

Statistical Methodology

預測臺北市信義區各捷運站
附近YouBike站點的自行車租借量

110426025 盧盈穎



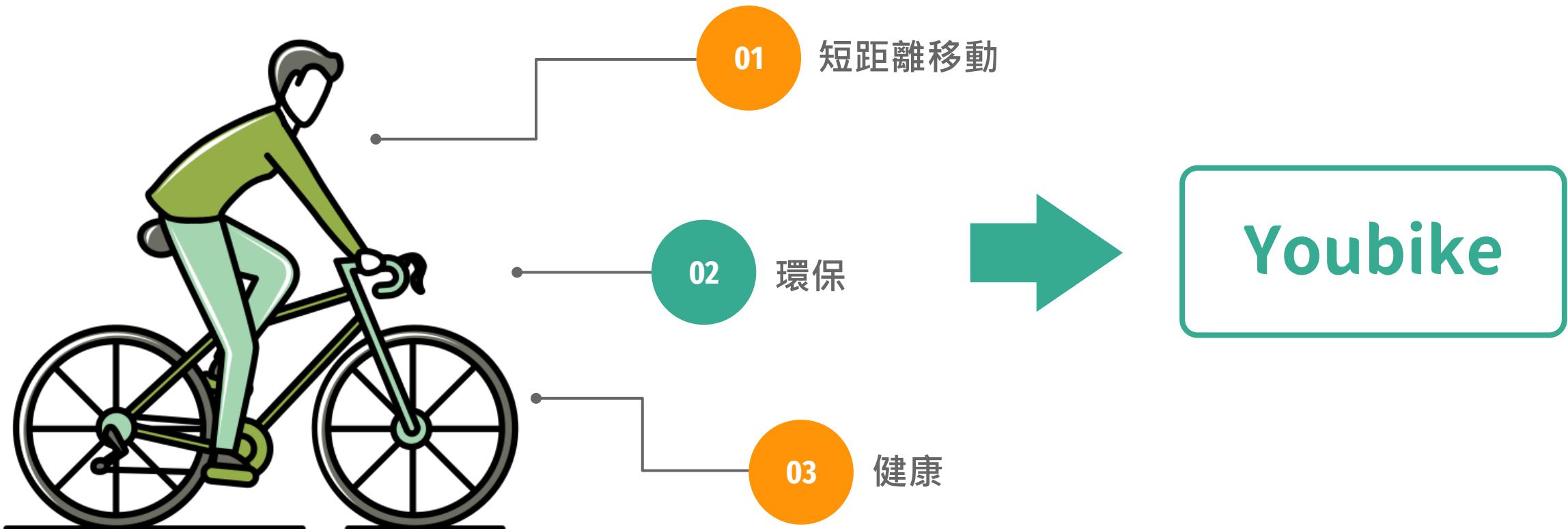
CONTENTS

- 1 研究背景
Research Background
- 2 研究動機
Research Motivation
- 3 研究目的
Research Purpose
- 4 預期結果
Expected Results
- 5 資料集介紹
Intro to Datasets
- 6 資料敘述
Data Statement
- 7 回歸分析
Regression
- 8 Neural Network Model
Model Application
- 9 模型應用
Model Application
- 結論
Conclusion

研究背景

Research Background

- 近兩年「零碳排」是最流行的幾個話題之一 → **節能減碳**
- 除了捷運、公車等大眾運輸，共享單車也是一個好的選擇



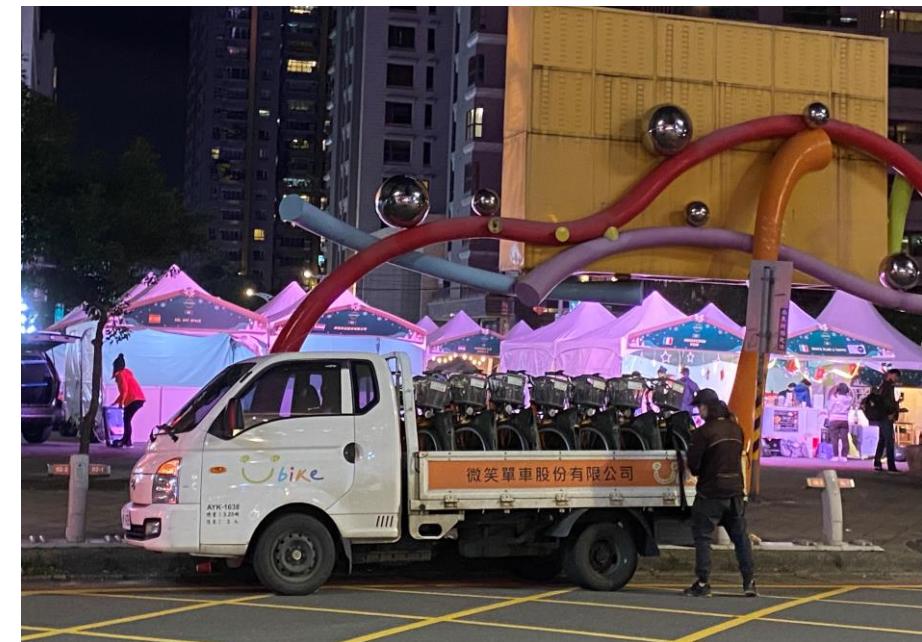


YouBike微笑單車

- 「YouBike 微笑單車」為提供 24小時「甲地租、乙地還」租賃服務的電子無人自動化管理公共自行車系統
- 目前是採取人工的方式派遣各站點的腳踏車

臺北市信義區

- 台灣的蛋黃區
- 新興商業區、最具指標性的商圈
- 上班族日常：交通繁忙、塞車
- 晚間、假日：購物商場



YouBike補車 2022/12/18 20:30

研究背景

Research Background

選用臺北市信義區六個捷運站點鄰近Youbike站點進行Youbike腳踏車租借量預測

- 台北101/世貿站
- 國父紀念館站
- 市政府站
- 松山站
- 永春站
- 象山站



研究動機

Research Motivation

根據HELLO TAIPEI 臺北市單一陳情系統統計，
在105~108年之間，無車可借及無位可還為
Youbike民眾陳情案件中佔最多的比例。

Youbike 目前遇到的困境：

- 無車可借
- 無位可還

<https://udn.com> 地方 · Translate this page ::

民眾借不到YouBike頻客訴議員曝北市3個月收252件 - 聯合報

Apr 25, 2022 — 台北市目前約有2萬多台YouBike，但議員王欣儀指出，民眾在尖峰時段時常**借不到車**、**還不了車**，甚至在近3個月1999列管案件中YouBike相關的申訴案就 ...

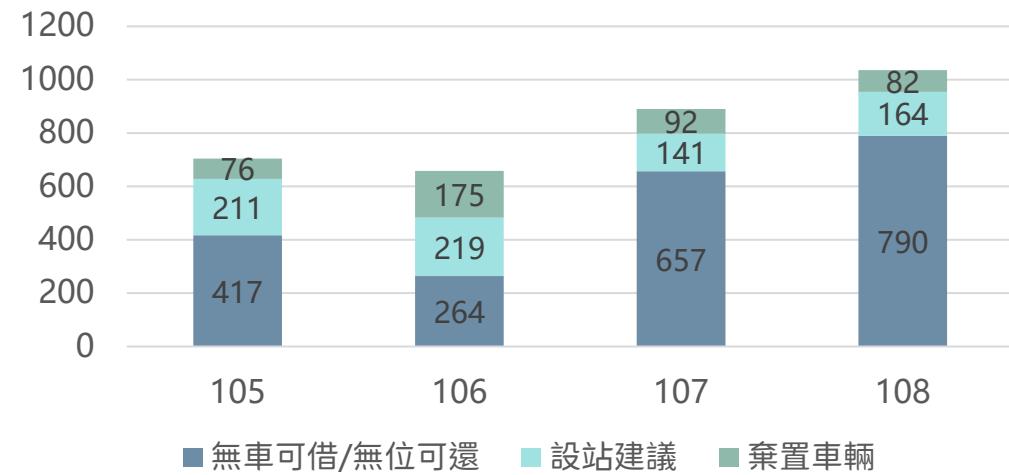
<https://udn.com> 地方 · Translate this page ::

北市YouBike常常借不到、還不了議員批調度出問題 - 聯合報

Aug 13, 2022 — 台北市推動公共運輸，將YouBike視為綠色交通的最後一里路，不過就有民眾抱怨，捷運站外的微笑單車不是**租不到**就是**還不了**，質疑調度模式有問題。



臺北市Youbike陳情案件



研究目的

Research Purpose

- ① 無車可借
- ② 無位可還

預測租借量

適時調整派遣人員
回收車輛或新增車輛

預期結果

Expected Results

嘗試不同模型的建置，期望找到適合預測腳踏車租借量的模型，以此提供於YouBike公司得知於下一階段是否會有缺少車輛或過多車輛的狀況，藉此進行人員的工作派遣回收及新增車輛。

資料集介紹 / 資料前處理

Intro to Datasets



YouBike2.0臺北市公共自行車即時資訊

交易序號	卡號	扣款金額	借車時間	借車站代號	借車站	借車車柱	自行車編號	還車時間	還車站代號	還車站	還車車柱	租用時間	手機序號	卡片種類	費率別	業者代碼	根據檔案是IPASS或ECC
0	172	4B5FE1D076C491F3F3BAD948933386B9	45	2021-07-22 19:10:10	210	æ, ç%4a%0(Eæ»å»è~å£	026B 0000003122	2021-07-22 21:14:45	1235	æ·éçæ°ç«TM(2èTM¥å†ºå£)	052B	2:04:35	D9073F36	1	L04	A209	IPASS
1	332	102A8AD91AFB281425C044C3D97BB99B	5	2021-07-22 19:05:53	210	æ, ç%4a%0(Eæ»å»è~å£	001A 00050019D0	2021-07-22 19:10:39	119	æ·éçæ°ç«TM(2èTM¥å†ºå£)	015A	0:04:46	6F84D336	1	L04	A209	IPASS
2	533	7C0D4E3F3F11636F89056806C93DD5BE	15	2021-07-22 18:21:30	210	æ, ç%4a%0(Eæ»å»è~å£	019A 0000000D31	2021-07-22 18:58:48	78	æ·éçæ°ç«TM(2èTM¥å†ºå£)	005B	0:37:18	8FECB836	1	L04	A209	IPASS
3	2	2E7E0AF1B02A8F59885D30353824D5F1	5	2021-07-22 14:15:59	210	æ, ç%4a%0(Eæ»å»è~å£	019A 00000008EF	2021-07-22 14:30:23	184	å§æ½ç¤¾å	001B	0:14:24	19ACE53A	1	L04	A209	IPASS
4	227	FBE7159C9F3B8BC2	5	2021-07-22 10:32:19	210	æ, ç%4a%0(Eæ»å»è~å£	014A 00050001C9	2021-07-22 10:39:27	119	æ·éçæ°ç«TM(2èTM¥å†ºå£)	017A	0:07:08	62A2583A	1	L04	A209	IPASS
...
1928342	2502	A0C12DC46883B26D9A1DE4BD4F394456	5	2021-07-11 19:27:39	343	å, åæ±è~å, æ@µ185å~å£	007A 00000035E7	2021-07-11 19:41:08	1	æ·éçæ°ç«TM(3èTM¥å†ºå£)	069A	0:13:29	39083537	1	L01	96	ECC
1928343	455	4EF292D6D7E79D3460CA0D59F005AA22	5	2021-07-11 12:54:39	249	å, è~æ , æ~è~å£	004B 00050023D5	2021-07-11 13:03:49	186	è†ºåE~å, ç«å¤å	037B	0:09:10	39093537	1	L01	95	ECC
1928344	834	0323A13FC3BDC7BD95BC1EA72255A32D	15	2021-07-11 15:35:42	241	æ~å£å...åº	008B 0005001656	2021-07-11 16:15:33	241	æ~å£å...åº	001B	0:39:51	3909E236	1	L01	97	ECC
1928345	391	9FE6A91AA451080989056806C93DD5BE	25	2021-07-11 16:40:20	327	é‡æ...¶å~æµ~è~å£	003B 0000003737	2021-07-11 18:11:00	327	é‡æ...¶å~æµ~è~å£	010A	1:30:40	390EF436	1	L01	97	ECC
1928346	2338	93674D88D077AD50D672C956C661C7CB	5	2021-07-11 06:59:05	249	å, è~æ , æ~è~å£	013B 000500070B	2021-07-11 07:02:05	82	æ·éçæ°ç«TM(3èTM¥å†ºå£)	025B	0:03:00	39105B36	1	L01	96	ECC

1928347 rows × 18 columns

2021年7月 YouBike 租借詳細資訊

總共1,928,347筆資料

資料集介紹 / 資料前處理

Intro to Datasets



臺北捷運各站分時進出量統計

	日期	時段	進站	出站	人次
0	2021-07-01	0	松山機場	松山機場	0
1	2021-07-01	0	松山機場	中山國中	0
2	2021-07-01	0	松山機場	南京復興	0
3	2021-07-01	0	松山機場	忠孝復興	0
4	2021-07-01	0	松山機場	大安	2
...
7593259	2021-07-31	23	丹鳳	徐匯中學	0
7593260	2021-07-31	23	丹鳳	三和國中	0
7593261	2021-07-31	23	丹鳳	三重國小	0
7593262	2021-07-31	23	丹鳳	迴龍	1
7593263	2021-07-31	23	丹鳳	丹鳳	2

