書面報告

莊頌恩

6/17/2021

COVID-19對台灣國旅的影響:觀光局2018, 2019和2020三年數據分析

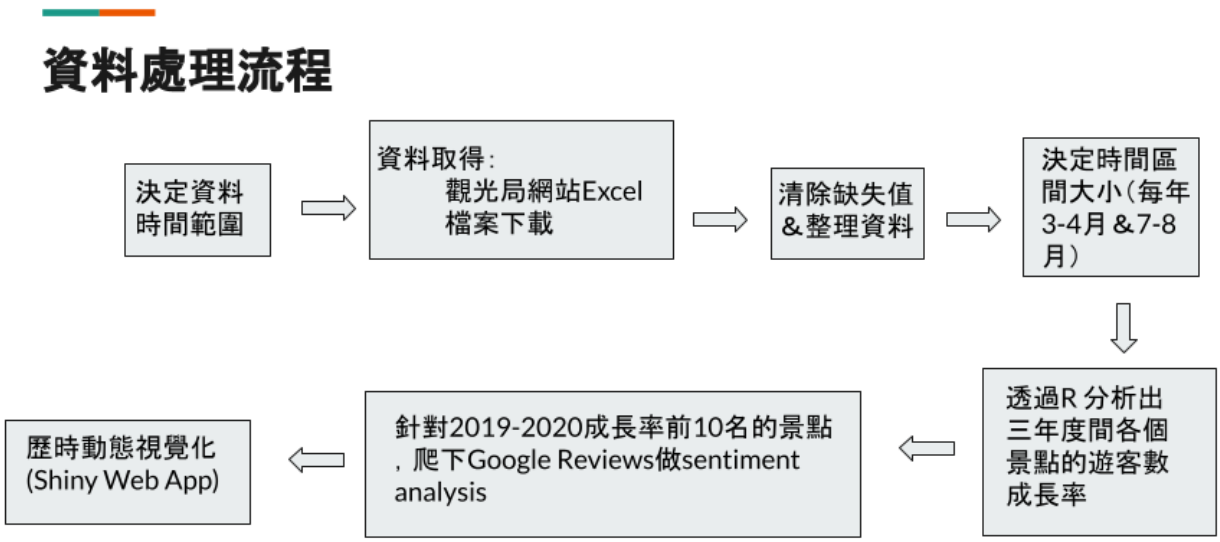
G01 Лосось (俄國鮭魚） 組員：莊頌恩、盧妍蓁、陳郁婕、朱嘉棟

1. 簡介

去年2020全球covid-19疫情慘重，大部分國家實施禁足，但台灣因為防疫政策優秀，在6月之後生活幾乎回復正常，因此國旅大爆發. 我們因此透過分析觀光局提供2018~2020年全台觀光景點每月觀光人數的紀錄，來找出國旅爆發受惠最多的景點，並透過情感分析，分析這些景點受到歡迎的原因.

2. 方法

```
* 2.1 資料取得：
    * 觀光局2018, 2019, 2020三個年分各個景點每月觀光人數統計
    * Google Reviews crawling
* 2.2 資料處理流程
```



* 2.3 原始碼運作說明（請參照各段code block’s註解）：

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.0.2
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.2
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──
```

```
## ✓ ggplot2 3.3.3     ✓ purrr   0.3.4
## ✓ tibble  3.1.0     ✓ dplyr   1.0.5
## ✓ tidyr   1.1.3     ✓ stringr 1.4.0
## ✓ readr   1.4.0     ✓ forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
## Warning: package 'tibble' was built under R version 4.0.2
```

```
## Warning: package 'tidyr' was built under R version 4.0.2
```

```
## Warning: package 'readr' was built under R version 4.0.2
```

```
## Warning: package 'purrr' was built under R version 4.0.2
```

```
## Warning: package 'dplyr' was built under R version 4.0.2
```

```
## Warning: package 'stringr' was built under R version 4.0.2
```

```
## Warning: package 'forcats' was built under R version 4.0.2
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

