

題目：2018-2023國內旅遊概況

壹、動機

2020年的COVID-19疫情對全球旅遊產業造成重大衝擊，國外旅遊受限，連帶影響國內旅遊。然而，隨著疫情趨緩，人們逐漸恢復旅遊習慣。自去年底起，國外旅遊也開始復甦，引發我們對旅遊行為變化的好奇。所以我們想了解近五年來國內旅遊的概況。

貳、目的

透過探索過去五年的台灣的旅遊景點資料，了解台灣國內旅遊的熱門景點或是地區並加以分析，歸納出近年的旅遊趨勢。

參、假說

近年來不論是新聞、旅遊頻道，或是好友們出遊的社群媒體貼文，可以看到越來越多人在出遊時選擇前往南部地區，而許多行程是包含了探訪名勝古蹟或者是在都市生活難以欣賞到的自然景觀。因此我們假設南部地區的旅遊人數有上升趨勢，而熱門景點的類型應該要有高比例是文化古蹟或是自然景觀。

肆、資料取得

一、交通部觀光數統計資料庫

（一）縣市及遊憩景點旅客人數月資料 [all_count.csv](#)

共包含65個欄位：

1. 縣市：中文的縣市名稱
2. Location：英文的縣市名稱
3. 遊憩據點：中文的遊憩景點名稱
4. Scenic Spots：英文的遊憩景點名稱
5. 107年1月- 111年12月：共60欄各月份遊客人數
6. 小計：各遊憩景點107年1月- 111年12月的旅客數總計

（二）類型及遊憩景點旅客人數月資料 [classification.csv](#)

共包含65個欄位：

1. 類型：中文的類型
2. Class：英文的類型
3. 遊憩據點：中文的遊憩景點名稱
4. Scenic Spots：英文的遊憩景點名稱
5. 107年1月- 111年12月：共60欄各月份遊客人數
6. 小計：各遊憩景點107年1月- 111年12月的旅客數總計

（三）中華民國臺灣旅遊狀況調查報告（旅遊天數）[旅遊天數.csv](#)

共包含6個欄位

	旅遊天數	2018	2019	2020	2021	2022
0	一天	66.40	66.40	66.40	71.90	71.40
1	二天	20.20	21.90	20.20	18.20	17.80
2	三天	9.90	8.90	9.90	7.30	8.10
3	四天及以上	3.50	2.70	3.50	2.60	2.70
4	平均旅遊天數	1.49	1.51	1.54	1.45	1.47

(四) 中華民國臺灣旅遊狀況調查報告(住宿方式) [住宿方式.csv](#)
共包含6欄

	住宿方式	2018	2019	2020	2021	2022
0	當日來回、沒有在外過夜	68.0	66.4	66.4	71.9	71.4
1	旅館	15.1	17.1	17.0	13.2	14.1
2	親友家(含自家)	7.8	7.0	7.1	7.1	5.7
3	民宿	7.0	7.8	7.8	6.0	7.1
4	露營	1.2	1.0	1.2	1.4	1.3
5	招待所或活動中心	0.7	0.6	0.4	0.3	0.3
6	其他	0.1	0.1	0.2	0.1	0.1

伍、方法與分析

一、前處理

(一) 縣市及遊憩景點旅客人數月資料 [all_count.csv](#)
載入[all_count.csv](#)，存成df後發現有兩列欄位名，且都有缺失值，因此刪掉第1列，保留第二列，結果如下：

	NaN	Location	NaN	Scenic Spots	2018/1	2018/2
0	臺北市	Taipei City	陽明山遊客中心	Yangmingshan Visitor Center	18832	52386
1	臺北市	Taipei City	陽明書屋	Yangmingshuwu	1616	2745
2	臺北市	Taipei City	陽明公園	Yangming Park	8750	247131
3	臺北市	Taipei City	大屯遊憩區	Datun Recreation Area	30588	52168
4	臺北市	Taipei City	龍鳳谷遊憩區	Longfonggu Recreation Area	27405	45304

第1欄與第3欄缺失的欄位名分別填入'chi_Location'、'chi_Scenic Spots'。

	chi_Location	Location	chi_Scenic Spots	Scenic Spots	2018/1	2018/2
0	臺北市	Taipei City	陽明山遊客中心	Yangmingshan Visitor Center	18832	52386
1	臺北市	Taipei City	陽明書屋	Yangmingshuwu	1616	2745
2	臺北市	Taipei City	陽明公園	Yangming Park	8750	247131
3	臺北市	Taipei City	大屯遊憩區	Datun Recreation Area	30588	52168
4	臺北市	Taipei City	龍鳳谷遊憩區	Longfonggu Recreation Area	27405	45304

接著檢查資料型態，發現數字被儲存成字串，將其更改為int。

再來新增一個欄位df['Area']，根據縣市名稱判斷，儲存該縣市所屬的地區。

for location in df['Location']:		
if location in ['Taipei City', 'New Taipei City', 'Keelung City', 'Hsinchu City', 'Taoyuan City', 'Hsinchu County', 'Yilan County']:	areas.append('北部')	Area
elif location in ['Taichung City', 'Miaoli County', 'Changhua County', 'Nantou County', 'Yunlin County']:	areas.append('中部')	北部
elif location in ['Kaohsiung City', 'Tainan City', 'Chiayi City', 'Chiayi County', 'Pingtung County']:	areas.append('南部')	北部
elif location in ['Hualien County', 'Taitung County']:	areas.append('東部')	北部
elif location in ['Penghu County', 'Kinmen County', 'Lienchiang County']:	areas.append('離島')	北部
else:		北部
areas.append(None) # 如果不符合以上狀況，設為None		

