

近兩年美籍聯合航空與其共享代碼 航司在美國境內航班延誤分析

報告者：蔡瑋城

2024年1月

商業大數據分析人才養成班

報告大綱

一、專案發想

- 動機
- 題目訂定

二、資料分析流程

三、結果說明

四、結論與建議

五、問題與解決



一、專案發想

動機：做一個與我們有關的資料分析



題目訂定：幫別人解決問題

航空產業

近2年，
航班穩定復甦逾九成

疫情後2022年與2023年之航班數
與疫情前2019年差距分別為9%與8%



數據統計

缺少跨月份比較
與個案討論

美國聯邦航空總署 (FAA)
(Aviation System
Performance Metrics, ASPM)



航班安排

聯航 台美唯一直飛
共享航班 抵達各機場

UAs



近兩年美籍聯合航空與其共享代碼
航司在美國境內航班延誤分析

以台灣人旅遊的角度，試著回答：

**Q. 從台灣出發搭機至美國境內機場，
是否需擔心航班延誤抵達影響行程？**

二、資料分析流程

二、資料分析流程

資料蒐集:

- 資料集 - Kaggle
- 數值來源 - FAA, ASPM

內容:

- 2019年至2023年8月31日
- 含3百萬筆資料, 32個欄位
- 欄位種類:
 - ✓ 日期
 - ✓ 航司與代碼
 - ✓ 出發抵達機場
 - ✓ 出發與抵達延誤時間
 - ✓ 表定與實際飛行時間
 - ✓ 取消、轉降、延誤歸因

資料分析:

一. 描述性:

- ✓ 資料處理並統計數值:
將日期欄位分為年,月,日
與年-月欄位, 篩選UAs

內容:

- 2022年至2023年8月31日
- 含約29萬筆資料, 36個欄位

二. 迴歸方程式: 後續詳述

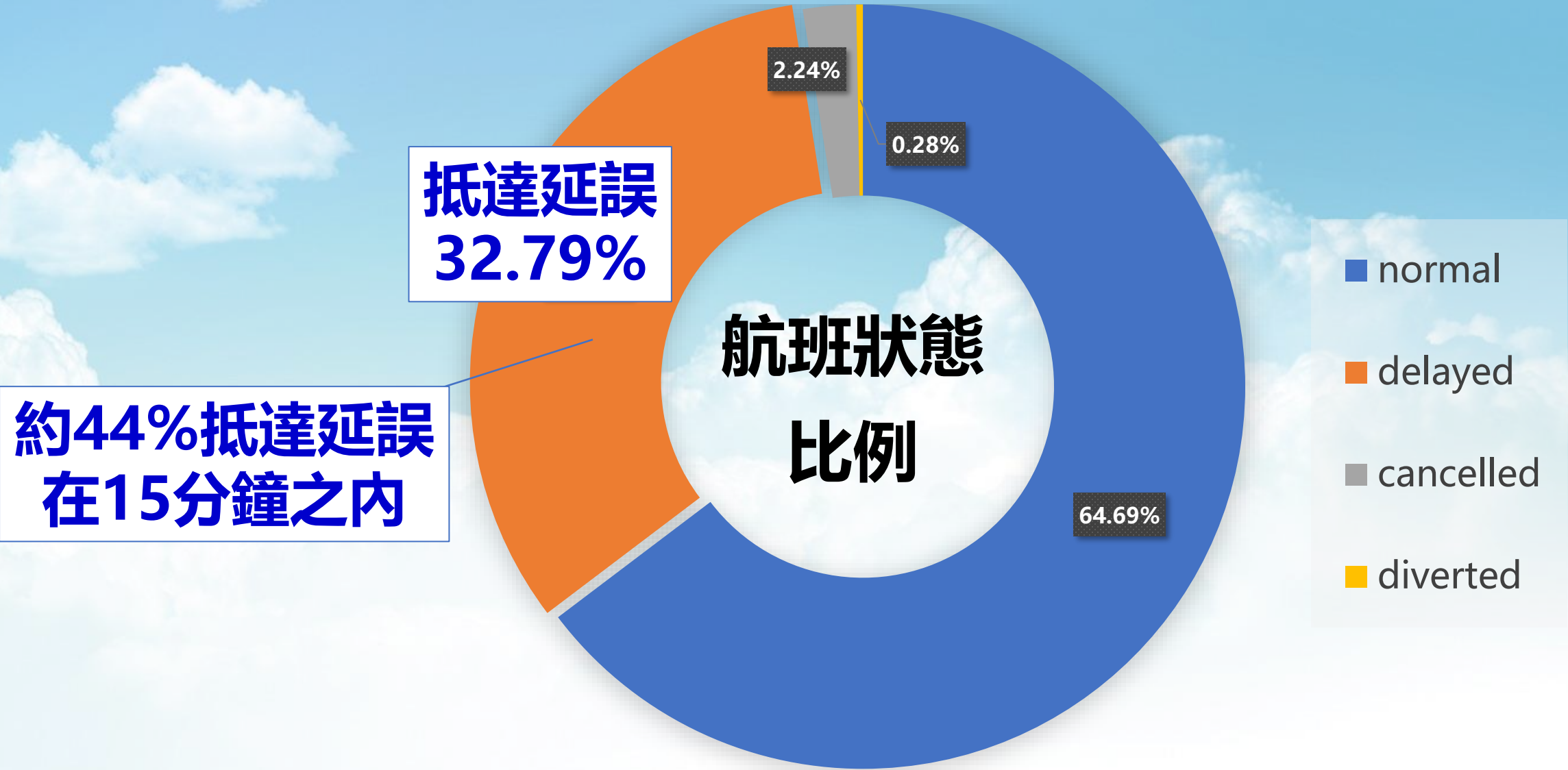
匯入資料庫:



