

DA幹部專案

李叡

Table of Contents

01
...

資料前處理

Python 程式碼

02
...

Q1

選擇兩個城市設點 & Why ?

03
...

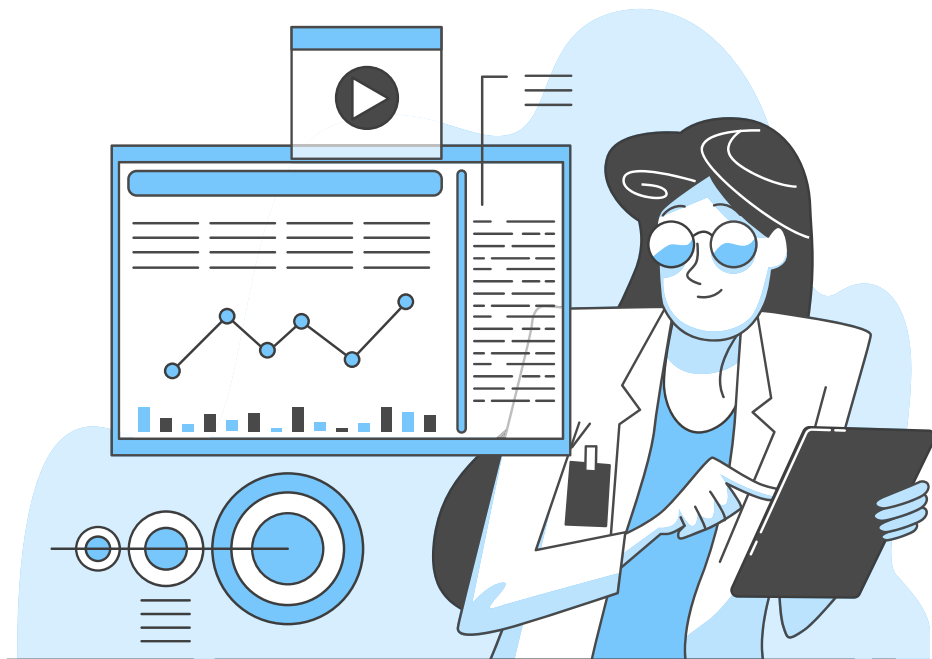
Q2

是否需要電子載具服務?

04
...

Q3

首選的離島 & Why ?





01

資料前處理



Step 1: 導入套件及讀取資料

```
1  import pandas as pd
2
3  #讀入csv
4  invoice_data_df = pd.read_csv("./DA幹部專案/taiwan_invoice_data.csv")
5  print(invoice_data_df.head())
6  print(invoice_data_df.info())
```

利用 head() 及 info() 先稍微瀏覽一下資料的概況

Step 2: 資料清洗及格式整理

```
7
8  #刪除沒用column
9  invoice_data_df = invoice_data_df.drop("Unnamed: 9" ,axis=1).drop('Unnamed: 10',axis=1)
10
11  #刪除Na值
12  invoice_data_df =invoice_data_df.dropna()
13
14  #將年的資料型態改成int
15  invoice_data_df["年"] = invoice_data_df["年"].astype("int")
16
17  #刪除重複值
18  invoice_data_df = invoice_data_df.drop_duplicates()
19
20  #確認年份正確
21  invoice_data_df = invoice_data_df.drop(invoice_data_df[invoice_data_df["年"]>2021].index)
22  invoice_data_df = invoice_data_df.drop(invoice_data_df[invoice_data_df["年"]<2018].index)
23
24  #確認月份正確
25  invoice_data_df = invoice_data_df.drop(invoice_data_df[invoice_data_df["月"]<1].index)
26  invoice_data_df = invoice_data_df.drop(invoice_data_df[invoice_data_df["月"]>12].index)
27
```