最後檢查空值與重複值的數量皆為0。

(二) 縣市及遊憩景點旅客人數月資料 classification.csv

載入classification.csv。

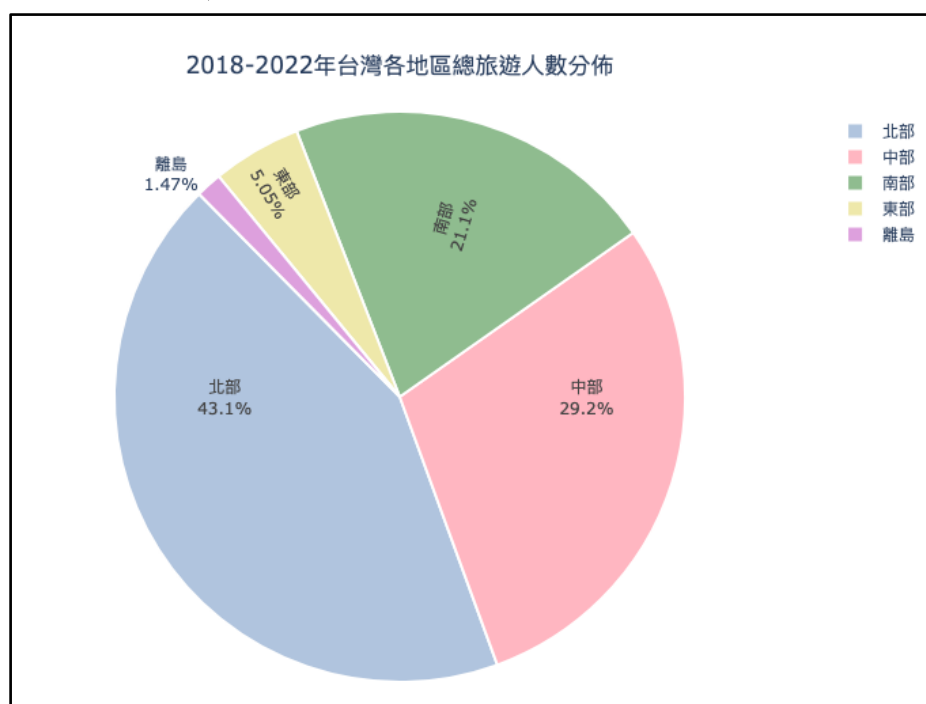
	類型	Unnamed: 1	遊憩據點	Unnamed: 3	107年1月	107年2月
0	NaN	Class	NaN	Scenic Spots	2018/1	2018/2
1	國家公園	National Parks	墾丁國家公園管理處遊客中心	Kenting National Park Headquarters Visitor Center	5806	9185
2	國家公園	National Parks	鵝鑾鼻公園	Eluanbi Park	39422	69925
3	國家公園	National Parks	貓鼻頭公園	Maobitou Park	53133	69205
4	國家公園	National Parks	佳樂水	Jialeshuei	6779	16057

刪除第四欄以後的不會用到的欄位，並在第1欄與第3欄缺失的欄位名分別填入'Class'、'Scenic Spots'。結果如下：

	類型	Class	遊憩據點	Scenic Spots
1	國家公園	National Parks	墾丁國家公園管理處遊客中心	Kenting National Park Headquarters Visitor Center
2	國家公園	National Parks	鵝鑾鼻公園	Eluanbi Park
3	國家公園	National Parks	貓鼻頭公園	Maobitou Park
4	國家公園	National Parks	佳樂水	Jialeshuei
5	國家公園	National Parks	社頂自然公園	Sheding Nature Park

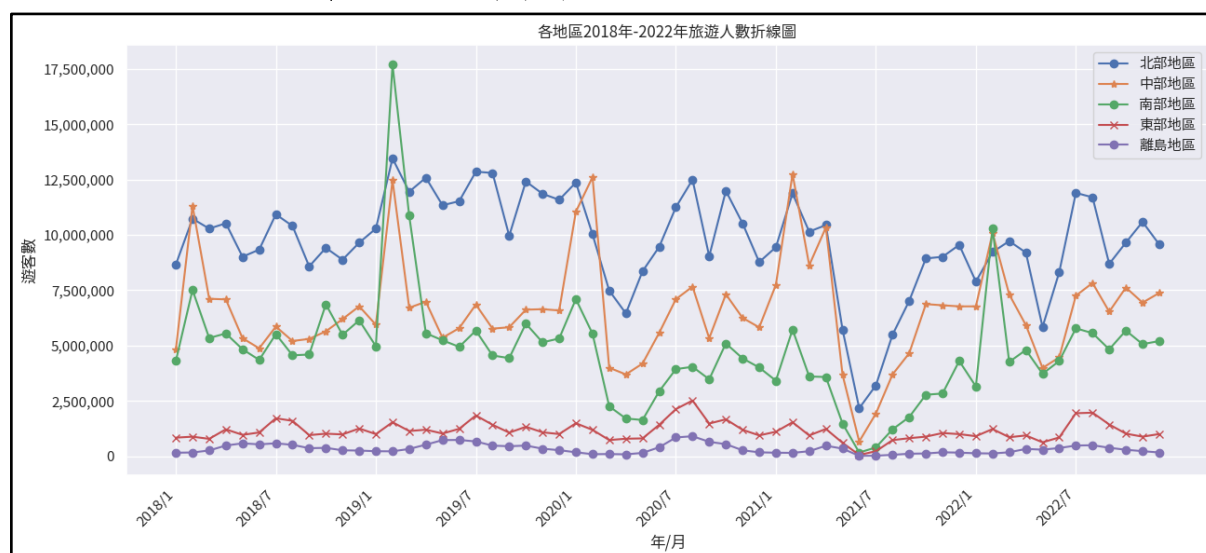
最後檢查空值與重複值的數量皆為0。

二、各地區旅遊人數分布比例



這裡我們使用了2018-2022年間的年資料做分析，其中我們先將地區設為我們的索引，接著進行每月的旅遊人數總和計算，再取得各地區的總和，最後再將此資料進行繪圖，由圖片我們清楚看出人數最多的地區為北部，其次是中部，再者才是我們所假設的南部地區，從這邊我們發現北部非但為累積人數最高的縣市，而且佔了將近一半的數量，所以我們再藉由後面的分析圖表去試試看能不能推導出相關原因。

