減少 1317 次、溫度每上升一度，借用量會增加 132 次。

若將一天中分成四時段進行 OLS 回歸分析，檢視其與車流量、天氣變數之間的關係，顯著水準設為 0.05 時可推得之結果如下：

1. 整體：溫度、車流量上升，借用量皆會增加
2. 凌晨時段 (1-6 時)：其他變數不顯著，推論凌晨時段使用者最不受天氣因素影響
3. 早晨時段 (7-12 時)：晴天借用量增加 570 次
4. 下午時段 (13-18 時)：晴天借用量增加 1173 次、陰天借用量增加 967 次、雨天借用量減少 1661 次，推論下午時段的使用者對於天氣因素的變化最敏感
5. 晚上時段(19-24 時)：雨天借用量減少 1760 次，推論晚上的使用者較擔心雨天視線不佳，而選擇不騎乘共享單車

- 每小時會員比例與天氣變數和車流量之間的關係（圖九）

	coef	std err	t	P> t	[0.025	0.975]
const	85.1151	4.324	19.684	0.000	76.626	93.604
average_vol	0.0278	0.002	16.017	0.000	0.024	0.031
Sunny	0.8338	3.794	0.220	0.826	-6.615	8.283
Cloudy	0.3316	3.758	0.088	0.930	-7.046	7.710
Rainy	-0.2181	1.176	-0.185	0.853	-2.527	2.091
Foggy	1.8067	3.771	0.479	0.632	-5.596	9.210
Temp	-0.7897	0.055	-14.263	0.000	-0.898	-0.681
Wind	-0.0009	0.040	-0.022	0.983	-0.080	0.078
Humidity	0.0546	0.016	3.379	0.001	0.023	0.086

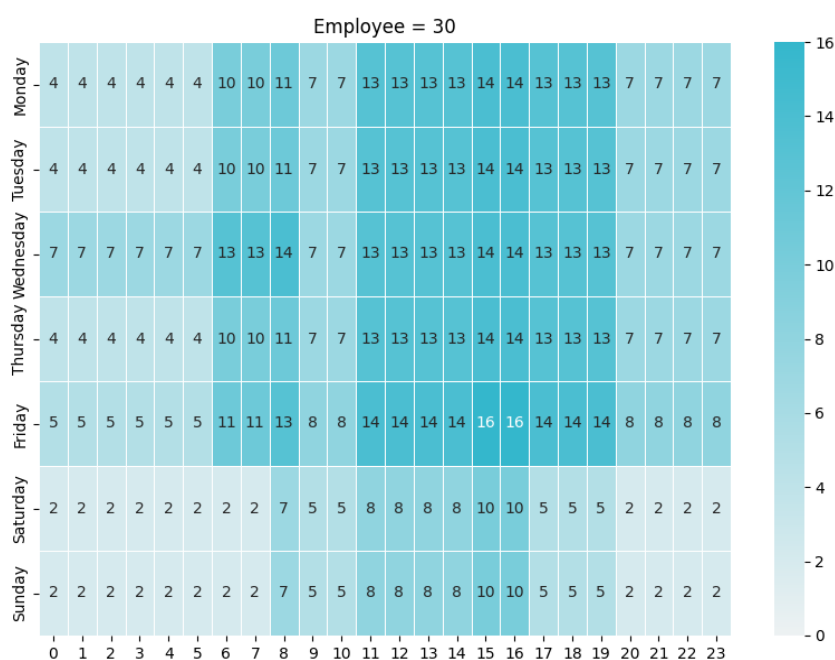
（圖九）

結果判讀：透過 OLS 回歸模型，並將顯著水準設為 0.05，可以推論當車流量越多、濕度越高時，會員的借用比例也會越高，而溫度越高時，會員借用的比例則越低。因此，對於會員而言，即使是溫度低、濕度高，仍會選擇騎乘共享單車，對於天氣的敏感度較低，而非會員的天氣敏感度較高。

模型之延伸應用

根據不同區域所需之正職員工人數，在考量不同時段和天氣下的借用需求後，規劃適合之員工的班表作為處方性分析的結果。(圖七)

- 當員工人數達 30 人時，各時段的員工人數



（圖十）

- 員工分配原則

整體：每位員工一週工作 40 小時，每天不超過 8 小時，且平日員工數量多於週末、白天員工數量多於夜晚，並在溫度高時增加員工數量、雨天時減少員工數量。

1. 平日：員工分流上班，以同時滿足上班和下班的調度需求，並於週三和週五增加夜間調度之員工數量，以滿足白天各站點的車輛借用需求，另外在週五安排最多員工，以提前調度車輛為週末準備。

