

# 한국에 방문한 외국인 데이터셋 분석



팀원 : 김현석, 이세현, 이수진

# Contents

---

001	002	003	004
데이터소개	방문 국가	상위 5개국 분석	회고
<ul style="list-style-type: none"><li>▪ 데이터 형성 목적</li><li>▪ 데이터 셋</li></ul>	<ul style="list-style-type: none"><li>▪ 60개국</li><li>▪ 상위 5개국 선정</li></ul>	<ul style="list-style-type: none"><li>▪ 예측 및 원인 제시</li><li>▪ 결론</li></ul>	<ul style="list-style-type: none"><li>▪ 데이터셋에 대해</li><li>▪ 프로젝트 결과 도출에 대해</li><li>▪ 프로젝트 진행과정과 마무리에 대해</li></ul>

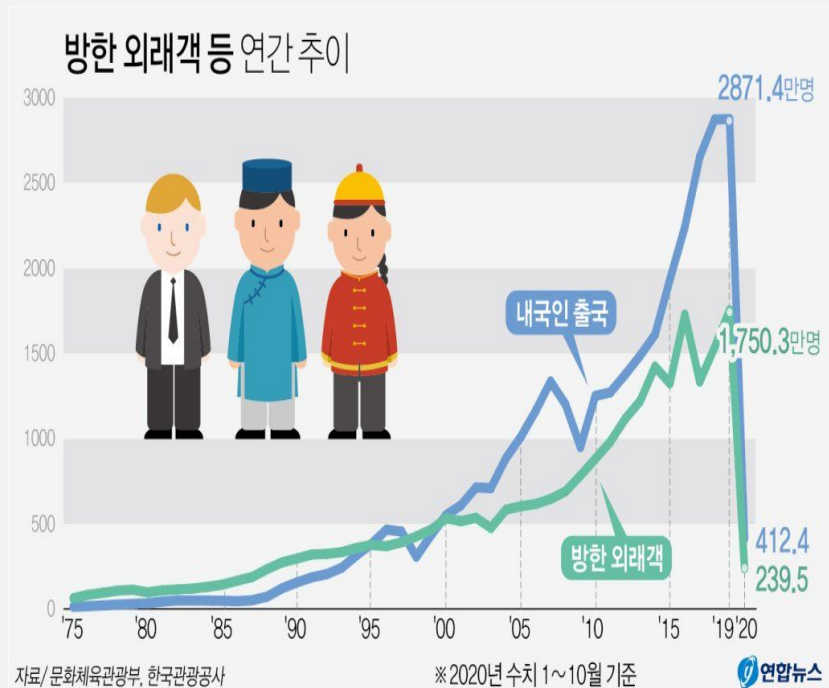
# 001

---

## 데이터 소개



# 연도별 방한 추이

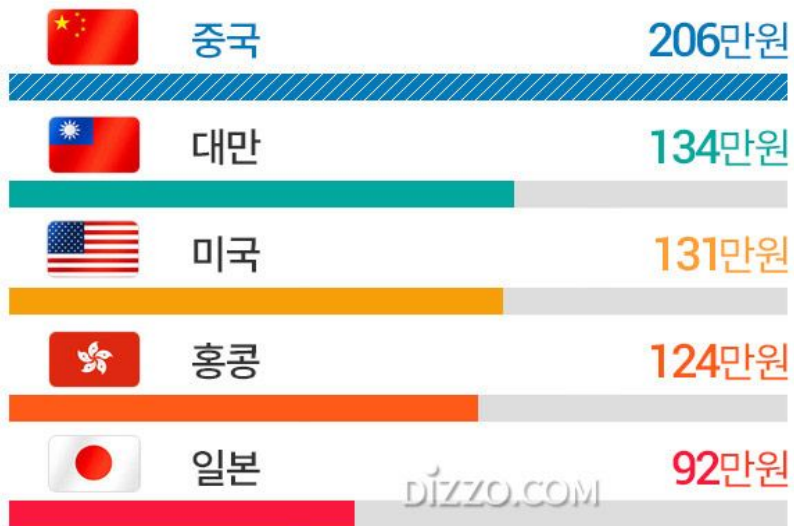


김계리 인턴 / 20201215 트위터 @yonhap\_graphics 페이스북 luney.kr/LeYN1

외국인이 다양한 연령과 국가에서 여러 목적으로 입국을 해오고 있습니다. 이로 인해서 얻어지는 여러 경제적인 효과를 무시할 수는 없을 것입니다.

# 데이터 생성 목적

Q 주요 국가별 외국인 1인 평균 지출 경비는?



그래픽 DIZZO

자료 문화체육관광부, 한국관광공사

- 방한 외국인을 위한 마케팅 전략 구축
- 최대 기대효과 구축

	date	nation	visitor	growth	share	age0-20	age21-30	age31-40	age41-50	age51-60	age61
0	2019-1	China	392814	28.737870	35.555117	36520	108591	103657	48574	40893	40998
1	2019-1	Japan	206526	23.606830	18.693468	18015	57921	34165	39811	33857	20330
2	2019-1	Taiwan	87954	16.003693	7.961057	18888	17927	18595	18862	8169	4566
3	2019-1	Hong Kong	35896	3.533212	3.249086	3890	11384	7400	5461	4629	2137
4	2019-1	Macao	2570	-12.376406	0.232621	223	1013	762	264	181	92
...	...	...	...	...	...	...	...	...	...	...	...
955	2020-4	Oceania others	13	-97.239915	0.044195	0	0	0	0	0	0
956	2020-4	South Africa	22	-98.101812	0.074792	0	1	4	0	0	0
957	2020-4	Africa others	177	-95.412131	0.601734	3	9	16	6	2	0
958	2020-4	Stateless	3	-95.312500	0.010199	0	0	0	0	0	0
959	2020-4	Overseas Korean	698	-96.540101	2.372939	13	73	72	102	176	262

960 rows × 11 columns

	date	nation	visitor	growth	share	tourism	business	official affairs	studying	others
0	2019-1	China	392814	28.737870	35.555117	320113	2993	138	8793	60777
1	2019-1	Japan	206526	23.606830	18.693468	198805	2233	127	785	4576
2	2019-1	Taiwan	87954	16.003693	7.961057	86393	74	22	180	1285
3	2019-1	Hong Kong	35896	3.533212	3.249086	34653	59	2	90	1092
4	2019-1	Macao	2570	-12.376406	0.232621	2506	2	0	17	45
...	...	...	...	...	...	...	...	...	...	...
955	2020-4	Oceania others	13	-97.239915	0.044195	0	0	0	0	13
956	2020-4	South Africa	22	-98.101812	0.074792	2	0	0	0	20
957	2020-4	Africa others	177	-95.412131	0.601734	11	5	0	11	150
958	2020-4	Stateless	3	-95.312500	0.010199	0	0	0	0	3
959	2020-4	Overseas Korean	698	-96.540101	2.372939	0	0	0	0	698

