天氣與捷運運量分析預測

巨量資料分析導論期末專題提案

107590026 資工三 賴璟霆
107590035 資工三 許哲維
107590036 資工三 藍裕翔

Team members

許哲維 X 賴璟霆 X 藍裕翔

Team Members

許哲維
資料前處理
資料視覺化

賴璟霆
資料整合
模型訓練

藍裕翔
資料前處理
資料收集

Introduction

藉由天氣狀況分析預測捷運運量

Introduction

透過不同的天氣狀況來分析、觀察捷運運量是否會因而有所改變，並藉由此分析來預測當天捷運運量以調節捷運之班距、降低不必要之成本，以達最高效益。
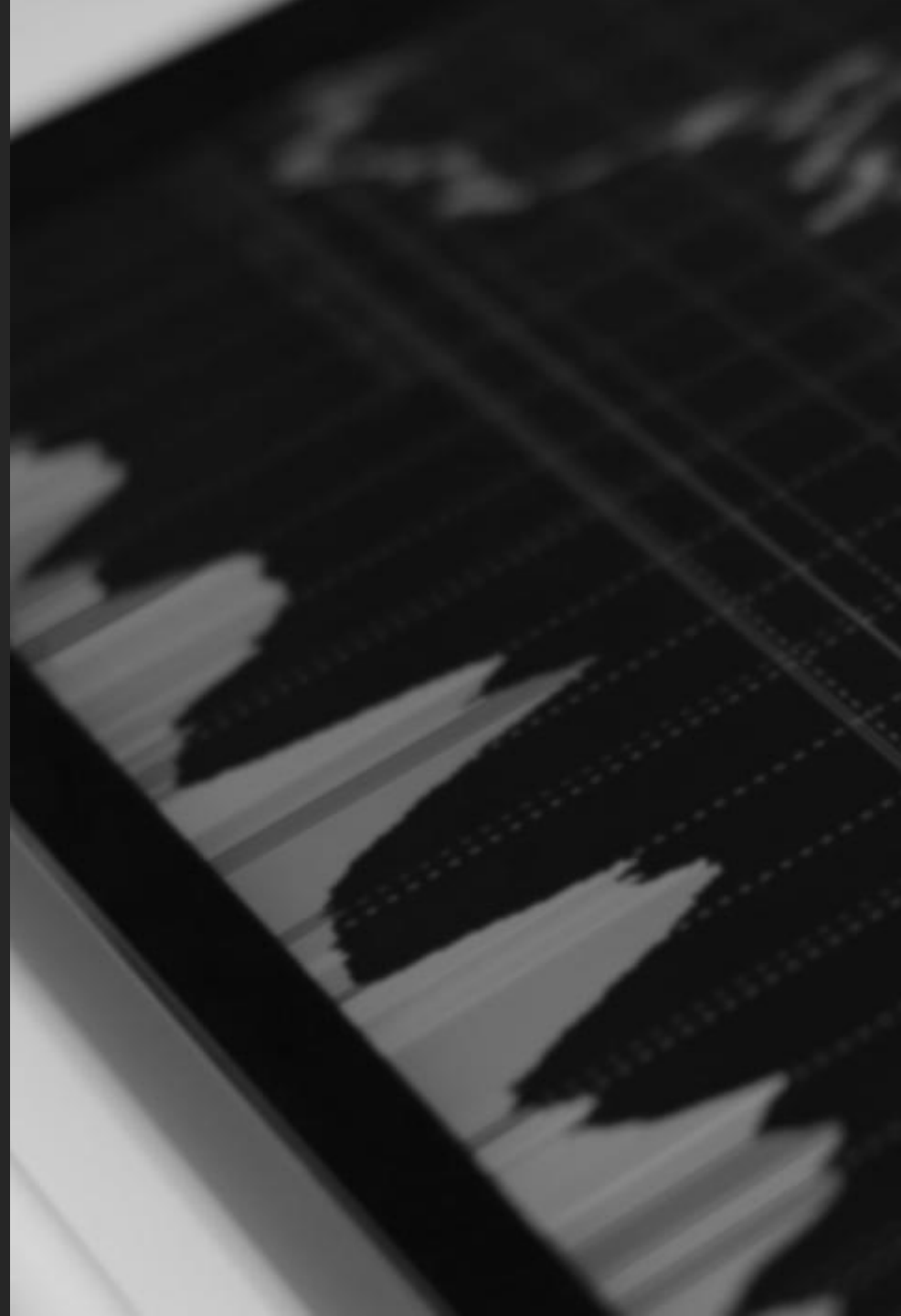
Get Dataset

