

第M組 MDS期末報告

組員：許捷翔、古森允、朱柏諺、吳政諒、蔡可亮

Agenda

1

議題樹發想

2

資料探索及問題定義

3

資料集說明

4

模型解釋及視覺化呈現

5

專案總結

提升 Citibike 的利潤

平衡站點供給需求

分析地區因素
對需求量的影響

預測下一季
各地區停車柱及
常駐調度人員數量

提升營收

降低成本

降低人力調度成本

降低會員調度
回饋點數機制成本

分析時間對需求量的影響

預測下一季各時段
常駐人員配置以及
臨時人員的需求量

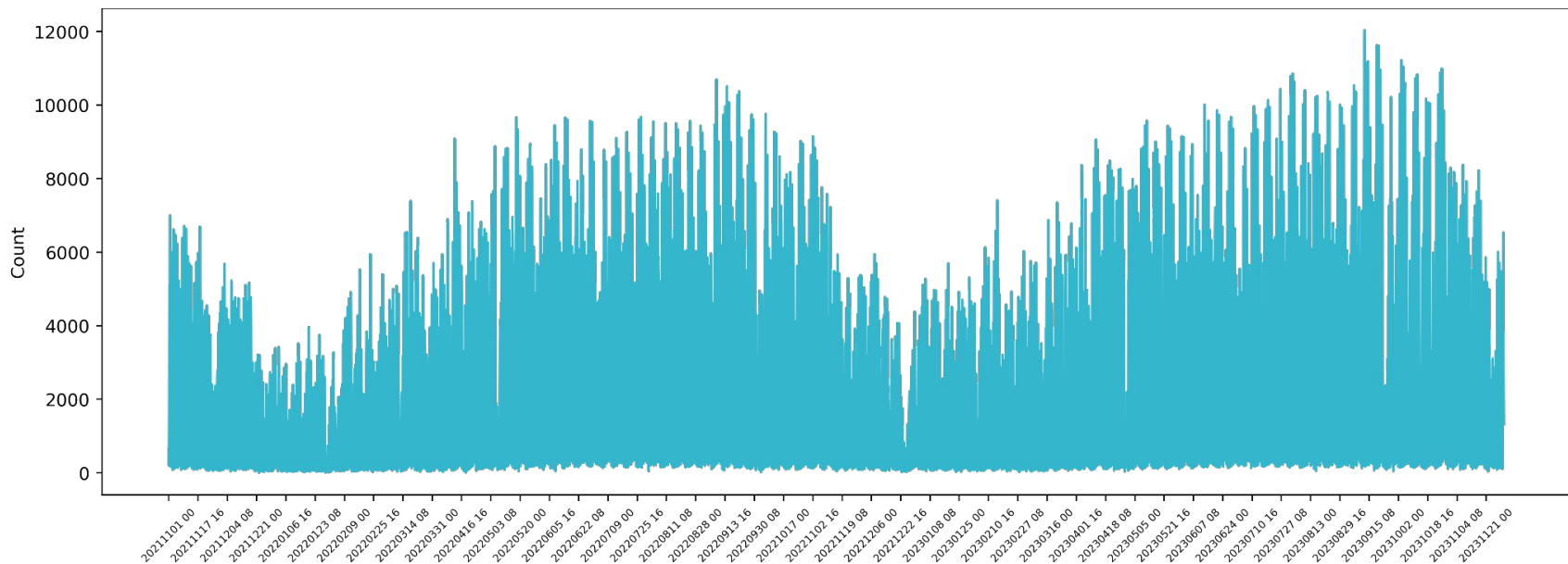
分析天氣因素
對需求量的影響

建立小區間預測模型

動態會員調度
回饋金機制

Citibike的借用量變化量大，因此有定期重新配置停車柱數量的需求。