2. 週末：週末下午安排最多員工工作，以滿足大眾使用習慣，並安排員工於溫度高、濕度低的晴朗天氣時，向非會員推廣會員方案。

利用天氣資料預測短期各小時借用量

LSTM 模型

使用資料概述

X：天氣資料以及 isHoliday（是否為國定假日）

Y：['Count']（借用量）

Training Scope：2021/11~2023/10

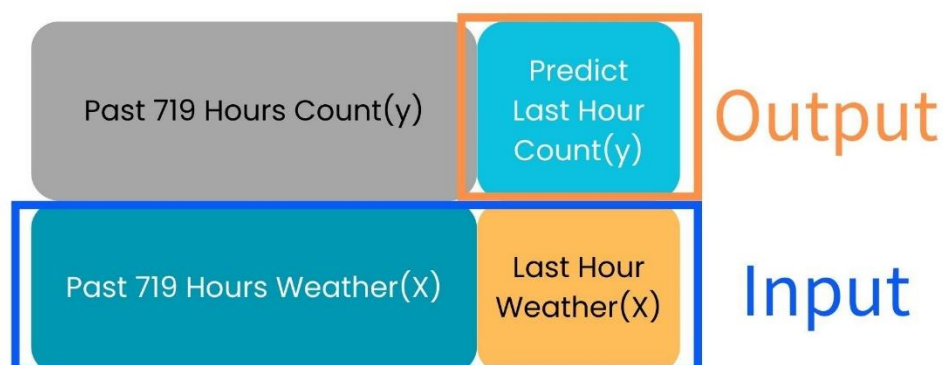
Testing Scope：2023/11

其中以「Weather_」當作 Prefix 的是敘述性的三十種天氣狀況，且部分時刻會包含兩種天氣資料。

Time Series LSTM

Constructing Phase

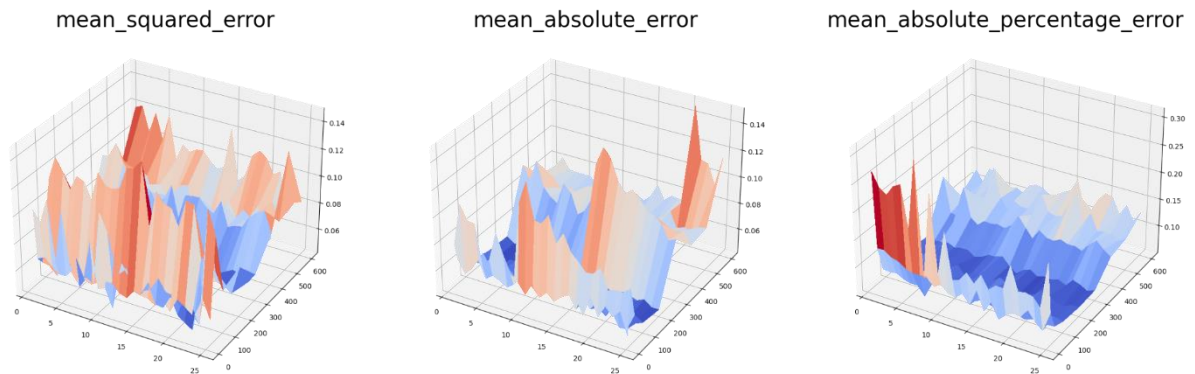
LSTM 的設計為輸入前 720 小時的 X 可以得到這 720 小時中最後一個小時的 y，即 sequence 的 timestep 為 720（如圖）。有別於典型 LSTM 使用前面預測後面的做法，由於被預測的 X 皆為可預期的資料（天氣資料可以混合使用歷史天氣資料與天氣預報、是否為國定假日也已事先決定），此模型把被預測小時的 X 放入輸入。