資料集(Dataset)

Get Dataset

天氣資料集
觀測資料查詢系統(台北測站)

捷運運量資料集
台北捷運旅運量各站進出
量統計(台北測站方圓2公里之捷運站)

About Dataset

資料內容與趨勢

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 7 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   Date         1095 non-null    object
 1   Total_IN     1095 non-null    float64
 2   Total_OUT    1095 non-null    float64
 3   Temperature  1095 non-null    float64
 4   RH           1095 non-null    float64
 5   WS           1095 non-null    float64
 6   Preci        1095 non-null    float64
dtypes: float64(6), object(1)
memory usage: 60.0+ KB
```
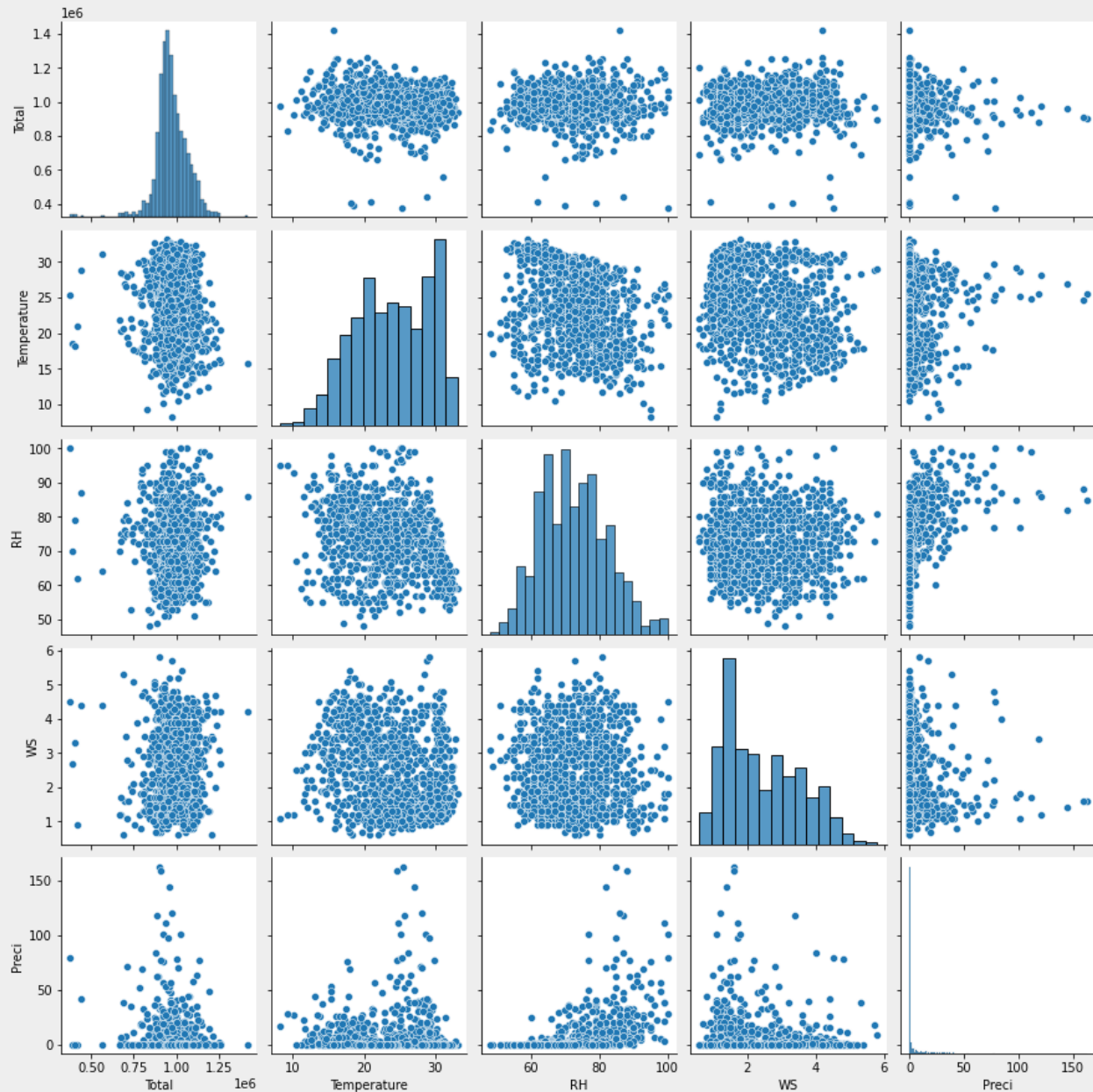
	Date	Total_IN	Total_OUT	Temperature	RH	WS	Preci
0	2017/1/1	482619.0	490231.0	22.6	66.0	3.4	0.0