三、各地區2018-2022年旅遊人數折線圖



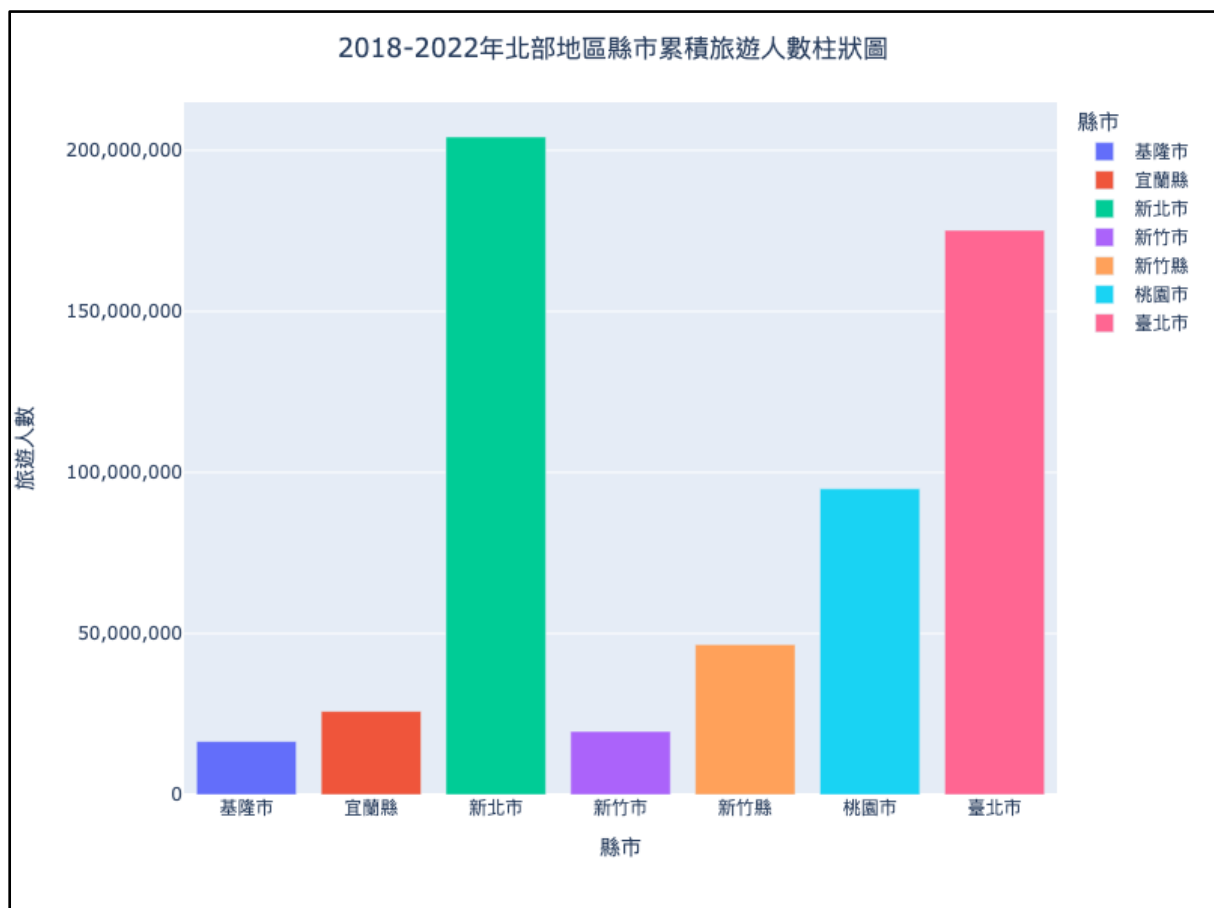
我們使用了2018年至2022年的月資料，以年份與月份為x軸，遊客數為y軸，繪製旅遊人數折線圖。其中又以統計資料庫的分類：北部、中部、南部、東部、離島分組，因此可以在圖表中觀察到不同地區的旅遊人數趨勢。



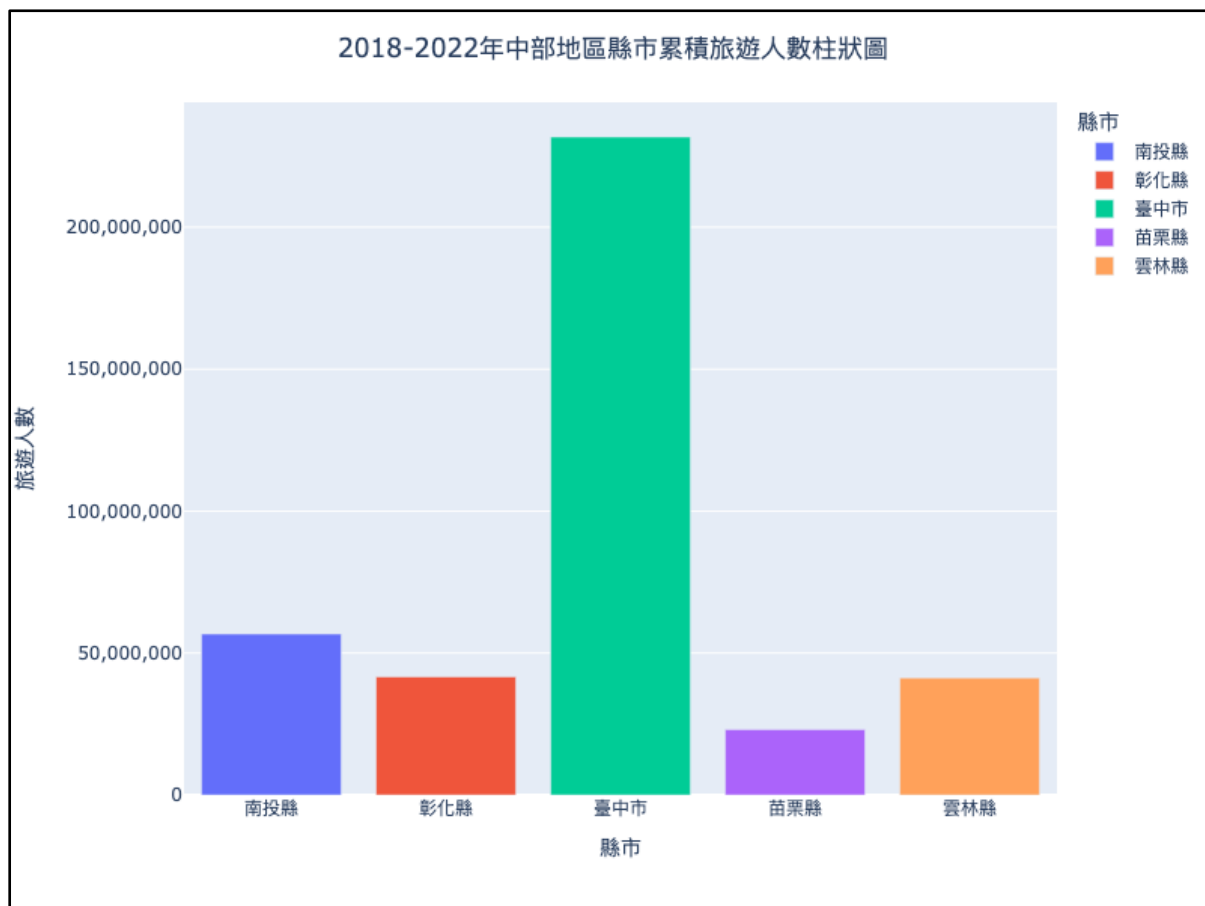
受到了疫情的影響，在圖中我們可以看到在2020年4月左右、2021年6月左右以及2022年5月左右都出現顯著的遊客人數下跌。