（圖十一）

Tuning Phase

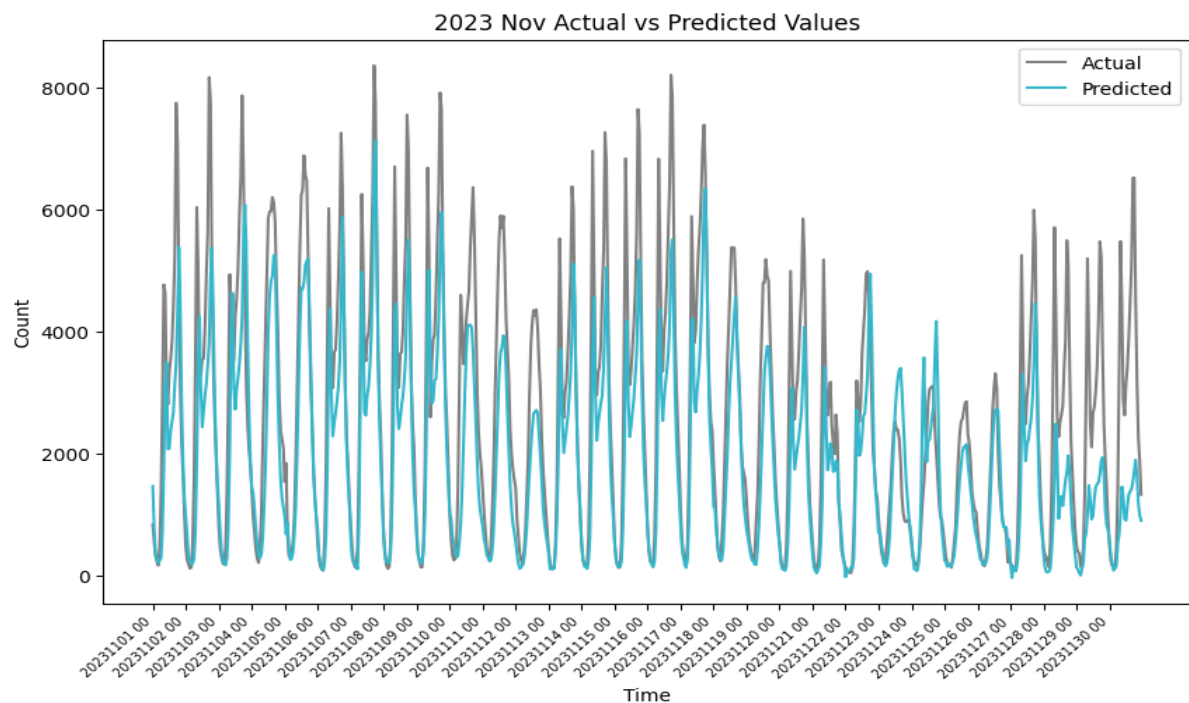
本模型使用的 optimizer 為 Adam，Loss functions 則有經過測試選擇，見下圖：



(圖十二)

上圖的 x 軸為 epoch、y 軸為 batch、z 軸為 Testing Data 的 MSE（三張圖的 z 軸刻度不一樣），可見使用 MAE 作為 Loss Function、batch = 300 時有最低且相對穩定的 MSE，因此選擇這些超參數調整我們的模型。

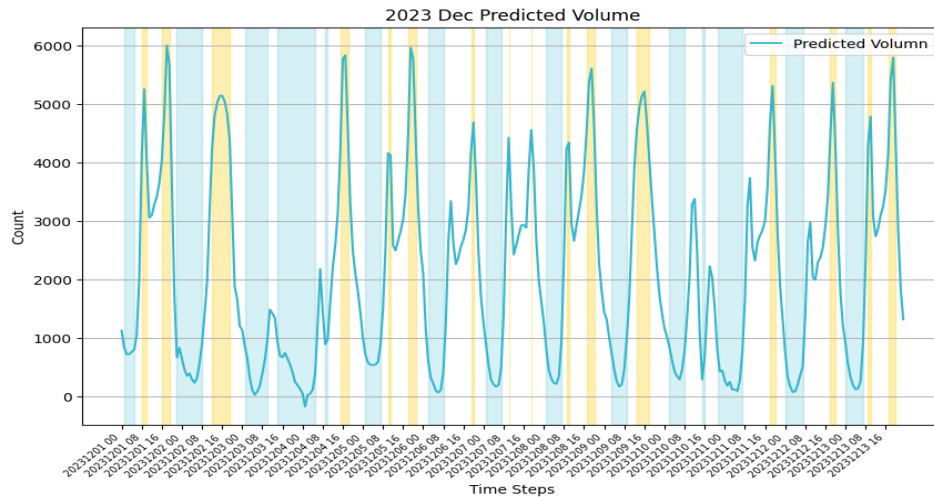
使用此模型的 Testing Data（2023 年 11 月）表現如下（MSE_before_inverse_transform = 0.050742034）：(圖十三)



(圖十三)

由上圖可見，本模型對於每小時的預測有一定準確度，依據本模型以及天氣預報的特性，企業可以用來作為曼哈頓地區總借用量的短期預測，因此，本研究推出的營運改善建議為 **Bike Angel 動態回饋點數調整機制**。