資料視覺化:



三、結果説明

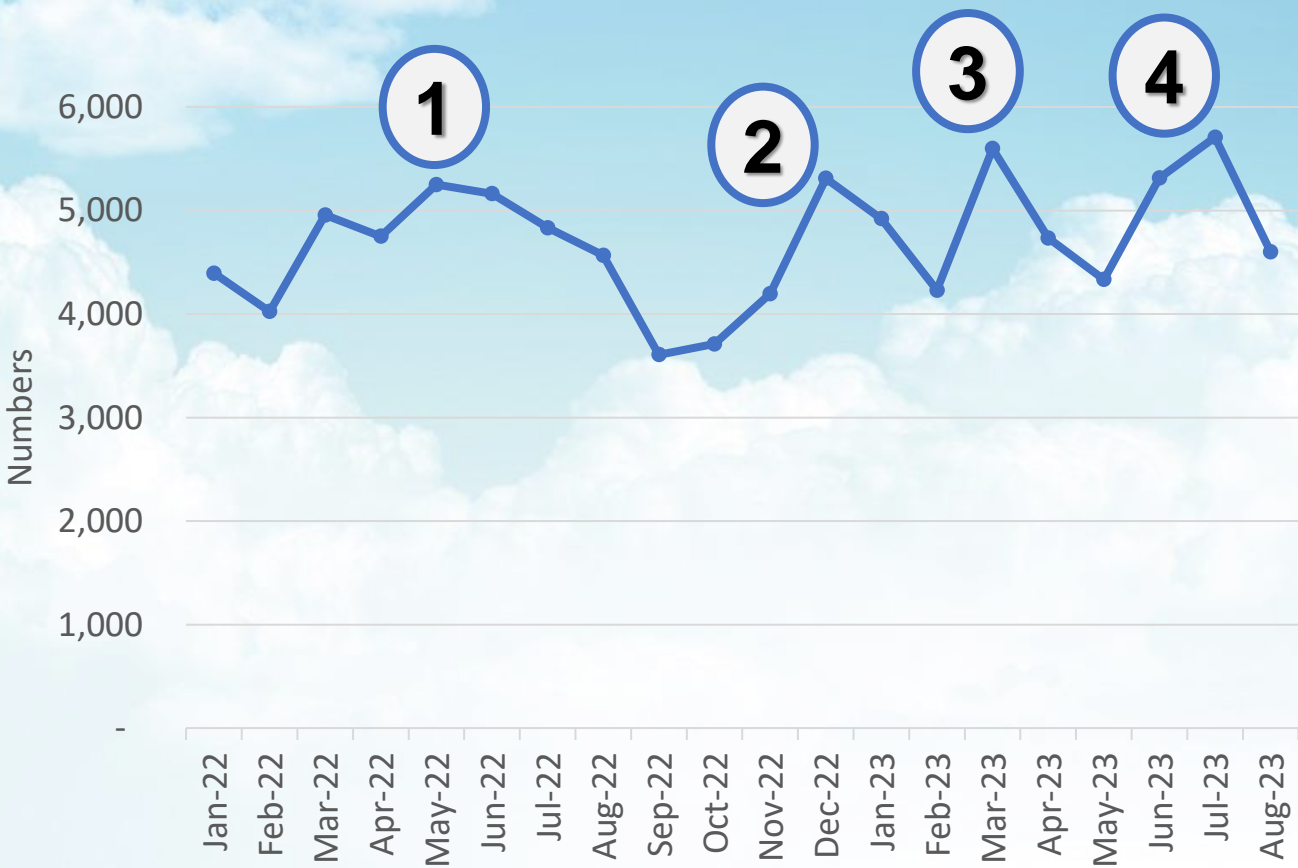


■ 約65%的航班準時或提早抵達，其餘取消或轉降的航班不到3%

備註：統計2022至2023年8月美國境內航班狀態。

回溯新聞，航班延誤多因天氣與國定假日

各月分抵達延誤時序圖



備註：以月份為單位統計2022至2023年8月美國境內航班抵達延誤時序變化。

新聞報導

1 2022年5月 悼念日

United States

U.S airlines cancel 2,500 flights over Memorial holiday weekend

By David Shepardson

May 31, 2022 6:15 AM GMT+8 · Updated 2 years ago