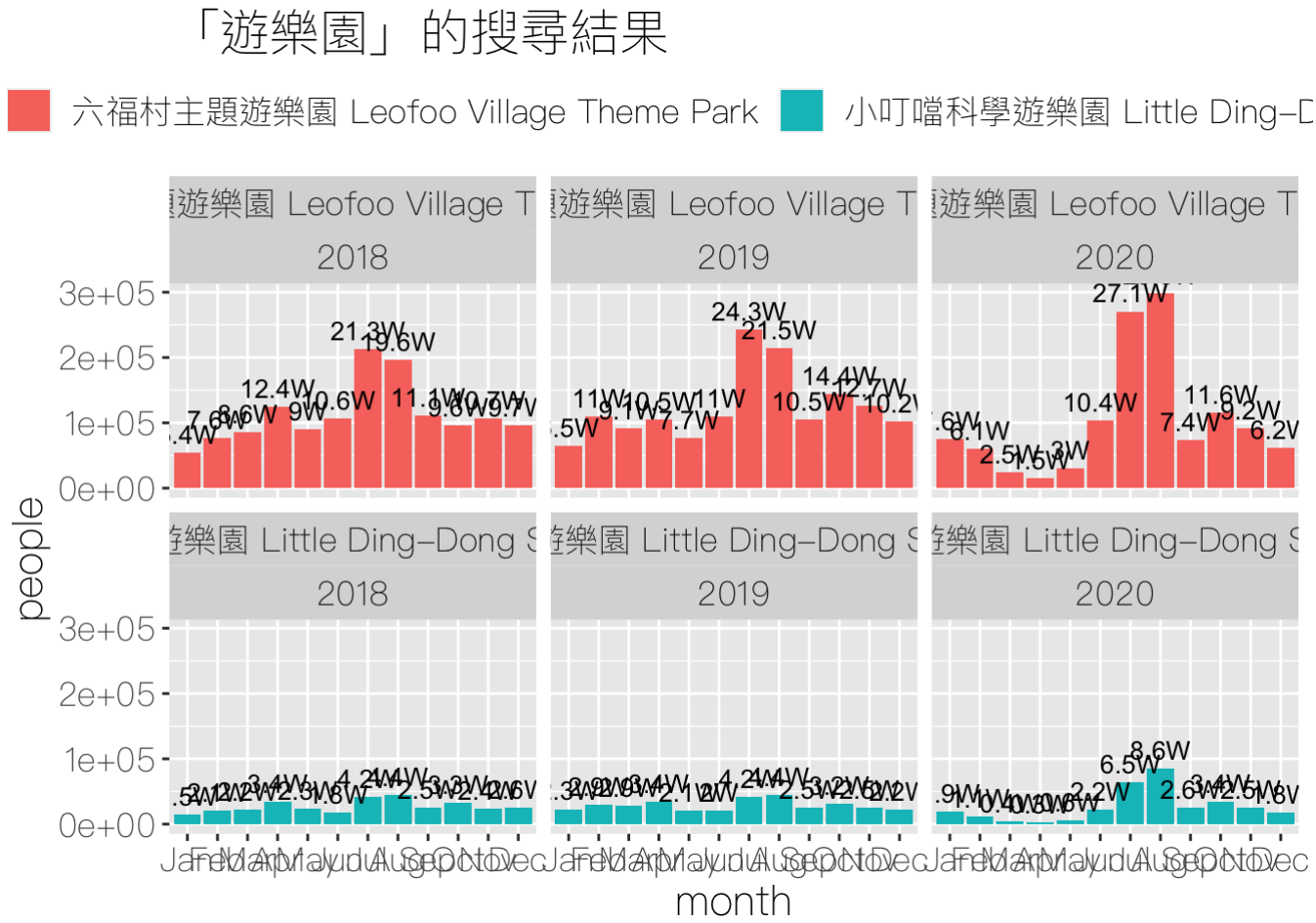
```
library(repr)
```

```
## Warning: package 'repr' was built under R version 4.0.2
```

```
# 有哪幾個年份的要讀進來
tourist__files <- c("tourist_18", "tourist_19", "tourist_20")


# 設定字體
theme <- theme_get()
theme$text$family <- "PingFangTC-Thin"
theme_set(theme)

# 讀檔案進來
for (i in tourist__files) {
    assign(i, readxl::read_excel(
        paste0(
            "./src/20",
            strsplit(i, split = "")[[1]][9],
            strsplit(i, split = "")[[1]][10],
            "_tourist.xlsx"
        ),
        skip = 1
    ))
}


# Change names and add years and bind the 3 files to create "tourist"
revised_names <- c("Type","Spots","City","Jan","Feb","Mar","Apr","May","Jun","Jul","A
ug","Sep","Oct","Nov","Dec","Total","Counting_method")
names(tourist_18) <- revised_names
tourist_18$year <- 2018
names(tourist_19) <- revised_names
tourist_19$year <- 2019
names(tourist_20) <- revised_names
tourist_20$year <- 2020
tourist <- bind_rows(tourist_18, tourist_19, tourist_20) %>%
    select(Spots, City, year, Jan:Total)
# 若新加檔案進來, 只要新增code到以上為止


# 移除缺失值的列
tourist <- na.omit(tourist)


# 把景點名稱中間的多餘空白清理乾淨(mutate出一行)
spots_zh <- c()
spots_en <- c()
for (spot in tourist$Spots) {
    spots_zh <- c(spots_zh, str_split(spot, "\r\n| +")[[1]][1])
}
for (spot in tourist$Spots) {
    spots_en <- c(spots_en,
                  trimws(
                      paste(str_split(spot, "\r\n| +")[[1]][-1], collapse = " "),
                      which = "both",
                      whitespace = "[ \t\r\n]"
                  ))
}
tourist <- tourist %>%
    mutate(Spots_zh = spots_zh,
```

```
            Spots_en = spots_en) %>%
    unite(spots, c("Spots_zh", "Spots_en"), sep = " ")


# 把City名稱只保留中文部分(直接取代City), 並整理欄位排序
cities <- c()
for (city in tourist$City) {
    cities <- c(cities, str_split(city, "\r\n| +")[[1]][1])
}
tourist$City <- cities
tourist <- tourist %>%
    select(City, spots, year:Total)


# 看各縣市的總遊客量
# tourist %>%
#     group_by(City, year) %>%
#     summarise(num_spots = n(),
#               total_tourist = sum(Total)) %>%
#     arrange(desc(total_tourist))

get_height <- function(place) {
    return (length(unique(tourist$spots[str_detect(tourist$spots, place)])))}


# 對特定地點的觀光人數對月份作圖, 並以年份做facet_wrap (包成function)
show_plot <- function(place = "101", save = FALSE, width = 10) {
    data <- tourist %>% filter(str_detect(spots, place)) %>%
        pivot_longer(cols = Jan:Dec,
                     names_to = "month",
                     values_to = "people") %>%
        select(!Total) %>%
        mutate(W_people = paste0(round(people/10000, 1), "W"))
    # cat(paste0("「", place, "」", "的搜尋結果:"), unique(data$spots), sep = "\n")
    data$month <-
        factor(as.character(data$month), levels = unique(as.character(data$month)))
    options(repr.plot.width= 20, repr.plot.height=7*length(unique(data$spots)))

    p <- ggplot(data, aes(month, people, fill = spots, label = W_people)) +
        geom_bar(stat = "identity") +
        theme(text = element_text(size=16), legend.position="top") +
        geom_text(color="black", size=3.5, angle=0, position = position_dodge(1), vju
st=-0.5)

    if (length(data$City) > 0){
        p <- p + facet_wrap(vars(spots, year), ncol = length(tourist__files)) +
            ggtitle(paste0("「", place, "」", "的搜尋結果")) # + scale_fill_brewer(palet
te = "Spectral")
    }
    else{
        p <- p + ggtitle(paste0("「", place, "」", "無搜尋結果"))
    }

    if (save) {
        ggsave(
            "plot.png",
            width = width,
            height = sum(c(3.5 * length(
```

```
                unique(data$spots)
            ), 0.5)),
            dpi = 320,
            device = "png"
        )
    }
    return (p)
}


# sample(tourist$spots, size = 30)
# options(repr.plot.width= 20, repr.plot.height=9)
show_plot("遊樂園", save = F, width = 12) #可自行調整width(有些地名太長塞不下)
```



```
# 看年成長率
sort_growth_rate <-
    function(focused_months = 1:12,             # 選擇月份
             tourists_threshold = 50000,        # 設定選擇月份內總觀光人次的門檻
             which_year_to_look_up = 2019) {    # 想觀察第n年到n+1年間的成長率

        # 將觀光景點以名字粗分
        new_tourist <- tourist %>% arrange(spots)
        spot_id <- c()
        temp <- c()
        id <- 1

        # 同一景點在不同年份可能有不同名字, 因此創建一新變項spot_id來對真正的景點編號
        for (i in 1:(length(new_tourist$spots) - 1)) {
            spot_id <- c(spot_id, id)
            if (new_tourist$year[i] + 1 != new_tourist$year[i + 1] |
                # 利用相同景點年份相連的特性
                new_tourist$City[i] != new_tourist$City[i + 1]) {
                # 利用相同景點縣市相同的特性
                id <- id + 1
                temp <- c(temp, NA) # 明年的數據還沒出來, 因此無法計算成長率
            }
            else {
                temp <- c(temp, 1)
            }
        }

        spot_id <- c(spot_id, id)
        temp <- c(temp, NA)
        new_tourist$spot_id <- spot_id # 以spot_id對景點編號

        # 將表格的月份collapse成一新變數month, 方便稍後統計
        new_tourist <-
            new_tourist %>% select(City, spot_id, spots, year:Total) %>%
            pivot_longer(cols = Jan:Dec,
                         names_to = "month",
                         values_to = "people") %>%
            select(!Total)

        mon <- c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","D
ec")

        # 留下指定月份的資料, 計算總觀光人次
        selected_new_tourist <- new_tourist %>%
            filter(month %in% mon[focused_months]) %>%
            group_by(spot_id, year) %>%
            summarise(total = sum(people))

        # 計算觀光人次年增量
        annual_growth <- c(diff(selected_new_tourist$total), NA) * temp

        selected_new_tourist$annual_growth <- annual_growth

        # 計算觀光人次年增率
        selected_new_tourist <- selected_new_tourist %>%
            mutate(annual_growth_rate = annual_growth / total)
```

```
        # 選定要觀察的年份，設定總人次門檻，並依照觀光人次年增率由高至低排列
        focused_new_tourist <- selected_new_tourist %>%
            filter(year == which_year_to_look_up, total >= tourists_threshold) %>%
            arrange(desc(annual_growth_rate))

        # 回原本dataframe抓spot_id對應的景點名稱
        get_name <- function(id) {
            return(paste0(unique(new_tourist$spots[new_tourist$spot_id == id]), colla
pse = " / "))
        }

        # 回原本dataframe抓spot_id對應的景點名稱所在縣市
        get_city <- function(id) {
            return(paste0(unique(new_tourist$City[new_tourist$spot_id == id]), collap
se = " / "))
        }

        # 把觀光人次年增率以百分比形式呈現
        percent <- function(x,
                            digits = 2,
                            format = "f") {
            paste0(formatC(100 * x, format = format, digits = digits), "%")
        }

        # 最後整理dataframe並回傳
        focused_new_tourist <- focused_new_tourist %>%
            mutate(
                City = get_city(spot_id),
                spot_name = get_name(spot_id),
                this_year_total = total,
                next_year_total = total + annual_growth,
                growth_rate = percent(annual_growth_rate)
            ) %>%
            ungroup() %>%
            select(City:next_year_total,
                   annual_growth,
                   growth_rate) %>%
            drop_na()

        return(focused_new_tourist)
    }

table1 <- sort_growth_rate(
    focused_months = 1:12,
    tourists_threshold = 1000000,
    which_year_to_look_up = 2019
)
```

```
## `summarise()` has grouped output by 'spot_id'. You can override using the `.groups
` argument.
```

```
# 檢查名字有沒有改變或混淆
check_names <- function() {
    new_tourist <- tourist %>% arrange(spots)
    spot_id <- c()
    temp <- c()
    id <- 1
    for (i in 1:(length(new_tourist$spots) - 1)) {
        spot_id <- c(spot_id, id)
        if (new_tourist$year[i] + 1 != new_tourist$year[i + 1] |
            # 利用相同景點年份相連的特性
            new_tourist$City[i] != new_tourist$City[i + 1]) {
            # 利用相同景點縣市相同的特性
            id <- id + 1
            temp <- c(temp, NA) # 明年的數據還沒出來, 因此無法計算成長率
        } else {
            temp <- c(temp, 1)
        }
    }

    spot_id <- c(spot_id, id)
    temp <- c(temp, NA)
    new_tourist$spot_id <- spot_id # 以spot_id對景點編號
    new_tourist <-
        new_tourist %>% select(City, spot_id, spots, year)

    spot_name <- c()
    spot_count <- c()
    dif_name <- c()
    cnt <- 1
    name <- new_tourist$spots[1]
    for (i in 2:length(new_tourist$spot_id)) {
        if (new_tourist$spot_id[i] == new_tourist$spot_id[i - 1]) {
            cnt <- cnt + 1
            name <- c(name, new_tourist$spots[i])
        } else{
            spot_count <- c(spot_count, cnt)
            cnt <- 1
            dif_name <- c(dif_name, length(unique(name)))
            spot_name <-
                c(spot_name, paste0(unique(name), collapse = " / "))
            name <- new_tourist$spots[i]
        }
    }
    return (tibble(
        name = spot_name,
        count = spot_count,
        n_name = dif_name
    ))
}

(check_names() %>% filter(n_name == 2))$name
```

```
##  [1] "五峰旗瀑布 Wufongci Waterfall / 五峰旗風景區 Wufongci Waterfall"
##  [2] "卑南文化公園 Beinan Cultural Park / 卑南遺址公園 Peinan Site Park"
##  [3] "國民革命忠烈祠 National Revolutionary Martyr' Shrine / 國民革命忠烈祠 National Re
volutionary Martyr’ Shrine"
##  [4] "愛河 Love River / 愛河(五福四路至博愛一路) Love River(Wufu 4th Rd.to Bo-ai 1st R
d.)"
##  [5] "杉林溪森林生態渡假園區 Sun Link Sea Forest Recreation Area / 杉林溪森林生態渡假園區
Sun-Link-Sea Forest and Nature Resort"
##  [6] "林本源園邸( 林家花園)  The Lin Family Mans / 林本源園邸( 林家花園) The Lin Family
Mans"
##  [7] "清境農場 Cingjing Farm / 清境農場 Qingjing Farm"
##  [8] "角板山行館 Jiaobanshan Resort / 角板山遊憩區 Jiaobanshan Resort"
##  [9] "陽明山遊客中心 Visitor Center / 陽明山遊客中心 Yangmingshan Visitor Center"
## [10] "鹽寮海濱公園 Yanliao Beach Park / 鹽寮福隆遊憩區 Yanliao Fujong Recreation Area"
## [11] "鼻頭角步道 Bitou Cape Trail / 鼻頭龍洞遊憩區 Bitou Longdong Recreation Area"
```

```
# 爬七、八月2019-20成長率前十名景點的Google Map評論各25則-----------------------
# July-August
jul_aus <- sort_growth_rate(7:8, 300000, 2019) %>% head(10)
```

```
## `summarise()` has grouped output by 'spot_id'. You can override using the `.groups
` argument.
```

```
jul_aus$spot_name
```

```
##  [1] "北港朝天宮 BeiGang ChaoTian Temple"
##  [2] "麗寶樂園 LihPaoLand"
##  [3] "虎頭山風景特定區 Houtou Mountain Scenic Area"
##  [4] "日月潭風景區 The Sunmoonlake Scenic Area"
##  [5] "新竹漁港 Hsinchu Fishing Port"
##  [6] "鹿野高臺 Luyeh High Terrace"
##  [7] "六福村主題遊樂園 Leofoo Village Theme Park"
##  [8] "溪頭自然教育園區 Xitou Nature Education Area"
##  [9] "國立海洋科技博物館 National Museum of Marine Science & Technology"
## [10] "國立科學工藝博物館 National Science and Technology Museum"
```

```
apify_input <- c("BeiGang ChaoTian Temple",
                 "Houtou Mountain Scenic Area",
                 "The Sunmoonlake Scenic Area",
                 "Hsinchu Fishing Port",
                 "Luyeh High Terrace",
                 "Leofoo Village Theme Park",
                 "Xitou Nature Education Area",
                 "National Museum of Marine Science & Technology",
                 "National Science and Technology Museum",
                 "Lion's Head Mountain Scenic Area")
apify_input_zh <- c("北港朝天宮", "虎頭山風景特定區",
                    "日月潭風景區", "新竹漁港",
                    "鹿野高臺", "六福村主題遊樂園",
                    "溪頭自然教育園區", "國立海洋科技博物館",
                    "國立科學工藝博物館", "獅頭山風景區")

reviews <- readr::read_csv("./src/google_map_reviews.csv")
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   .default = col_logical(),
##   address = col_character(),
##   categoryName = col_character(),
##   city = col_character(),
##   countryCode = col_character(),
##   locatedIn = col_character(),
##   `location/lat` = col_double(),
##   `location/lng` = col_double(),
##   neighborhood = col_character(),
##   phone = col_character(),
##   placeId = col_character(),
##   plusCode = col_character(),
##   postalCode = col_double(),
##   rank = col_double(),
##   `reviews/0/name` = col_character(),
##   `reviews/0/publishAt` = col_character(),
##   `reviews/0/publishedAtDate` = col_datetime(format = ""),
##   `reviews/0/responseFromOwnerDate` = col_datetime(format = ""),
##   `reviews/0/reviewerNumberOfReviews` = col_double(),
##   `reviews/0/stars` = col_double(),
##   `reviews/0/text` = col_character()
##   # ... with 180 more columns
## )
## ℹ Use `spec()` for the full column specifications.
```

```
# 把reviews的名稱裡的"/"改成"_"以便之後使用
for (i in seq_along(names(reviews))) {
    names(reviews)[i] <- str_replace_all(names(reviews)[i], "/", "_")
}

# 只挑出reviews中的文本們 (長相：review_texts[[3]] 就是每一個景點的第三則評論放在一起)
review_texts <- c()
for (i in seq(25)) {
    review_texts[i] <- reviews[paste0("reviews_", i-1, "_text")]
}

review_texts[[3]]
```

```
##  [1] "一年一度的媽祖聖誕慶典，真人藝閣登場……\n已經來畫好幾年藝閣了，大家有空一定要過來參加一次這個活動，只有北港有今年是4/30-5/4\n藝閣車大概有3-40輛每台車的造型與主題都不同，真人藝閣大大小小的年齡層都有，除了打扮化妝外還有特別主題的服裝搭配，每個上車的藝閣都會準備各式各樣的禮物從車上丟下來給遊行隊伍的朋友們，各式各樣的工具都上場接禮物...\n熱鬧的遊街中還有鞭炮炮城，炸過好運旺旺來……\n趁著還有3天的時間撥個空來北港逗熱鬧"
##  [2] "桐花季節，可以爬上又可以賞桐花一舉兩得，山上還有帥氣關公廟，路旁有各式小吃物品爬完山還可以逛逛街。"
##  [3] "虎頭山位在玉井竹圍裡境，因山勢酷似虎頭而得名，山標高239公尺。\n是玉井區域內非常有名的咖啡賞景勝地，山頂的幾間咖啡館每到假日總是一位難求，每家咖啡館都各有特色，所擁有的景觀條件也都不相同，各個角度都有不同的美，但唯一相同的是都位於玉井市區的至高處，白天可以眺望山中小鎮的景緻，夜晚則是可以欣賞玉井的精緻夜景，是許多年輕族群喜愛的浪漫約會處。\n山頂豎立了「抗日烈士余清芳紀念碑」是為紀念在此抗日殉役的余清芳。\n除了是約會賞景放空的好去處外，也是一處富有歷史教育意義的場址。"
##  [4] "晚上不用錢，適合晚上散步走走的好去處"
##  [5] "奇岩怪石這夜間可拍銀河或星軌"
##  [6] "美麗的日月潭 \U0001f44d\U0001f44d\U0001f44d\n祈禱降雨\n讓日月潭 再次壯麗❤️\n祈盼人民\n真心向善 敬畏天地❤️"
##  [7] NA
##  [8] NA
##  [9] "純粹樸實的鄉村漁港"
## [10] "如果能好好規劃一下，相信是一個很不錯的地方，看著落日，喝著啤酒，吃著小吃，在夕陽的餘暉之下享受大自然那是一件非常美好的事情"
## [11] "陽光🌞燦爛，\n風很大！\n路上有風箏\U0001fa81；\n海上有衝浪風箏\U0001f3c4♀，\n漁會漁獲強強滾……"
## [12] NA
## [13] "鹿野高台擁有絕佳視野，能夠一覽整個高台地區與卑南溪谷底的田野景色，也是臺灣東部一處優良的天然空域活動場地。每當6月至8月時，正是鹿野高台進行熱氣球活動的好時機，搭乘緩緩升空起飛的熱氣球，徜徉在花東縱谷的美景之中，親身體驗熱氣球起飛的這一份躍動。\n\n從龍田村沿著指標一路往高台走，地勢逐漸升高，經過高台茶葉展示中心後不遠，好大一片綠草如茵的寬廣草坡隨即映入眼簾，登上草坡最高處的高眺亭，整個高台地區與縱谷的景色，盡在眼前，一覽無遺。\n\n每逢暑假期間，盛大舉辦的臺灣國際熱氣球嘉年華以及光雕音樂會，都吸引來自世界各地的旅人前來參加，熱氣球繫留體驗、熱氣球自由飛行表演、絢麗燦爛的光雕音樂會，或是曙光光雕音樂會等活動，讓人永生難忘。"
## [14] "1. 設施的刺激程度相當完整！從很友善、很安全、適合幼童跟著家長一起玩的，到很令人膽怯、很刺激、很高壓的都有！要高度有高度（大怒神）、要速度有速度（笑傲飛鷹）、想玩水也有水（火山歷險），整體來說這邊的遊樂設施真的是涵蓋了非常全面的需求。\n2. 園區的地圖、分區規劃相當清晰明瞭，總共就四、五個區域，每個區域大致上來說都是一個圓形，而這些圓形都共同依靠著一個圓心形成了一個整齊的全貌。\n3. 有規模頗大的動物園區！算是一大特色。各種平常較難以見到的生物都能在這裡觀察到，尤其是大型動物（犀牛、水牛、長頸鹿…）\n4. 食物的部分只吃到一家漢堡店，雖然價位不是說很友善但漢堡是意外的好吃。\n\n總結來說，很推薦各個年齡層、各種成員組合可以來六福村度過一天假日！"
## [15] "很適合帶幼兒小孩來學習物理和科學知識。放假期間人會比較多，有地方玩水，不用換泳裝，但部分水上設施要7/1-8-31才有開放。"
## [16] "溪頭自然教育園區位於南投縣鹿谷鄉鳳凰山谷，因位於北勢溪源頭而得名，園區三面環山佔地廣闊，海拔約在800~2000m之間，目前也是臺灣大學農學院實驗林七個營林區之一，種植紅檜、銀杏、扁柏等珍貴樹種，還有上千種草本植物。 園區內更設有多條森林步道與登山步道可深入如人間仙境般的氤氳樹林中健行，也讓台灣人多年來提到森林浴，往往第一個就想起溪頭的蒼鬱樹林。"
## [17] "連綿而上的步道階梯，都是粉白色的 河津櫻 \U0001f338，好天氣藍天白雲的陽光襯托下，拍起來就一整片的粉白花道很不錯。\n228連假 第二天來 南投 鹿谷 賞櫻，前一兩週去台中新社看的亮紅色 八重櫻 已經都凋謝了，這兩天有關注網路花況訊息，所以選擇來這邊看的是粉紅色的 河津櫻\U0001f338。一大早就出發，抵達這邊還不到八點開園已經都在排隊了....人潮很恐怖，整個南投往日月潭跟九族文化村的路全部塞爆，還好我們的賞櫻景點都有避開，且很早就出發了。\n台大實驗林 鳳凰自然教育園區 2/28/2021"
## [18] "入館看展需要付門票，場地很大又乾淨。如果不想花費，附近也有免費鯊魚館可參觀，相當適合帶小孩來走走！"
## [19] "南館有一片大廣場 但天氣太熱 沒什麼人 北館內部展覽我不喜歡 一大堆都是要另外收費的特展 走那麼多路沒幾個展能看 二樓更是令人耗費元神 機械展入口閘門的阿桑 看到我沒蓋外出章就一直追究為什麼沒蓋 我哪知道？我也是第一次來啊？一再解釋自己從B1的農業展上來的她也不聽 女人何苦為難女人？不是有票根出示給她看就好嗎 ？？？碎念有夠久 還沒看展心先累一半 後來的機械展根本壓倒我的最後一根稻草 懂的人可能覺得很有趣 但我腦袋不夠好看不太懂直接心累到高速離開 呆坐在出口的椅子休息 剩下的展都沒心力看了...給到兩顆星是因為有看到可愛的尼莫而且南館的廁所很乾淨"
```

```
## [20] "竹藝博物館原來就在南投縣立文化中心的一樓\n第一次來還繞了一圈\n繞了南投高中好大一圈\n\n一開始只看到文化中心，接著竟然沒看到竹藝博物館\n後來停在文化中心的免費停車場後\n不甘願再去文化中心問一次，入門右手邊就寫著南投縣竹藝博物館\n\n原來博物館在文化中心的1樓 學到了\n展覽品還蠻多的"
## [21] "2020/12/26(六)造訪時，大溪分有七個木藝生態博物館，大多以日據時期留下來的場館為基礎改建，分別展示各種木頭相關的工藝及文化，此為藝師館。"
## [22] NA
## [23] "爬山遊客都很親切，雖然不認識但是會互相打招呼！"
## [24] "有多條步道可供民眾選擇，算輕鬆不難走，適合親子出遊踏青。\n\n其實路線入口不只大坑，我個人習慣從潭子領袖山莊這邊進來，人少很多，停車也很方便，經風動石再接往觀音亭，沿途景色優美，視野遼闊，很多蜥蜴小動物，帶小孩出來放電的好去處。\n\n最終一樣走回潭子，在潭子用午餐，避免在大坑那邊人擠人，這樣的走法，大概都是在地人比較知道走法，外地遊客跟住台中市區的人就比較不會這樣走。"
```

```
# 整理成每個景點的25則評論為一欄
nine_spots_reviews <- tibble(北港朝天宮 = vector("character", length = 25), 虎頭山風景特定區 = vector("character", length = 25), 日月潭風景區 = vector("character", length = 25),
                                    新竹漁港 = vector("character", length = 25), 鹿野高臺 = vector("character", length = 25), 六福村主題遊樂園 = vector("character", length = 25),
                                    溪頭自然教育園區 = vector("character", length = 25), 國立海洋科技博物館 = vector("character", length = 25), 國立科學工藝博物館 = vector("character", length = 25))
for (i in seq(25)) {
    nine_spots_reviews$北港朝天宮[i] <- review_texts[[i]][1]
    nine_spots_reviews$虎頭山風景特定區[i] <- review_texts[[i]][2]
    nine_spots_reviews$日月潭風景區[i] <- review_texts[[i]][6]
    nine_spots_reviews$新竹漁港[i] <- review_texts[[i]][11]
    nine_spots_reviews$鹿野高臺[i] <- review_texts[[i]][13]
    nine_spots_reviews$六福村主題遊樂園[i] <- review_texts[[i]][14]
    nine_spots_reviews$溪頭自然教育園區[i] <- review_texts[[i]][16]
    nine_spots_reviews$國立海洋科技博物館[i] <- review_texts[[i]][18]
    nine_spots_reviews$國立科學工藝博物館[i] <- review_texts[[i]][19]
}

#清數字
for (i in seq(9)) {
    nine_spots_reviews[[i]] <- gsub('[0-9]+', '', nine_spots_reviews[[i]])
}

# 讀sentiment analysis辭典
sent <- readr::read_csv("./src/ch.senti.lex.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   X1 = col_double(),
##   lemma = col_character(),
##   Polarity = col_character(),
##   Emotion_Category = col_number(),
##   Personality = col_double()
## )
```

```
# 大家一起情緒分析-----------------------------------------------------------------
sentiment_plot <- function(site) {
    # 斷詞
    library(jiebaR)
    seg <- worker()
    seg_site <- tibble(word = segment(nine_spots_reviews[[site]], seg))

    # 去掉stopwords
    stop_words <- readLines("./src/stopwords.txt", encoding = "UTF-8")
    stop_df <- tibble(word = stop_words)
    site_remove <- anti_join(seg_site, stop_df)

    # inner join 評論 和 辭典
    site_sent <- inner_join(site_remove, sent, by = c("word" = "lemma"))
    site_count <- site_sent %>%
        count(word, Polarity) %>%
        arrange(desc(n))

    # 景點評論負向詞 把n變-n是為了之後畫圖漂亮用的而已
    site_negative <- site_count %>%
        filter(Polarity == "N") %>%
        top_n(10) %>%
        mutate(n = -n)

    # 景點評論正向詞
    site_positive <- site_count %>%
        filter(Polarity == "P") %>%
        top_n(10)

    # 把正向跟負向的前十名合併
    site_pos_neg <- rbind(site_negative, site_positive)

    ggplot(site_pos_neg, aes(reorder(word, n), n, fill = Polarity, label = abs(n))) +
        geom_col() +
        coord_flip() +
        labs(title = paste0(site, "情緒分析 正向／負向 前十名"),
             y = "出現次數",
             x = "字詞") +
        theme(text = element_text(family="STHeiti", size=16), legend.position="top")
 +
        geom_text(color="black", size=4, angle=0, position = position_dodge(1), hjust
=-0.5)
}
```