7593264 rows × 5 columns



中央氣象局

臺北市24小時天氣資料

	時間	觀測時間	氣溫	相對溼度	風速	降水量	紫外線指數
0	2021-07-01 00:00:00	24	27.3	75	0.9	0	0
1	2021-07-01 01:00:00	1	27.4	74	1.2	0	0
2	2021-07-01 02:00:00	2	27.2	74	1.6	0	0
3	2021-07-01 03:00:00	3	27	75	0.1	0	0
4	2021-07-01 04:00:00	4	27.1	73	1.3	0	0
...
739	2021-07-31 19:00:00	19	30	62	1.3	0	0
740	2021-07-31 20:00:00	20	29.4	69	0.3	0	0
741	2021-07-31 21:00:00	21	29.1	72	1.5	0	0
742	2021-07-31 22:00:00	22	28.5	73	1.1	0	0
743	2021-07-31 23:00:00	23	28.1	71	1.1	0	0

744 rows × 7 columns

資料集介紹 / 資料前處理-捷運站人流量

Intro to Datasets

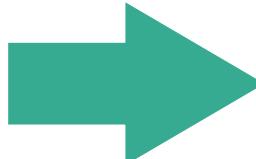
進站以捷運站-松山為例

挑選出進站為松山的資料

	日期	時段	進站	出站	人次
0	2021-07-01	0	松山	松山機場	0
1	2021-07-01	0	松山	中山國中	1
2	2021-07-01	0	松山	南京復興	3
3	2021-07-01	0	松山	忠孝復興	5
4	2021-07-01	0	松山	大安	4
...
70303	2021-07-31	23	松山	徐匯中學	7
70304	2021-07-31	23	松山	三和國中	11
70305	2021-07-31	23	松山	三重國小	19
70306	2021-07-31	23	松山	迴龍	8
70307	2021-07-31	23	松山	丹鳳	2

70308 rows × 5 columns

將在相同時段下，進入的站點
皆為松山的資料加總



整理後資料

	捷運站	時間	進站人數
0	松山	2021-07-01 0:00:00	222
1	松山	2021-07-01 1:00:00	0
2	松山	2021-07-01 5:00:00	0
3	松山	2021-07-01 6:00:00	144
4	松山	2021-07-01 7:00:00	373
...
646	松山	2021-07-31 19:00:00	3175
647	松山	2021-07-31 20:00:00	1807
648	松山	2021-07-31 21:00:00	1552
649	松山	2021-07-31 22:00:00	1496
650	松山	2021-07-31 23:00:00	735

651 rows × 3 columns

資料集介紹 / 資料前處理-捷運站人流量

Intro to Datasets

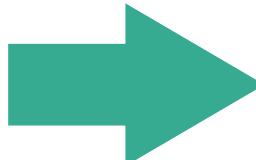
出站以捷運站-松山為例

挑選出進站為松山的資料

	日期	時段	進站	出站	人次
0	2021-07-01	0	松山機場	松山	0
1	2021-07-01	0	中山國中	松山	0
2	2021-07-01	0	南京復興	松山	7
3	2021-07-01	0	忠孝復興	松山	2
4	2021-07-01	0	大安	松山	2
...
70303	2021-07-31	23	徐匯中學	松山	0
70304	2021-07-31	23	三和國中	松山	3
70305	2021-07-31	23	三重國小	松山	1
70306	2021-07-31	23	迴龍	松山	0
70307	2021-07-31	23	丹鳳	松山	0

70308 rows × 5 columns

將在相同時段下，進入的站點
皆為松山的資料加總



整理後資料

	捷運站	時間	出站人數
0	松山	2021-07-01 0:00:00	148
1	松山	2021-07-01 1:00:00	0
2	松山	2021-07-01 5:00:00	0
3	松山	2021-07-01 6:00:00	139
4	松山	2021-07-01 7:00:00	355
...
646	松山	2021-07-31 19:00:00	2368
647	松山	2021-07-31 20:00:00	1467
648	松山	2021-07-31 21:00:00	1188
649	松山	2021-07-31 22:00:00	780
650	松山	2021-07-31 23:00:00	261

651 rows × 3 columns

資料集介紹 / 資料前處理-捷運站人流量

Intro to Datasets

出站以捷運站-松山為例

由於捷運凌晨1點至5點為停駛期間，故將這些期間以0補值進入資料

	時間	捷運站	出站人數	進站人數
0	2021-07-01 00:00:00	松山	148.0	222.0
1	2021-07-01 01:00:00	松山	0.0	0.0
2	2021-07-01 02:00:00	松山	0.0	0.0
3	2021-07-01 03:00:00	松山	0.0	0.0
4	2021-07-01 04:00:00	松山	0.0	0.0
...
739	2021-07-31 19:00:00	松山	2368.0	3175.0
740	2021-07-31 20:00:00	松山	1467.0	1807.0
741	2021-07-31 21:00:00	松山	1188.0	1552.0
742	2021-07-31 22:00:00	松山	780.0	1496.0
743	2021-07-31 23:00:00	松山	261.0	735.0

744 rows × 4 columns

補值後成果
使所有資料的筆數維持在744筆

資料集介紹 / 資料前處理-YouBike租借量

檢查資料是否有空值並刪除空值

```
data.isnull().sum()
```

交易序號	0
卡號	0
扣款時間	0
扣款金額	0
借車時間	0
借車站代號	0
借車站	0
借車車柱	0
自行車編號	0
還車時間	0
還車站代號	0
還車站	0
還車車柱	0
租用時間	0
手機序號	22
卡片種類	0
費率別	0
業者代碼	0
根據檔案是IPASS或ECC	0
整批資料更新時間	
存入DB時機	0
該筆資料更新時間	
該筆資料更新日期	

刪除沒有用到欄位

```
data = data.drop('交易序號', axis=1)
data = data.drop('卡號', axis=1)
data = data.drop('扣款時間', axis=1)
data = data.drop('卡片種類', axis=1)
data = data.drop('業者代碼', axis=1)
data = data.drop('根據檔案是IPASS或ECC', axis=1)
data = data.drop('整批資料更新時間', axis=1)
data = data.drop('存入DB時機', axis=1)
data = data.drop('該筆資料更新時間', axis=1)
data = data.drop('該筆資料更新日期', axis=1)
```

資料集介紹 / 資料前處理-YouBike租借量

Intro to Datasets

刪除異常租借時間

將每天租借時間大於8小時，與小於1分鐘的值刪除

```
S = S[(S['借車時間'] >= '2017-07-01 00:00:00') & (S['借車時間'] <= '2017-07-31 23:59:59')]  
S = S.reset_index(drop=True)  
S
```

	扣款金額	借車時間	借車站代號	借車站	借車車柱	自行車編號	還車時間	還車站代號	還車站	還車車柱	租用時間	手機序號	費率別
0	45	2017-07-22 19:10:10	210	æ, ç¾¤äºŒæ»œ¥å»è·å£	026B	0000003122	2017-07-22 21:14:45	1235	æ·é·æ·jæ°ç«™	052B	2:04:35	D9073F36	L04
1	5	2017-07-22 19:05:53	210	æ, ç¾¤äºŒæ»œ¥å»è·å£	001A	00050019D0	2017-07-22 19:10:39	119	æ·é·åšå—è·ç«™(2èTM¥å‡ºå£)	015A	0:04:46	6F84D336	L04
2	15	2017-07-22 18:21:30	210	æ, ç¾¤äºŒæ»œ¥å»è·å£	019A	0000000D31	2017-07-22 18:58:48	78	æ·é·åœ“å±±ç«™(2èTM¥å‡ºå£)	005B	0:37:18	8FECB836	L04
3	5	2017-07-22 14:15:59	210	æ, ç¾¤äºŒæ»œ¥å»è·å£	019A	00000008EF	2017-07-22 14:30:23	184	åœœ½çœ¾å	001B	0:14:24	19ACE53A	L04
4	5	2017-07-22 10:32:19	210	æ, ç¾¤äºŒæ»œ¥å»è·å£	014A	00050001C9	2017-07-22 10:39:27	119	æ·é·åšå—è·ç«™(2èTM¥å‡ºå£)	017A	0:07:08	62A2583A	L04
...
1928342	5	2017-07-11 19:27:39	343	å› åœ±è·å...æ@pu185å·å£	007A	00000035E7	2017-07-11 19:41:08	1	æ·é·å„ æ”ç‰œç«™(3èTM¥å‡ºå£)	069A	0:13:29	39083537	L01
1928343	5	2017-07-11 12:54:39	249	ä, è—æj, æ—è·å£	004B	00050023D5	2017-07-11 13:03:49	186	è‡ºåŒ—å„ç«åœ§å„	037B	0:09:10	39093537	L01
1928344	15	2017-07-11 15:35:42	241	æ—å£å..¬åœ‘	008B	0005001656	2017-07-11 16:15:33	241	æ—å£å..¬åœ‘	001B	0:39:51	3909E236	L01
1928345	25	2017-07-11 16:40:20	327	é‡æ...¶å—æµ·è·å£	003B	0000003737	2017-07-11 18:11:00	327	é‡æ...¶å—æµ·è·å£	010A	1:30:40	390EF436	L01
1928346	5	2017-07-11 06:59:05	249	ä, è—æj, æ—è·å£	013B	000500070B	2017-07-11 07:02:05	82	æ·é·è¥ç—ç«™(3èTM¥å‡ºå£)	025B	0:03:00	39105B36	L01

1928347 rows × 13 columns

資料集介紹 / 資料前處理-YouBike租借量

Intro to Datasets

篩選各捷運站附近的YouBike站 - 以松山站為例

```
#松山站附近(松山車站)
content = ['松山車站']
new=content[0].encode('utf-8').decode('latin-1')
g1 = aa[aa['還車站']==new]
g1 = g1.reset_index(drop=True)
g1
```

	扣款金額	借車時間	借車站代號	借車站	借車車柱	自行車編號	還車時間	還車站代號	還車站	還車車柱	租用時間	手機序號	費率別
0	5	2021-07-01 00:08:25	33	å,å™é«ä,	020A	0000002490	2021-07-01 00:16:40	22	æ%å±è»šç«™	013B	0:08:15	FD17D53A	L01
1	5	2021-07-01 00:10:52	261	å¥å°æ-°å¥	001B	0000002543	2021-07-01 00:28:56	22	æ%å±è»šç«™	009B	0:18:04	BCF26737	L01
2	5	2021-07-01 00:11:55	261	å¥å°æ-°å¥	003A	0000001FFC	2021-07-01 00:28:53	22	æ%å±è»šç«™	010B	0:16:58	81A85837	L01
3	5	2021-07-01 00:20:53	26	æ·éæ·†é™½ç«™(1è™¥å‡ºå£)	020A	00070002F0	2021-07-01 00:41:10	22	æ%å±è»šç«™	006B	0:20:17	E562F938	G03
4	15	2021-07-01 01:54:19	294	æœ±å™å†å‘	015B	0005001684	2021-07-01 02:45:26	22	æ%å±è»šç«™	014A	0:51:07	6874F239	L04
...
15378	5	2021-07-31 22:41:58	379	æ·¥å...‰å...¬åç'	008B	0005001B58	2021-07-31 22:49:34	22	æ%å±è»šç«™	003A	0:07:36	78AAE93A	L01
15379	5	2021-07-31 22:43:36	42	æ·éå¾Œå±å¥åç«™(1è™¥å‡ºå£)	019B	0000003950	2021-07-31 22:48:48	22	æ%å±è»šç«™	013B	0:05:12	4712FC3A	L01
15380	5	2021-07-31 22:55:03	40	ç‰æ^å...¬åç'	017A	00050018AC	2021-07-31 23:05:19	22	æ%å±è»šç«™	016B	0:10:16	B293FE35	L01
15381	5	2021-07-31 22:58:08	1	æ·éå„, æ"å°æç«™(3è™¥å‡ºå£)	014A	000700042F	2021-07-31 23:07:22	22	æ%å±è»šç«™	011A	0:09:14	80385B3A	L01
15382	5	2021-07-31 23:12:49	241	æ—å£å...¬åç'	002B	0005000036	2021-07-31 23:19:03	22	æ%å±è»šç«™	013A	0:06:14	EF65B436	L01