另外在2019年初可以發現遊客人數大增，在查閱相關新聞報導，且輔以統計資料庫的來台資料後，我們推測此現象是因為境外遊客大量增加，因為2019年是境外旅客來台人數創歷史新高的一年。不過在原資料中，並無紀錄境內境外遊客人數的欄位，因此無法確認我們的推理是否能完全解釋此現象，有或是有更多未考量的因素。

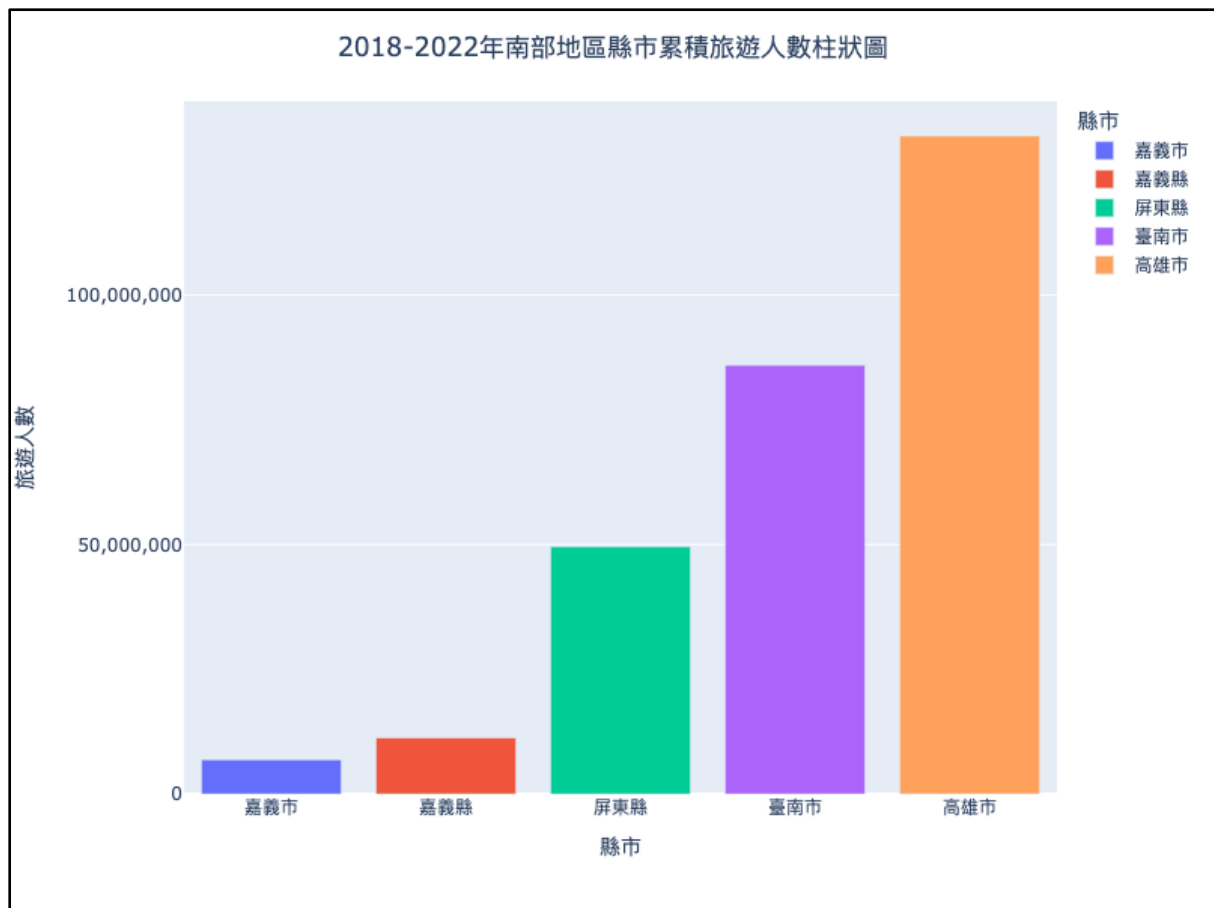
四、各地區內各縣市旅遊人數比例



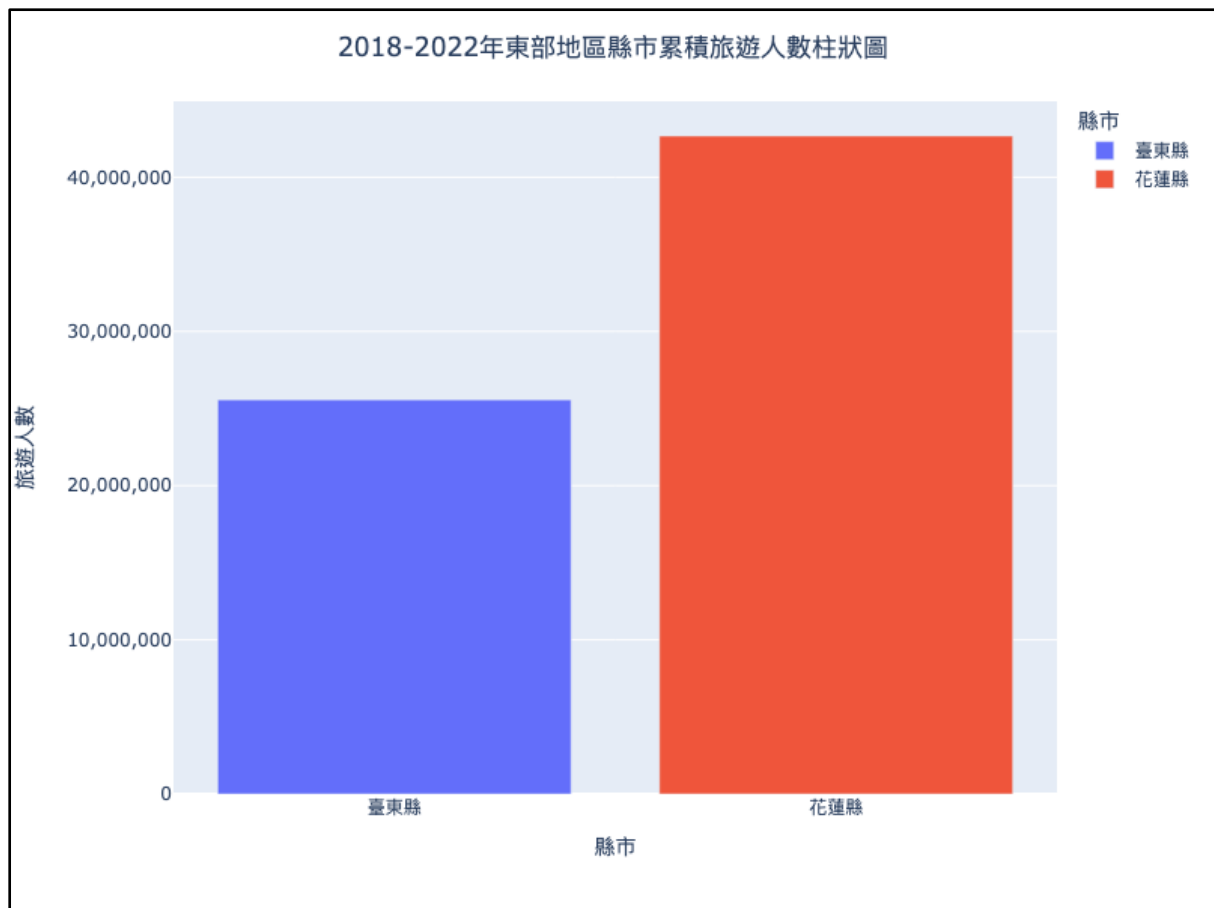
在這裡簡介縣市的繪圖過程，我們先對待分析地區做統計總旅遊人數，接著將x軸設為縣市，y軸設為旅遊人數，再進行圖表標題的設定與調整，即得到該圖。此圖為北部地區的縣市旅遊人數分佈柱狀圖，從這裡我們發現北部地區的縣市為五個地區當中最多的地區，且旅遊累積人數有突破一億的縣市有兩個，甚至新北市有達到兩億，在此似乎能先對我們的假設做呼應：景點類型應為自然景觀或文化古蹟，因為在新北市有很多著名的自然觀光景點，像是野柳地質公園，或是龍洞等等，揭示台灣十分著名的自然景點觀光勝地；而文化古蹟也有不少，像是淡水紅毛城，或是黃金博物館，都是觀光客十分喜愛的地點！



接著是中部地區的縣市累積旅遊人數柱狀圖，可以看到台中市一支獨秀，甚至超過兩億蠻多的，比新北市還多，在此我們認為可能是因為台中腹地面積大，且無論是藝文活動的推廣，或是都市慢活的魅力，皆為台中帶來許多的觀光人潮！而其他縣市明顯遜色許多，我們是推測為雖有一些自然景點吸引遊客，但整體城市規劃沒有台中來的便民，導致遊客前往意願降低。



再來是南部地區的縣市累積旅遊人數，在此地區沒有任何一個縣市突破一億，而高雄市為此地區最高之縣市，我們認為是此市擁有機場，且也有捷運輕軌系統，擁有便利的交通外，也充滿許多吸引遊客的觀光景點，像是可以觀賞美麗的落日美景的西子灣，或是有一堆展演空間的棧貳庫；而身為第二名的台南市，為充斥著美食的文化古都，也是遊客們喜愛前往的地區，但這個縣市較為可惜的是交通沒有那麼的理想，首先其高鐵站的位置太過偏避，遊客不易抵達鬧區景點，第二大眾運輸工具不夠完善，就如同我們在中部地區所述之原因，城市的便利性會影響遊客前往的意願。