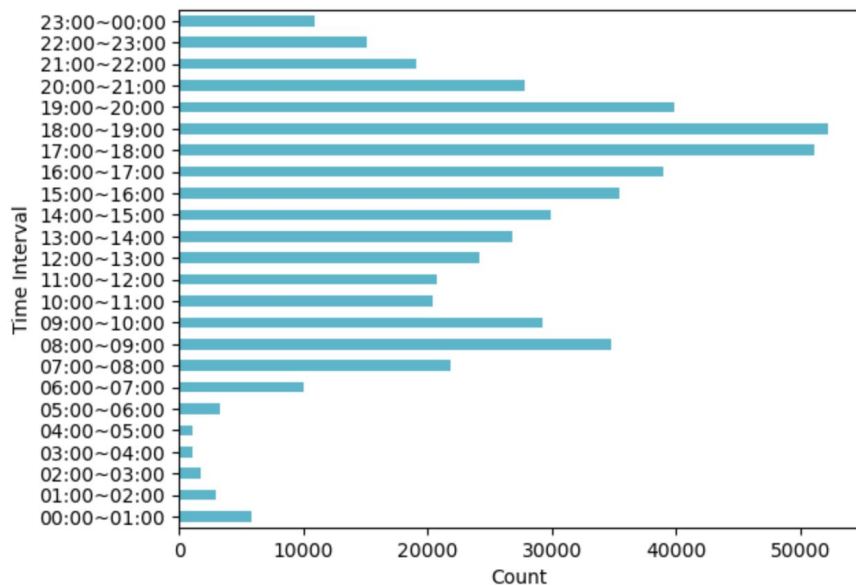
2021/11 ~ 2023/11 的 Citibike 借用量趨勢折線圖



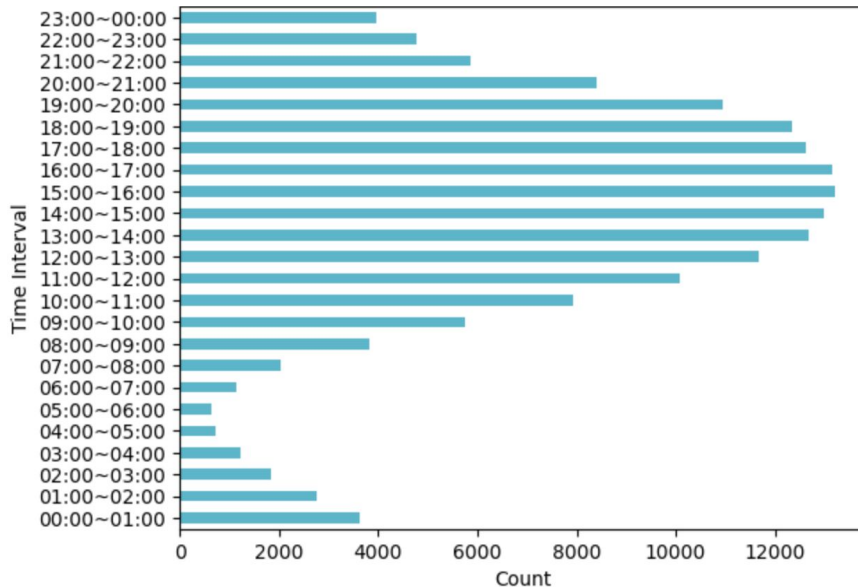
期望透過定期調整各地區的停車柱配置數量，以符合當地居民的借用需求，讓itibike的利益最大化。

無論是平日或假日離尖峰時段的使用量差異大, 因此衍生出調度需求

Citibike於平日各時段的平均借用量



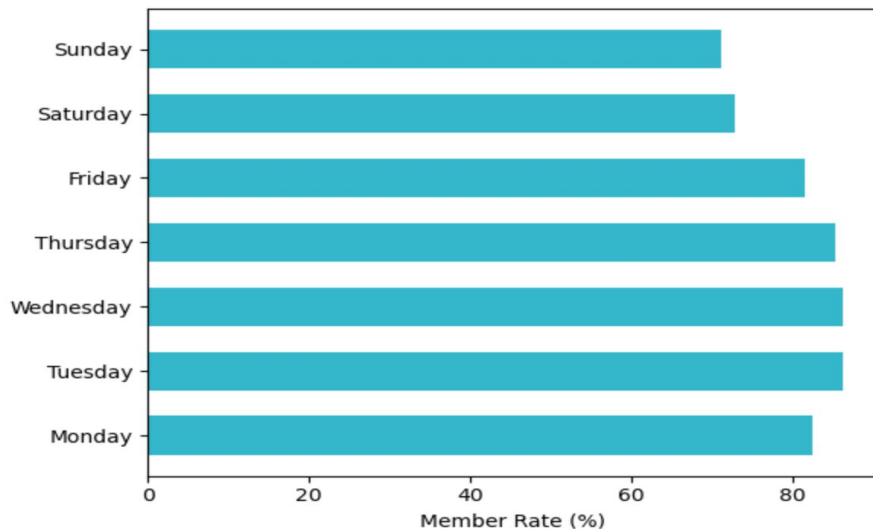
Citibike於假日各時段的平均借用量



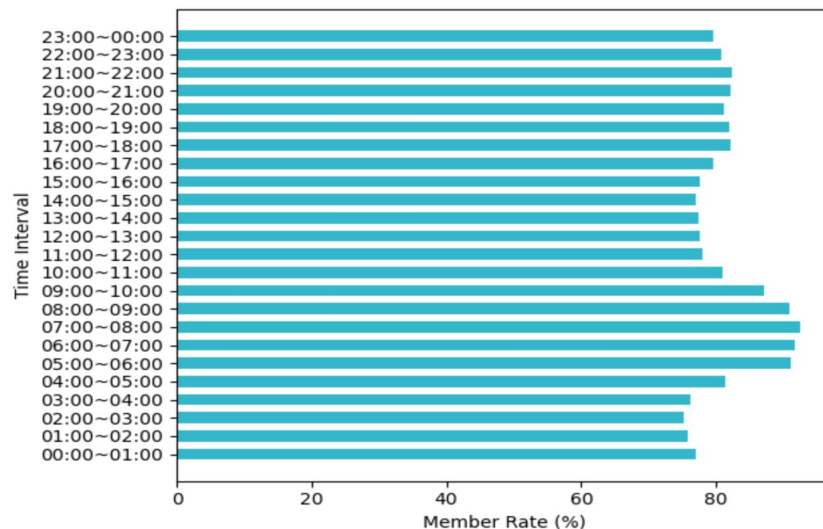
期望透過為尖峰時段借用量較高的站點安排所需要的調度人員數量, 讓itibike的利益最大化。