15383 rows × 13 columns

借車站及還車站有亂碼

資料集介紹 / 資料前處理-YouBike租借量

Intro to Datasets

篩選各捷運站附近的YouBike站 - 以松山站為例

將借車站及還車站亂碼轉為中文

	扣款金額	借車時間	借車站代號	借車站	借車車柱	自行車編號	還車時間	還車站代號	還車站	還車車柱	租用時間	手機序號	費率別
0	5	2021-07-22 23:07:44	22	松山車站	015A	0000002647	2021-07-22 23:24:54	170	新東公園	016B	0:17:10	6D753136	L04
1	5	2021-07-22 23:00:03	22	松山車站	012B	0005001325	2021-07-22 23:08:18	40	玉成公園	001B	0:08:15	675DF23A	L04
2	5	2021-07-22 21:42:14	22	松山車站	002B	00050003A9	2021-07-22 21:45:03	28	五常公園	004A	0:02:49	C2100D37	L04
3	25	2021-07-22 21:32:05	22	松山車站	020A	00050026AC	2021-07-22 22:48:20	130	羅斯福寧波東街口	001A	1:16:15	846F9B3A	L04
4	5	2021-07-22 21:22:22	22	松山車站	018B	0007000348	2021-07-22 21:37:59	113	仁愛逸仙路口	018B	0:15:37	14BAD03A	L04
...
17346	5	2021-07-24 18:11:09	22	松山車站	012A	00050027FA	2021-07-24 18:40:27	156	東湖國小	001B	0:29:18	FE63D936	L01
17347	15	2021-07-24 05:55:21	22	松山車站	016B	00000025D6	2021-07-24 06:54:13	291	捷運松山站(3號出口)	017A	0:58:52	FEC9E637	L01
17348	5	2021-07-24 01:23:57	22	松山車站	001B	0000000BAD	2021-07-24 01:30:36	26	捷運昆陽站(1號出口)	001B	0:06:39	FF414336	L01
17349	35	2021-07-24 20:31:25	22	松山車站	003B	0007000A31	2021-07-24 22:12:47	22	松山車站	015B	1:41:22	FFD74A37	L01
17350	35	2021-07-24 20:35:01	22	松山車站	012B	0000002D05	2021-07-24 22:13:06	22	松山車站	011A	1:38:05	FFD74A37	L01

17351 rows × 13 columns



資料集介紹 / 資料前處理 - YouBike租借量

Intro to Datasets

篩選各捷運站附近的YouBike站 - 以松山站為例

計算在每1個小時，在YouBike借車、還車人數

借車人數

	站點	借車數	時間
0	松山車站	15.0	2021-07-01 00:00:00
1	松山車站	2.0	2021-07-01 01:00:00
2	松山車站	2.0	2021-07-01 02:00:00
3	松山車站	1.0	2021-07-01 03:00:00
4	松山車站	1.0	2021-07-01 04:00:00
...
739	松山車站	59.0	2021-07-31 19:00:00
740	松山車站	50.0	2021-07-31 20:00:00
741	松山車站	40.0	2021-07-31 21:00:00
742	松山車站	33.0	2021-07-31 22:00:00
743	松山車站	6.0	2021-07-31 23:00:00

744 rows × 3 columns

還車人數

	站點	還車數	時間
0	松山車站	4.0	2021-07-01 00:00:00
1	松山車站	4.0	2021-07-01 01:00:00
2	松山車站	3.0	2021-07-01 02:00:00
3	松山車站	1.0	2021-07-01 03:00:00
4	松山車站	4.0	2021-07-01 04:00:00
...
739	松山車站	66.0	2021-07-31 19:00:00
740	松山車站	28.0	2021-07-31 20:00:00
741	松山車站	31.0	2021-07-31 21:00:00
742	松山車站	32.0	2021-07-31 22:00:00
743	松山車站	4.0	2021-07-31 23:00:00

744 rows × 3 columns

資料集介紹 / 資料前處理-YouBike租借量

Intro to Datasets

合併各捷運站附近的YouBike站 - 以臺北101/世貿為例

合併基準：

找尋捷運站附近步行10分鐘內可到達的YouBike站點進行合併。

YouBike站與捷運站距離

- 捷運臺北101/世貿站：43mm(步行1分鐘)
- 信義基隆路口：300mm(步行3分鐘)
- 三張犁路口：600mm(步行8分鐘)
- 市民廣場路口：450mm(步行6分鐘)



綠色：捷運站鄰近的YouBike站

藍色：捷運站



資料集介紹 / 資料前處理-天氣資料

Intro to Datasets

篩選7月份天氣資料

	時間	觀測時間	氣溫	相對溼度	風速	降水量	紫外線指數
0	2021-07-01 00:00:00	24	27.3	75	0.9	0	0
1	2021-07-01 01:00:00	1	27.4	74	1.2	0	0
2	2021-07-01 02:00:00	2	27.2	74	1.6	0	0
3	2021-07-01 03:00:00	3	27	75	0.1	0	0
4	2021-07-01 04:00:00	4	27.1	73	1.3	0	0
...
739	2021-07-31 19:00:00	19	30	62	1.3	0	0
740	2021-07-31 20:00:00	20	29.4	69	0.3	0	0
741	2021-07-31 21:00:00	21	29.1	72	1.5	0	0
742	2021-07-31 22:00:00	22	28.5	73	1.1	0	0
743	2021-07-31 23:00:00	23	28.1	71	1.1	0	0

744 rows × 7 columns

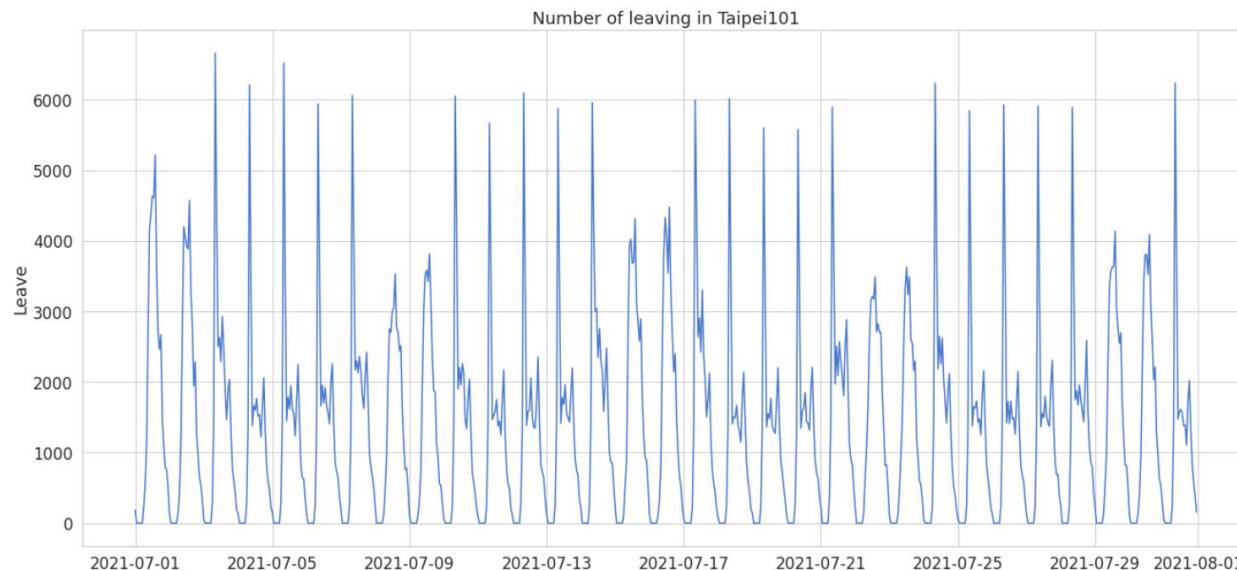
以時間、氣溫、相對濕度、風速、降水量、紫外線指數等因素進行預測

資料敘述 - 捷運站出站人流量

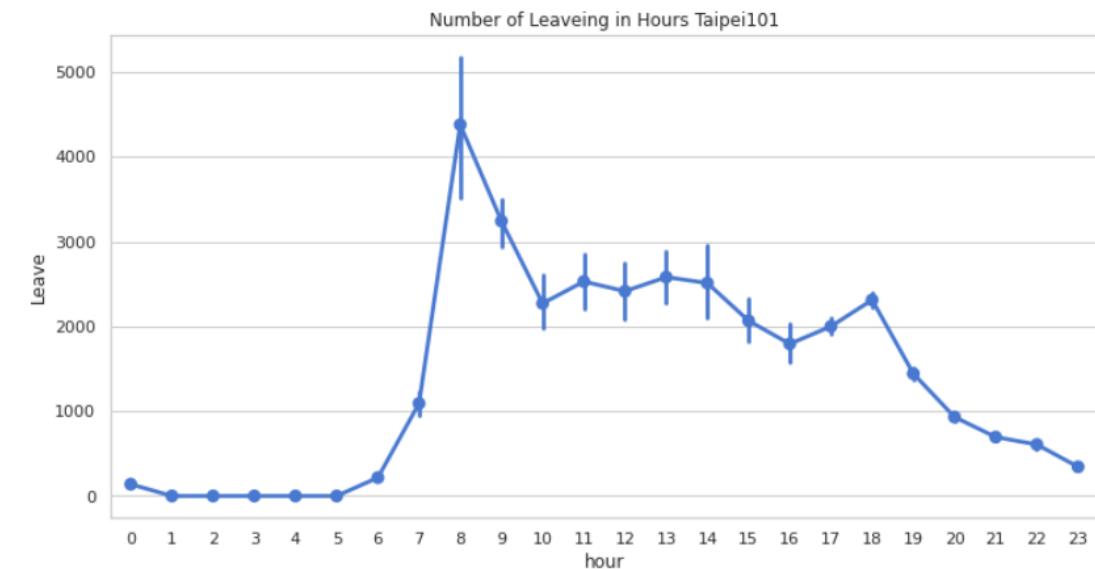
Data Statement

以臺北101/世貿站為例

每日出站人數



每小時出站人數



左圖：看出在7月份的資料中每日出站人數最高落於6000人左右，且每日都依循這個規律。

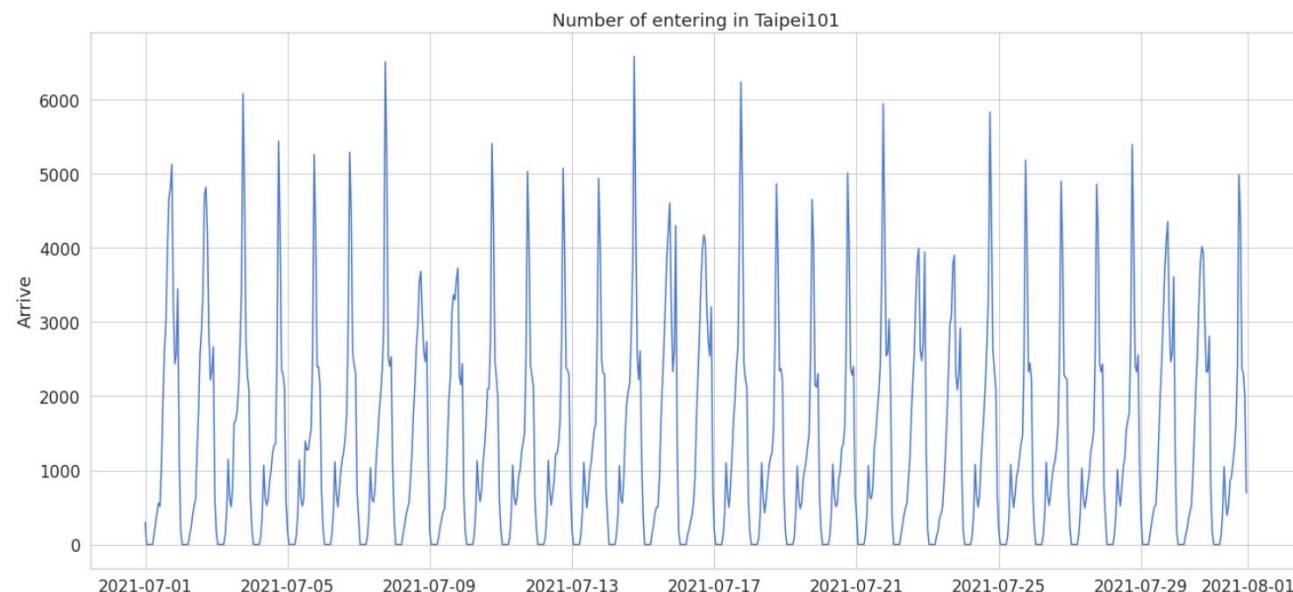
右圖：7/1出站人數大概在中午過後到下午3點間會到達高峰→中午過後大家較容易來到臺北101/世貿站進行活動。