Step 2: 資料清洗及格式整理

```
28  """
29  #離群值檢驗
30  data_amount = invoice_data_df["電子發票張數"].to_numpy()
31  data_total = invoice_data_df["電子發票金額"].to_numpy()
32  q1_amount,q3_amount = np.percentile(data_amount, (25, 75))
33  IQR_amount = q3_amount - q1_amount
34  q1_total,q3_total = np.percentile(data_total, (25, 75), interpolation='midpoint')
35  IQR_total = q3_total - q1_total
36  amount_MAX = q3_amount + 1.5 * IQR_amount
37  amount_min = q1_amount - 1.5 * IQR_amount
38  total_MAX = q3_total + 1.5 * IQR_total
39  total_min = q1_total - 1.5 * IQR_total
40
41  print(invoice_data_df["電子發票張數"].sort_values(ascending=False,ignore_index=True).head(20))
42  print(invoice_data_df["電子發票金額"].sort_values(ascending=False,ignore_index=True).head(20))
43  """
```

本來有利用1.5倍四分位距來挑出離群值但發現山選出來的資料過多，再進一步列出離群值後發現並無特別明顯的極端出現，推測是由不同行業的性質導致在電子發票的金額及張數上有相當大的落差

Step 3: 資料合併及輸出 CSV 檔

```
44 #讀入行業別對照
45 industry_category_data = pd.read_csv("./DA幹部專案/industry_category.csv")
46
47 #合併
48 final_data = pd.merge(invoice_data_df, industry_category_data,on="行業別")
49
50 #輸出
51 final_data.to_csv("final_data.csv")
```

處理後成果

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 65510 entries, 0 to 65546
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   發票年月    65510 non-null object
1   縣市        65510 non-null object
2   縣市代號    65510 non-null object
3   行業別      65510 non-null object
4   載具類別    65510 non-null object
5   電子發票張數 65510 non-null int64
6   電子發票金額 65510 non-null int64
7   年          65510 non-null int64
8   月          65510 non-null int64
dtypes: int64(4), object(5)
memory usage: 7.0+ MB
None
```

處理後



處理前



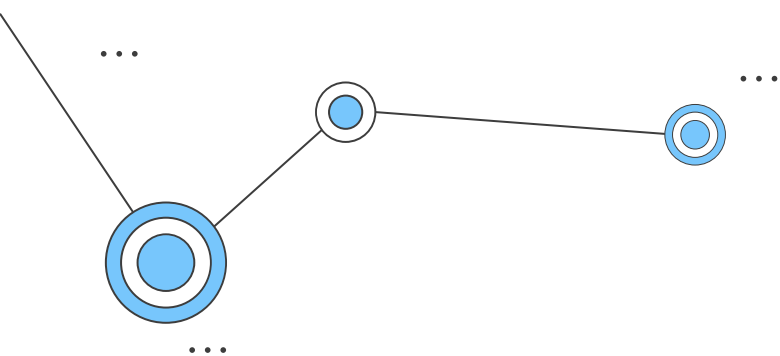
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 65549 entries, 0 to 65548
Data columns (total 11 columns):
#   Column      Non-Null Count  Dtype
---  -
0   發票年月    65548 non-null object
1   縣市        65548 non-null object
2   縣市代號    65548 non-null object
3   行業別      65536 non-null object
4   載具類別    65546 non-null object
5   電子發票張數 65549 non-null int64
6   電子發票金額 65549 non-null int64
7   年          65546 non-null float64
8   月          65549 non-null int64
9   Unnamed: 9  0 non-null      float64
10  Unnamed: 10 0 non-null      float64
dtypes: float64(3), int64(3), object(5)
memory usage: 5.5+ MB
None
```