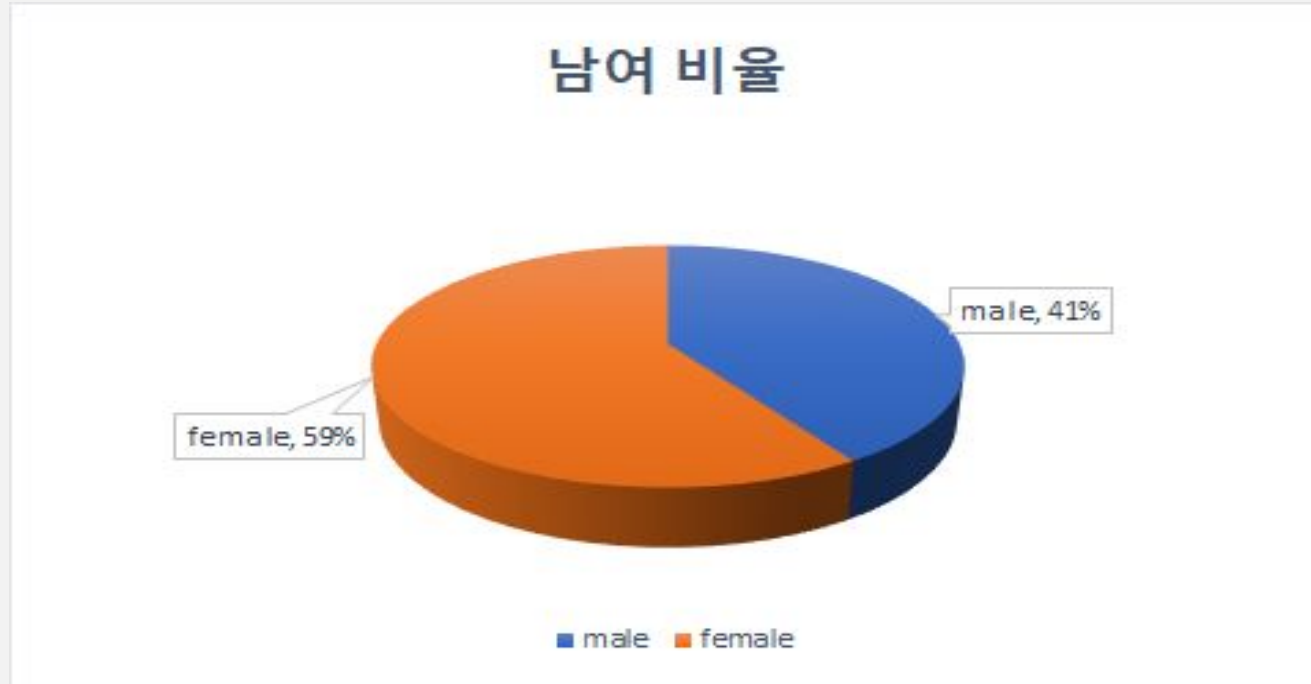
960 rows × 10 columns

	date	nation	visitor	growth	share	male	female	crewman
0	2019-1	China	392814	28.737870	35.555117	147511	231722	13581
1	2019-1	Japan	206526	23.606830	18.693468	75070	129029	2427
2	2019-1	Taiwan	87954	16.003693	7.961057	30805	56202	947
3	2019-1	Hong Kong	35896	3.533212	3.249086	12172	22729	995
4	2019-1	Macao	2570	-12.376406	0.232621	748	1787	35
...	...	...	...	...	...	...	...	...
955	2020-4	Oceania others	13	-97.239915	0.044195	0	0	13
956	2020-4	South Africa	22	-98.101812	0.074792	3	2	17
957	2020-4	Africa others	177	-95.412131	0.601734	30	6	141
958	2020-4	Stateless	3	-95.312500	0.010199	0	0	3
959	2020-4	Overseas Korean	698	-96.540101	2.372939	287	411	0