資料敘述 – 捷運站進站人流量

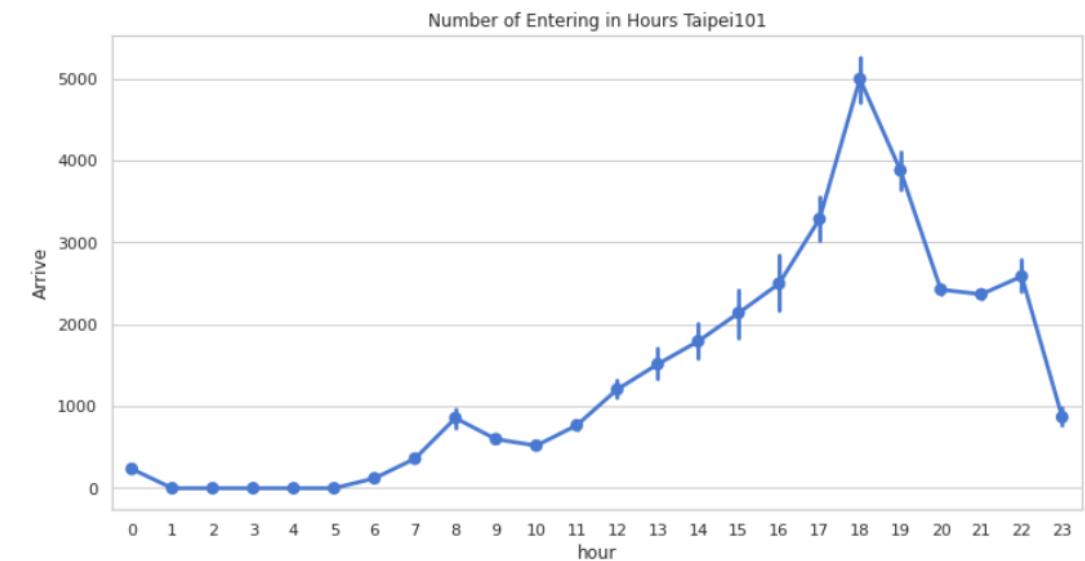
Data Statement

以臺北101/世貿站為例

每日進站人數



每小時進站人數



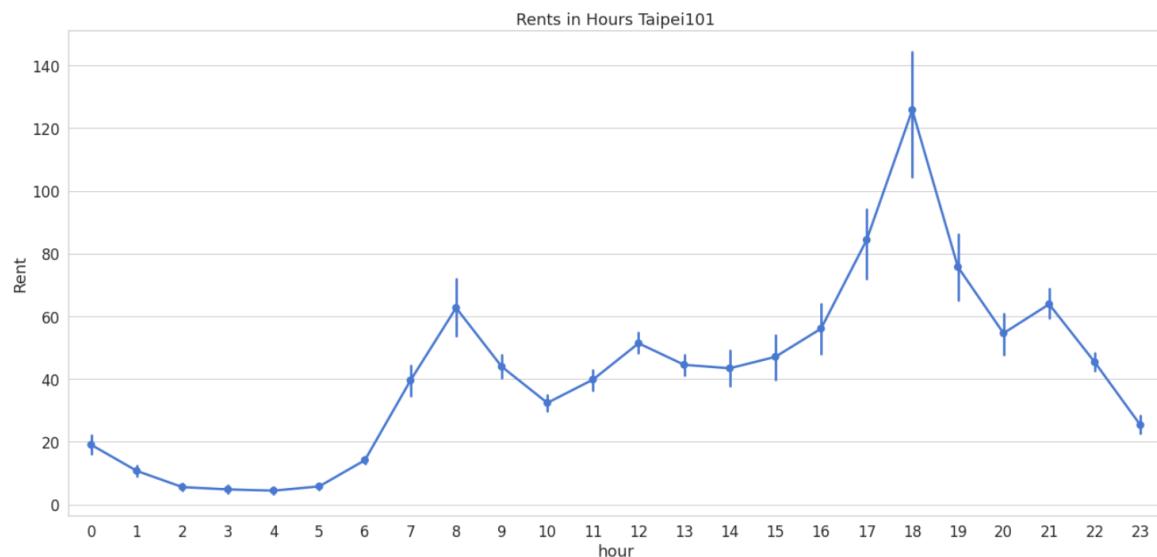
右圖：進站人數大概在中午過後到下午 6 點間持續上升，在晚間 9 點過後有出現第 2 次進站高峰，可以由此推斷由於臺北101/世貿站鄰近信義商圈，大多數店家營業時間到晚間 10 點結束，故使用量增加。