平假日和一天中會員佔整體的比例不同，在分析後鎖定週末下午推廣借用者加入會員

Citibike於週一到週日會員占整體借用者的比例



Citibike於一天中各時段會員占整體借用者的比例



分析與推論：上班族是會員的比例高，而週末使用者是會員的比例低，故鎖定後者為推廣之目標群體
行動：透過了解使用者需求（如站點增設、車輛調度）並推廣其加入會員，以提高使用者的顧客終生價值

動態資料

Citibike 借還車紀錄及平均車流量資料

資料內容: 2023 年 5 月的每小時曼哈頓借車數量總合與曼哈頓車流量資料

欄位名稱	曼哈頓站點借車數量	時間	車流量	是否為會員
單位	次 / 時	xxxx 年 xx 月 xx 日 xx 時	輛 / 小時	Member/Casual

曼哈頓每小時天氣資料

資料內容: 2023 年 5 月的天氣資料

欄位名稱	溫度	風速	濕度	氣壓	能見度	30 種天氣	時間
單位	度C	km/h	%	mbar	km	晴、雨、霧...	xxxx 年 xx 月 xx 日 xx 時



動態資料預測曼哈頓整體借用量的變化以安排不同時段人力調度

靜態資料

曼哈頓各區人口資料

資料內容: 人口數、人口密度、家庭數等

欄位名稱	區名	總人口數	人口密度	家庭總數	各人種比例	65歲以上人口
單位	無	人	人 / mi^2	戶	%	%

曼哈頓各區房屋資料

資料內容: 建築數量、設施數量等

欄位名稱	區域內各類建築數	街道各類設施數	各類建築數量	街道里程	各式交通工具普及率
單位	棟	個	棟	mi	%



靜態資料了解不同曼哈頓地區的借用量以配置地區間的(常駐)總人力與車柱數量

建立線性模型預測不同地區特徵下曼哈頓各個地區 Citibike 借用量

執行步驟

細項說明及結果呈現

資料前處理

- 變數初步挑選: 從 125 個地區資料中篩選出 44 項影響使用量的關鍵因素
- 數據轉換: 將資料中的資料以地區組成群組, 以歸納各地區特徵變數
- 刪除高相關變數: 刪除同時與其他3個變數相關性達 0.7 以上之變數
- 重要變數挑選: 用 Lasso Regression 挑選重要變數同時降低多重共線性

建立線性迴歸模型

- 迴歸模型建立: 用 Forward Selection 建立線性模型並將顯著門檻設為 0.05

迴歸模型結果 (變數: 係數)