再來是Shiny Web App的部份（但因為是pdf檔案，無法顯示出Shiny App動態介面，改成在下面結果部分，使用截圖展示）：

```
library(shiny)
```

```
## Warning: package 'shiny' was built under R version 4.0.2
```

```
library(shinyWidgets)
```

```
## Warning: package 'shinyWidgets' was built under R version 4.0.2
```

```
library(gapminder) # data
```

```
## Warning: package 'gapminder' was built under R version 4.0.2
```

```
library(DT)
```

```
## Warning: package 'DT' was built under R version 4.0.2
```

```
##
## Attaching package: 'DT'
```

```
## The following objects are masked from 'package:shiny':
##
##     dataTableOutput, renderDataTable
```

```
library(shinythemes)
```

```
## Warning: package 'shinythemes' was built under R version 4.0.2
```

```
library(ggplot2)
library(rsconnect)
```

```
## Warning: package 'rsconnect' was built under R version 4.0.2
```

```
##
## Attaching package: 'rsconnect'
```

```
## The following object is masked from 'package:shiny':
##
##     serverInfo
```

```
source("./src/system_setup.R")
```

```
## Warning in dir.create("~/.fonts"): '/Users/seanchuang/.fonts' already exists
```

```
## Warning in system("fc-cache -f ~/.fonts"): error in running command
```

```
data <- source("./src/final_project.R", local = TRUE)
```

```
## `summarise()` has grouped output by 'spot_id'. You can override using the `.groups
` argument.
```

```
## `summarise()` has grouped output by 'spot_id'. You can override using the `.groups
` argument.
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   .default = col_logical(),
##   address = col_character(),
##   categoryName = col_character(),
##   city = col_character(),
##   countryCode = col_character(),
##   locatedIn = col_character(),
##   `location/lat` = col_double(),
##   `location/lng` = col_double(),
##   neighborhood = col_character(),
##   phone = col_character(),
##   placeId = col_character(),
##   plusCode = col_character(),
##   postalCode = col_double(),
##   rank = col_double(),
##   `reviews/0/name` = col_character(),
##   `reviews/0/publishAt` = col_character(),
##   `reviews/0/publishedAtDate` = col_datetime(format = ""),
##   `reviews/0/responseFromOwnerDate` = col_datetime(format = ""),
##   `reviews/0/reviewerNumberOfReviews` = col_double(),
##   `reviews/0/stars` = col_double(),
##   `reviews/0/text` = col_character()
##   # ... with 180 more columns
## )
## ℹ Use `spec()` for the full column specifications.
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   X1 = col_double(),
##   lemma = col_character(),
##   Polarity = col_character(),
##   Emotion_Category = col_number(),
##   Personality = col_double()
## )
```