資料敘述 – YouBike租借量

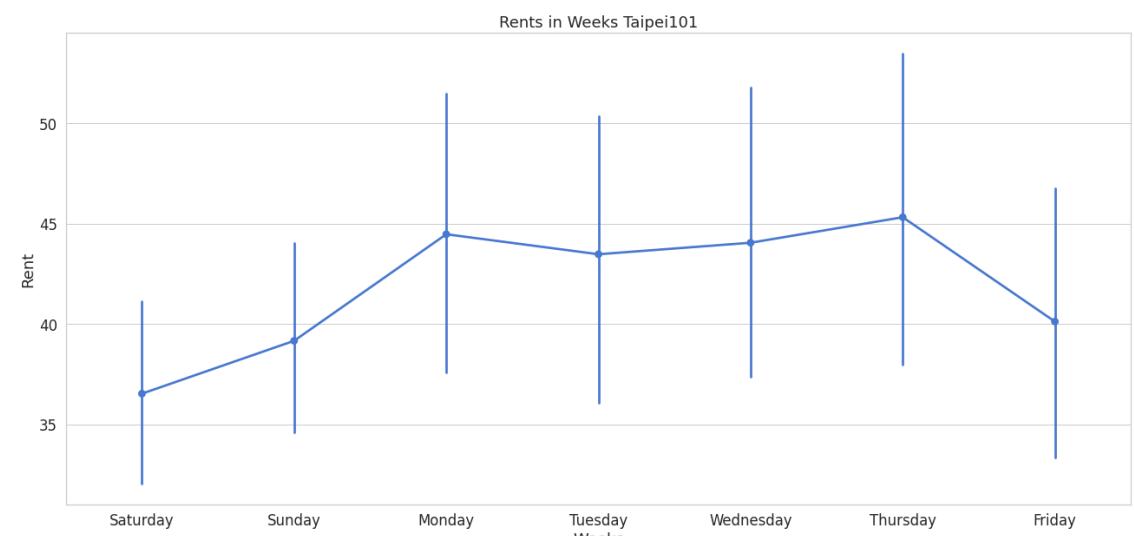
Data Statement

以臺北101/世貿站附近YouBike站為例

每小時租借腳踏車人數



每星期租借腳踏車人數



左圖：看出在上午6點到8點間租車人數有上升，在晚上6點時達到整日租借量最高，在晚間9點時，正好是商場關閉時間，因此有些民眾租借腳踏車離開。

右圖：看出平日維持著穩定的租借量，但到星期五以後，租借量有下降的情況。

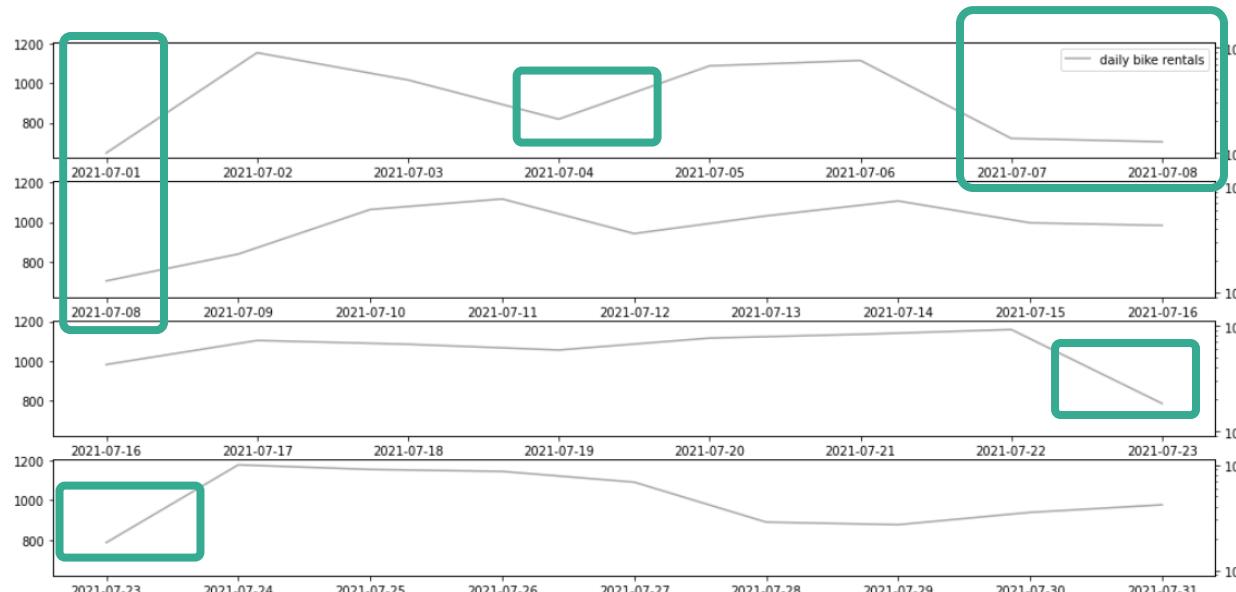
資料敘述 - 天氣

Data Statement

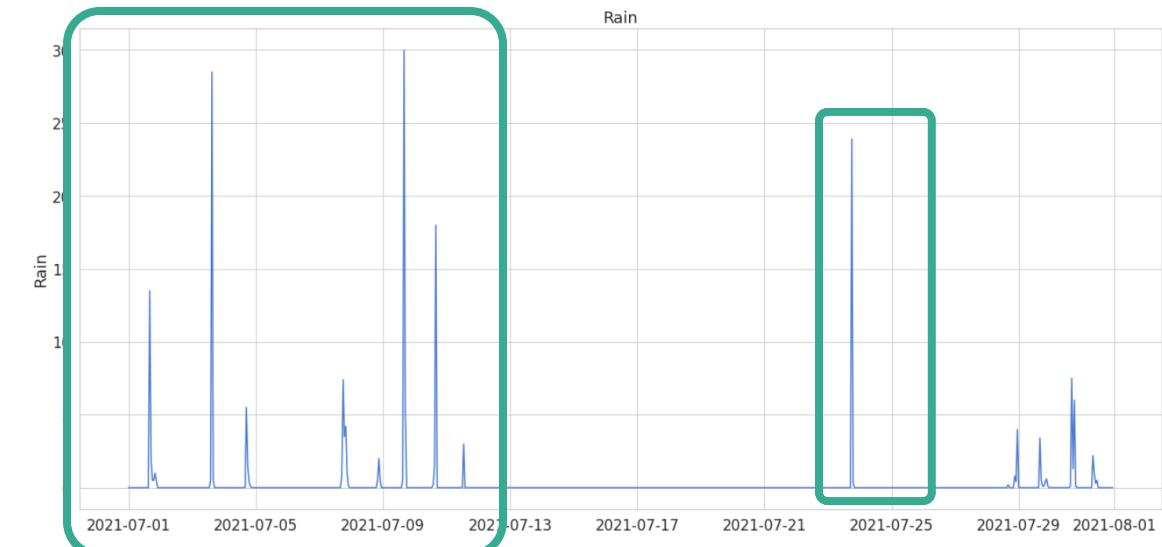
以臺北101/世貿站附近YouBike站為例

YouBike租借量

Daily bike rentals and number of hours in Taipei101



雨量圖



可以看出雨量會影響到租借量，在7/1、7/4、7/7、7/8、7/23這幾天租借量有下降。

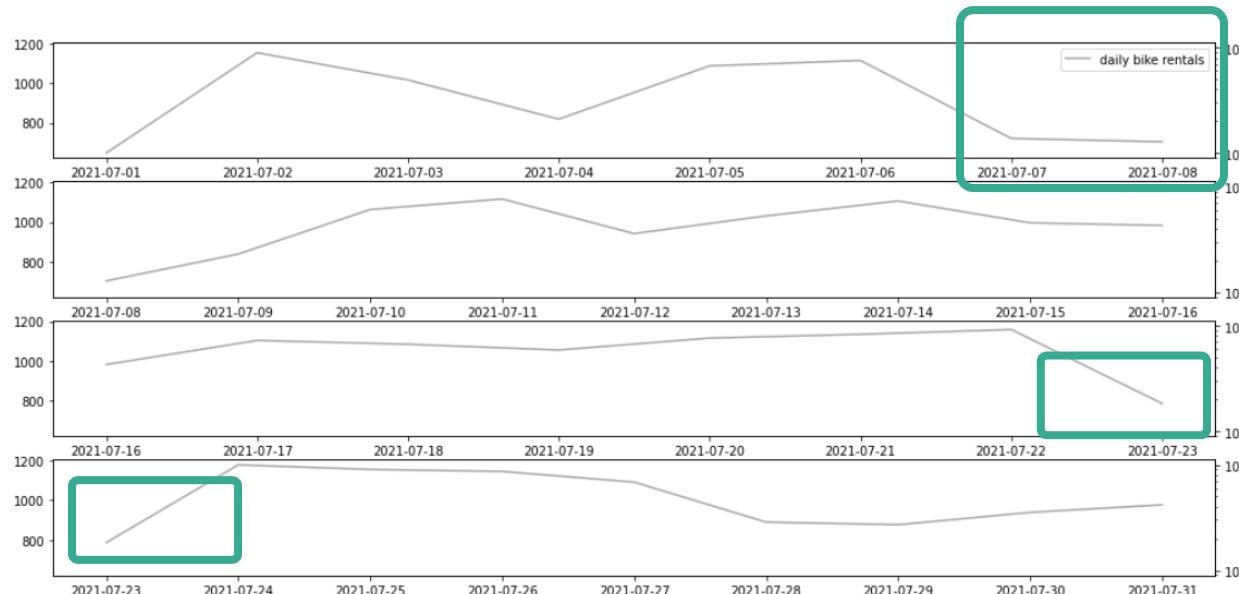
資料敘述 - 天氣

Data Statement

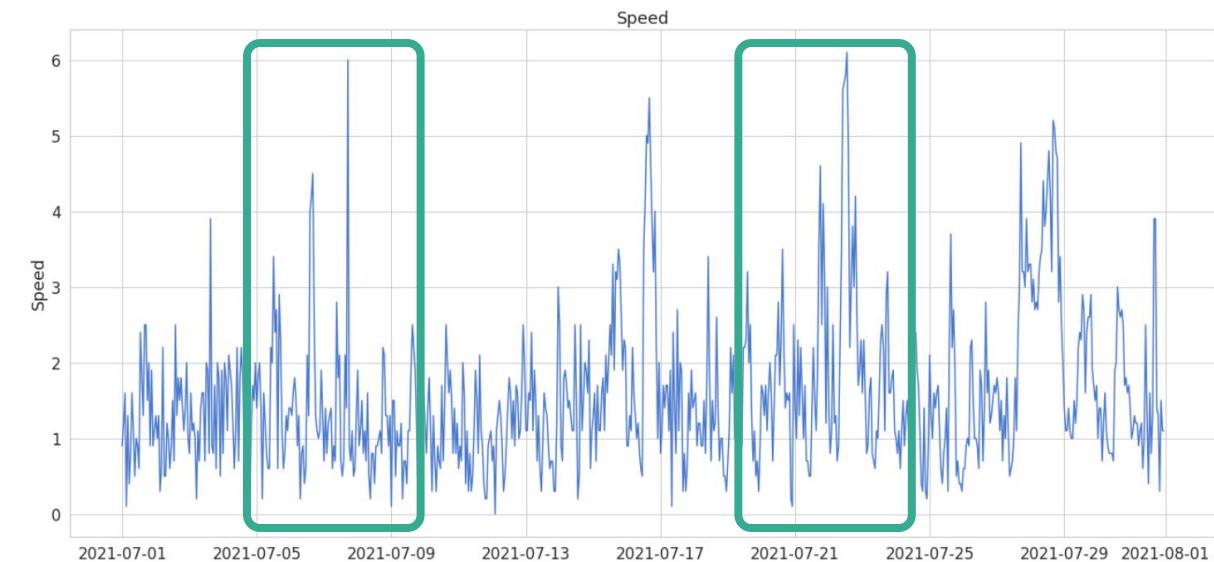
以臺北101/世貿站附近YouBike站為例

YouBike租借量

Daily bike rentals and number of hours in Taipei101



風速



可以看出風速會影響到租借量，在7/7、7/8、7/23這三天有因為風速較高，而租借量有下降的情況。

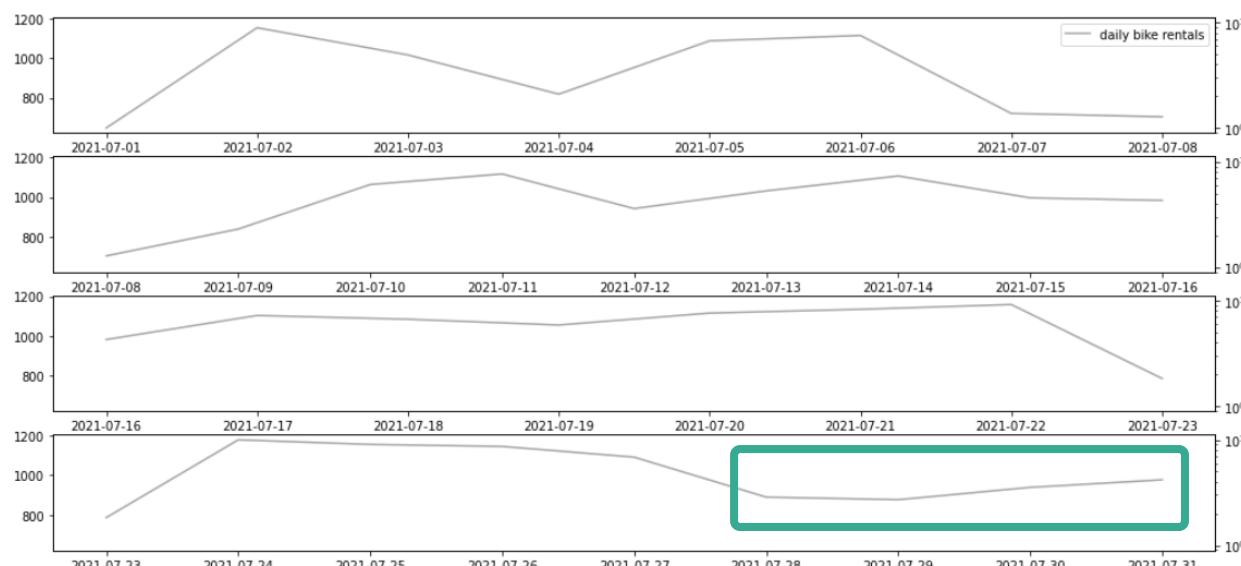
資料敘述 - 天氣

Data Statement

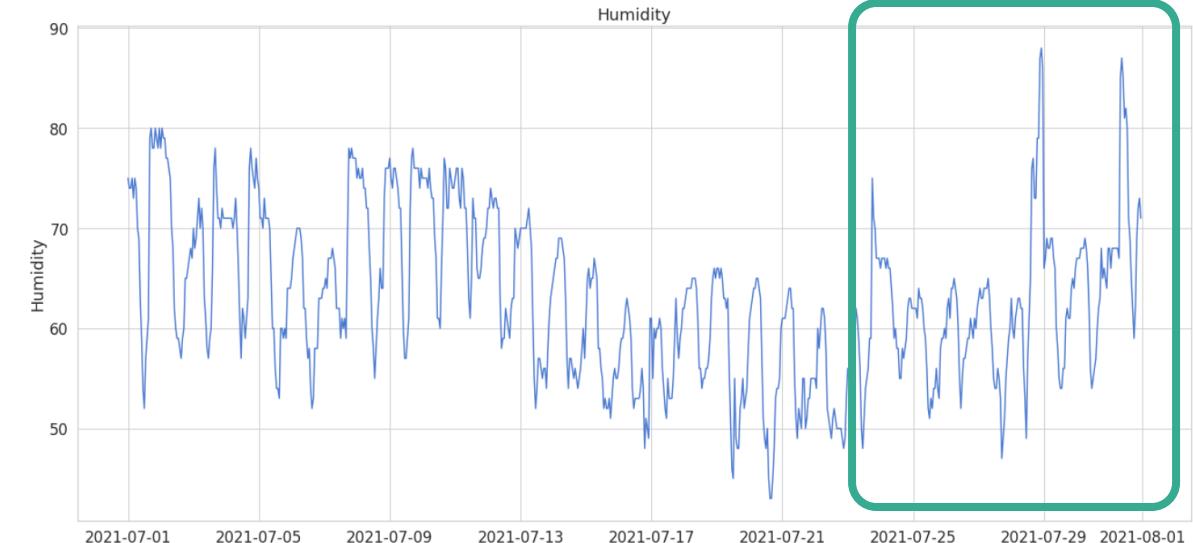
以臺北101/世貿站附近YouBike站為例

YouBike租借量

Daily bike rentals and number of hours in Taipei101



相對溼度



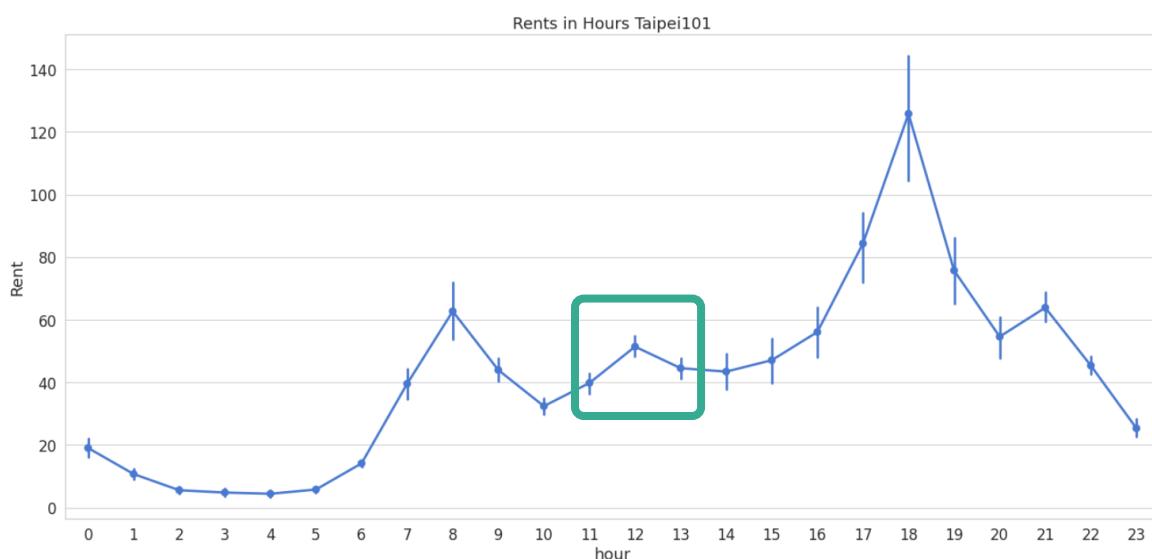
可以看出相對溼度較高的情況下，會導致有些天數的租借量相較先前的其他天低。

資料敘述 - 天氣

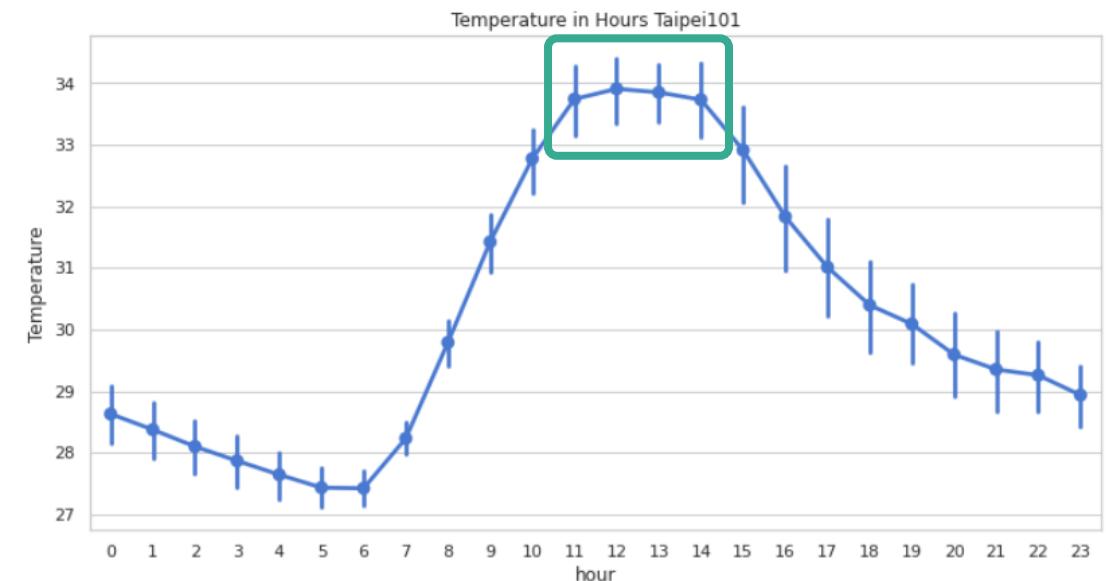
Data Statement

以臺北101/世貿站附近YouBike站為例

YouBike每小時平均租借量



氣溫



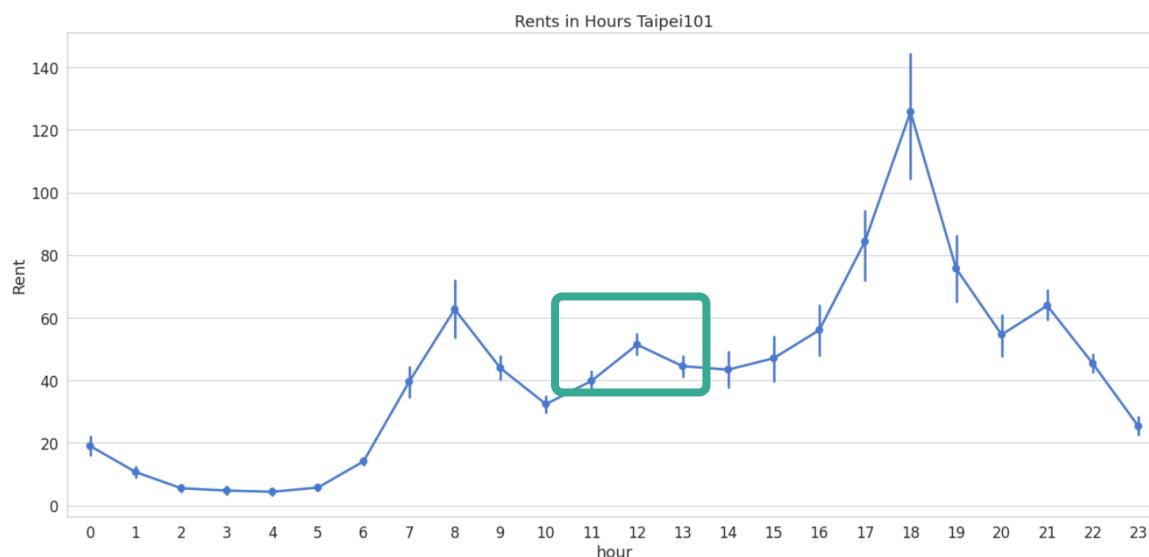
7月的氣溫維持在27到34度之間，但當每日氣溫來到最高溫，租借量反而上升不多，雖然中午為上班族的休息期間或吃飯時間，但租借量卻沒有如同晚餐時段一樣高租借量。