建築物總數: 0.0271

地鐵覆蓋率: 40650

付費停車場數: 137.5426

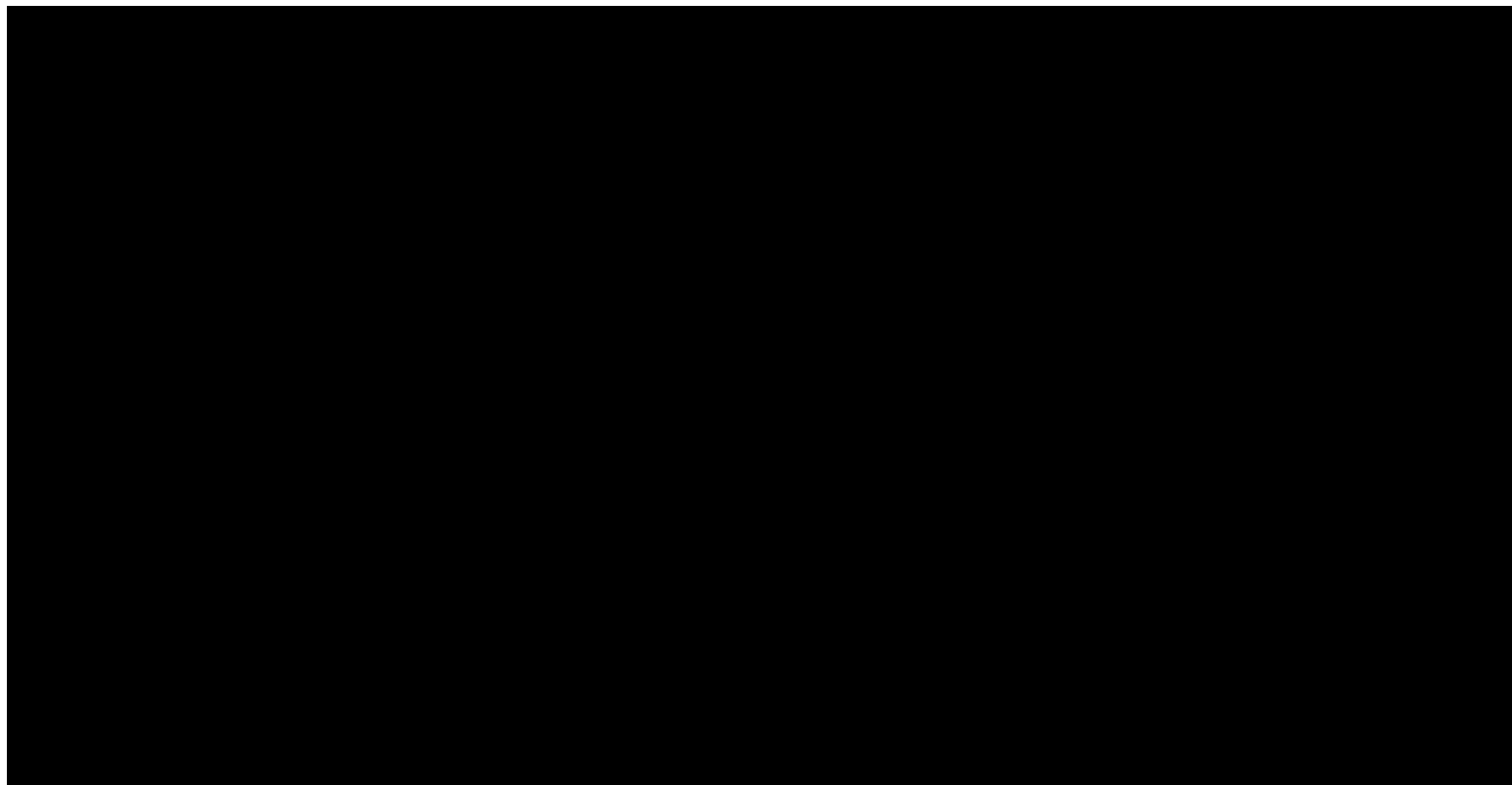
低收入戶數: 1.3834

教育機構數: - 643.7581

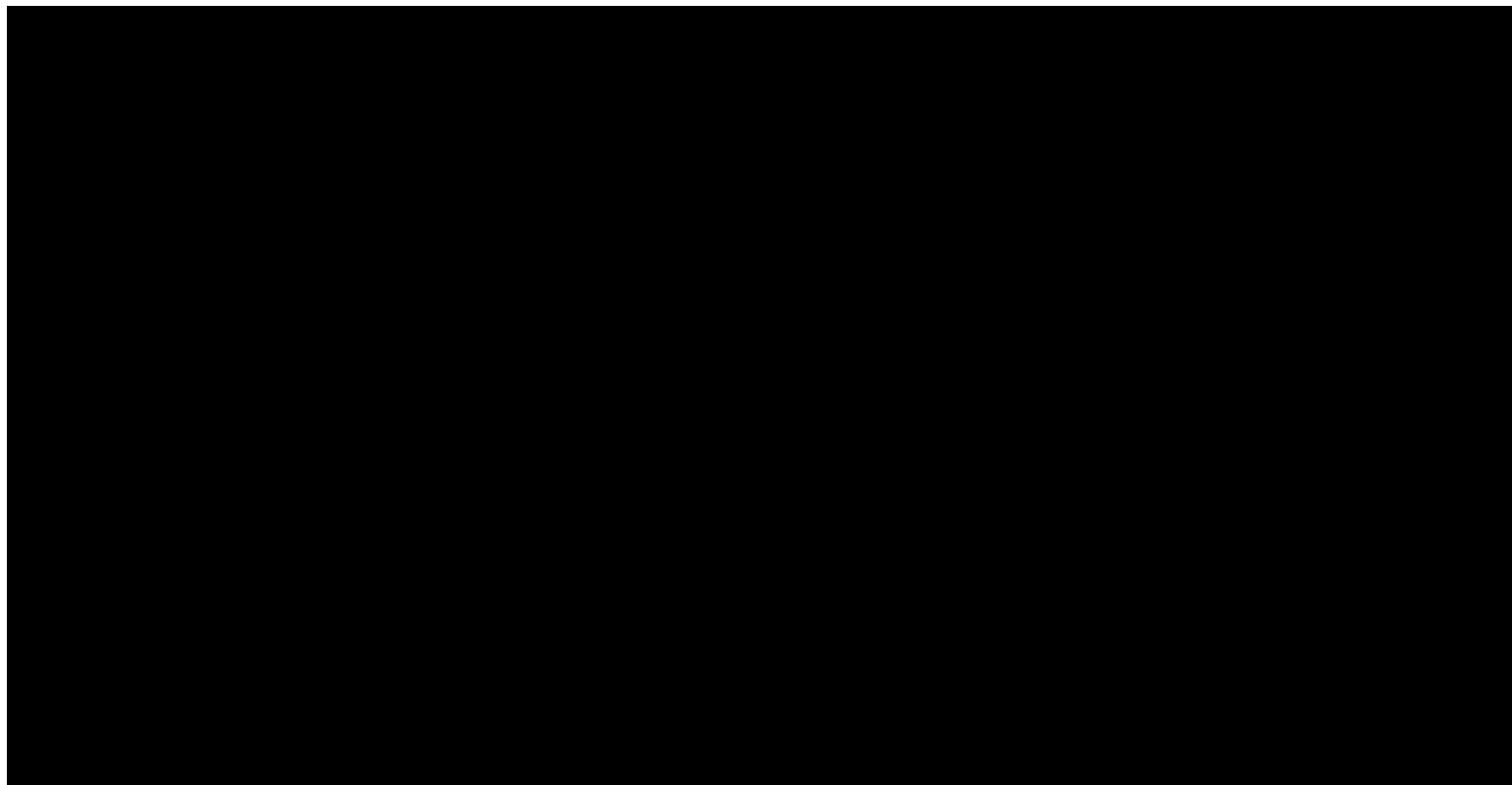
戶外休閒設施數: - 313.8403

常數項:
- 8516.5871

預測各地區下一季共享單車借用量, 並最佳化停車柱數量及調度人員數量



預測各地區下一季的借用量變化量、停車柱變化量及調度人員變化量



分析影響每小時借用量、會員比例的因素以及天氣對不同時段借用量的影響

資料前處理

- 特徵整合: 將天氣變數進行獨熱編碼後, 再將特徵整合成晴天、陰天、雨天、霧天
- 數據轉換: 將所有樣本依照每天每小時進行分群, 車流量則取群內樣本的平均值
- 遺漏值填補: 部分缺漏的風力資料透過取前後一小時的平均值進行填補
- 共線性檢定: 刪除 VIF 的值大於 10 之變數, 包含氣壓、能見度