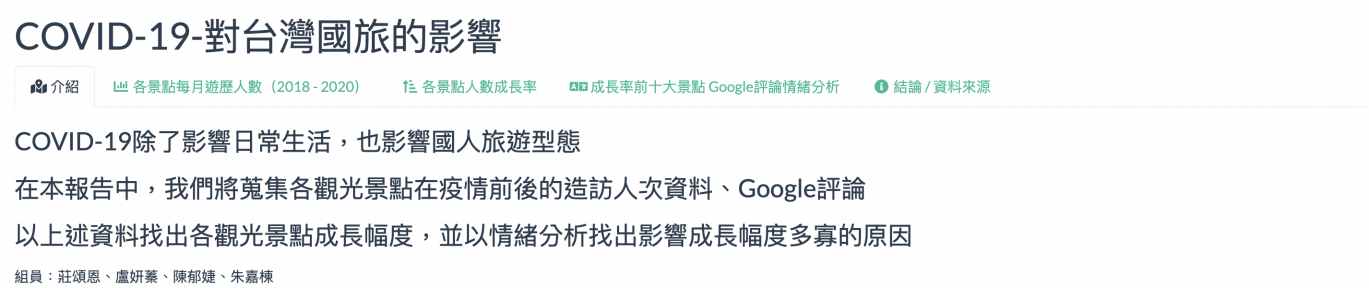
```
ui <- fluidPage(
  theme = shinytheme("flatly"),
  h1("COVID-19-對台灣國旅的影響"),
  # Create a container for tab panels
  tabsetPanel(
      tabPanel(
      title = "介紹",
      h3("COVID-19除了影響日常生活，也影響國人旅遊型態"),
      h3("在本報告中，我們將蒐集各觀光景點在疫情前後的造訪人次資料、Google評論"),
      h3("以上述資料找出各觀光景點成長幅度，並以情緒分析找出影響成長幅度多寡的原因"),
      h3("\n"),
      h5("組員：莊頌恩、盧妍蓁、陳郁婕、朱嘉棟")
    ),

    # Create a "Plot" tab
    tabPanel(
      title = "各景點每月遊歷人數（2018 - 2020）",
      searchInput("place",
                  label = "輸入一個觀光景點(包含片面字詞 / regex)",
                  placeholder = "例: 日月潭, 墾丁, 遊樂園, Zoo",
                  btnSearch = icon("search"),
                  btnReset = icon("remove"),
                  width = "450px"
                  ),
      plotOutput("showplot")
      ),
    # Create "Table" tab
    tabPanel(
      title = "各景點人數成長率",
      selectInput("period", "請選擇時間區段", choices = c("全年度", "暑假（7-8月)", "春假
（3-4月)")),
      selectInput("year", "請選擇年份", choices = c("2018-2019", "2019-2020")),
      DT::dataTableOutput("growth")
    ),
    tabPanel(
      title = "成長率前十大景點 Google評論情緒分析",
      selectInput("site", "請選擇景點", choices = c("北港朝天宮", "虎頭山風景特定區",
                                                    "日月潭風景區", "新竹漁港",
                                                    "鹿野高臺", "六福村主題遊樂園",
                                                    "溪頭自然教育園區", "國立海洋科技博物館",
                                                    "國立科學工藝博物館")),
      plotOutput("sentimentplot")
    ),
    tabPanel(
      title = "結論 / 資料來源",
      h4("在疫情爆發年（2020），台灣實施嚴格邊境管制，以致國際觀光客無法自由往來，使國內3-4月"),
      h4("各觀光景點造訪人次大幅流失。然而隨國內疫情平穩及政府推出的安心旅遊補助政策，國內各大"),
      h4("景點除了恢復觀光熱潮，更相較前兩年呈現大幅的觀光成長。其中室外景點的成長幅度又高於室"),
      h4("內景點。")
    )
  )
)

server <- function(input, output) {
  output$showplot <- renderPlot(
    height = reactive(get_height(place = input$place)*400 + 50),
    width = 1500,
```

```
    {
      req(input$place)
      show_plot(place = input$place)
    }
  )

  growth_data <-  reactive({if (input$period == "全年度") {
    if (input$year == "2018-2019") {
      sort_growth_rate(1:12, 1000000, 2018)
    } else {sort_growth_rate(1:12, 1000000, 2019)}
  } else if (input$period == "暑假（7-8月)" ) {
    if (input$year == "2018-2019") {
      sort_growth_rate(7:8, 300000, 2018)
    } else {sort_growth_rate(7:8, 300000, 2019)}
  } else {
    if (input$period == "春假（3-4月)") {
      if (input$year == "2018-2019") {
        sort_growth_rate(3:4, 150000, 2018)
      } else {sort_growth_rate(3:4, 150000, 2019)}
    }
  }})
  output$growth <- DT::renderDataTable({
    data <- growth_data()
    data
  })

  output$sentimentplot <- renderPlot(
    height = 700,
    width = 1300,
    {
      sentiment_plot(input$site)
      })
}

shinyApp(ui, server)
```