資料敘述 - 天氣

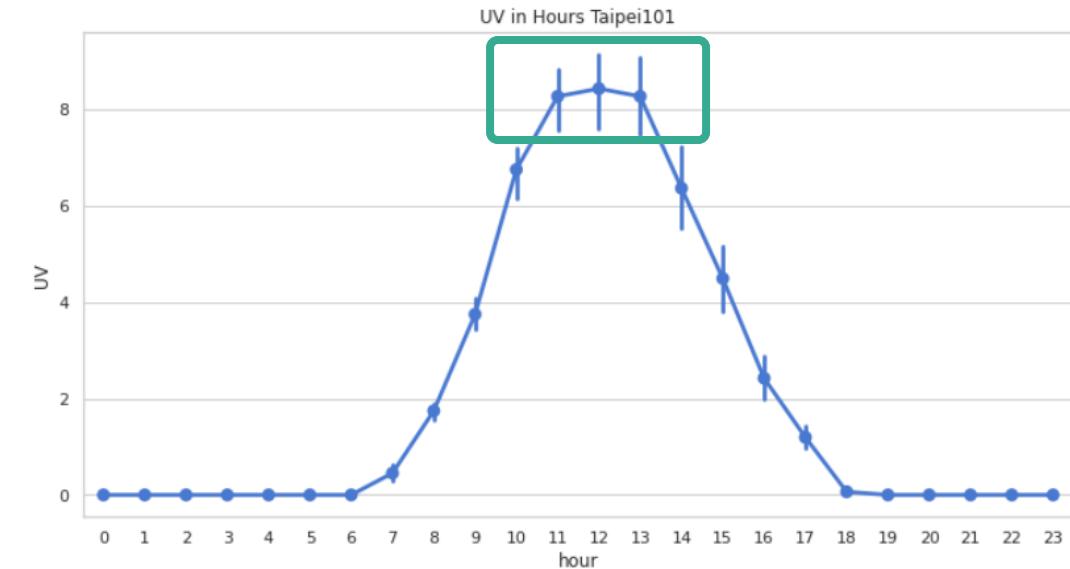
Data Statement

以臺北101/世貿站附近YouBike站為例

YouBike每小時平均租借量



紫外線指數



可以看出紫外線指數高的情況下，中午時段雖然是上班族中午休息或吃飯時間，但借車數並沒有像晚間時那麼高，原因可能是中午的紫外線指數很高。

回歸分析

Regression

劃分訓練集及測試集

```
X_train, X_test, y_train, y_test = train_test_split(X2, y2, test_size=0.1\  
                                                 , random_state=20)
```

```
print('Train size:', len(X_train))  
print('Test size:', len(X_test))
```

Train size: 669
Test size: 75

訓練集：699筆

測試集：75筆

回歸分析

Regression

以臺北101/世貿站附近YouBike站為例

回歸式

$$\hat{y} = -38.72 + 2.46x_1 - 0.24x_2 - 2.45x_3 - 2.23x_4 - 2.46x_5 + 0.01x_6 + 0.01x_7$$

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error
lr_rmses = []
lr = LinearRegression()
lr.fit(X_train, y_train)
lr_prediction = lr.predict(X_test)
residual = (y_test - lr_prediction)
Rsq = lr.score(X_train, y_train)
w_0 = lr.intercept_
w_1 = lr.coef_

print('Interception : ', w_0)
print('Coeficient : ', w_1)
print("R_sq : ", round(r2_score(y_test, lr_prediction), 3))
print("MSE : ", round(mean_squared_error(y_test, lr_prediction), 3))
print("RMSE : ", round(mean_squared_error(y_test, lr_prediction)**0.5))
```

```
Interception : -38.723603648781435
Coeficient : [ 2.4623341 -0.23827925 -2.44671071 -2.23444 -2.46183454  0.00721457
 0.01521158]
R_sq : 0.533
MSE : 515.725
RMSE : 23
```

x_1 : 氣溫	x_5 : 紫外線指數
x_2 : 相對溼度	x_6 : 捷運站進站人數
x_3 : 風速	x_7 : 捷運站出站人數
x_4 : 降雨量	

準確率：53.3%

負相關：相對溼度、風速、雨量、紫外線

正相關：氣溫、進站人數、出站人數

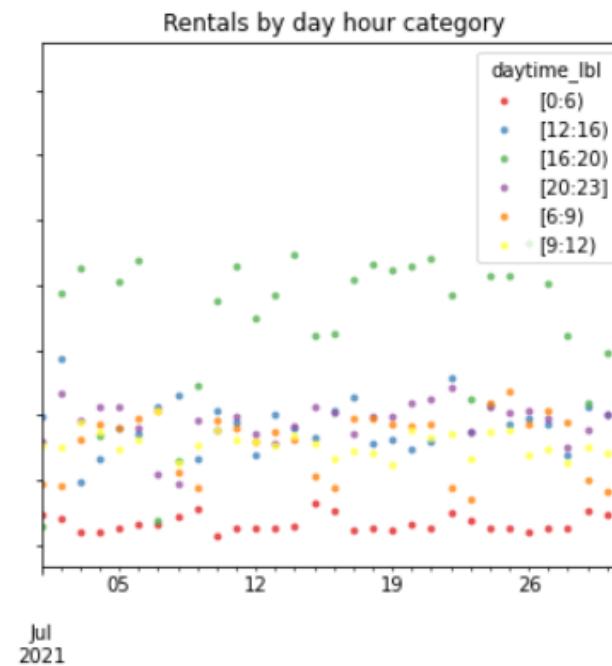
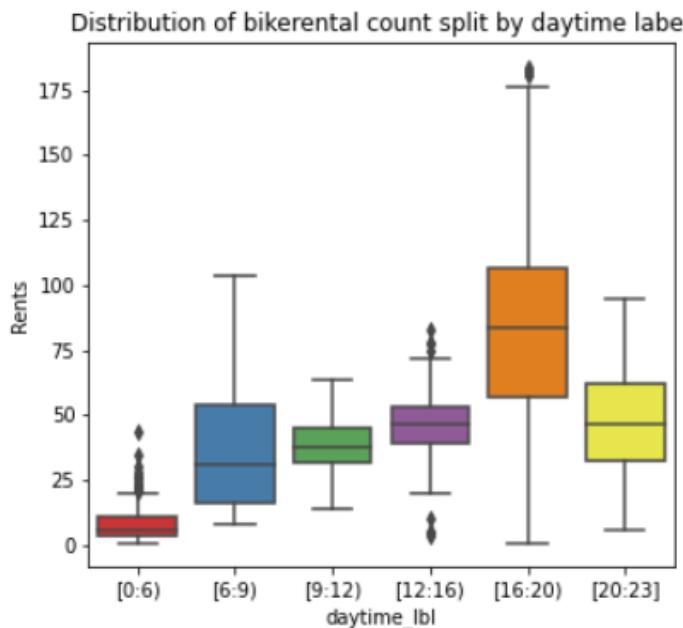
加入時間因素進行訓練模型，期望提升模型準確率

回歸分析

Regression

以臺北101/世貿站附近YouBike站為例

加入時間因素 → 將每小時劃分區間，轉換為dummy variable



總共分為6組

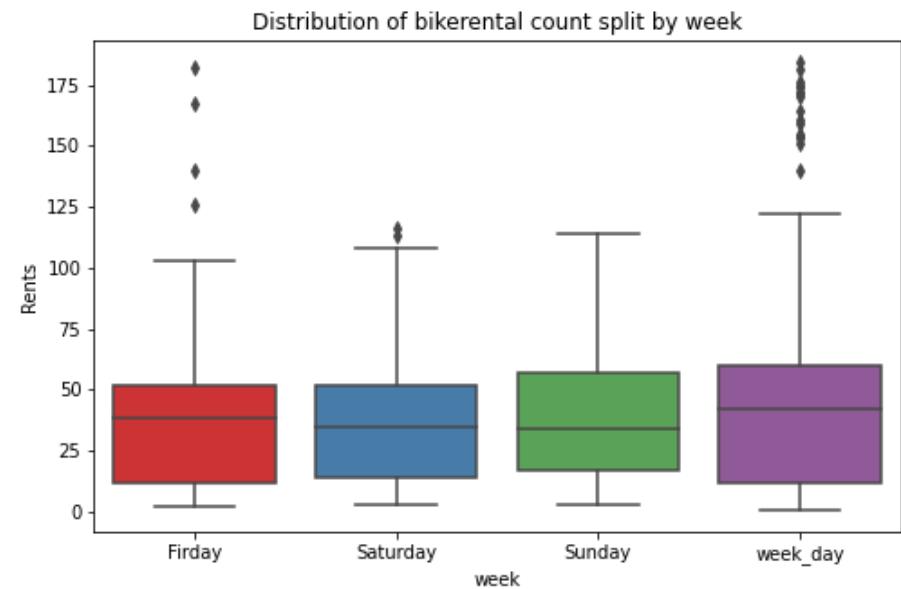
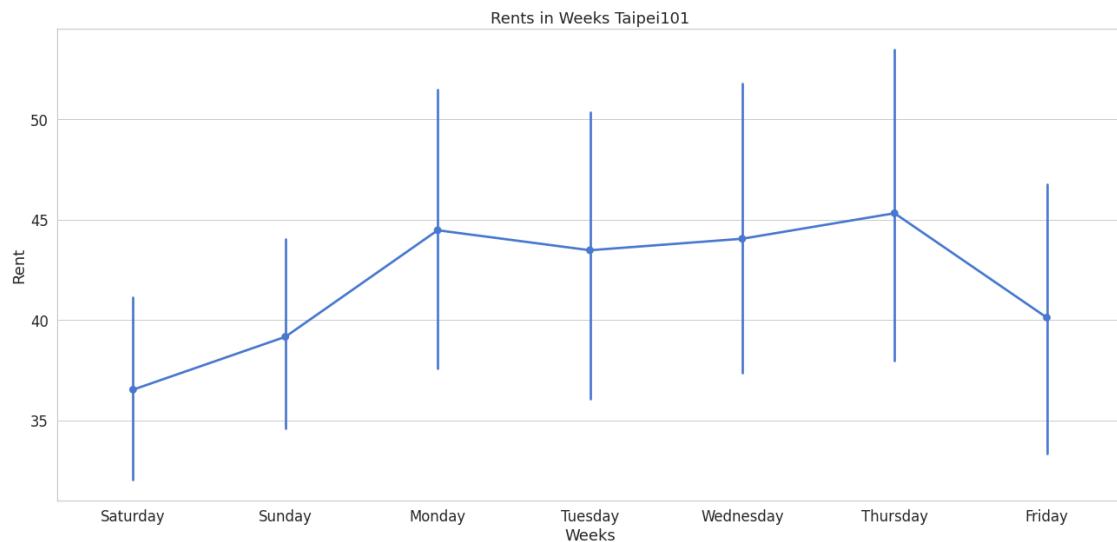
- 0~5點：凌晨，較少人租借腳踏車
- 6~8點：上午通勤時間，學生、上班族租借腳踏車，為全天第二多租借量的時段
- 9~11點：上午上班/課時間
- 12~15點：中午時段，為全天第三多的借車量
- 16~19點：下午通勤時間、晚餐時間，為全天第一多的借車量
- 20~23點：夜間活動時間

回歸分析

Regression

以臺北101/世貿站附近YouBike站為例

加入時間因素 → 將星期劃分區間，轉換為dummy variable



- 星期五：比平日租借量低
- 星期六：為一周內最低租借量
- 星期日：比起星期六有開始回升的趨勢
- 平日：維持在一定的租借量

回歸分析

Regression

以臺北101/世貿站附近YouBike站為例

回歸式

$$\hat{y} = -23.15 + 1.79x_1 - 0.32x_2 - 1.3x_3 - 2.32x_4 - 1.42x_5 + 0.005x_6 + 0.02x_7 \\ - 2.21x_8 + 8.89x_9 + 1.4x_{10} + 13.31x_{11} + 4.1x_{12} - 3.16x_{13} + 1.59x_{14} + 7.8x_{15}$$