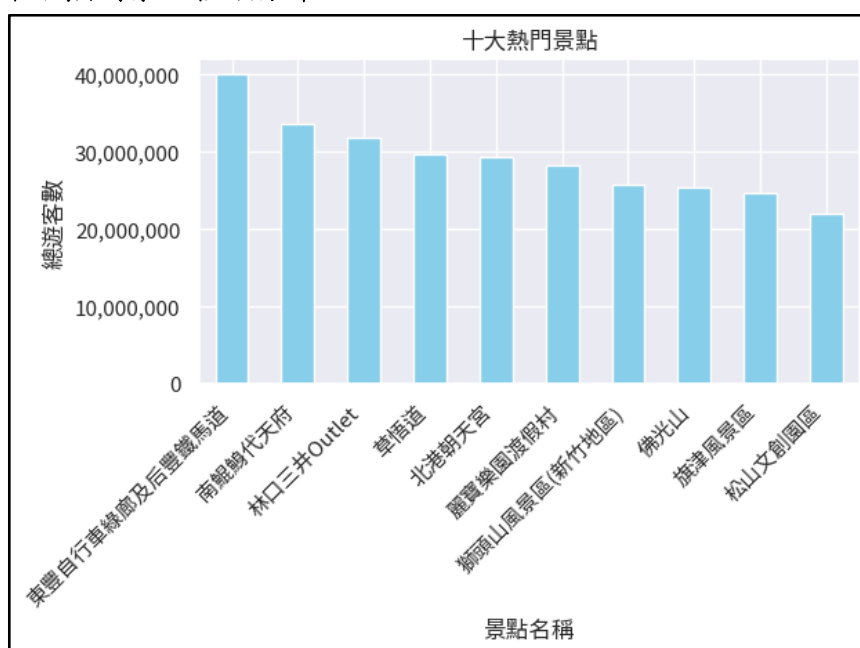


第四個地區是東部地區，這個地區只有兩個縣市：台東縣以及花蓮縣，可以看到y軸的部分，其座標只到四千萬，明顯地比起前三個地區來得少很多。而花蓮縣又是將近台東縣的兩倍，在此我們一樣是先推測為交通的便利性，接著我們認為與縣市的大小也有關，花蓮縣的面積為4629平方公里，台東縣為3515平方公里。

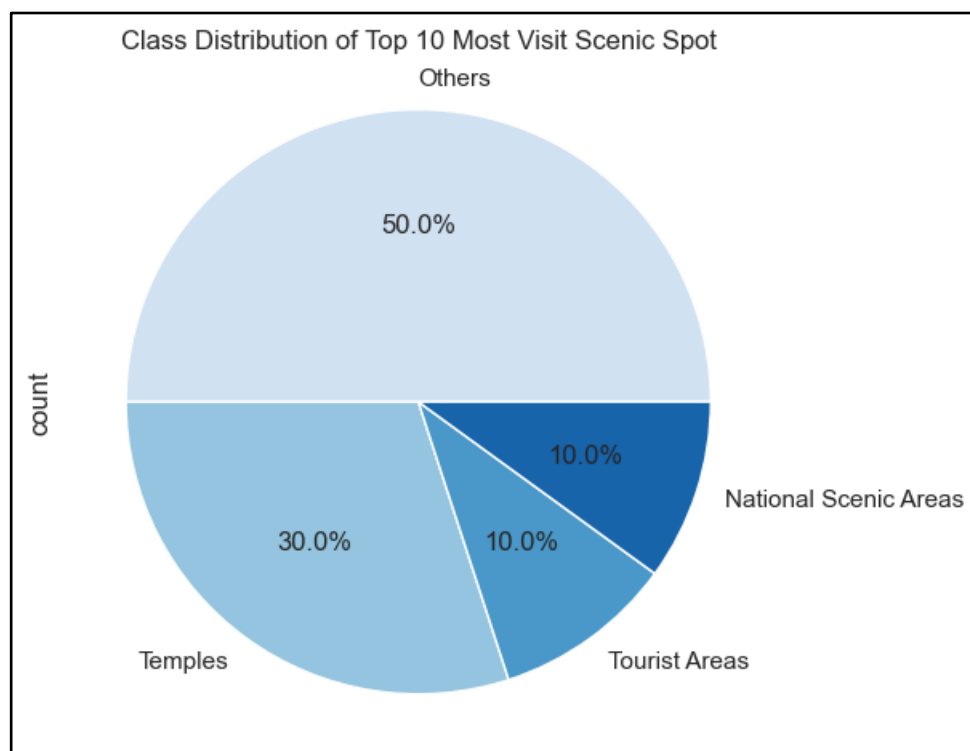


最後是離島地區的縣市累積旅遊人數，再y軸座標軸我們觀察到其最高值只標註到一千萬，又比東部地區的更少，而在此地區金門縣為最高，澎湖其次，但差距並不大，那我們認為這個分佈可能與其他縣市的原因有所雷同，像是交通、腹地面積，或者是景點著名程度等等皆有關聯。

五、全台前十名熱門景點種類分布



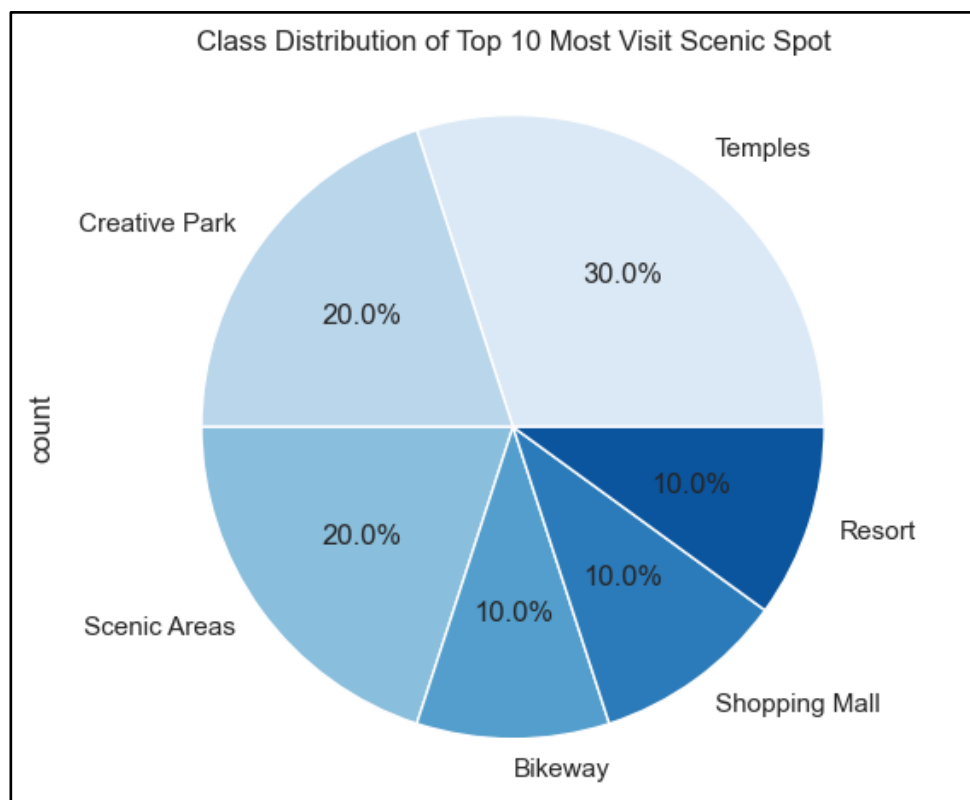
我們使用2018年至2022年共五年的累計旅客人數，選擇前十大旅客數的景點作為十大熱門景點，並且儲存在名為`top_10_spots_df`的Dataframe中。接著以長條圖呈現這十個景點與累計遊客人數。