960 rows × 8 columns



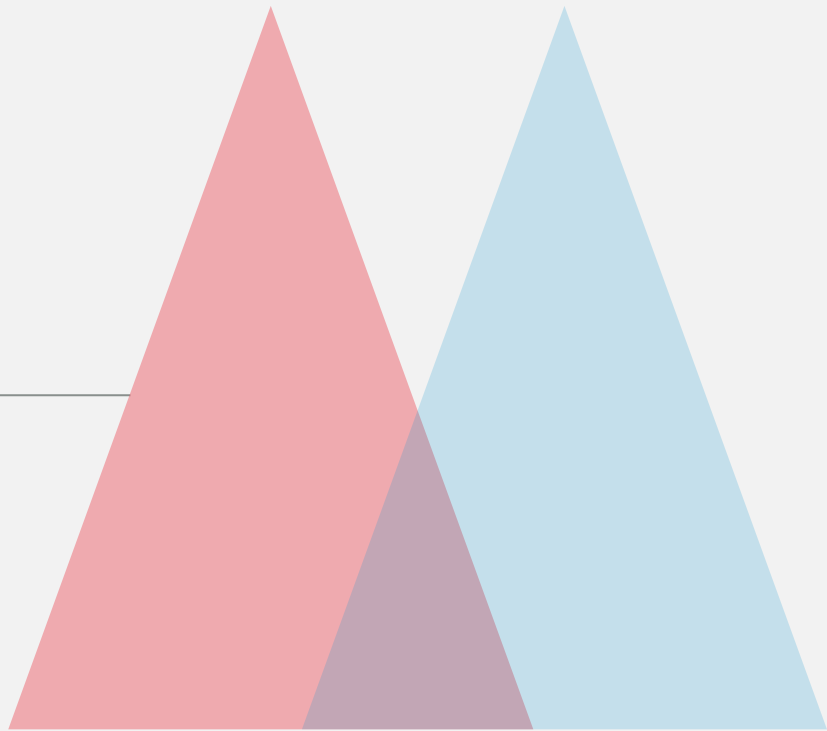
# Percentage of Gender

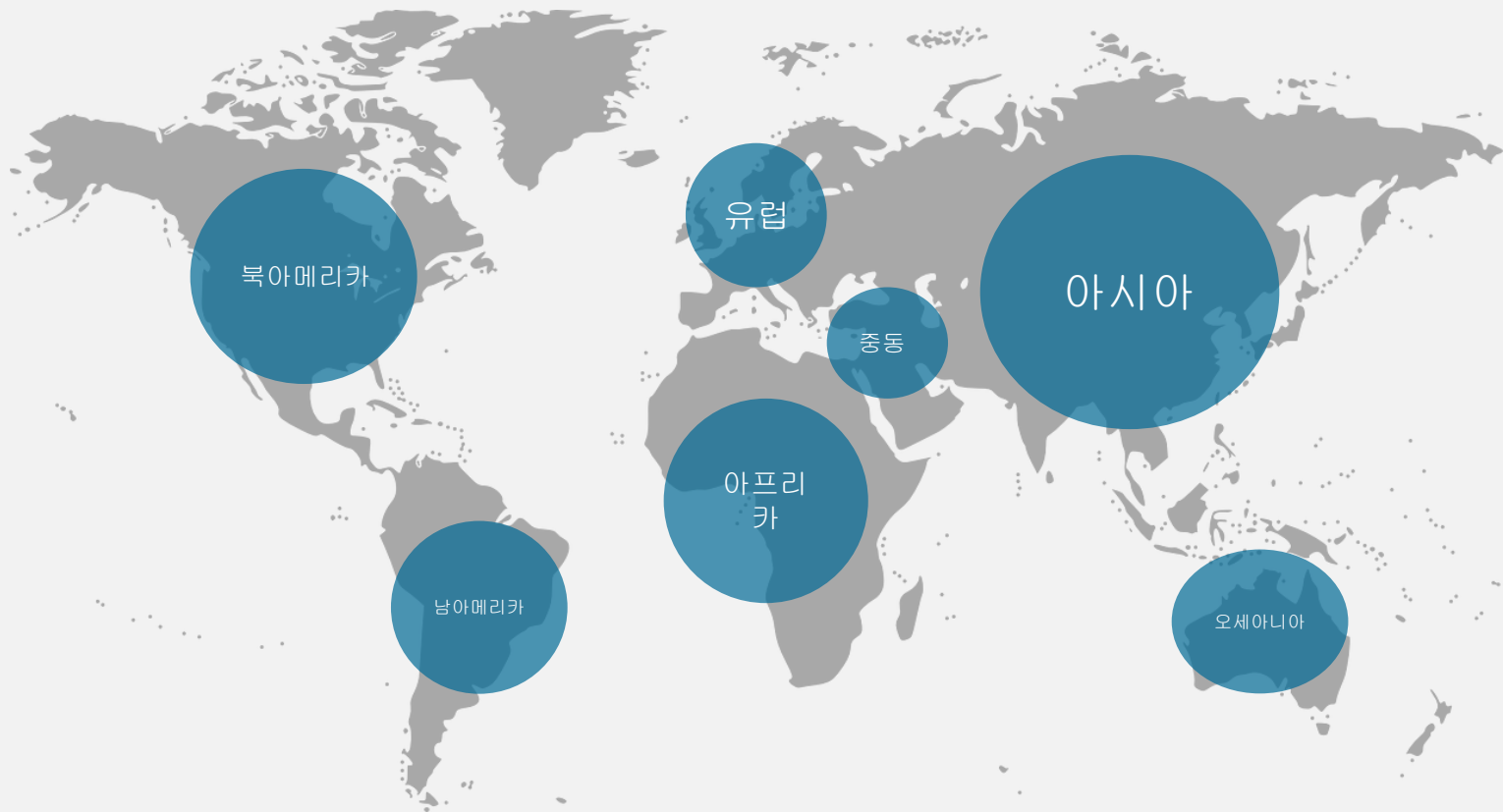


# 002

---

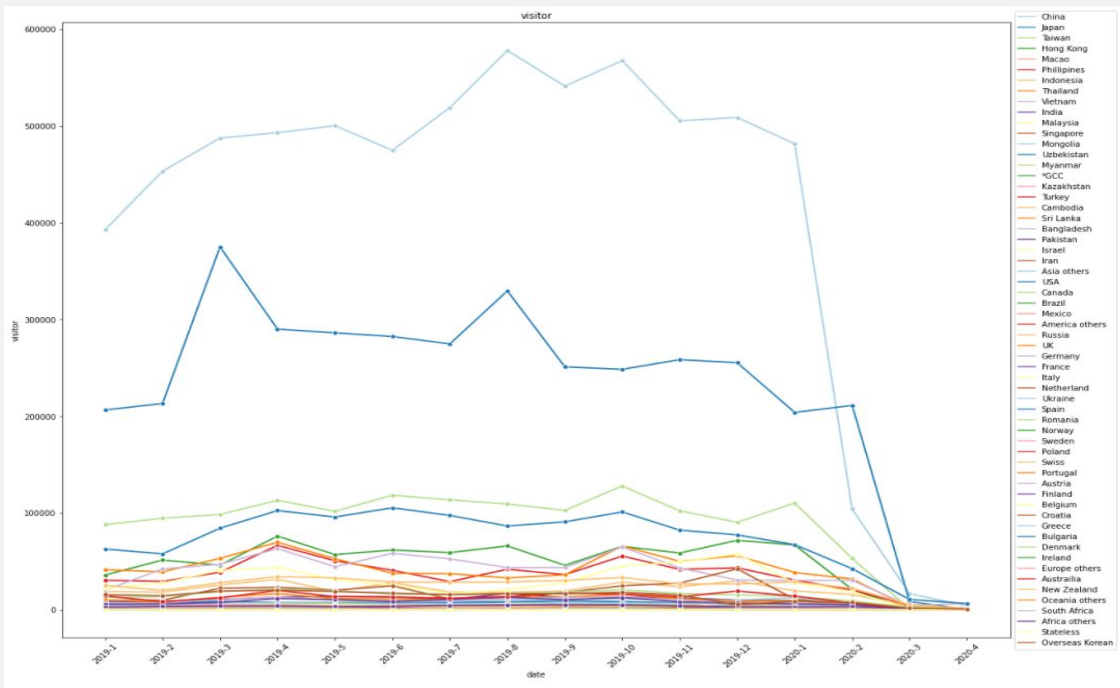
방문 국가





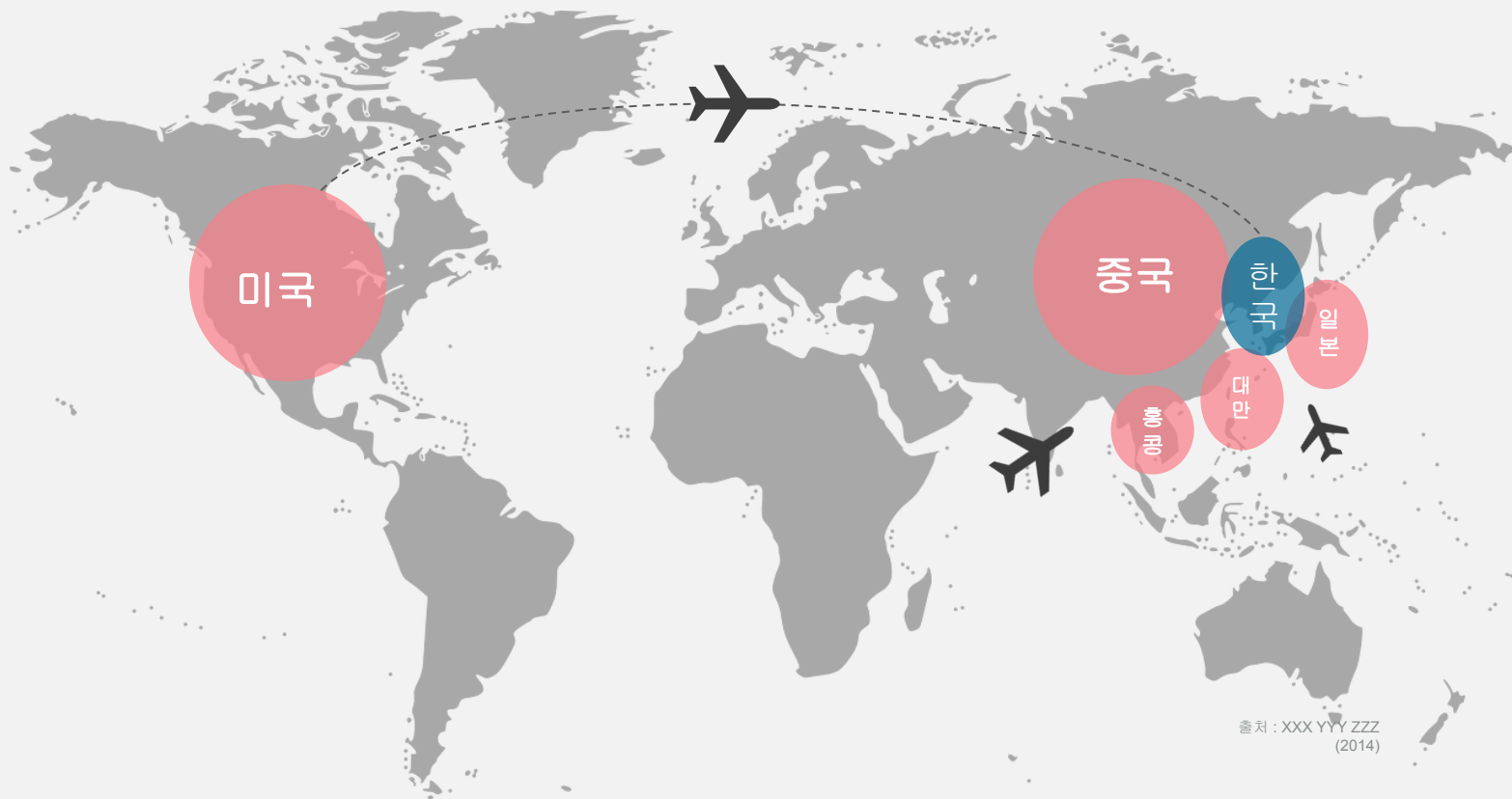
# 세계 60개국

Africa others	America others	Asia others	Australia	Austria	Bangladesh
Belgium	Brazil	Bulgaria	Cambodia	Canada	China
Croatia	Denmark	Europe others	Finland	France	Germany
Greece	Hong Kong	India	Indonesia	Iran	Ireland
Israel	Italy	Japan	Kazakhstan	Macao	
Malaysia	Mexico	Mongolia	Myanmar	Netherland	New