處理後成果

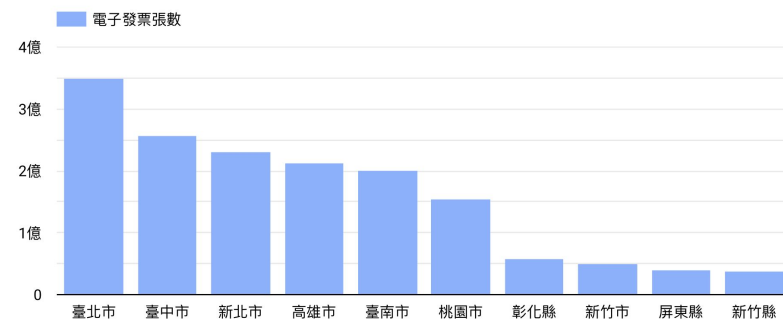
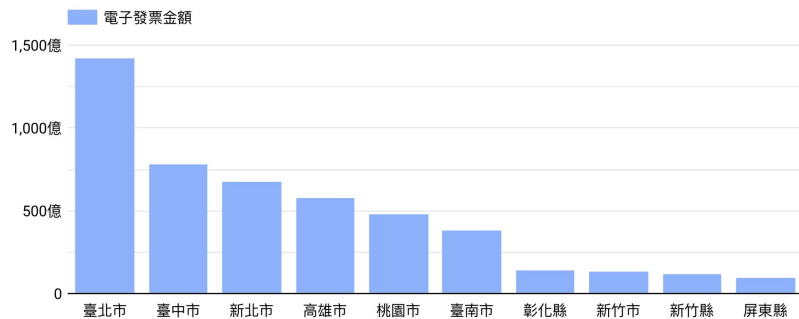
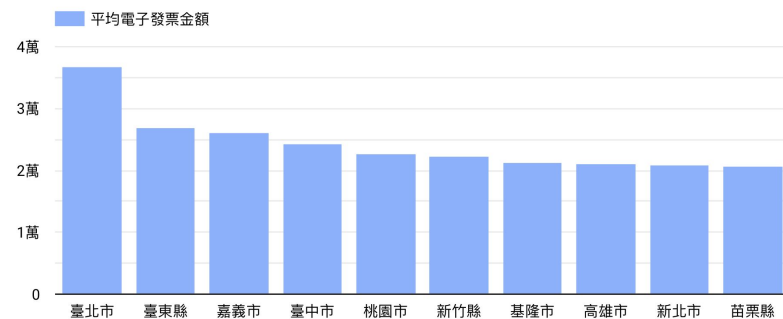
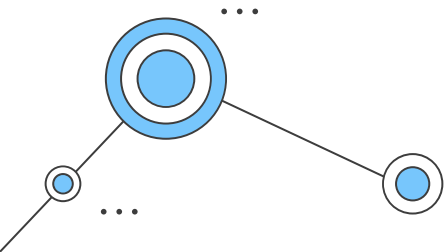
▼	發票年月 ▼	縣市 ▼	縣市代號 ▼	行業別 ▼	載具類別 ▼	電子發票張數	電子發票金額	年 ▼	月 ▼	產業分類 ▼
0	2018/1/1	臺北市	A	不動產開發業	無載具	911490	1824251744	2018	1	不動產業
1	2018/1/1	高雄市	E	不動產開發業	無載具	165881	482368856	2018	1	不動產業
2	2018/1/1	新北市	F	不動產開發業	無載具	297524	436508466	2018	1	不動產業
3	2018/1/1	臺中市	B	不動產開發業	無載具	126352	172296821	2018	1	不動產業
4	2018/1/1	臺南市	D	不動產開發業	無載具	1839	54786501	2018	1	不動產業
5	2018/1/1	桃園市	H	不動產開發業	無載具	20080	23720784	2018	1	不動產業
6	2018/1/1	新竹市	O	不動產開發業	無載具	4027	19383861	2018	1	不動產業
7	2018/1/1	臺北市	A	不動產開發業	使用載具	15368	17640810	2018	1	不動產業
8	2018/1/1	新北市	F	不動產開發業	使用載具	3588	2998886	2018	1	不動產業
9	2018/1/1	南投縣	M	不動產開發業	無載具	1251	1951635	2018	1	不動產業
10	2018/1/1	高雄市	E	不動產開發業	使用載具	2659	908990	2018	1	不動產業
11	2018/1/1	臺中市	B	不動產開發業	使用載具	2333	801981	2018	1	不動產業
12	2018/1/1	新竹縣	J	不動產開發業	無載具	1739	637271	2018	1	不動產業
13	2018/1/1	桃園市	H	不動產開發業	使用載具	1212	89866	2018	1	不動產業
14	2018/1/1	臺南市	D	不動產開發業	使用載具	1068	-5724	2018	1	不動產業
15	2018/1/1	嘉義縣	Q	不動產開發業	無載具	1071	-8743	2018	1	不動產業
16	2018/2/1	臺北市	A	不動產開發業	無載具	914881	1979922316	2018	2	不動產業
17	2018/2/1	新北市	F	不動產開發業	無載具	369284	593331464	2018	2	不動產業