Aa

2 2022年12月 聖誕節+暴風雪

在聖誕節假期前後，一場大規模冬季風暴席捲了美國大部分地區，使航空公司紛紛取消了數千個航班。據報導，目前至少有65人因為這場風暴而喪生。

3 2023年3月 暴風雪

暴風雪肆虐美國東岸 將近5千航班延誤或取消

曾惠敏 / 編譯 發布時間：2023-03-15 18:58 更新時間：2023-03-15 19:19

4 2023年7月 美國國慶

國慶天空大亂…4500航班取消或誤點、旅客受困機場 砲轟布塔朱吉

2023-07-03 14:29 世界日報 / 記者顏伶如 / 綜合報導

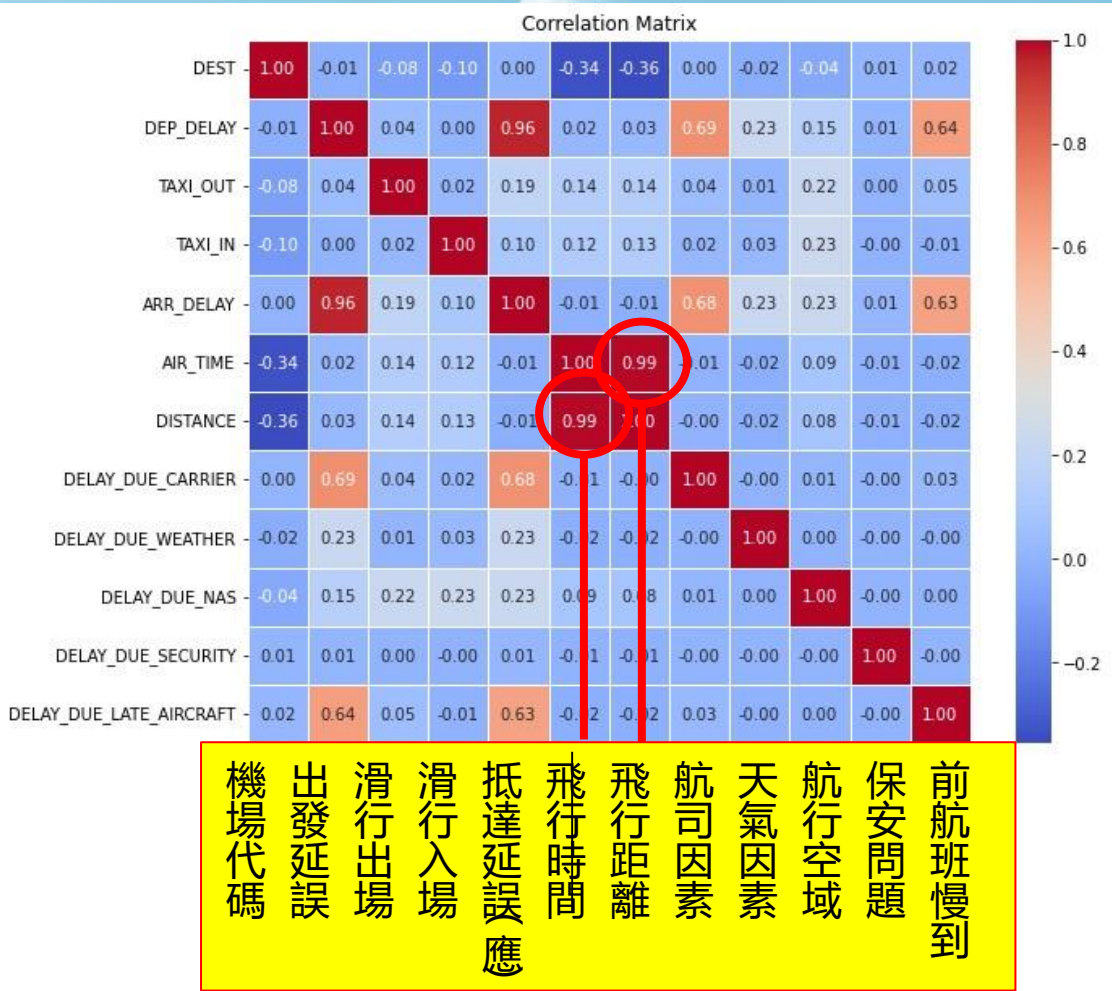
個案討論-UAs從 SFO起飛至美國境內各機場(線性回歸方程式估算抵達延誤時間)