Zealand	Norway	Oceania others	Overseas Korean	Pakistan	Phillipines
Poland	Portugal	Romania	Russia	Singapore	South Africa
Spain	Sri Lanka	Stateless	Sweden	Swiss	Taiwan
Thailand	Turkey	UK	USA	Ukraine	Uzbekistan
Vietnam					

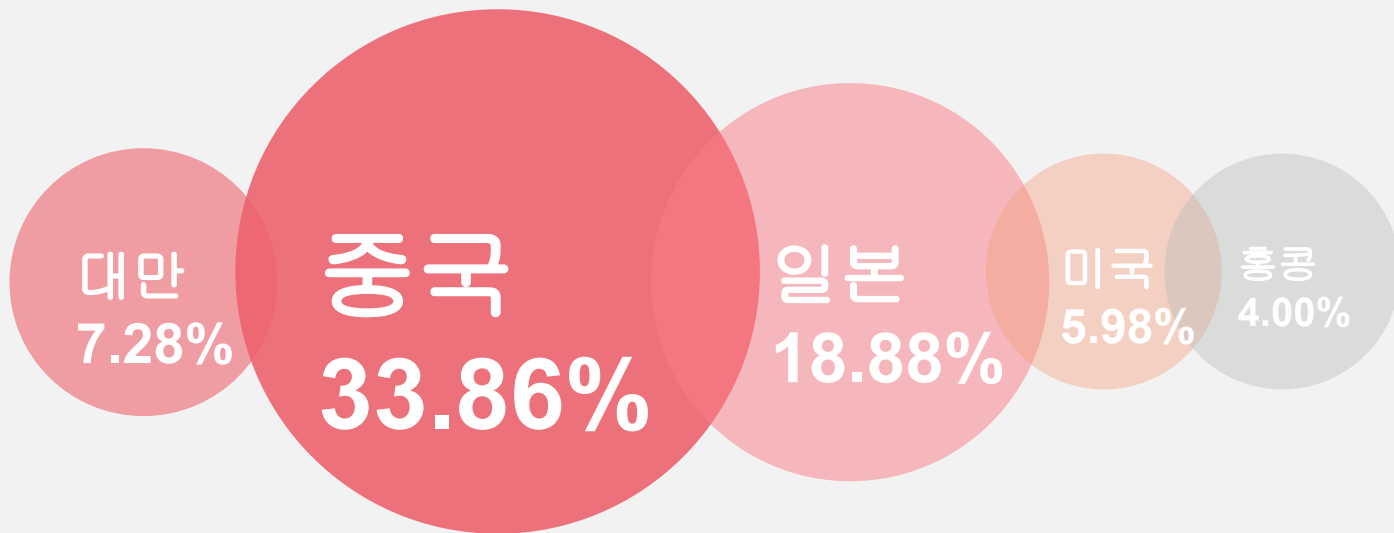


## 60개국 방문자 그래프

## 상위 5개국



## 왜 상위 5개국인가??

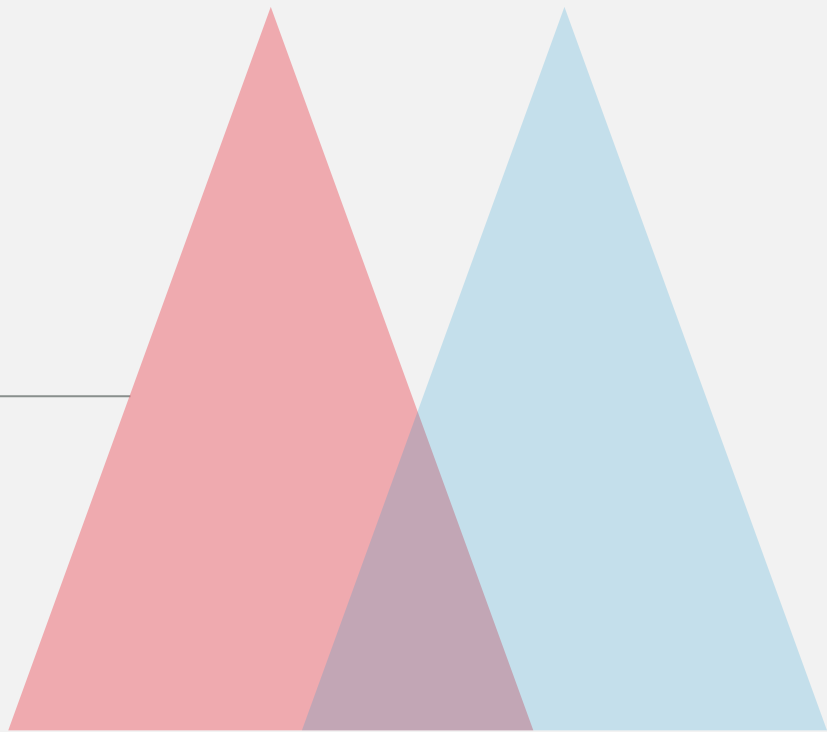


전체 방문객 수의 약 70%를 차지

# 003

---

## 상위 5개국 분석







외국인들의 한국 방문에 영향을 미칠 수 있는 국내외 이슈 조사

상위 5개국 실제 데이터에서 방문자 수 증감 부분 조사

가설과 데이터와의 상관관계

무슨 이유로 특정 기간에 방문이 급증, 급감했을까?