Bike Angel 動態回饋點數調整機制（圖十四）



（圖十四）

LSTM 模型可以預測未來一週的每小時借用量，藉此 Citi Bike 可以提前做好以下準備：首先，預測結果可以用來動態調整 Bike Angels 的回饋點數比率。在預測將面臨較大借用量的尖峰時段，Citi Bike 可以提高回饋點數比率，以 Bike Angels 制度激勵用戶協助車輛調度，降低額外調度成本。而在預測較低借用量的離峰時段，則可以降低點數比率，減少不必要的回饋成本。

其次，預測結果也可以用來評估未來一週臨時工的人力需求。在預期尖峰時段，Citi Bike 可以提前擴聘更多臨時工；而在離峰時段，則可以減少臨時工的預算，如此一來可以更精準地控制人力成本。

為了提高 LSTM 模型的預測準確度，Citi Bike 在模型訓練階段透過調整參數找到誤差最小的模型。更準確的預測結果，可以讓上述營運決策更加可靠。

總結而言，運用 LSTM 模型預測每小時借用量變化，不僅可以幫助 Citi Bike 準確掌握人力和資源需求，也可以透過動態調整點數回饋等機制提高車輛重新調度的效率，從而增進營運效能並降低成本。

五、研究結論

（一）專案總結

本專案旨在以多元線性回歸模型以及 LSTM 模型提升共享單車之人員與車輛調度效率，以此提升 Citi Bike 利潤。

以多元線性回歸模型分別建立了符合曼哈頓各地區的停車柱與常駐人員數量需求的預測模型，以及符合曼哈頓整體各時段的常駐人員調度的班表，分別減少車輛的閒置時間以及提升車輛的調度效率。以 LSTM 模型實施動態點數回饋調整機制以及每週評估臨時調度人員需求，達到精準掌握 Bike Angel 會員調度車輛點數回饋成本以及臨時調度人員預算。

透過以上解決方案幫助 Citi Bike 提升營收、降低成本，最終達到 Citi Bike 提高利潤的目標。

(二) 未來研究建議

資料侷限性：

- 由於缺乏站點停車柱空位資訊，在每一個站點停車柱數量都不同的狀況下，無法確認借用需求與實際需要調度的需求的關係。
- 天氣資料涵蓋範圍：本研究只能尋找到紐約中央公園為觀測站的資料，本模型往後要拓展到紐約其他區域是否能適用有待商榷，且由於天氣資料的本身的侷限性，使得不同地區間若有些微天氣因素差別會影響模型準確度。
- 資料不易維護、預測：以目前天氣預報科技而言，擁有高準確率的時效很短，影響到 LSTM 模型的時效性。此外，由於人口等地區資料為多筆資料集合併而成，原始資料集的更新頻率若不一致將導致不易維護，且各地區因素的預測（例如該地區人口成長率）通常也需要嚴謹研究。

研究未涵蓋的議題：

- 曼哈頓以外的 Citi Bike 資料：由於原始資料集過於龐大，在硬體、計算能力受到限制的情況下，我們只採樣曼哈頓地區的資料分析，未來可以擴大研究範圍探討更多可能的影響因子。
- 研究各站點地理位置的網路關係
- 本資料集有包含每個旅程的起訖點、起訖時間，未來可以使用 GNN 等資料分析這類因子。
- 可以結合本研究提過的方案，使用作業研究的方法規劃調度問題（如調度最小化成本）。
- 亦可將曼哈頓的地形資料併入資料集，分析地形（例如上坡、平均坡度）是否會影響騎乘意願等問題。
- 導入自動化工具：若要應用於實務上，企業可以將本研究的模型、分析等方法組合成自動化的工具，增加營運效率，提升利潤。

六、參考資料

- <https://reurl.cc/yYneWy>
- <https://data.cityofnewyork.us/Transportation/Automated-Traffic-Volume-Counts/7ym2-wayt>
- <https://CitiBikenyc.com/system-data>
- [Citi Bike May 2023 \(kaggle.com\)](https://www.kaggle.com/datasets/citibikenyc/citi-bike-may-2023)
- [How we're rebalancing the Citi Bike system | Citi Bike NYC](https://www.citibikenyc.com/stations-and-bikes)
- [Weather in April 2023 in New York, New York, USA \(timeanddate.com\)](https://timeanddate.com/weather/new-york-usa/april2023)
- [NYC Building Classifications - DOF](https://www.dof.gov/ny/buildings)
- [May 2023 Citi Bike Monthly Report.docx \(mot-marketing-whitelabel-prod.s3.amazonaws.com\)](https://prod.s3.amazonaws.com/mot-marketing-whitelabel-prod.s3.amazonaws.com/May2023CitiBikeMonthlyReport.docx)
- <https://worldpopulationreview.com/states/new-york-population>