02

Q1: 選擇兩個城市

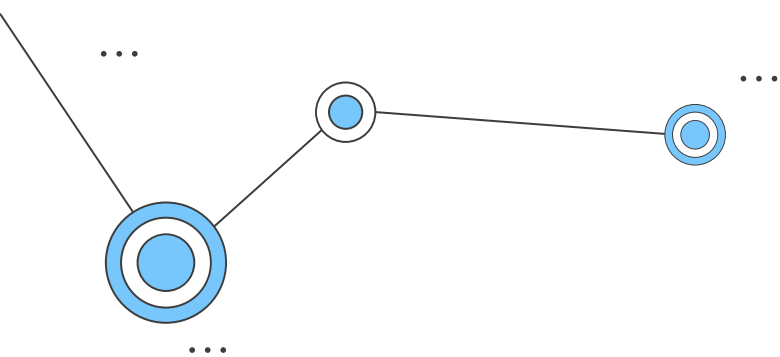


第一張圖顯示出各縣市之間的平均單位電子發票的金額，用以表示客單價，而下面兩張圖的金額及發票張數用來評斷市場規模，雖說台東的客單價位居第二，但加上市場規模做綜合的考慮後，除台北外，台中作為第二選擇更為合適

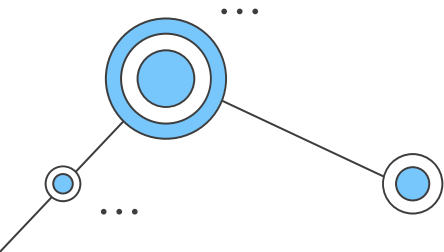


03

Q2: 電子載具服務？

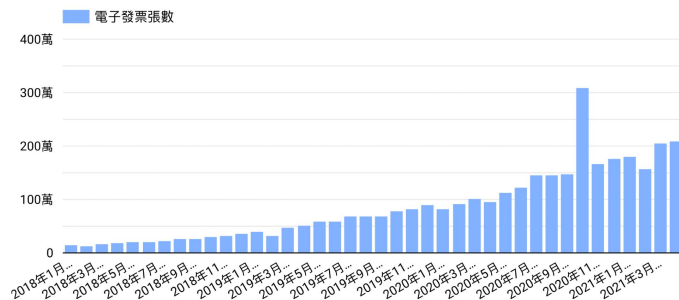


從時間變化來看兩座城市的載具使用人數都在不斷成長(其中 2020 年 10 月張數特別突出, 推測與政府振興券發放有關)