整體模型解釋

車流量及天氣對借用量的影響

- 車流量越多借用量越高
- 溫度每升 1 度, 借用量增加 133 次
- 雨天借用量減少 1317 次

每小時會員佔所有借用者的比例

- 車流量越多會員借用比例增加
- 溫度越高會員借用比例減少
- 濕度越高會員借用比例增加

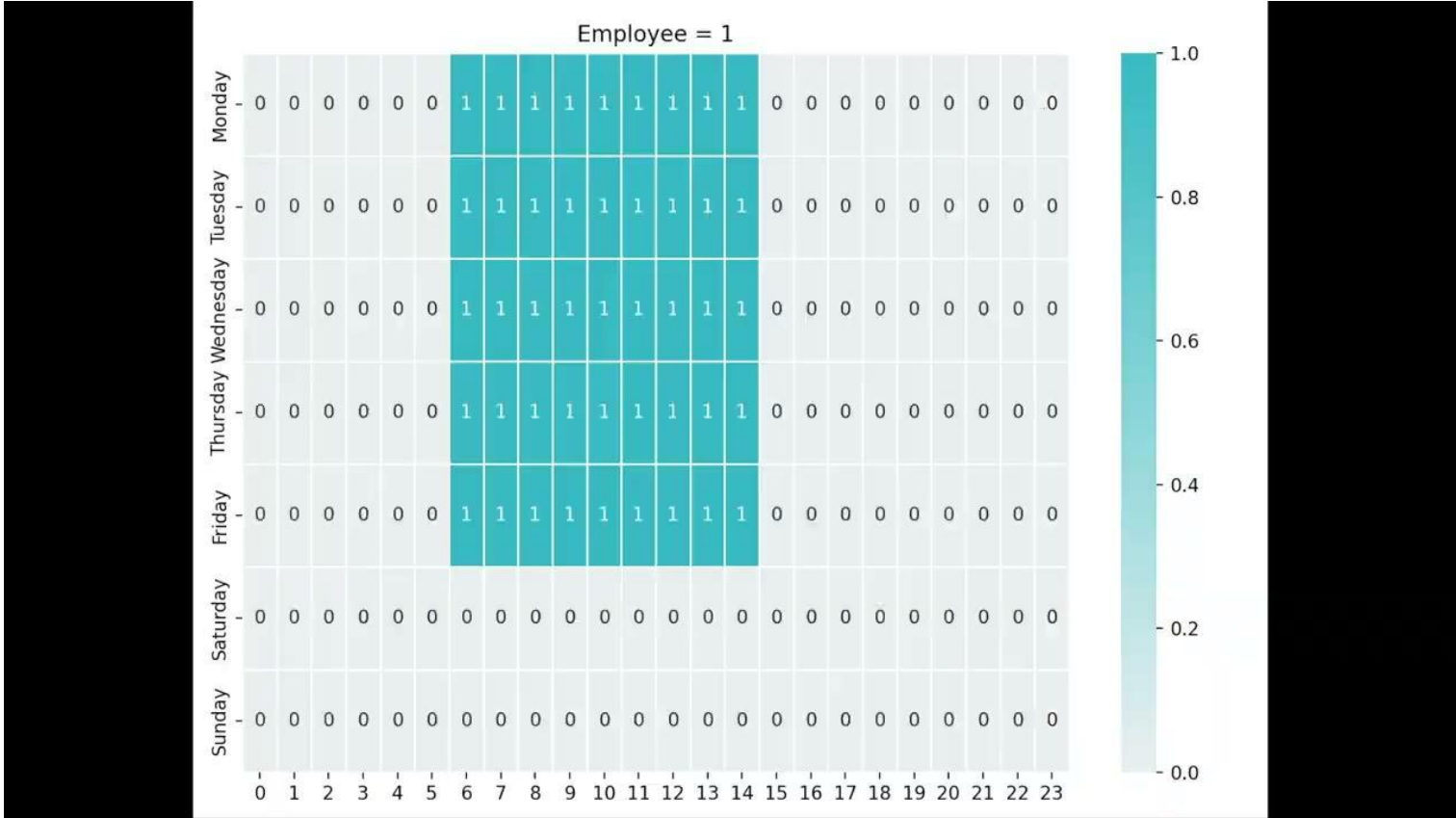
天氣對不同時段借用量的影響

所有時段溫度越高, 借用量皆越多

- 凌晨 (1-6點)
其他變數皆不顯著
- 早上 (7-12點)
晴天借用量增加 570 次
- 下午 (13-18點)
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
- 晚上 (19-24點)
雨天借用量減少 1760 次

線性迴歸模型 (顯著水準0.05)