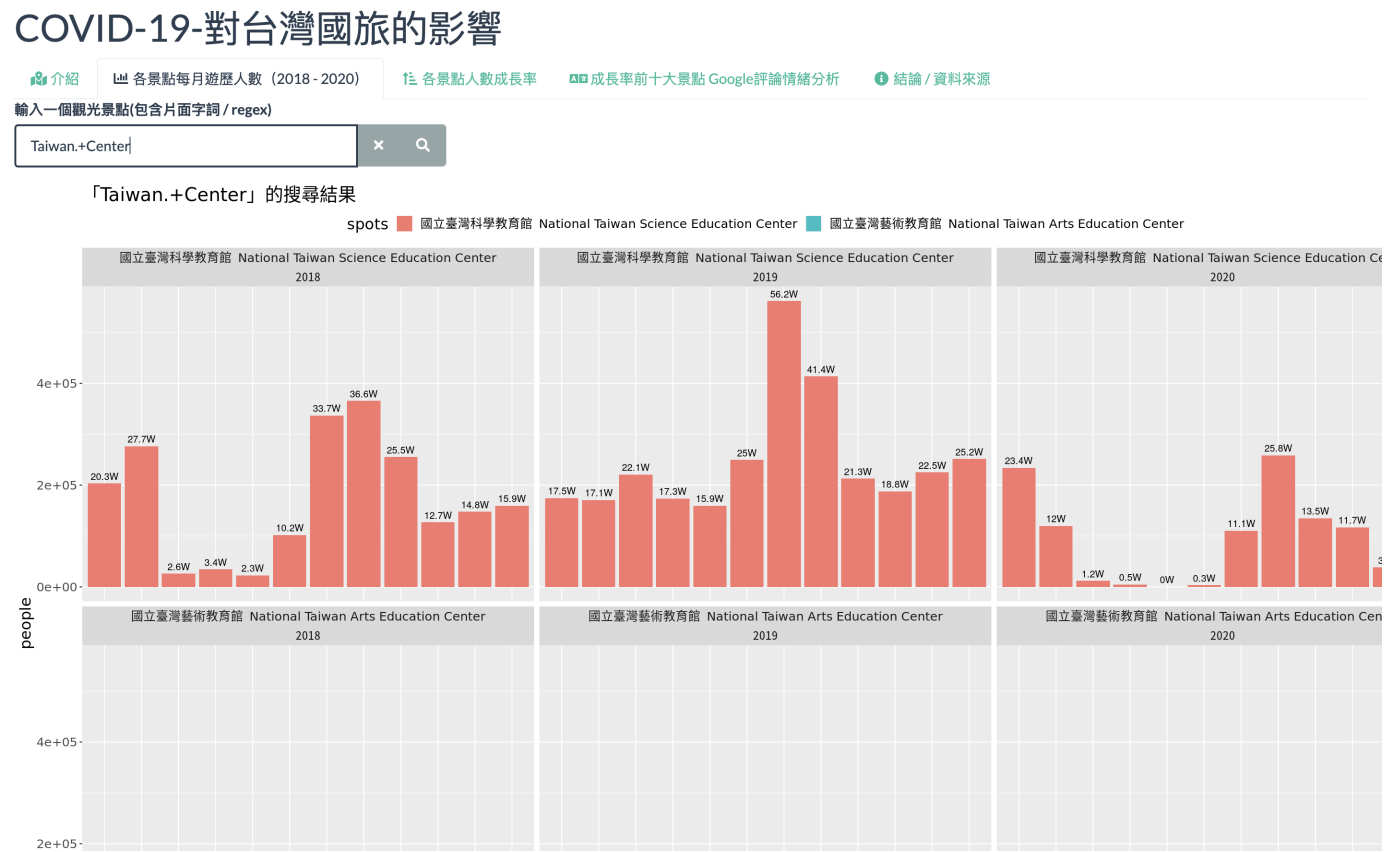
Shiny applications not supported in static R Markdown documents

3. 結果呈現與討論

由於將我們的資料整理好之後，運用Shiny Web App的方式，建立好幾個tab，讓使用者可以自行選擇查詢，得到想要的資訊。 下圖是我們Shiny App的介面截圖：