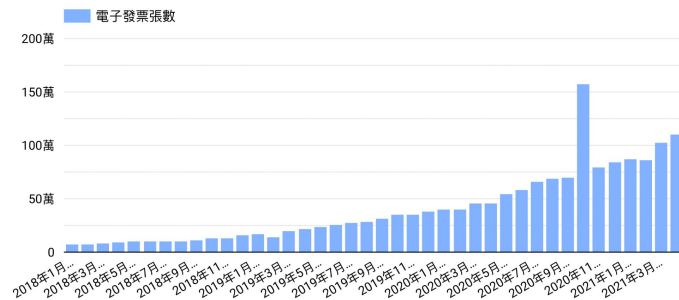
行業別: 餐飲業	(1) ▾
縣市: 臺北市	(1) ▾
載具類別: 使用載具	(1) ▾

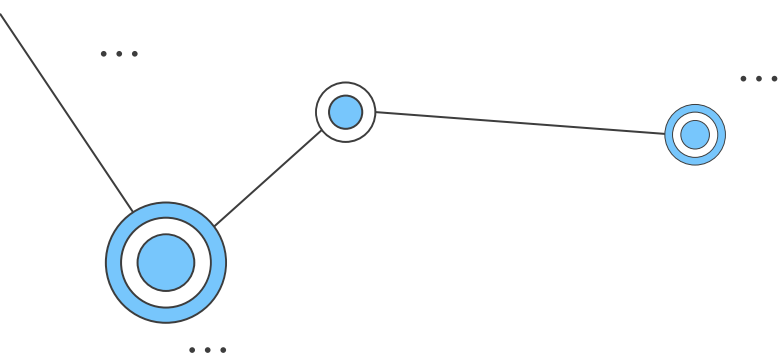
🔍 | ☰ | ⋮



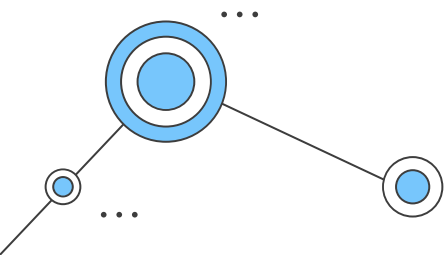
行業別: 餐飲業	(1) ▾
縣市: 臺中市	(1) ▾
載具類別: 使用載具	(1) ▾