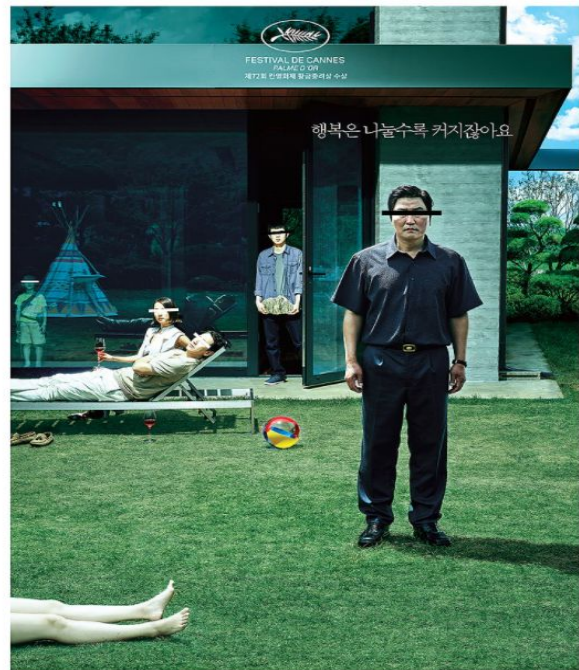
## 북미 정상회담 결렬



이미지 출처 : 전북일보

북미 정상회담 결렬

2019.02.28



송강호 이선균 조여정 최우식 박소담 황제민  
제작/배급 CJ엔터테인먼트 제작 (주)파라다임엔터테인먼트 134.98.0000

기생충

2019 봉준호 감독 작품 | 5월 30일 대개봉

영화 기생충  
칸영화제 황금종려상 수상

2019.05.25

이미지 출처 : 위키백과



이미지 출처 : 위키백과

일본의 수출 규제 및 불매운동

2019.07.01

## BTS 콘서트

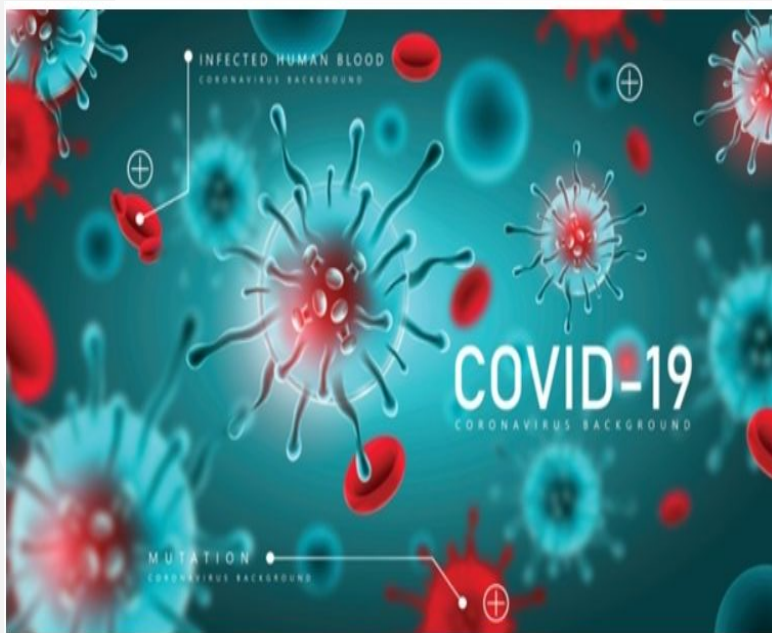


이미지 출처 : 나무위키

BTS 팬미팅 2019.06.15~23

BTS 서울 콘서트 2019.10.26~27, 29

## COVID-19



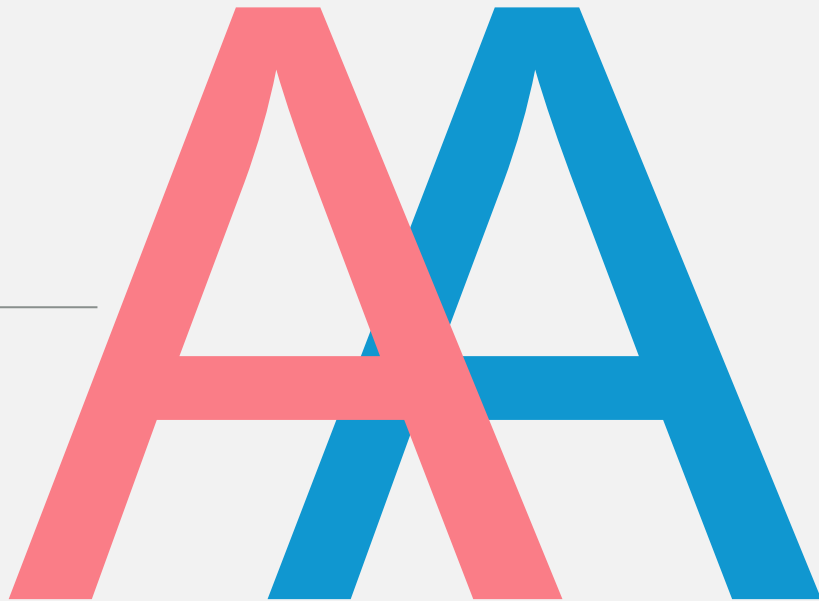
이미지 출처 : 포토파크닷컴

코로나19 유행

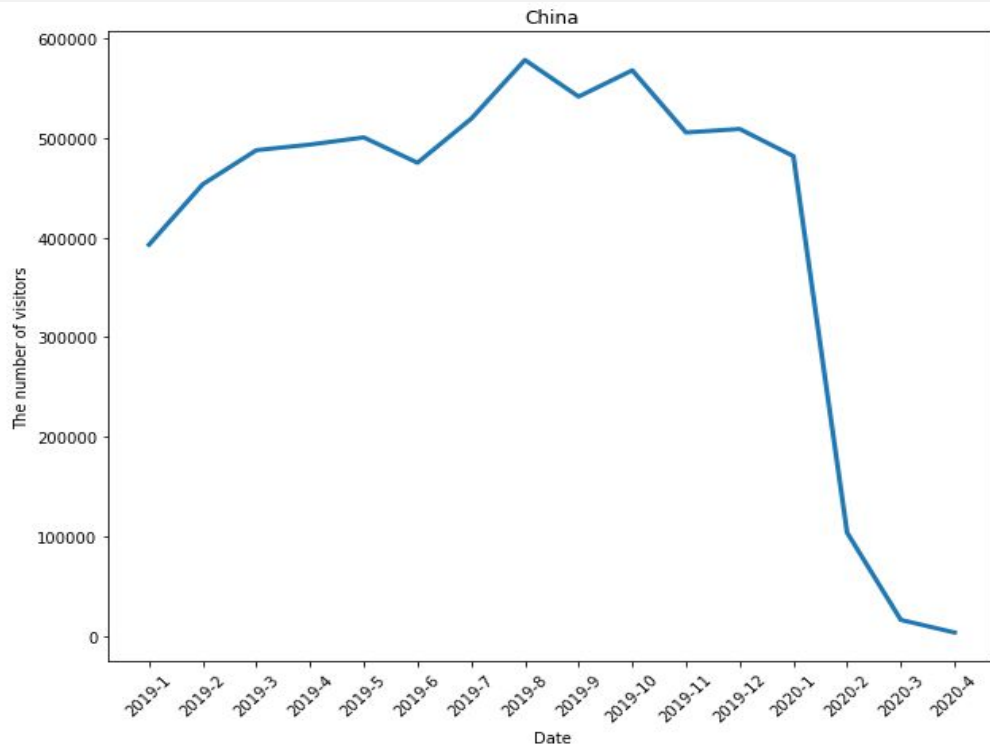
2020.01.21~

---

# 데이터 분석







## 주요 증감 구간 분석

**2019년 6월~8월, 증가:** 여름방학(7월-8월), 휴가

**2019년 8월 말~9월 말, 감소:** 방학, 휴가 종료

**2019년 9월 말~10월 말, 증가:**  
중추절(9월), 국경절(10월)

**2019년 10월~11월, 감소:**  
중추절, 국경절 방문 인원 귀국

**2020년 1월 말 이후, 감소:** 코로나19 유행

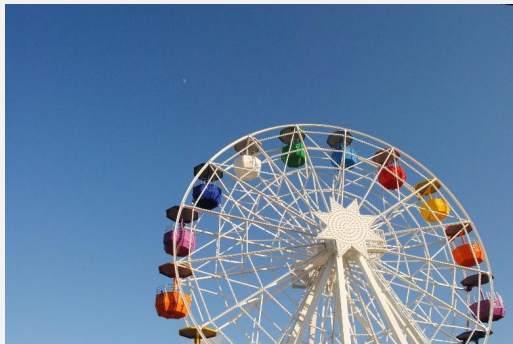
## 원인 도출



### 중국의 공휴일

<http://kr.people.com.cn/n3/2021/0115/c414430-9810061.html>

중국 국무원 판공청은 2021년 공식적인 공휴일을 지정했다. 중국인들은 발표된 내용에 따라 일을 하고 공휴일에는 휴식을 취한다. 한국과 비슷하지만 다른 중국의 2021년 공식 공휴일은 언제인지 알아보자.



### 중국의 여름방학

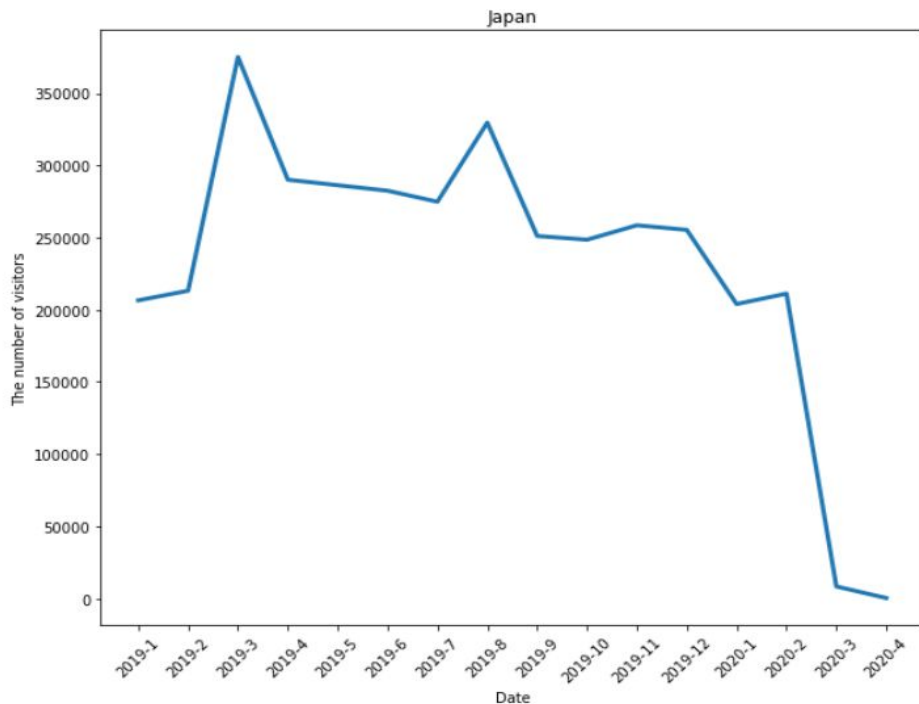
<http://m.joongdo.co.kr/view.php?key=20210713010002877>

여행을 떠난다. 중국에서 '만권의 책을 읽는 것보다 만리의 길을 걷는 것이 낫다'라는 말이 있다. 그래서 부모들이 애들을 데리고 여행을 많이 다니려고 한다. 중국뿐만 아니라 해외로도 많이 나간다.