```
lr_rmses = []
lr = LinearRegression()
lr.fit(X_train,y_train)
lr_prediction = lr.predict(X_test)
residual = (y_test - lr_prediction)
Rsq = lr.score(X_train,y_train)
w_0 = lr.intercept_
w_1 = lr.coef_

print('Interception : ', w_0)
print('Coeficient : ', w_1)
print("R_sq : ",round(r2_score(y_test,lr_prediction),3))
print("MSE : ",round(mean_squared_error(y_test,lr_prediction),3))
print("RMSE : ",round(mean_squared_error(y_test,lr_prediction)**0.5))

Interception : -23.146736137858923
Coeficient : [ 1.79148324e+00 -3.18021451e-01 -1.30150616e+00 -2.31633497e+00
 -1.42337481e+00 5.70632625e-03 1.53562659e-02 -2.21264615e+00
 8.89496789e+00 1.40419826e+00 1.33062636e+01 4.10041851e+00
 -3.15807790e+00 1.58576825e+00 7.80106614e+00]
R_sq : 0.62
MSE : 419.814
RMSE : 20
```

x_1 : 氣溫
 x_2 : 相對溼度
 x_3 : 風速
 x_4 : 降雨量
 x_5 : 紫外線指數
 x_6 : 捷運站進站人數
 x_7 : 捷運站出站人數

x_8 : $\begin{cases} 1: 6\text{點到}8\text{點} \\ 0: \text{非}6\text{點到}8\text{點} \end{cases}$

x_9 : $\begin{cases} 1: 9\text{點到}11\text{點} \\ 0: \text{非}9\text{點到}11\text{點} \end{cases}$

x_{10} : $\begin{cases} 1: 12\text{點到}15\text{點} \\ 0: \text{非}12\text{點到}15\text{點} \end{cases}$

x_{11} : $\begin{cases} 1: 16\text{點到}19\text{點} \\ 0: \text{非}16\text{點到}19\text{點} \end{cases}$

x_{12} : $\begin{cases} 1: 20\text{點到}23\text{點} \\ 0: \text{非}20\text{點到}23\text{點} \end{cases}$

x_{13} : $\begin{cases} 1: \text{星期六} \\ 0: \text{非星期六} \end{cases}$

x_{10} : $\begin{cases} 1: \text{星期日} \\ 0: \text{非星期日} \end{cases}$

x_{11} : $\begin{cases} 1: \text{平日} \\ 0: \text{非平日} \end{cases}$

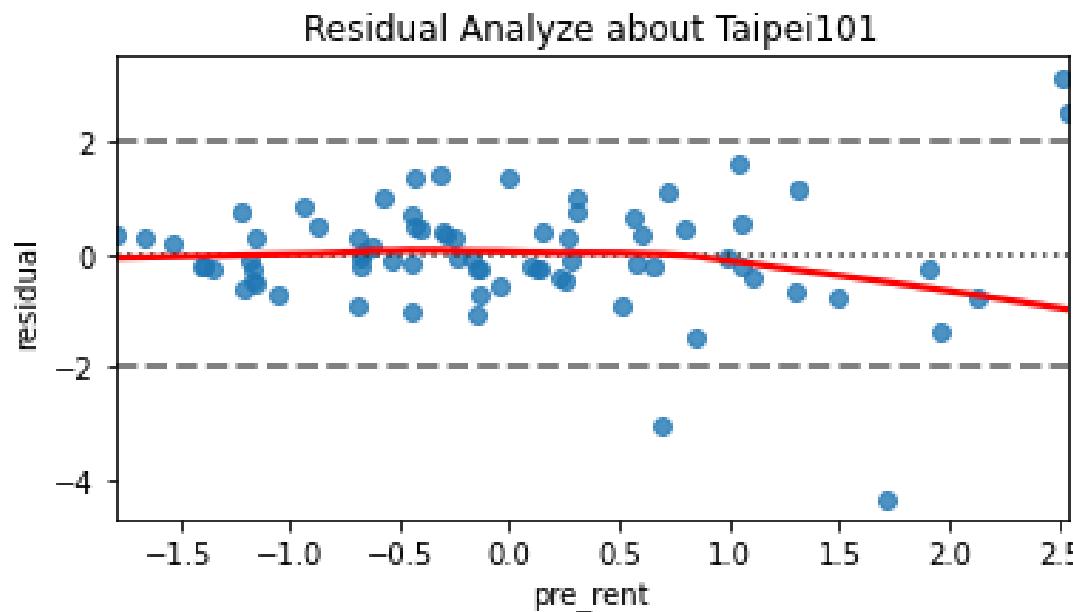
準確率：62%

回歸分析

Regression

以臺北101/世貿站附近YouBike站為例

殘差分析



可以看出殘差的分布大概落在2~ -2之間



回歸分析

Regression

各個站點的準確率比較

站點	加入時間因素前準確率	加入時間因素後準確率
臺北101/世貿站	53.3%	62%
國父紀念館站	66.3%	68%
市政府站	53%	55.5%
松山站	61.8%	65.5%
永春站	71.3%	73.9%
象山站	66.4%	65.2%

大部分的站點加入時間因素進行模型的訓練準確率都有提高

NN Model – Bidirectional LSTM

Bidirectional LSTM

劃分訓練集及測試集

```
train_size = int(len(df)*0.9)
test_size = len(df) - train_size
train, test = df.iloc[0:train_size], df.iloc[train_size:len(df)]
print(train.shape, test.shape)
```

(669, 8) (75, 8)

訓練集：699筆

測試集：75筆

NN Model – Bidirectional LSTM

Bidirectional LSTM

以臺北101/世貿站附近YouBike站為例

模型建置

```
model = keras.Sequential()
model.add(keras.layers.Bidirectional(keras.layers.LSTM(units=256, activation='relu', return_sequences=True),
                                     input_shape=(X_train.shape[1], X_train.shape[2])))
model.add(Dropout(0.2))
model.add(keras.layers.Bidirectional(keras.layers.LSTM(units=64, activation='relu', return_sequences=True)))
model.add(Dropout(0.2))
model.add(keras.layers.Bidirectional(keras.layers.LSTM(units=16, activation='relu')))
model.add(Flatten())
model.add(Dense(10, activation='linear'))
model.add(keras.layers.Dense(units=1))

opt = optimizers.Adam(learning_rate=0.0001)
model.compile(loss='MSE', optimizer=opt)
```

透過三層雙向LSTM進行模型堆疊，最後透過兩層全連接層進行維度的下降，在優化器的部分，透過學習速率設定調整模型收斂程度，MSE進行損失函數的計算。

Layer (type)	Output Shape	Param #
=====		
bidirectional (Bidirectional)	(None, 1, 512)	542720
1)		
dropout (Dropout)	(None, 1, 512)	0
bidirectional_1 (Bidirectional)	(None, 1, 128)	295424
na1)		
dropout_1 (Dropout)	(None, 1, 128)	0
bidirectional_2 (Bidirectional)	(None, 32)	18560
na1)		
flatten (Flatten)	(None, 32)	0
dense (Dense)	(None, 10)	330
dense_1 (Dense)	(None, 1)	11
=====		
Total params:	857,045	
Trainable params:	857,045	
Non-trainable params:	0	

NN Model – Bidirectional LSTM

Bidirectional LSTM

以臺北101/世貿站附近YouBike站為例

損失函數

```
history = model.fit(X_train, y_train, epochs=200, batch_size=32, validation_split=0.1, shuffle=False)
```