🔍 | ☰ | ⋮





近一步去看兩座城市使用載具各年的比例可以發現台北的比例上升的幅度還滿大的, 所以台北的分店建議選擇與業者合作提供載具服務



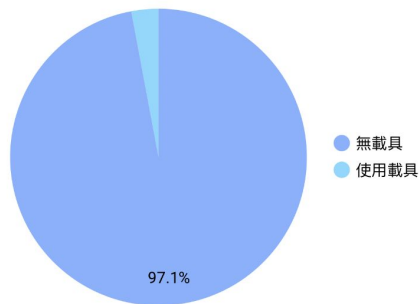
行業別: 餐飲業

(1) ▾

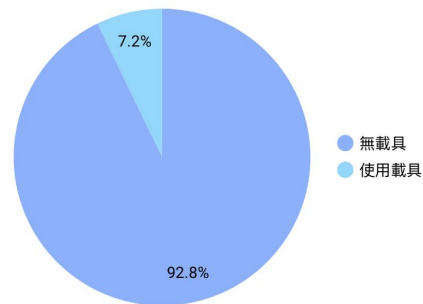
縣市: 臺北市

(1) ▾

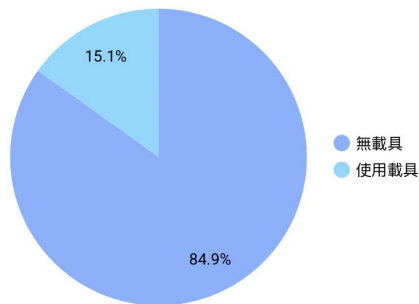
2018



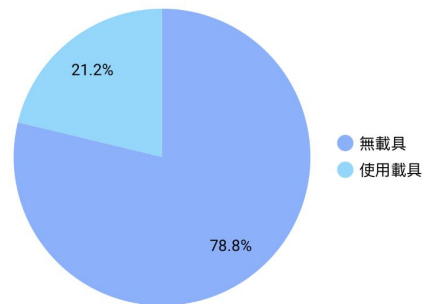
2019

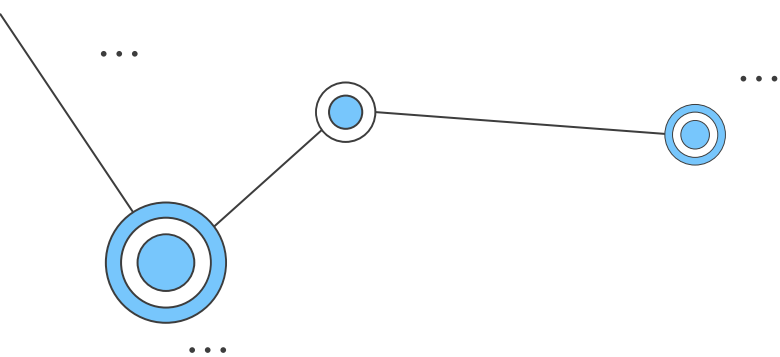


2020

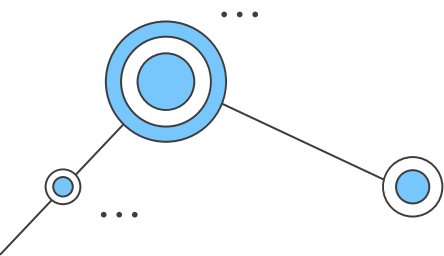


2021





相較台北到2021年接近四分之一的人使用載具，在台中這邊的成長率較為平緩，但考慮未來趨勢及環保等因素，還是可以提供載具服務