1	2017/1/2	413764.0	402311.0	21.4	73.0	1.9	0.0
2	2017/1/3	458880.0	451012.0	20.7	83.0	3.5	0.6
3	2017/1/4	449018.0	449439.0	23.2	73.0	3.5	0.0
4	2017/1/5	453509.0	456262.0	22.4	75.0	0.7	0.0

資料內容

1. 資料量1095筆資料

2. 資料項目

3. 收集時間(2017-2019)

資料趨勢

1. 由圖表探索資料

2. 運量與天氣之趨勢

Methods

決策樹(Decision Tree)

	precision	recall	f1-score	support
3.0	0.00	0.00	0.00	2
4.0	0.00	0.00	0.00	3
5.0	0.00	0.00	0.00	1
6.0	0.00	0.00	0.00	5
7.0	0.00	0.00	0.00	14
8.0	0.14	0.05	0.07	106
9.0	0.51	0.78	0.62	409
10.0	0.29	0.11	0.16	214
11.0	0.06	0.05	0.05	61
12.0	0.00	0.00	0.00	7
accuracy			0.43	822
macro avg	0.10	0.10	0.09	822
weighted avg	0.35	0.43	0.36	822

	precision	recall	f1-score	support
3.0	0.00	0.00	0.00	2
4.0	0.00	0.00	0.00	3