接著將`class_df`將中的類型，即統計資料庫的分類，填入在`top_10_spots_df`中新增的'`Class`'欄位。並且利用新的'`Class`'欄位繪製十大熱門景點類型分佈的圓餅圖。

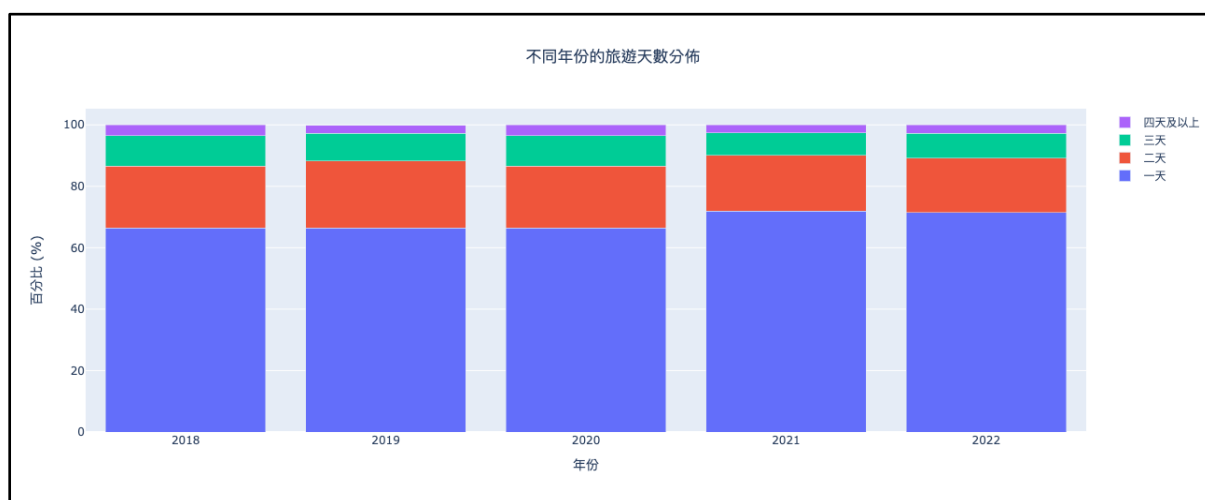
從圖表中我們可以看到類型屬於Temples的有3成，符合我們對熱門景點的類型應該要有高比例是文化古蹟或是自然景觀的假設。

然而我們也發現如果使用官方分類，會導致圖表中一半資料被劃分為Others，而Others中又包含了多樣類型的景點，如此一來可能導致我們無法看到數據中更細緻的訊息。因此我們決定手動幫`top_10_spots_df`添加更多類別，並將其命名為`adj_top_10_spots_df`。



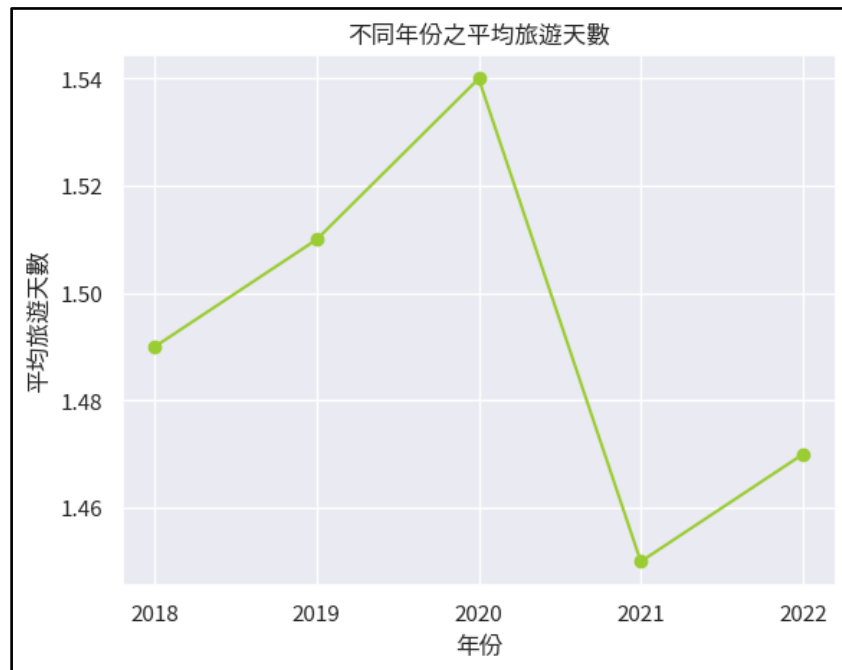
根據adj_top_10_spots_df繪製的圓餅圖如上，可以看到自然景觀類的Scenic Areas佔有兩成，也有符合我們原先的假設，另外逛街類型的景點（Shopping Mall和Creative Park）相加也有三成，也算是大宗。