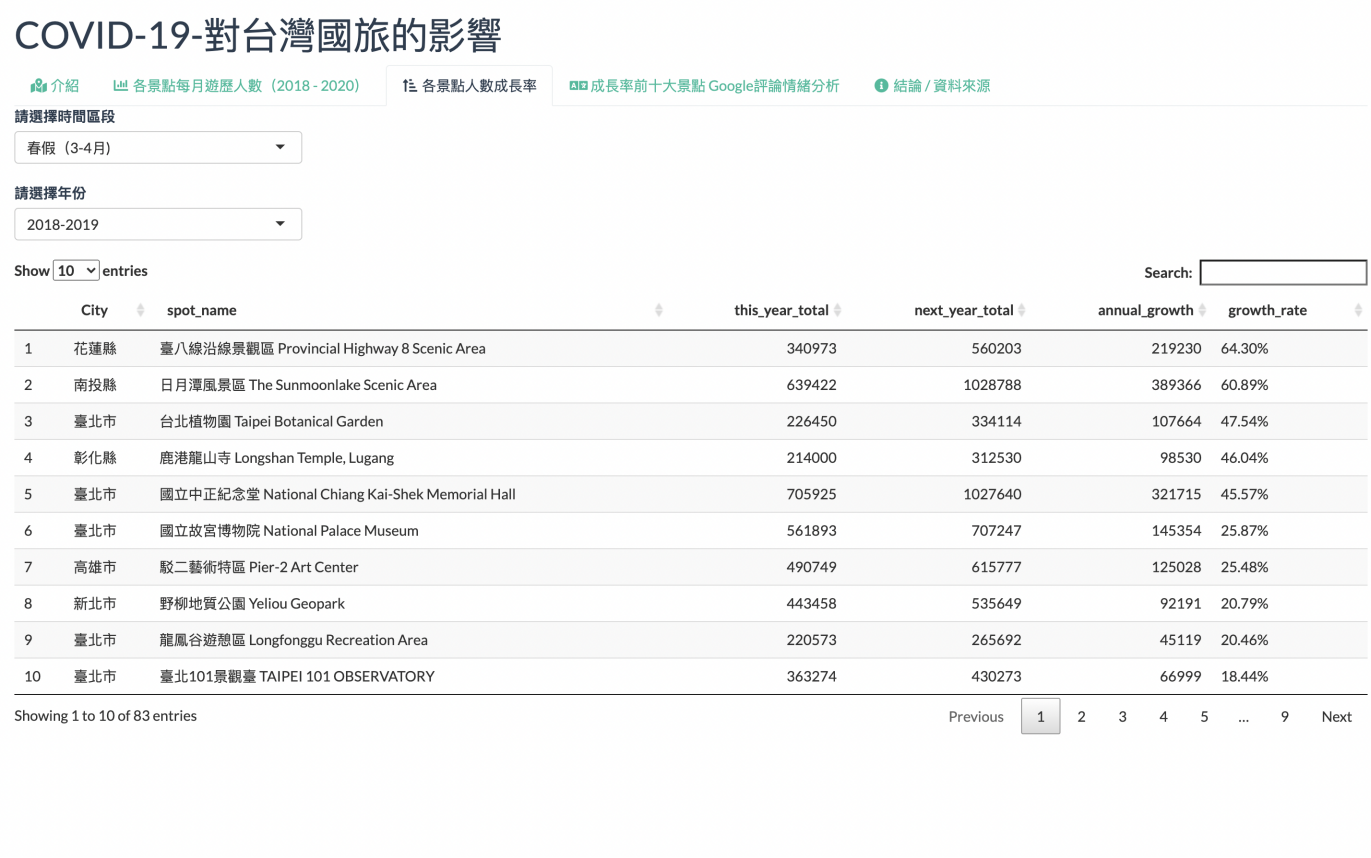
在『各景點每月遊歷人數』這個tab中，使用者可以輸入關鍵字，系統便會提供符合關鍵字的景點資料(2018-2020每個月遊客數)。這裏我們的function是設計成可以接受中文/英文(case-sensitive)/regular expressions查詢，這裏示範使用英文+regex。



下一個tab『各景點人數成長率』，則是透過讓使用者輸入時間區間，和想要查詢的年份，獲得以遊客數成長率作為排名的表單。

首先，這邊在解說一下我們對各個景點所定義的成長率。 *成長率 = 今年度時間區間遊客數 - 去年度時間區間遊客數 / 去年度時間區間遊客數*

這邊由截圖來展示由2018-2019春假假遊客數成長率來排名的結果：



COVID-19-對台灣國旅的影響

介紹 | 各景點每月遊歷人數（2018 - 2020） | 各景點人數成長率 | 成長率前十大景點 Google評論情緒分析 | 結論 / 資料來源

請選擇時間區段

春假（3-4月）

請選擇年份

2018-2019

Show 10 entries Search:

	City	spot_name	this_year_total	next_year_total	annual_growth	growth_rate
1	花蓮縣	臺八線沿線景觀區 Provincial Highway 8 Scenic Area	340973	560203	219230	64.30%
2	南投縣	日月潭風景區 The Sunmoonlake Scenic Area	639422	1028788	389366	60.89%
3	臺北市	台北植物園 Taipei Botanical Garden	226450	334114	107664	47.54%
4	彰化縣	鹿港龍山寺 Longshan Temple, Lugang	214000	312530	98530	46.04%
5	臺北市	國立中正紀念堂 National Chiang Kai-Shek Memorial Hall	705925	1027640	321715	45.57%
6	臺北市	國立故宮博物院 National Palace Museum	561893	707247	145354	25.87%
7	高雄市	駁二藝術特區 Pier-2 Art Center	490749	615777	125028	25.48%
8	新北市	野柳地質公園 Yeliou Geopark	443458	535649	92191	20.79%
9	臺北市	龍鳳谷遊憩區 Longfonggu Recreation Area	220573	265692	45119	20.46%
10	臺北市	臺北101景觀臺 TAIPEI 101 OBSERVATORY	363274	430273	66999	18.44%

Showing 1 to 10 of 83 entries Previous 1 2 3 4 5 ... 9 Next

最後一個tab，則是我們針對2019和2020年暑假（因為根據新聞媒體報導，國旅大爆發在2020暑假達到最高峰），遊客數成長率最高的10個景點，去爬他們的Google Reviews，做情感分析，並找出正負面字彙最常被使用的一些詞彙。

這邊要說的是，因為我們是使用第三方crawler，有免費額度的限制，因此只能爬大約10個景點的Google Reviews。但我們後面還是會討論成長率倒數的景點。

先來看前10名的部分

COVID-19-對台灣國旅的影響

介紹 各景點每月遊歷人數（2018 - 2020） 各景點人數成長率 成長率前十大景點 Google評論情緒分析 結論 / 資料來源

請選擇時間區段

暑假（7-8月）

請選擇年份

2019-2020

Show 10 entries Search:

	City	spot_name	this_year_total	next_year_total	annual_growth	growth_rate
1	雲林縣	北港朝天宮 BeiGang ChaoTian Temple	400600	1106500	705900	176.21%
2	臺中市	麗寶樂園 LihPaoLand	1184536	2166279	981743	82.88%
3	桃園市	虎頭山風景特定區 Houtou Mountain Scenic Area	332860	578788	245928	73.88%
4	南投縣	日月潭風景區 The Sunmoonlake Scenic Area	759187	1190743	431556	56.84%
5	新竹市	新竹漁港 Hsinchu Fishing Port	636757	876265	239508	37.61%
6	臺東縣	鹿野高臺 Luyeh High Terrace	317141	412211	95070	29.98%
7	新竹縣	六福村主題遊樂園 Leofoo Village Theme Park	457430	568887	111457	24.37%
8	南投縣	溪頭自然教育園區 Xitou Nature Education Area	342118	423557	81439	23.80%
9	基隆市	國立海洋科技博物館 National Museum of Marine Science & Technology	510068	607302	97234	19.06%
10	高雄市	國立科學工藝博物館 National Science and Technology Museum	453355	491727	38372	8.46%