5.0	0.00	0.00	0.00	1
6.0	0.00	0.00	0.00	5
7.0	0.00	0.00	0.00	14
8.0	0.17	0.05	0.07	106
9.0	0.50	0.73	0.59	409
10.0	0.25	0.13	0.17	214
11.0	0.07	0.08	0.08	61
12.0	0.00	0.00	0.00	7
14.0	0.00	0.00	0.00	0
accuracy			0.41	822
macro avg	0.09	0.09	0.08	822
weighted avg	0.34	0.41	0.35	822

模型判斷(全部)

1. 決策樹

2. 隨機森林

	precision	recall	f1-score	support
3.0	0.00	0.00	0.00	1
4.0	0.00	0.00	0.00	1
6.0	0.11	1.00	0.20	2
7.0	0.11	0.50	0.18	2
8.0	0.22	0.09	0.13	22
9.0	0.36	0.47	0.41	74
10.0	0.35	0.24	0.29	70
11.0	0.00	0.00	0.00	24
12.0	0.00	0.00	0.00	6
14.0	0.00	0.00	0.00	1
accuracy			0.28	203
macro avg	0.12	0.23	0.12	203
weighted avg	0.28	0.28	0.27	203

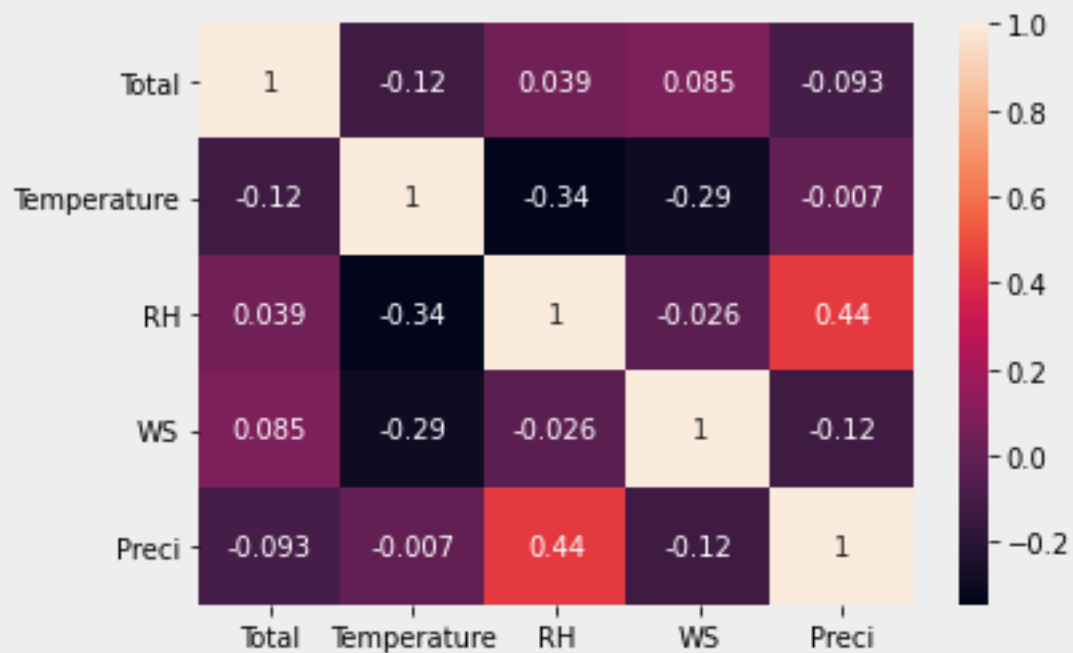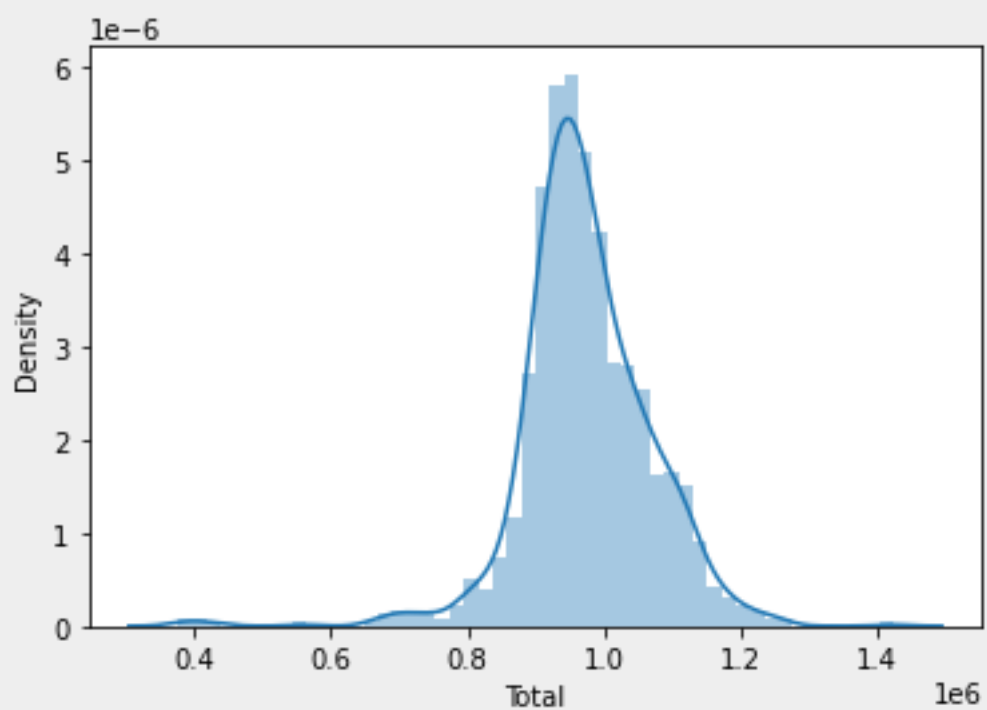|	precision	recall	f1-score	support