步驟1: 資料前處理

1. **篩選資料:** 使用pandas的loc, filter, 篩選SFO起飛與自變數欄位
2. **編碼化:** OrdinalEncoder編碼化抵達機場代碼
3. **刪除NA與補0:** 用pandas drop(axis=1), fillna, 刪除轉降和取消的航班資料, 把延誤航班歸因空白欄位補0

備註: 經資料前處理, 建模資料為13,479筆。

步驟2: 挑選自變數欄位: 檢視相關性



個案討論-UAs從 SFO起飛至美國境內各機場(線性回歸方程式估算抵達延誤時間)

■ **選擇模型切割6:4：** 考量該方程式MSE是第二低， 且其測試R-Squared是各組中最大， 故此方程式對未見過的資料有較佳解釋能力。

	訓練：測試 8:2	訓練：測試 7:3	訓練：測試 7.5 :2.5	訓練：測試 6:4
R-Squared_訓練	0.9698	0.9703	0.9694	0.9682
R-Squared_測試	0.9500	0.9570	0.9587	0.9656
MSE_測試	63.5148	64.8981	64.1570	64.0452

備註：
不切割資料集：R-squared: 0.9672 MSE: 63.5078
測試資料比例，參考探討航班延誤之文獻，測試資料比例介於0.2~0.4

四、結論與建議

了解航班延誤情形有利於行前規劃

延誤 > 15mins

< 20%

假日
&
惡劣天氣

個案討論：從台灣出發搭聯航至美國SFO，再轉機美國各地機場

- 建立線性回歸方程式，訓練測試為6:4之方程式做為預測抵達延誤時間之參考。

五、問題解決

Q. 如何將2019/1/1的年月日分開成不同欄位

- 日期格式轉換: 使用pd.to_Datetime將日期字串轉換為日期變數

✓ %Y/%m/%d指定日期字串格式。

```
rawdata["FL_DATE"]=pd.to_datetime(rawdata["FL_DATE"],format="%Y/%m/%d")
```

✓ 將日期變數拆分為年、月、日三個欄位，和年-月欄位。

```
rawdata["Year"]=rawdata["FL_DATE"].dt.year  
rawdata["Month"]=rawdata["FL_DATE"].dt.month  
rawdata["Day"]=rawdata["FL_DATE"].dt.day  
rawdata["Year-month"]=rawdata["FL_DATE"].dt.strftime("%Y-%m")
```

Q. 計算延誤, 取消, 轉降之航班個數

- 使用df_delay.value_counts().get(True,0)計算個數
 - ✓ 如果遇到True就算一個否則算0個。

```
filter_delay=data22_23_UAs["ARR_DELAY"]>0  
filter_can=data22_23_UAs["CANCELLED"]==1  
filter_div=data22_23_UAs["DIVERTED"]==1  
delay_counts=filter_delay.value_counts().get(True,0)  
can_counts=filter_can.value_counts().get(True,0)  
div_counts=filter_div.value_counts().get(True,0)
```

THANKS FOR LISTENING!

Feel free to ask questions~