```
Epoch 1/200  
19/19 [=====] - 9s 74ms/step - loss: 0.6927 - val_loss: 0.8796  
Epoch 2/200  
19/19 [=====] - 0s 23ms/step - loss: 0.6857 - val_loss: 0.8699  
Epoch 3/200
```



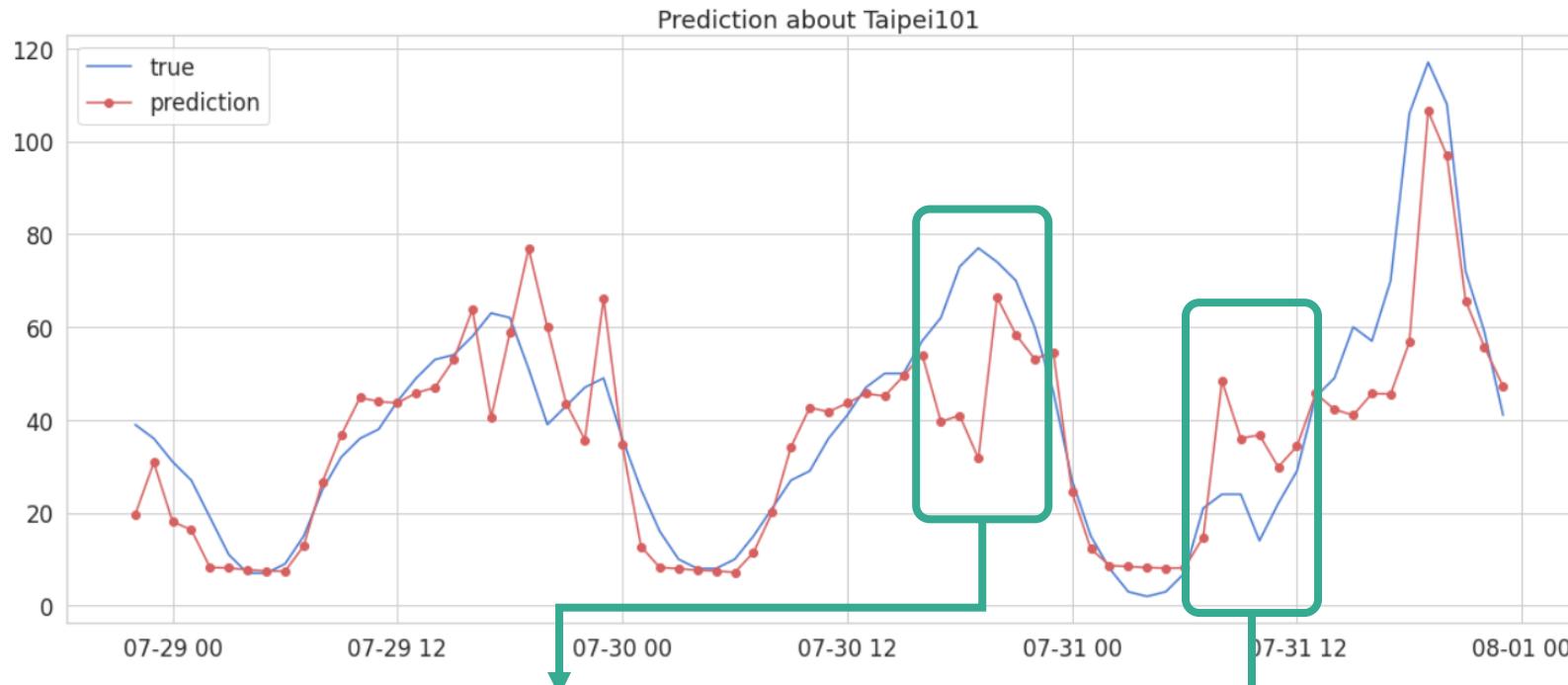
紀錄模型執行200次的損失函數，在batch_size大小都維持在32時，損失函數確實有下降，且有逐漸靠攏的情況。

NN Model – Bidirectional LSTM

Bidirectional LSTM

以臺北101/世貿站附近YouBike站為例

模型執行後結果



在前面訓練的時候可能沒有訓練在這個時段有什麼特別的因素使預測整體的租借量下降或者是上升

MAE 0.22364962708859423
MSE 0.10894894870846035
RMSE 0.33007415637771514
R2 0.723110427805925

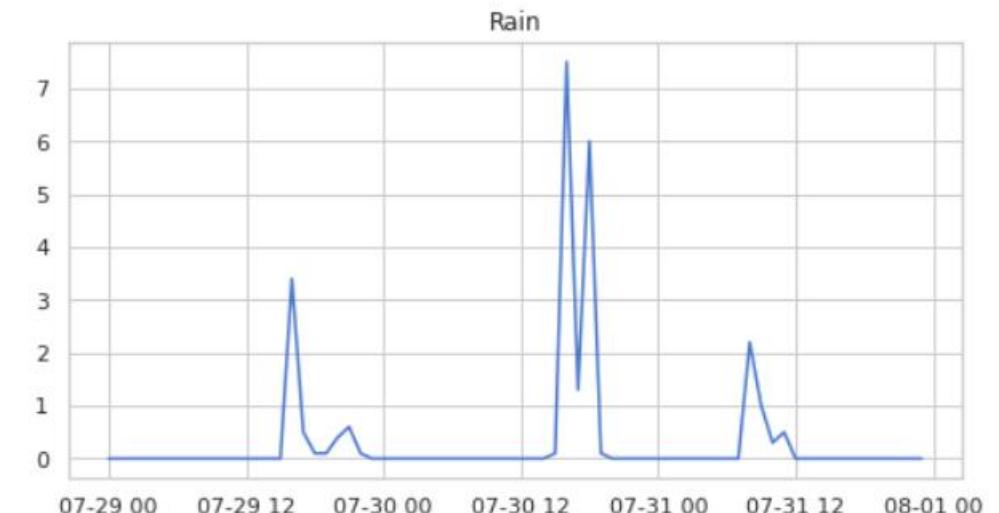
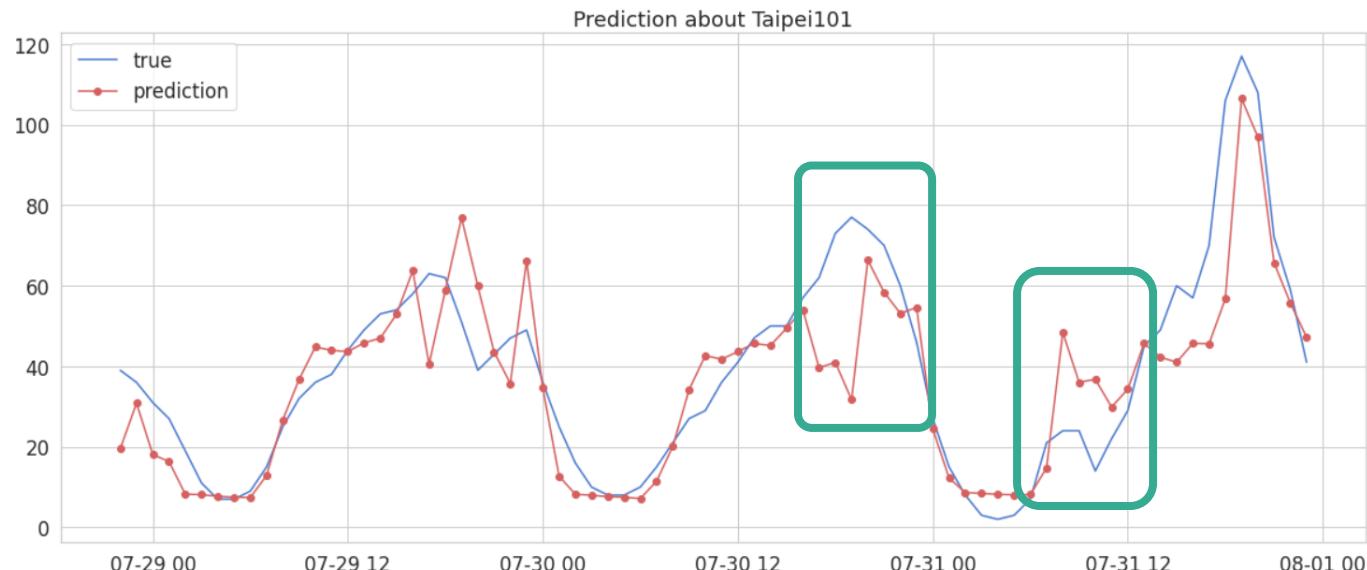
藍線：實際值，
紅線：預測值。
準確率= 72.31 %。

NN Model – Bidirectional LSTM

Bidirectional LSTM

以臺北101/世貿站附近YouBike站為例

執行後借車結果與雨量比較



右邊2021/07/29~2021/07/31的雨量，可以看到7/30和7/31都是有下雨的，但下雨的大小不同，可以看到模型有預測出雨量會造成租借量下降，但下降的幅度是有誤的。



NN Model – Bidirectional LSTM

Bidirectional LSTM

各個站點的準確率比較

站點	回歸模型	NN Model
臺北101/世貿站	62%	72.31%
國父紀念館站	68%	74.99%
市政府站	55.5%	75.47%
松山站	65.5%	75.64%
永春站	73.9%	77.56%
象山站	65.2%	69.96%

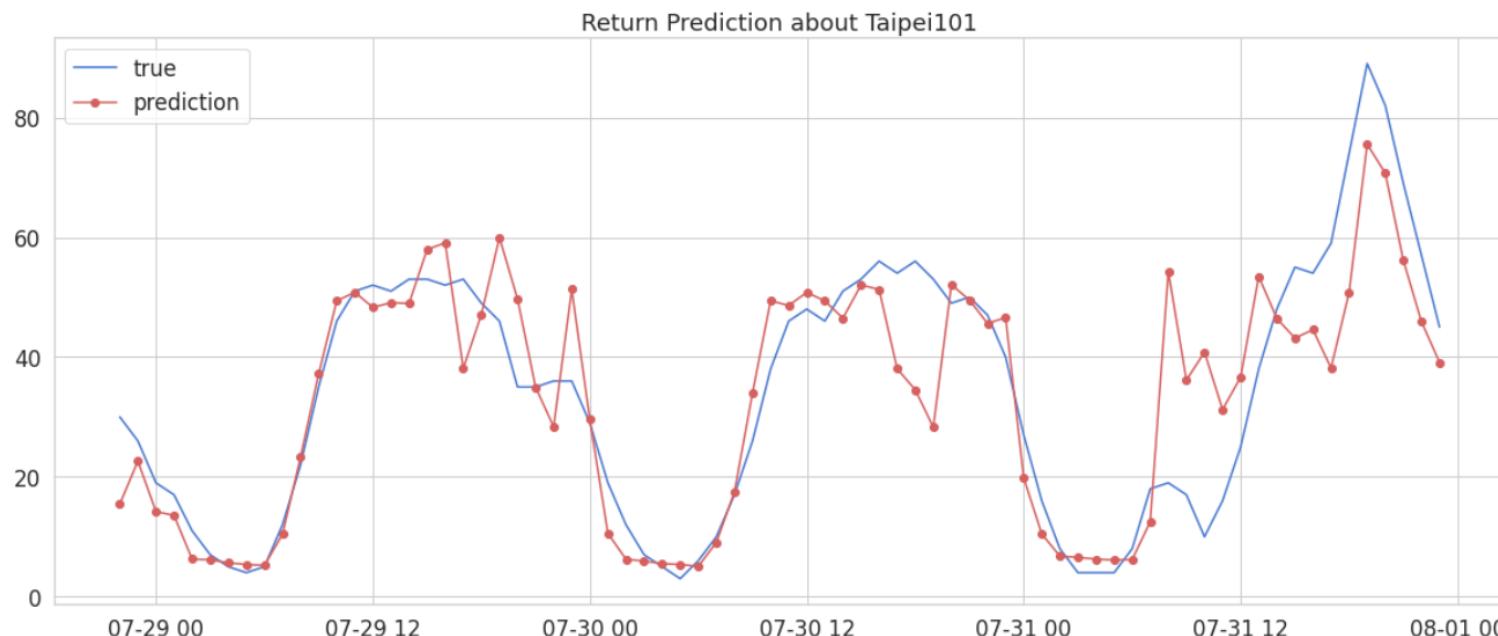
使用LSTM所訓練出的預測模型準確率比起回歸模型來說高，且整體來說都有超過70%的準確率

模型應用

Model Application

以臺北101/世貿站附近YouBike站為例

模型還車執行後結果



MAE 0.1790541484445486
MSE 0.06624171294597392
RMSE 0.25737465482439004
R2 0.7510186434787062

藍線：實際值，
紅線：預測值。
準確率 = 75.1%。

在預測還車量時，整體的準確率比預測借車輛時高，但仍有些地方預測不準確。

模型應用

Model Application

各個站點的預測借、還車準確率比較

站點	借車	還車
臺北101/世貿站	72.31%	75.1%
國父紀念館站	74.99%	76.39%
市政府站	75.47%	79.66%
松山站	75.64%	77.64%
永春站	77.56%	75.18%
象山站	69.96%	70.24%

借、還車的準確率沒有相差很大，因此，接下來綜合兩者訓練出的模型，進行
8/1~8/3的租借量預測。

模型應用

Model Application

8/1~8/3的氣溫及捷運人流資料

時間	氣溫	相對溼度	風速	降水量	紫外線指數	出站人數	進站人數
2021-08-01 00:00:00	27.4	76	1.9	0.0	0	48.0	50.0
2021-08-01 01:00:00	27.3	74	1.2	0.0	0	4.0	0.0
2021-08-01 02:00:00	27.1	75	0.8	0.0	0	0.0	0.0
2021-08-01 03:00:00	26.9	76	0.9	0.0	0	0.0	0.0
2021-08-01 04:00:00	26.8	76	1.0	0.0	0	0.0	0.0
...
2021-08-03 19:00:00	27.1	85	0.9	4.0	0	555.0	1511.0
2021-08-03 20:00:00	27.5	85	0.7	0.0	0	419.0	703.0
2021-08-03 21:00:00	27.7	85	0.9	0.0	0	485.0	505.0
2021-08-03 22:00:00	27.6	86	1.0	0.0	0	403.0	439.0
2021-08-03 23:00:00	27.7	84	1.5	0.0	0	148.0	253.0

72 rows × 7 columns

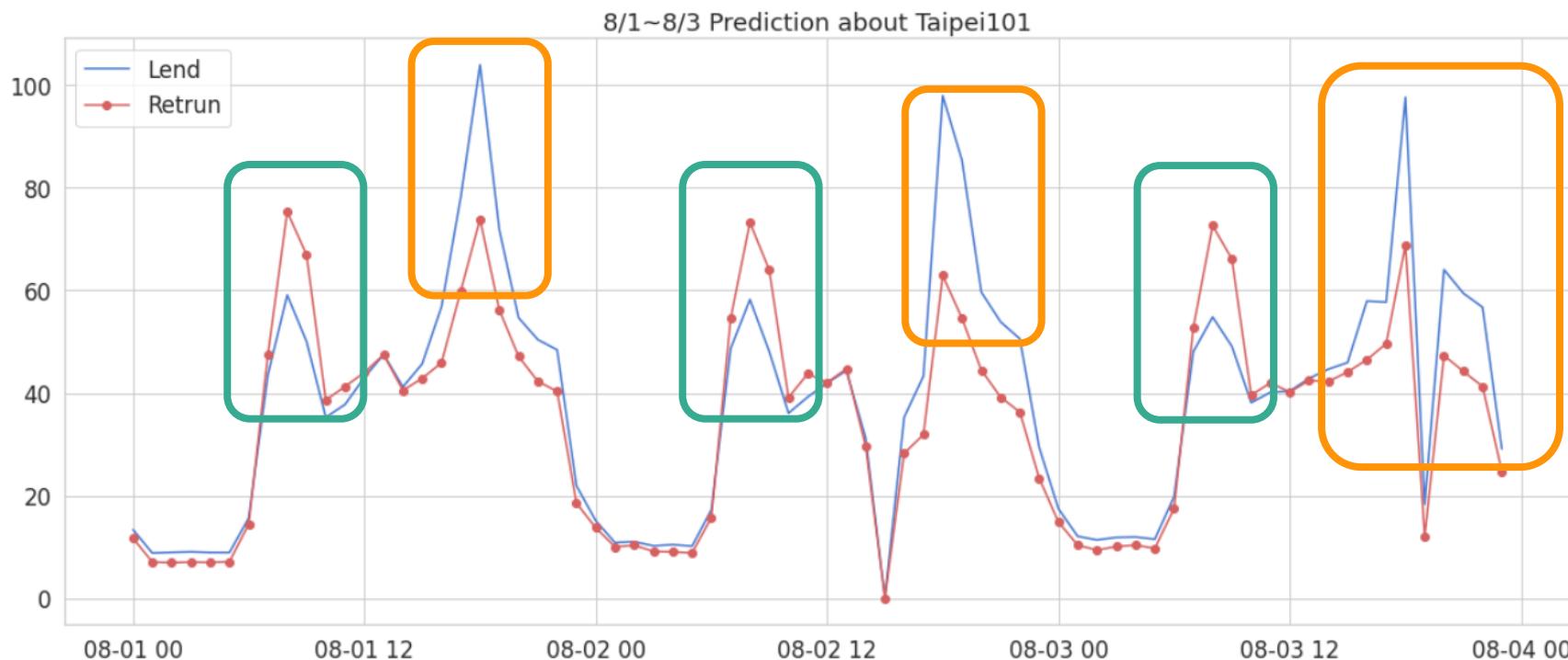
將此資料放入訓練好的LSTM模型中
進行預測。

模型應用

Model Application

以臺北101/世貿站附近YouBike站為例

8/1~8/3的預測結果



藍線：借車數

紅線：還車數

綠色圈起處有發生腳踏車
無位可停的狀況

橘色圈起處有發生無腳踏
車可以借的狀況



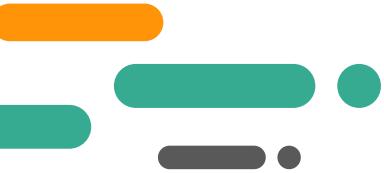
模型應用

Model Application

以臺北101/世貿站附近YouBike站為例

8/1~8/3的預測結果_方案建議

- 無位可停：在 7 ~ 8 點期間可能會發生，可以在 7 點多時增派員工將車輛進行回收，以確保在 8 點時，每位民眾較少發生無位可停的情況
- 無車可借：在 15 ~ 20 點時，建議在這段期間可以增加補車頻率。



結論

Conclusion

- 在預測租借量時，比較兩種模型的準確率，以時間為主的資料類型來說，LSTM的準確率確實比起回歸模型好，但LSTM需要花費比較多的時間進行模型的訓練。
- 在預測結果中，大多數的站點預測都能達到70%以上，在預測結果中，可以看到模型大該都是會低估借車數，這表示其實有更多人有借車需求，因此，YouBike公司可以多設置腳踏車柱進行歸還，或者多增派人員進行補車。

Thank you!