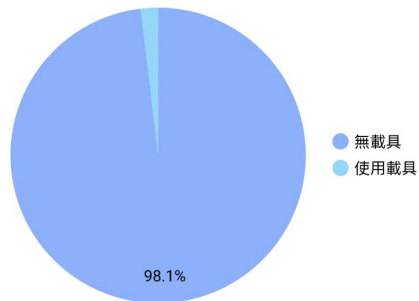
行業別: 餐飲業

(1) ▾

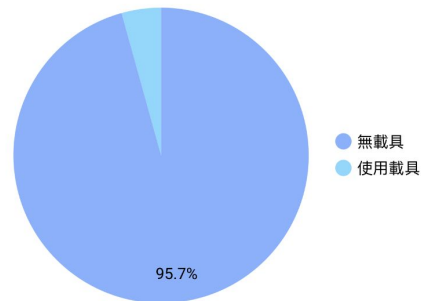
縣市: 臺中市

(1) ▾

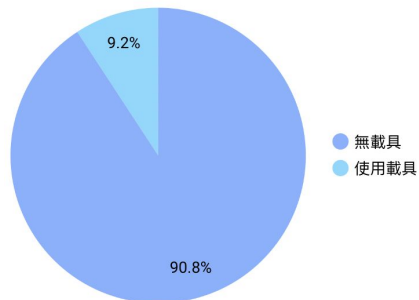
2018



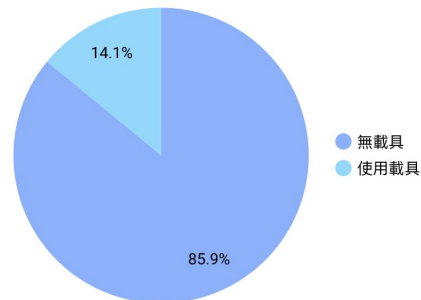
2019



2020

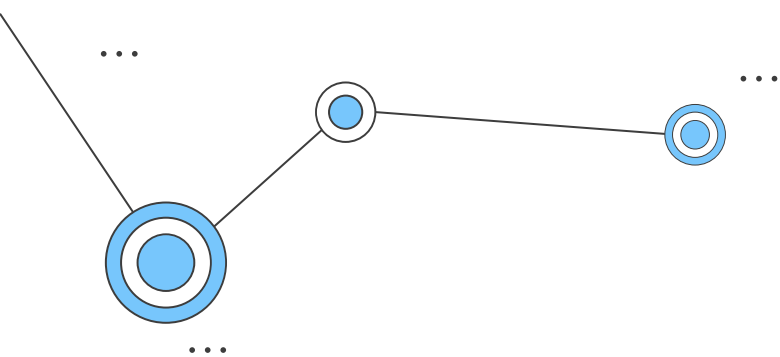


2021

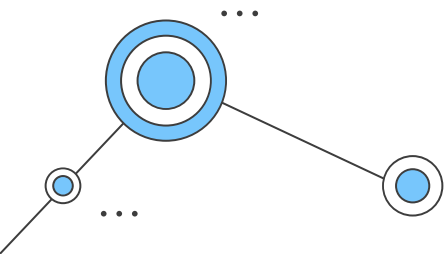


04

Q3: 首選的離島



單就三個離島比較，客單價方面其實並無顯著差異，但澎湖在市場規模方面有優勢

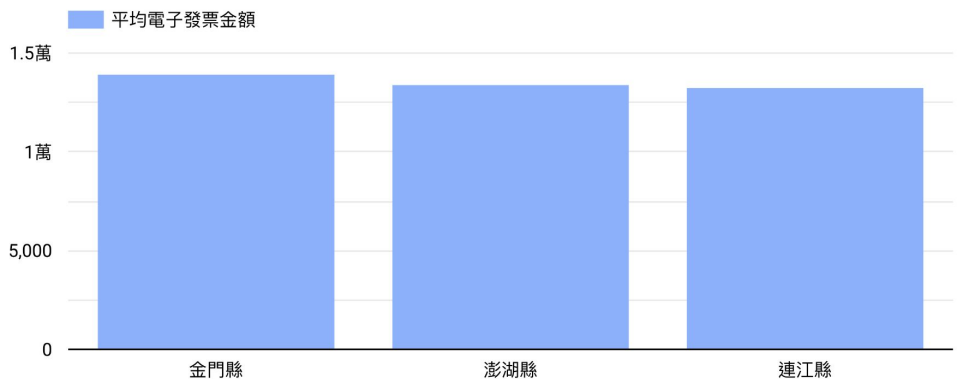
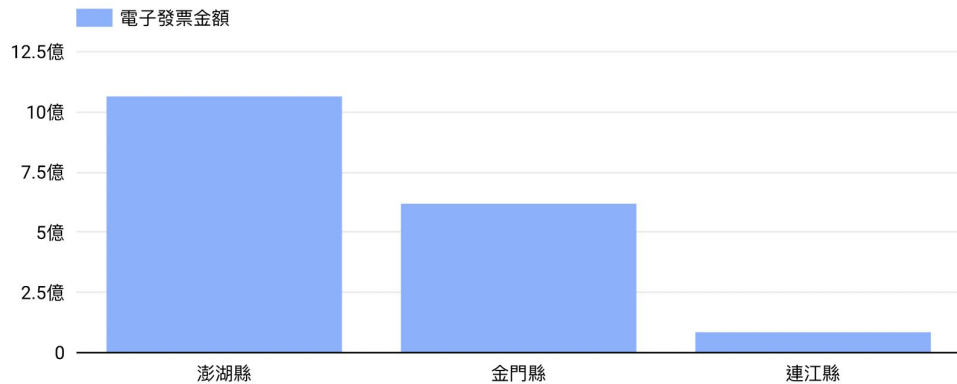