針對各地區的人力需求，規劃員工的工作時間和各時段的員工數量



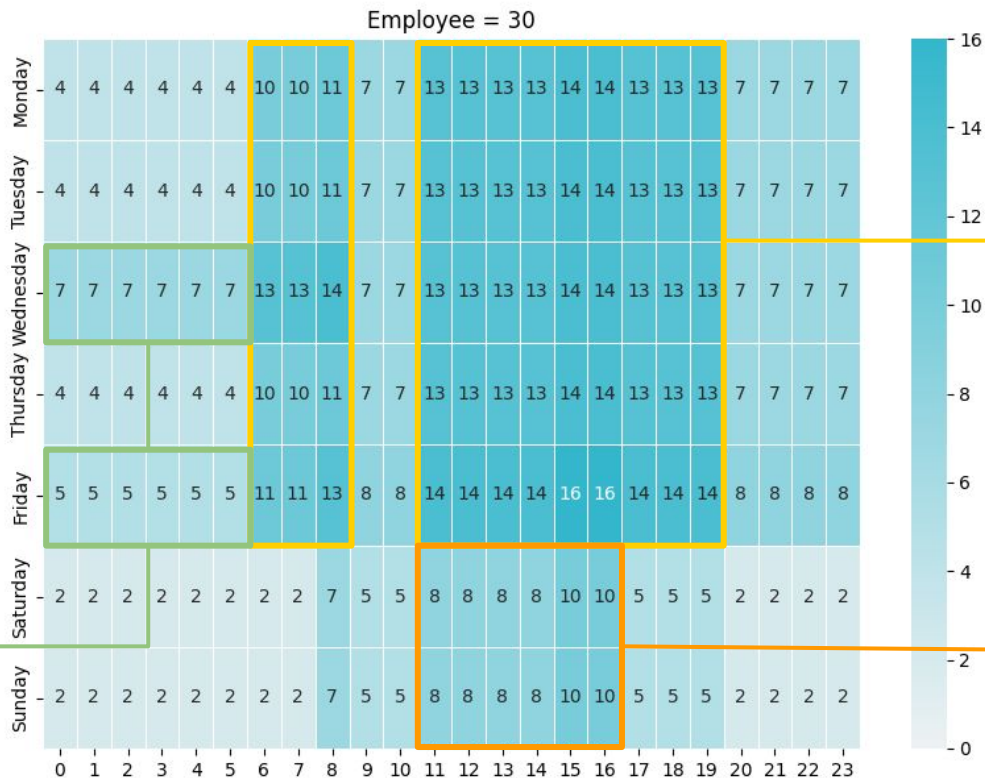
剖析當某區員工人數達30人時，各時段的員工數量之分配原則

總覽

每位員工一週工作 40 小時，每天不超過 8 小時，並根據使用需求安排不同時段的員工數量：

- 平日 > 週末
- 白天 > 夜晚
- 溫度高時增加人力
- 雨天時減少人力

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

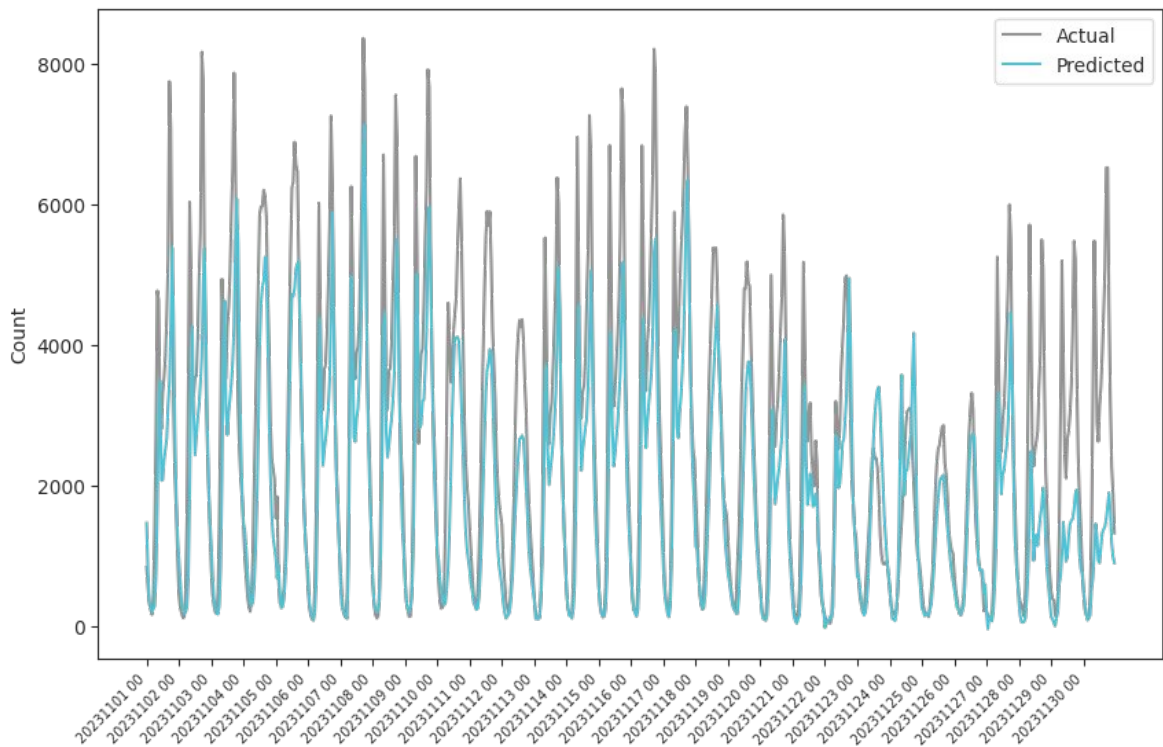


1. 平日員工分流上班，滿足上班和下班的調度需求

2. 於週五設置最多人力，提前調度車輛為週末準備

週末時下午有較多員工，以滿足大眾使用習慣，並安排員工於溫度高、濕度低時推廣會員方案

建立 LSTM 模型對於每天每小時的總借用量進行預測 (Testing)