3.0	0.00	0.00	0.00	1
4.0	0.00	0.00	0.00	1
6.0	0.00	0.00	0.00	2
7.0	0.00	0.00	0.00	2
8.0	0.00	0.00	0.00	22
9.0	0.39	0.53	0.45	74
10.0	0.36	0.43	0.39	70
11.0	0.09	0.04	0.06	24
12.0	0.00	0.00	0.00	6
14.0	0.00	0.00	0.00	1
accuracy			0.34	203
macro avg	0.08	0.10	0.09	203
weighted avg	0.28	0.34	0.31	203

模型判斷(分季)

1. 決策樹

2. 隨機森林

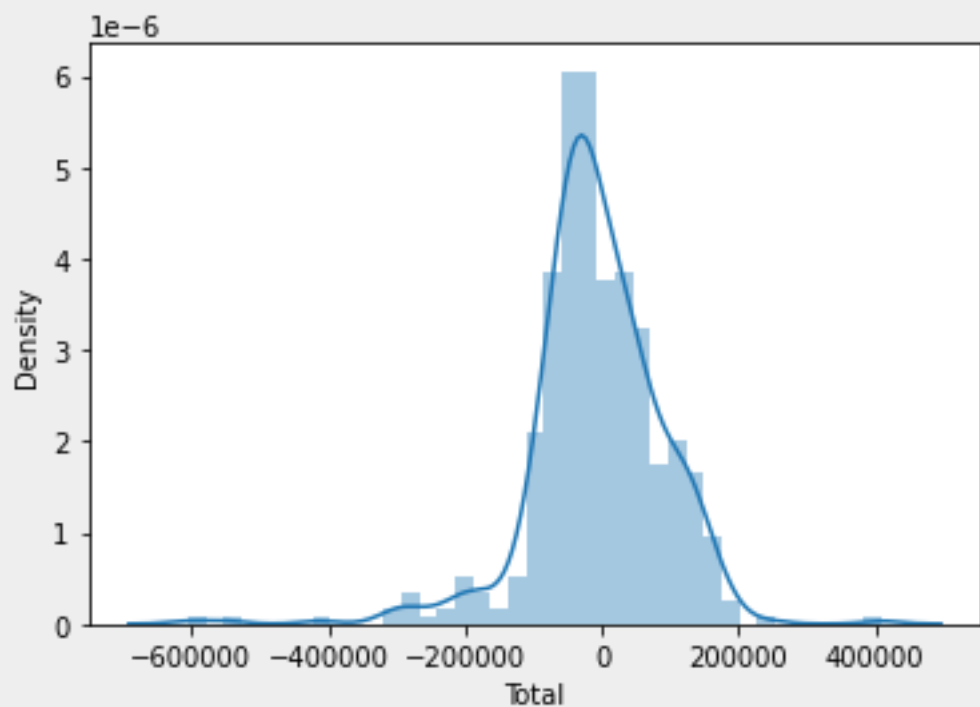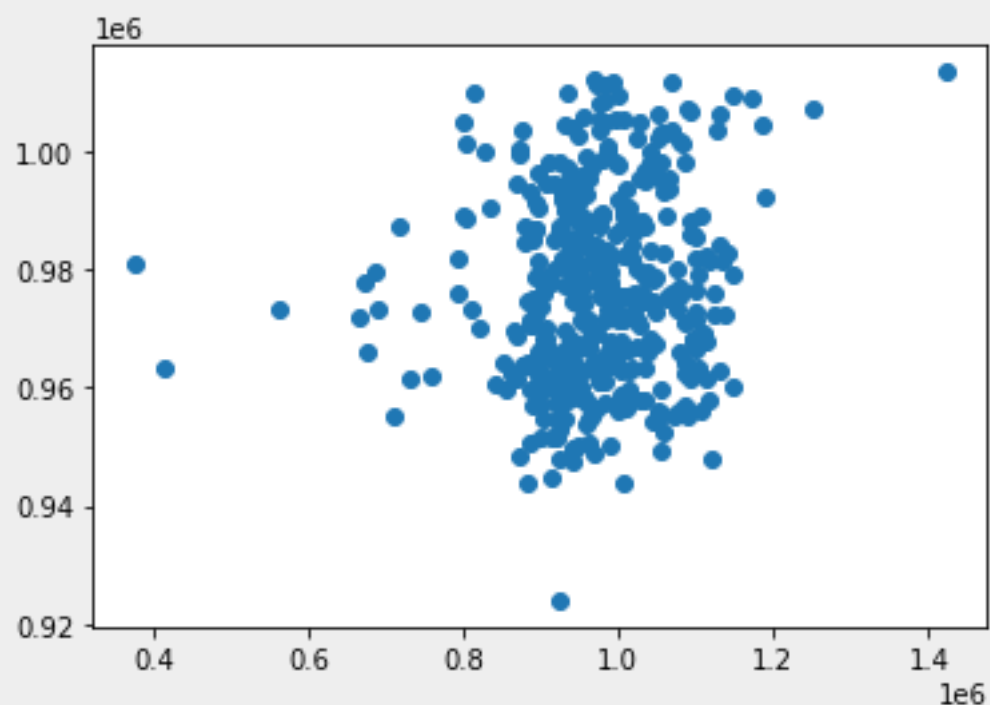# Methods

線性迴歸(Linear Regression)

資料趨勢

1. 運量之集中區間

2. 關係係數

資料預測

1. 實際與預測運量關係

2. 實際與預測值差值分布圖

Summary

決策樹的Accuracy為0.43

隨機森林的Accuracy為0.41

線性迴歸的均方根誤差(RMSE)為
97411.1144482189

Summary

透過上述分析，可知經過分季後accuracy有稍微的提升，而在預測中可知其資料較為集中，且可發覺其均方根誤差偏大、精確度較低。

THANK YOU

報告結束，謝謝大家