### 중국의 중추절과 국경절

[https://www.kotra.or.kr/kbc/KTMUIU100M.html?RowCountPerPage=10&SEARCH\\_TRADE\\_CD=&SEARCH\\_TYPE=&SEARCH\\_VALUE=&USER\\_ID=&USER\\_NM\\_ENG=&USER\\_NM\\_KOR=&USER\\_DEPT\\_CD=&USER\\_USER\\_TYPE=&TOP\\_MENU\\_CD=37175&LEFT\\_MENU\\_CD=37189&PARENT\\_MENU\\_CD=&MENU\\_CD=37189&SITE\\_CD=01005&SITE\\_SE\\_CD=1039149&AREA\\_CD=&NATION\\_CD=&TRADE\\_CD=&BBS\\_GUBUN=&BBS\\_ID=&ARTICLE\\_ID=3018314&Page=7](https://www.kotra.or.kr/kbc/KTMUIU100M.html?RowCountPerPage=10&SEARCH_TRADE_CD=&SEARCH_TYPE=&SEARCH_VALUE=&USER_ID=&USER_NM_ENG=&USER_NM_KOR=&USER_DEPT_CD=&USER_USER_TYPE=&TOP_MENU_CD=37175&LEFT_MENU_CD=37189&PARENT_MENU_CD=&MENU_CD=37189&SITE_CD=01005&SITE_SE_CD=1039149&AREA_CD=&NATION_CD=&TRADE_CD=&BBS_GUBUN=&BBS_ID=&ARTICLE_ID=3018314&Page=7)



## 주요 증감 구간 분석

**2019년 2월~3월, 증가:** 항공권 수급 개선 등

**2019년 3월~4월, 감소:** 항공권 수급 개선 등으로 일시적 방문 증가했던 인원 귀국

**2019년 7월~8월, 증가:** 여름방학(7월 말~8월 말), 휴가

**2019년 8월~9월, 감소:** 방학, 휴가 종료, 무역 규제 이후 불매운동 등 관계 악화

**2019년 12월~2020년 1월, 감소:** 법정 공휴일 없는 기간

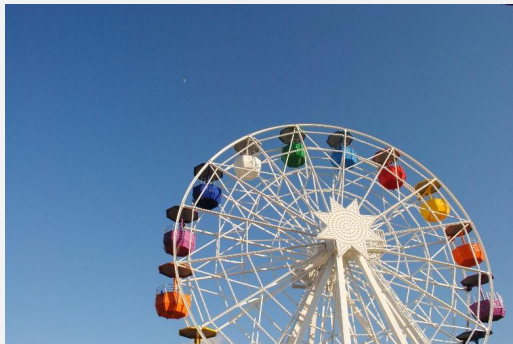
**2020년 2월 이후, 감소:** 코로나19 유행



## 항공권 수급 개선

<https://www.yna.co.kr/view/AKR20190423127900030>

일본인 관광객은 37만5천119명으로, 전년보다 27.4% 증가했다. 이는 3월 입국객 기준으로는 역대 최고 수치이기도 하다. 항공권 수급개선 등으로 30세 이하 젊은 층 관광객이 크게 늘어난 것이 이유로 분석된다.