目標 - 即時預測總借用量

每小時
時間序列分析

天氣資料

窗口: 前 720 hr
(含被預測小時)

天氣
預報

歷史
觀測

Training:
2021.11~2023.10

Testing:
2023.11

Tuning: batch / Loss functions

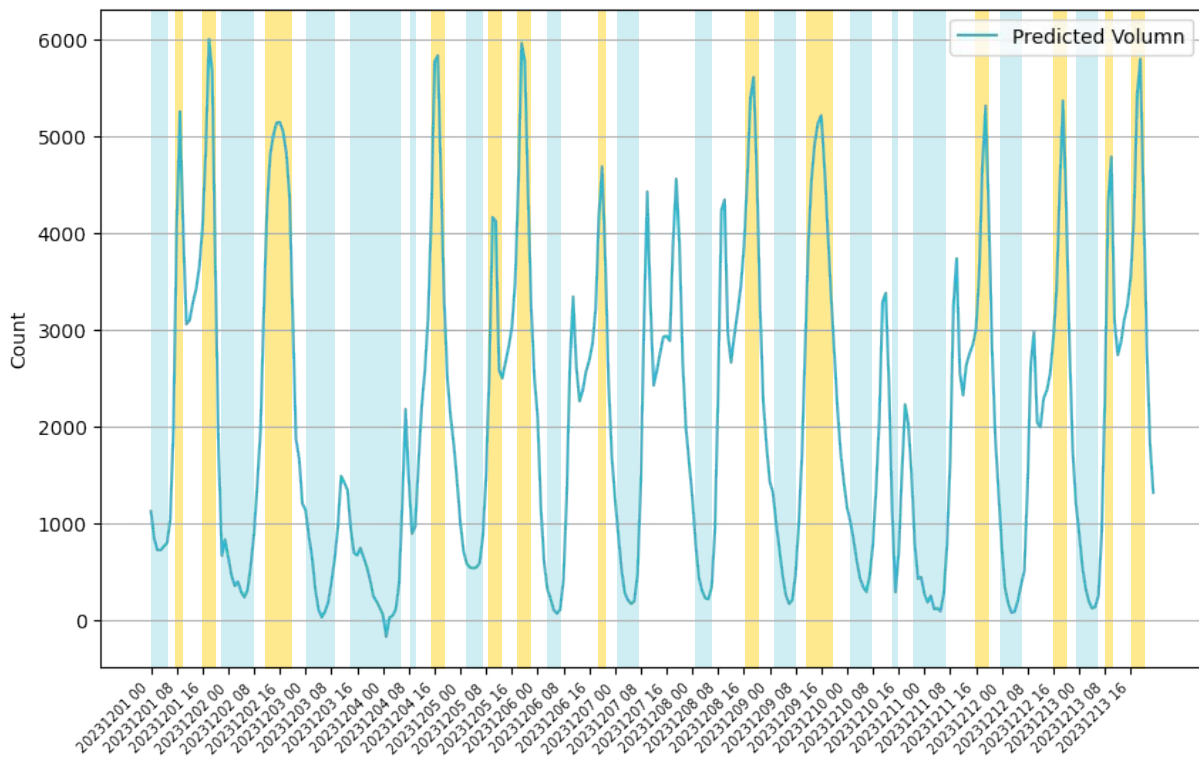
建立回饋點數調整機制、臨時工預算預測

運用 LSTM 模型預測結果動態調整 Bike Angels 回饋點數及臨時工預算

	動態點數回饋調整機制	配套措施	潛在風險
策略方向	Bike Angels 機制 鼓勵用戶將Citibike從需求量較高的域移到需求量較低的區域，並給予點數回饋(點數可用來兌換獎勵)，藉此借助用戶的力量，達成的調度需求。	Part-Time 預算控制 受限天氣預報的準確率，因此使用未來一週天氣預報資料與歷史資料，提前預測未來一週臨時工的工作時數。	<ul style="list-style-type: none">- 調度成本增加: 模型將使用量預測過低導致回饋點數不足，降低 Bike Angels 的誘因，恐增加額外的調度成本。- 回饋點數成本增加: 模型將使用量預測過高導致非必要的時段，過多 Bike Angels 協助調度，增加回饋成本- 天氣預報嚴重錯誤: 受限於天氣預報本身準確率，可能降低模型預測準確率。
模型預測結果	尖峰	增加回饋點數 降低調度成本	增加臨時工預算 擴聘臨時工
	離峰	降低回饋點數 降低回饋成本	減少臨時工預算 減聘臨時工