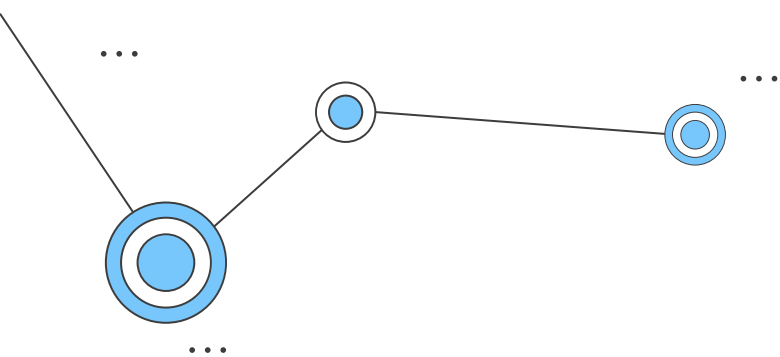
行業別: 餐飲業

(1) ▾

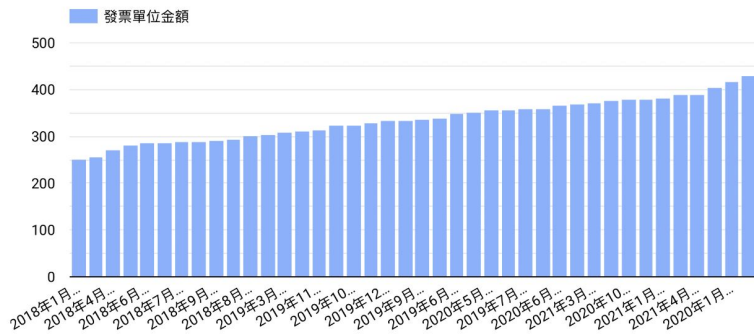
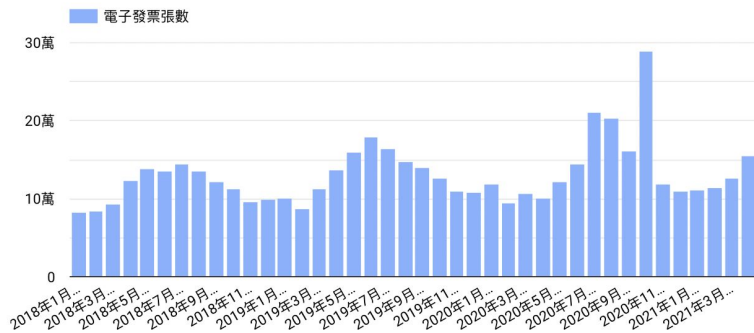
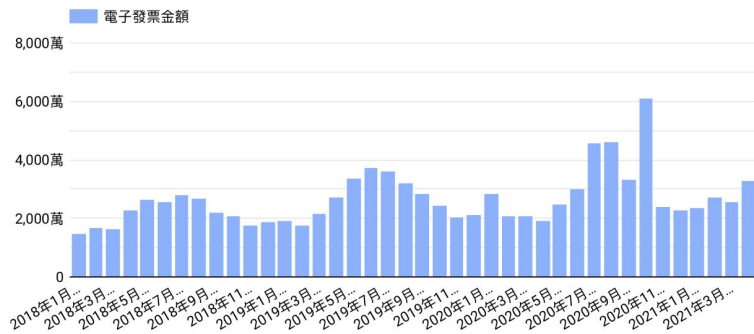
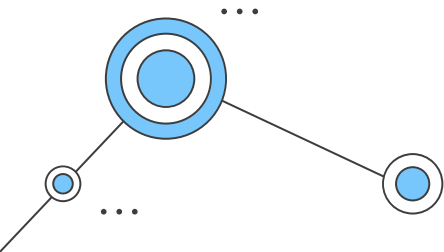
縣市: 澎湖縣, 金門縣, 連江縣

(3) ▾





從左圖可以看到離島市場的規模近年不
 段提升且是高度週期性的，通常暑期為
 旺季，綜合上述建議能透過快閃店的方
 式僅於每年6-9月於澎湖營業





Thanks!

Data Studio : <https://datastudio.google.com/s/jdFmPfMInAM>

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), infographics & images by [Freepik](#) and illustrations by [Stories](#)