## 일본의 여름방학

<https://blog.naver.com/dymg98/220869463838>

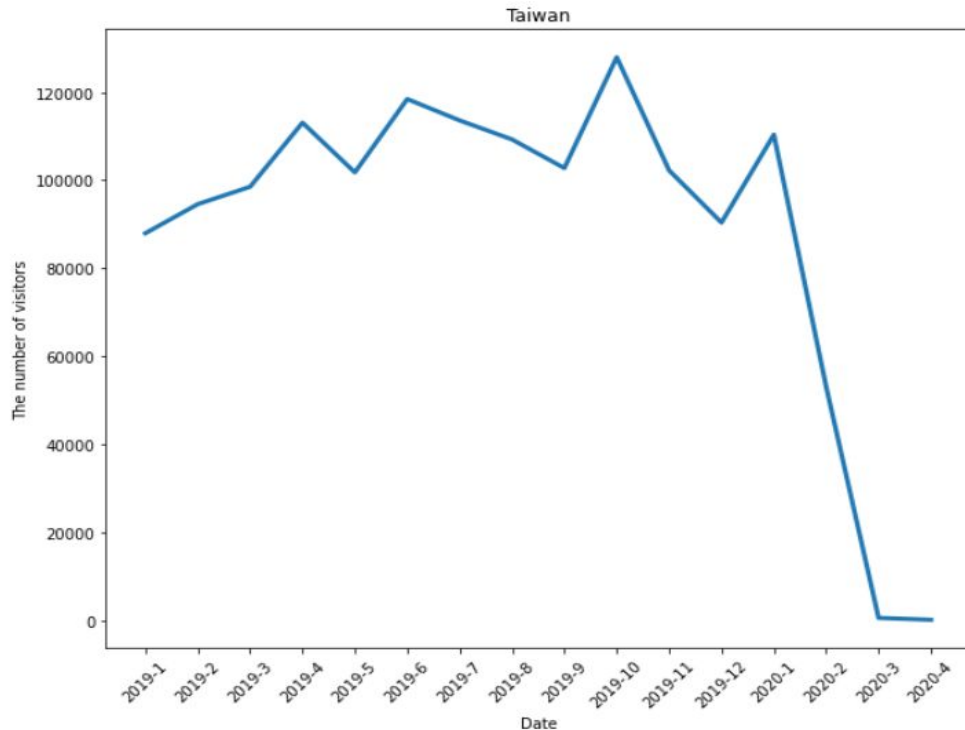
'초/중/고등학교'의 학기를 먼저 살펴보자면, 각 학기의 사이에는 총 세 번의 방학이 있어요. 7월 말~8월 말(약 30일)의 여름 방학, 12월 말~1월 초(약 2주)의 겨울 방학, 3월 말~4월 초(약 2주)의 봄 방학으로 나뉩니다.



## 불매운동

<https://www.sisajournal.com/news/articleView.html?idxno=188130>

일본 정부가 지난 7월1일 한국을 겨냥해 경제 제재를 내리자 시민들은 "강제징용 피해자 배상 판결에 대한 경제보복 조치"라며 "일본 제품 불매로 대응하겠다"고 나섰다.



## 주요 증감 구간 분석

**2019년 9월~10월, 증가:** 추석 등 공휴일(9월, 10월)

**2019년 10월~12월, 감소:** 추석 등 공휴일 방문했던 인원 귀국

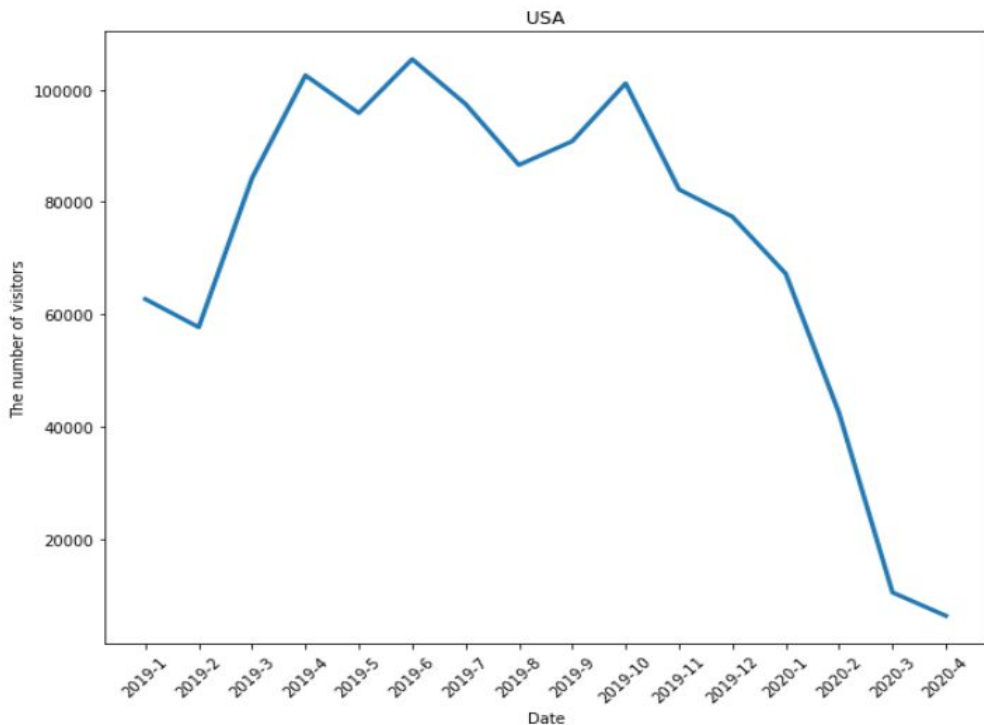
**2019년 12월~2020년 1월, 증가:** 춘절(1월)

**2020년 1월 이후, 감소:** 춘절 방문 인원 귀국, 코로나19 유행

일자	요일	공휴일
1월1일 ~ 1월3일	금요일 ~ 일요일	원단, 신년
2월10일 ~ 2월16일	수요일 ~ 화요일	음력 설
2월27일 ~ 3월1일	토요일 ~ 월요일	228 화평기념일
4월2일 ~ 4월5일	금요일 ~ 월요일	어린이날&청명절
6월12일 ~ 6월14일	토요일 ~ 월요일	단오절
9월18일 ~ 9월21일	목요일 ~ 일요일	추석
10월9일 ~ 10월11일	토요일 ~ 월요일	중화민국국경일 쌍십절
12월31일 ~ 2022년 1월2일	금요일 ~ 일요일	원단, 신년

## 대만의 공휴일

대만은 주5일 근무제로 토요일과 일요일은 휴무이며, 공휴일 및 기념일이 주말과 겹칠 경우 대체 휴무일을 지정해서 휴무한다.