運用 LSTM 模型預測結果動態調整回饋點數的發放比率

2023年12月的借用量預測圖及回饋比率訂定區間



動態回饋金調整機制

黃色區段
使用較高回饋比率，如增加10%

藍色區段
使用較低回饋比率，如減少10%

專案總結

目標定義	符合各地的借用需求	符合各時段的長期借用需求	符合各時段的短期借用需求
模型選擇	多元線性迴歸模型	多元線性迴歸模型	LSTM 模型
解決方案	每季調整各地 停車柱數 每季調整各地 常駐調度人員數	每季規劃各地 常駐調度人員的班表	實施動態點數 回饋調整機制 每週評估 臨時調度人員 需求
預期效益	符合各地的借用需求 減少共享單車的 閒置時間	增加 車輛調度的效率 平衡 尖峰時段 的車輛配置	精準掌握 點數回饋成本 精準掌握臨時 調度人員預算

透過以上三項解決方案幫助CitiBike提升營收、降低成本，最終**為CitiBike大幅提高利潤**的目標

Appendix

1

借用量跟停車柱數量之間的轉換方法

2

借用量跟常駐調度人員數量之間的轉換方法

3

LSTM Tuning Phase

4

未來研究建議

各地區預測借用量與停車柱數、常駐調度人員數之間的轉換

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

$(\text{調度人員平均薪資} / \text{各站點人力成本}) = \text{各站點的常駐調度人員數量}$

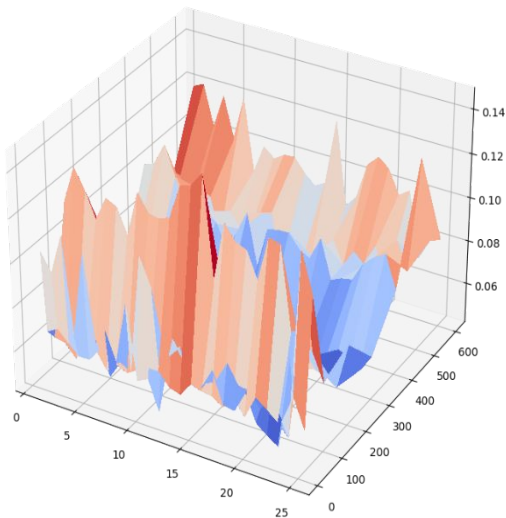
實務

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。

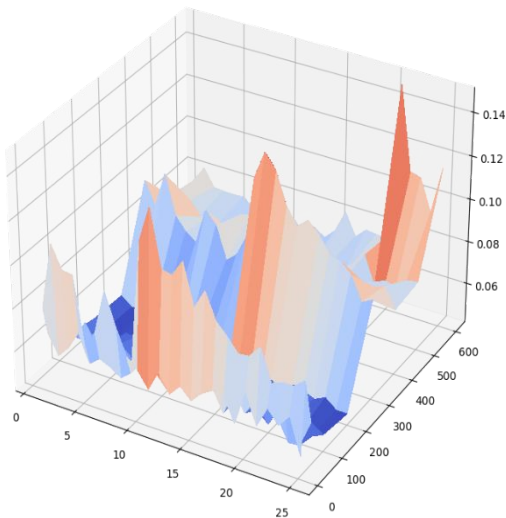
建立 LSTM 模型對於每天每小時的總借用量進行預測 (Tuning)

* Training Dataset: 2021/11/01-2023/10/31

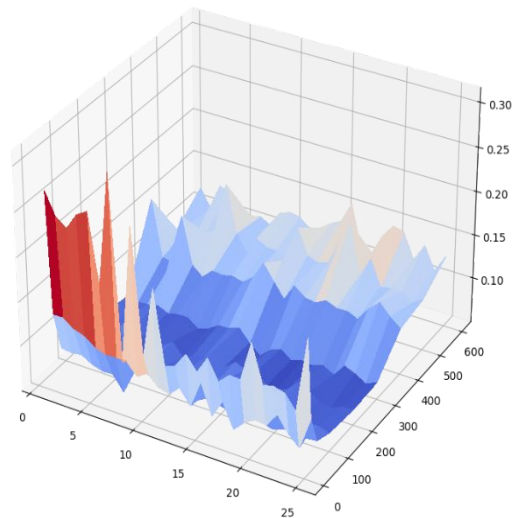
mean_squared_error



mean_absolute_error



mean_absolute_percentage_error



在 LSTM 的訓練階段我們採用不同的 loss functions, 找到 batch size 300、timestep 720、loss function = MAE, 使用測試集有最小、穩定的誤差項 (MSE)。(圖中 x = batch size, y = epoch, z = error term)

未來研究建議

本專案 局限性

缺乏剩餘停車柱空位資訊

無法確認借用量需求與實際需要調度的需求的關係

天氣資料涵蓋地區太大

天氣資料為整個紐約的天氣資料，無法結合地區因素

資料不易預測、維護

天氣預報準確率的時效很短、人口等地區資料不易維護

未分析 的問題

使用 GNN 進行分析

可以進一步了解如何降低人力調度成本

將範圍擴大到曼哈頓外

除了分析其他地區的資料，也可以將各大地區分類

考慮地形問題

導入地形資料，分析上下坡等因素是否影響騎乘意願

建議

導入自動化工具

將未來地區預測成長資料、天氣預報導入決策模型