Showing 1 to 10 of 49 entries Previous 1 2 3 4 5 Next

我們發現大概可將景點分為三個類別：自然美景、人文歷史、室內知識。

透過針對成長率前10名的景點進行Google Reviews爬蟲，並進行情感分析，結果如下（總共有10個，但這邊就只節選幾個來展示）：

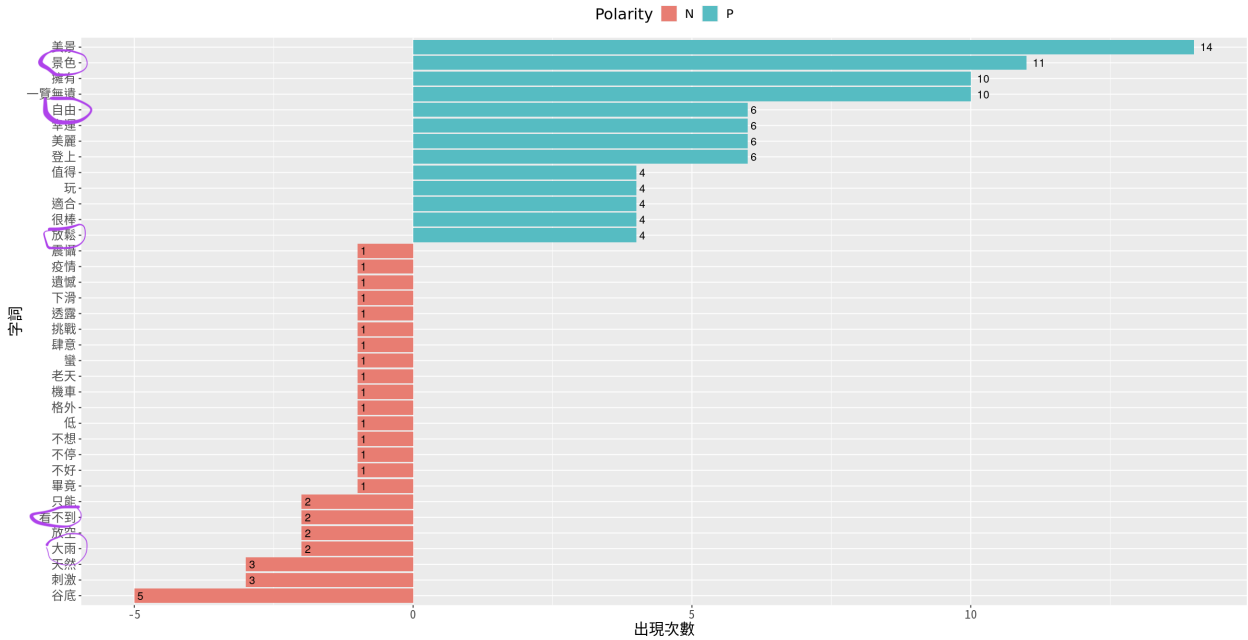
COVID-19-對台灣國旅的影響

介紹 | 各景點每月遊歷人數（2018 - 2020） | 各景點人數成長率 | 成長率前十大景點 Google評論情緒分析 | 結論 / 資料來源

請選擇景點

鹿野高臺

鹿野高臺情緒分析 正向／負向 前十名



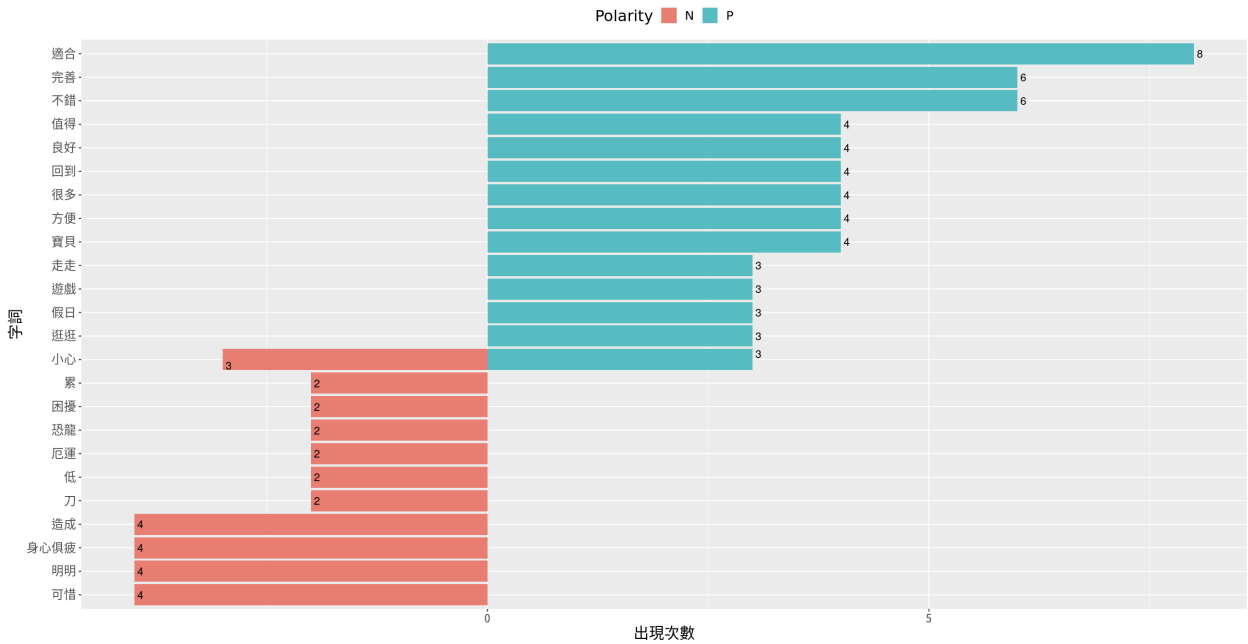
COVID-19-對台灣國旅的影響

介紹 | 各景點每月遊歷人數（2018 - 2020） | 各景點人數成長率 | 成長率前十大景點 Google評論情緒分析 | 結論 / 資料來源

請選擇景點

虎頭山風景特定區

虎頭山風景特定區情緒分析 正向／負向 前十名



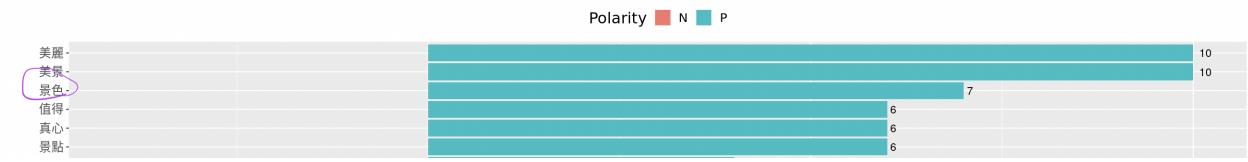
COVID-19-對台灣國旅的影響

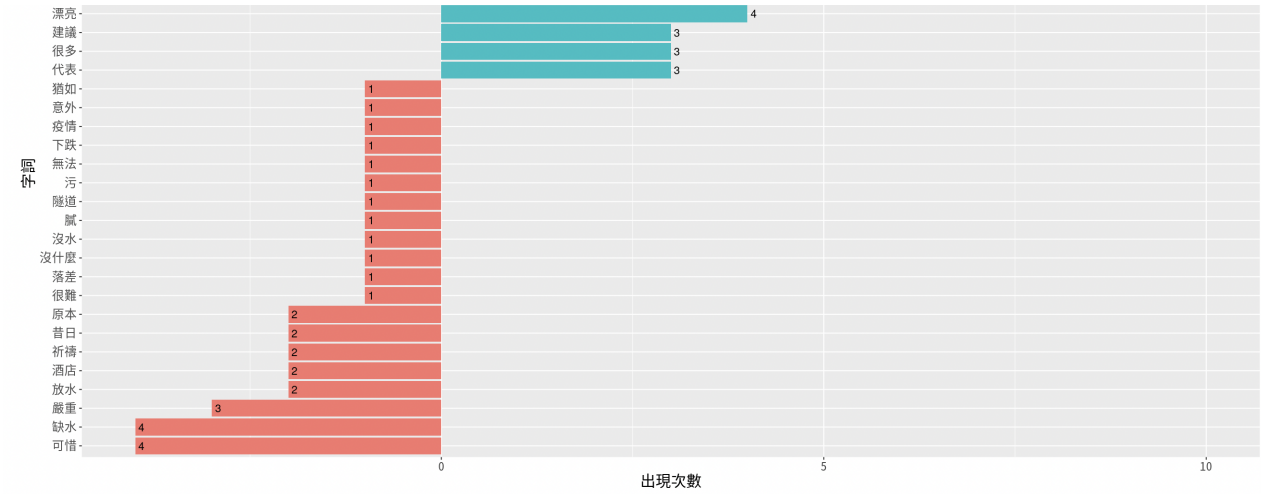
介紹 | 各景點每月遊歷人數（2018 - 2020） | 各景點人數成長率 | 成長率前十大景點 Google評論情緒分析 | 結論 / 資料來源

請選擇景點

日月潭風景區

日月潭風景區情緒分析 正向／負向 前十名




漂亮
建議
很多
代表
猶如
意外
疫情
下跌
無法
污
隧道
膩
沒水
沒什麼
落差
很難
原本
昔日
祈禱
酒店
放水
嚴重
缺水
可惜
字詞
4
3
3
3
1
1
1
1
1
1
1
1
1
1
1
1
2
2
2
2
2
3
4
4
0
5
10
出現次數