六、旅遊天數



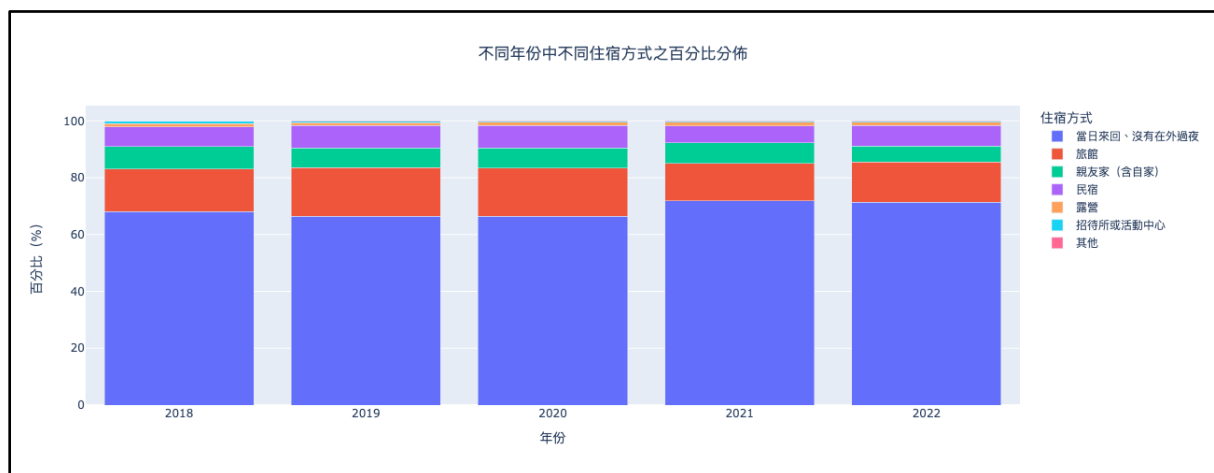
此圖為不同年份的旅遊天數分佈，因為此資料並沒有直接的csv檔可使用，所以我們是透過中華民國臺灣旅遊狀況調查去找到住宿天數，再統整出2018-2022的住宿天數狀況，並製成csv檔後再讀取使用。因為原始資料是有直接將平均天數也直接納入表格當中，所以我們在繪製此圖前有先使用melt函數去對此dataframe做整理，我們將 id_vars=['旅遊天數'], value_vars=['2018', '2019', '2020', '2021', '2022'], var_name='Year', value_name='Percentage'，接著篩選出不是平均旅遊天數的資料，最後

再去繪製成堆疊柱狀圖。其中我們可以看到2021年以及2022年明顯受到疫情的影響，導致一日遊的比例上升。



所以我們將剛剛先篩出來的平均旅遊天數再繪製成一張圖表做觀察，發現2021年為平均旅遊天數最短的一年，而2022年有稍微回升一點，感覺與疫情的趨緩有所關聯。

七、住宿方式



接著我們希望透過住宿方式去比較旅遊天數的分佈，在此一樣是透過中華民國臺灣旅遊狀況調查去找到住宿方式，接著再加以製成csv檔，在此也有將此dataframe使用melt函數做整理，我們將id_vars=['住宿方式'], var_name='Year', value_name='Percentage'，接著即可繪圖，在此圖表中我們觀察到當日來回的比例依樣在2021年以及2022年明顯增加，但其他住宿方式並沒有到特別大幅度更動，而是偏向等比例的縮減，比較明顯能看出變動的為2021年時的旅館以及民宿的比例似乎都較其他年的來的短許多，而親友家則有明顯的增加，推測是因為這樣的話即使確診也較容易釐清感染源，以及通知相關接觸人員自主健康管理。

我們使用套件WordCloud，並利用df['Location']和df['Scenic Spots']繪製文字雲。希望從縣市名稱與景點名稱中，觀察到哪些縣市，或是哪些種類的景點出現頻率高。



間而變化,近五年的數據可能無法充分反映當前的趨勢。但或許我們可以更改時限,去擴大探討。

捌、分工表

項目	劉綺栩	何嫻瑋
程式	<ul style="list-style-type: none"> ● 旅遊天數、住宿方式資料前處理 ● 各地區旅遊人數分佈比例圓餅圖 ● 各地區內各縣市旅遊人數柱狀圖 ● 旅遊天數之堆疊柱狀圖以及平均旅遊天數折線圖 ● 住宿方式之累積柱狀圖 	<ul style="list-style-type: none"> ● 遊憩景點月資料前處理 ● 旅遊人數折線圖 ● 十大景點類型分布圓餅圖 ● 文字雲
簡報	<ul style="list-style-type: none"> ● 動機 ● 各地區旅遊人數分佈比例圓餅圖 ● 各地區內各縣市旅遊人數柱狀圖 ● 旅遊天數之堆疊柱狀圖以及平均旅遊天數折線圖 ● 住宿方式累積柱狀圖 ● 研究限制 	<ul style="list-style-type: none"> ● 目的 ● 假說 ● 旅遊人數折線圖 ● 十大景點類型分布圓餅圖 ● 文字雲 ● 結論
書面報告	<ul style="list-style-type: none"> ● 動機 ● 各地區旅遊人數分佈比例圓餅圖 ● 各地區內各縣市旅遊人數柱狀圖 ● 旅遊天數之堆疊柱狀圖以及平均旅遊天數折線圖 ● 住宿方式之累積柱狀圖 ● 研究限制 	<ul style="list-style-type: none"> ● 目的 ● 假說 ● 資料取得 ● 遊憩景點月資料前處理 ● 旅遊人數折線圖 ● 十大景點類型分布圓餅圖 ● 文字雲 ● 結論

玖、參考文獻

<https://stat.taiwan.net.tw/>

<https://www.epochtimes.com/b5/20/1/6/n11772407.htm>

<https://admin.taiwan.net.tw/businessinfo/FilePage?a=14644>