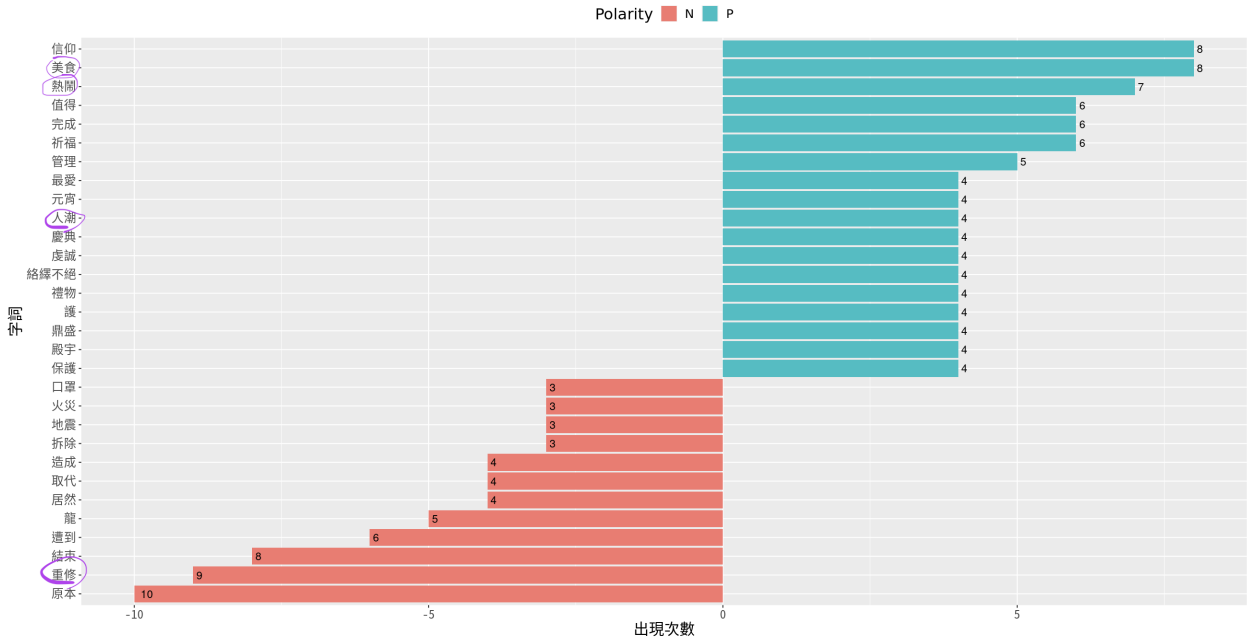
COVID-19-對台灣國旅的影響

介紹 | 各景點每月遊歷人數（2018 - 2020） | 各景點人數成長率 | 成長率前十大景點 Google評論情緒分析 | 結論 / 資料來源

請選擇景點

北港朝天宮

北港朝天宮情緒分析 正向／負向 前十名



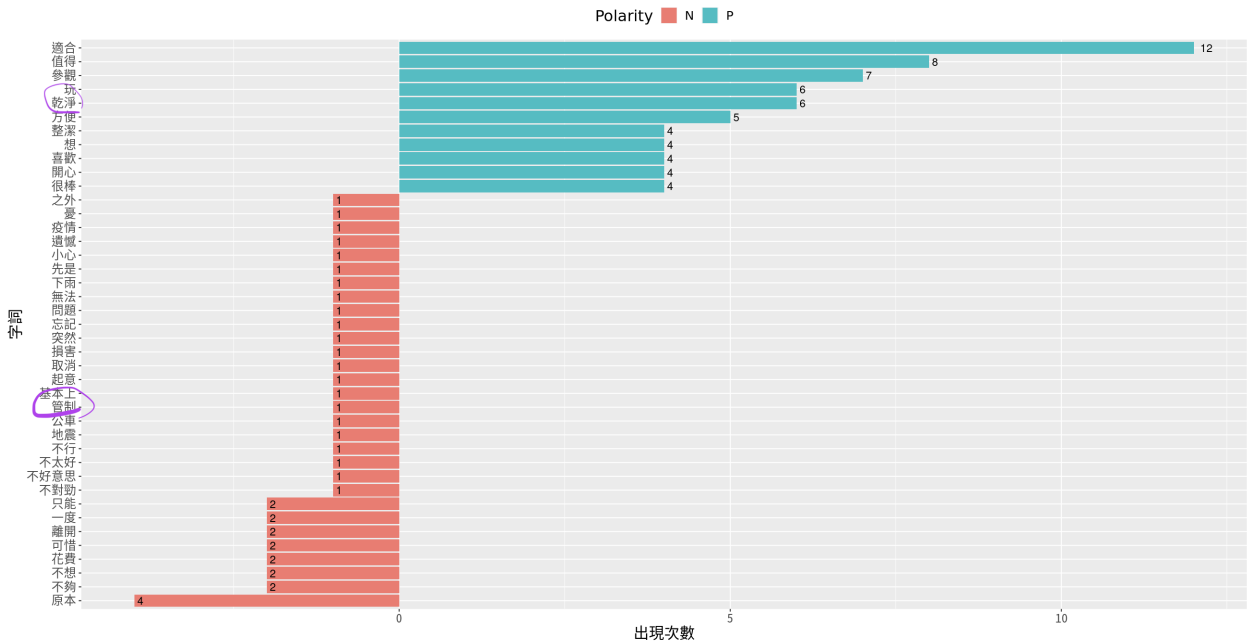
COVID-19-對台灣國旅的影響

介紹 | 各景點每月遊歷人數（2018 - 2020） | 各景點人數成長率 | 成長率前十大景點 Google評論情緒分析 | 結論 / 資料來源

請選擇景點

國立海洋科技博物館

國立海洋科技博物館情緒分析 正向／負向 前十名



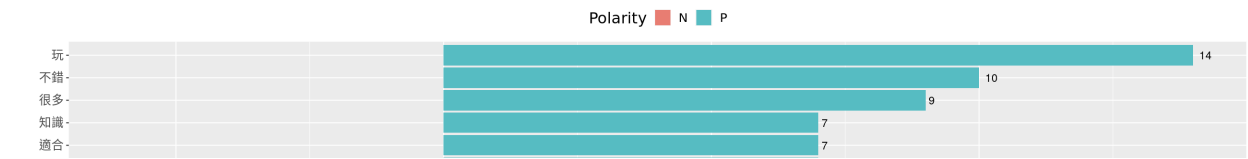
COVID-19-對台灣國旅的影響

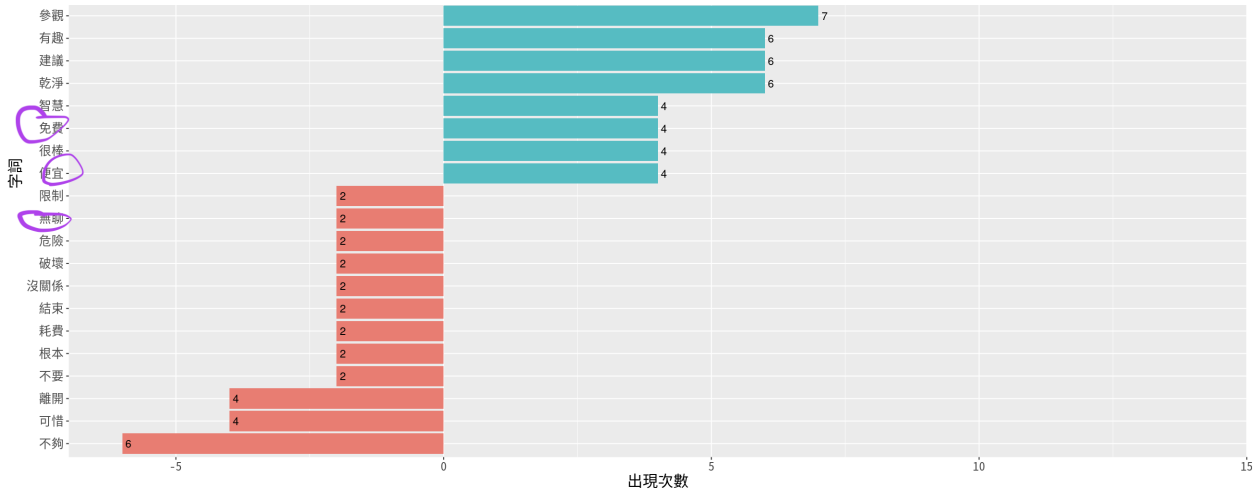
介紹 | 各景點每月遊歷人數（2018 - 2020） | 各景點人數成長率 | 成長率前十大景點 Google評論情緒分析 | 結論 / 資料來源

請選擇景點

國立科學工藝博物館

國立科學工藝博物館情緒分析 正向／負向 前十名





接著，我們將評論中重複出現關鍵字根據正負面意涵，建成景點類型表格如下：

景點類別	自然美景	人文歷史	室內知識
正面詞彙	景色、自由、放鬆	美食、人潮、熱鬧	免費、便宜、乾淨
負面詞彙	下雨、看不到	重修	管制、無聊

可以看出，針對不同類別的景點，遊客們會有不同的正負面觀感，但同一個類型下，大致上有一定類似的正負面意見。

再來是成長率倒數前10名的景點。

Show 10 entries Search:

	City	spot_name	this_year_total	next_year_total	annual_growth	growth_rate
41	臺北市	國父紀念館 National Dr. Sun Yat-Sen Memorial Hall	837950	355490	-482460	-57.58%
42	高雄市	旗津風景區 Cijin Scenic Aera	1214973	497831	-717142	-59.03%
43	臺北市	臺北101景觀臺 TAIPEI 101 OBSERVATORY	412037	162814	-249223	-60.49%
44	臺北市	國立臺灣科學教育館 National Taiwan Science Education Center	975550	368902	-606648	-62.19%
45	臺南市	麻豆代天府 Daitianfu Temple, Madou	657463	220445	-437018	-66.47%
46	新北市	野柳地質公園 Yeliou Geopark	423495	89486	-334009	-78.87%
47	高雄市	佛光山 Foguangshan	1081677	227758	-853919	-78.94%
48	臺北市	國立中正紀念堂 National Chiang Kai-Shek Memorial Hall	848970	146100	-702870	-82.79%
49	臺北市	國立故宮博物院 National Palace Museum	692871	111760	-581111	-83.87%

Showing 41 to 49 of 49 entries Previous 1 2 3 4 5 Next

成長率倒數前10名的景點

但就如前面所述，由於我們是使用第三方crawler軟件，在crawl完前10名景點的Google Reviews之後，就已將免費使用的額度用盡，因此無法去取得倒數10名的Google Reviews和相關分析。

因此我們主觀判斷這些景點主要是仰賴外國遊客。但因為疫情，沒有外國遊客，因此遊客數急降。但這個論點有需要更多資料去佐證。

總結一下，透過觀光局提供的資料分析，我們得知因為過內疫情平緩的緣故，台灣在2020暑假國旅大爆發，有些景點因此受惠。透過Google Reviews爬蟲，和情感分析，我們大致可以了解民眾為何喜歡去這些景點，也同時了解民眾不喜歡的點。另外，雖然沒有很直接的資料佐證，但我們猜測觀光人數成長率最差的景點，或許都是主要依賴外國觀光客，因此觀光人數驟降。

4. 分工狀況

工作內容	負責人

工作內容	負責人
集思廣義、擬定方向、結果討論	每個人
清理整理資料、提供技術支援	朱嘉棟
爬蟲Google Reviews	盧妍蓁＆陳郁婕＆朱嘉棟
Sentiment Analysis	盧妍蓁＆陳郁婕＆朱嘉棟
Shiny Web App	盧妍蓁＆陳郁婕＆朱嘉棟
口頭＆書面報告	莊頌恩