## 주요 증감 구간 분석

**2019년 2월~4월, 증가:** 봄방학(3월, 4월)

**2019년 6월~8월, 감소:** 아래의 유학생들 방학으로 귀국 추정

**2019년 8월~10월, 증가:** 개강으로 학업 목적 인원 증가 관찰됨.

**2019년 10월 이후, 감소:** 법정 공휴일 거의 없는 기간 (10월~12월 말. 여행 목적 인원 감소 관찰됨), 코로나19 유행(2020년 1월 이후. 특히 크게 감소함)

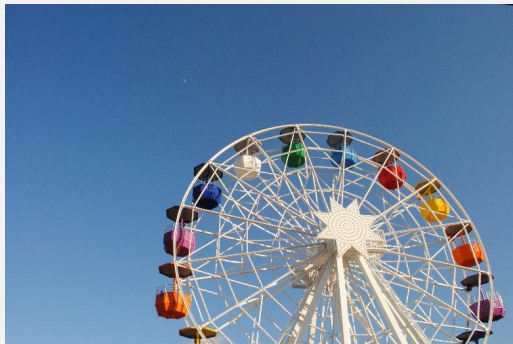
## 원인 도출



### 북미 정상회담 결렬

<https://www.yna.co.kr/view/GYH20190414000200044>

도널드 트럼프 대통령이 3차 북미정상회담의 사전 수순으로서 남북간 대화를 강조하고 이를 통해 비핵화 해법에 관한 김정은 북한 국무위원장의 의중을 확인하고 싶다는 입장을 표명한데 따른 것이다.



### 미국의 봄방학

<https://blackmoney.kr/49>

미국의 방학기간은 한국의 대학의 방학기간과 다릅니다. 우리나라와 다른점은 우리나라는 여름방학, 겨울방학 기간이 비슷한데 비해, 미국은 여름 방학이 훨씬 길합니다. 그리고 개학이 우리는 봄인데 여긴 가을이 개학이에요.

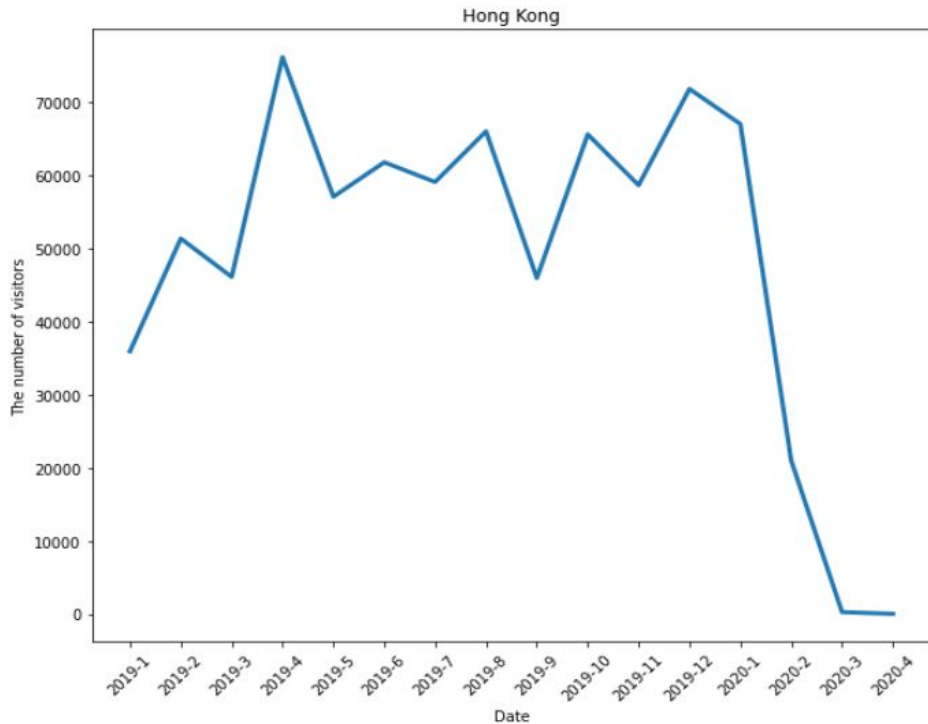


### 미국의 공휴일

<https://kr.usembassy.gov/ko/embassy-consulate-ko/holiday-ko/>

미국에는 연방(전국) 공휴일(federal holiday)이 없다. 모두 50개에 해당하는 미국의 각 주는 각각 자체적으로 공휴일을 정할 권한을 가지고 있기 때문이다.





## 주요 증감 구간 분석

**2019년 3월~4월, 증가:** 공휴일(4월)

**2019년 4월~5월, 감소:** 공휴일 방문 인원 귀국

**2019년 8월~9월 감소:** 공휴일 없음(8월-9월 중순)

**2019년 9월~10월 증가:** 추석, 가을방학, 국경일 등 (10월)

**2019년 12월~1월초 증가:** 연말 및 연초 휴가

**2020년 1월 이후, 감소:** 코로나19 유행

신년	1월1일	월요일
음력 설	2월12일-15일	금요일-월요일
성금요일(Good Friday)	4월2일-3일	금요일-월요일
청명절	4월4일	일요일
Easter Monday	4월5일	월요일
노동절	5월1일	토요일
부처님 오신 날	5월19일	수요일
단오절	6월14일	월요일
홍콩특별행정구 성립 기념일	7월1일	목요일
추석 이튿날	9월22일	수요일
국경절	10월1일	금요일
중양절	10월14일	목요일
크리스마스	12월25일	토요일
박싱데이	12월27일	월요일

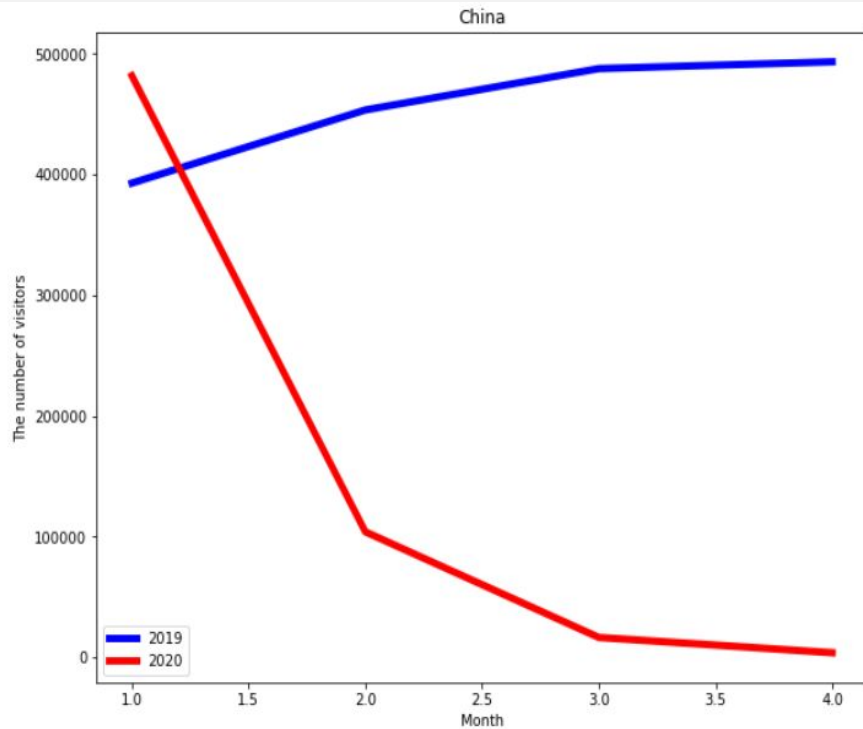
## 홍콩의 공휴일

홍콩은 1997년 중국에 반환되기 전까지 장기간 영국의 통치 하에 있었기 때문에 홍콩에는 동서양의 풍습이 혼재되어 있다. 공휴일도 마찬가지로 동서양의 공휴일이 혼재되어 휴무한다.

---

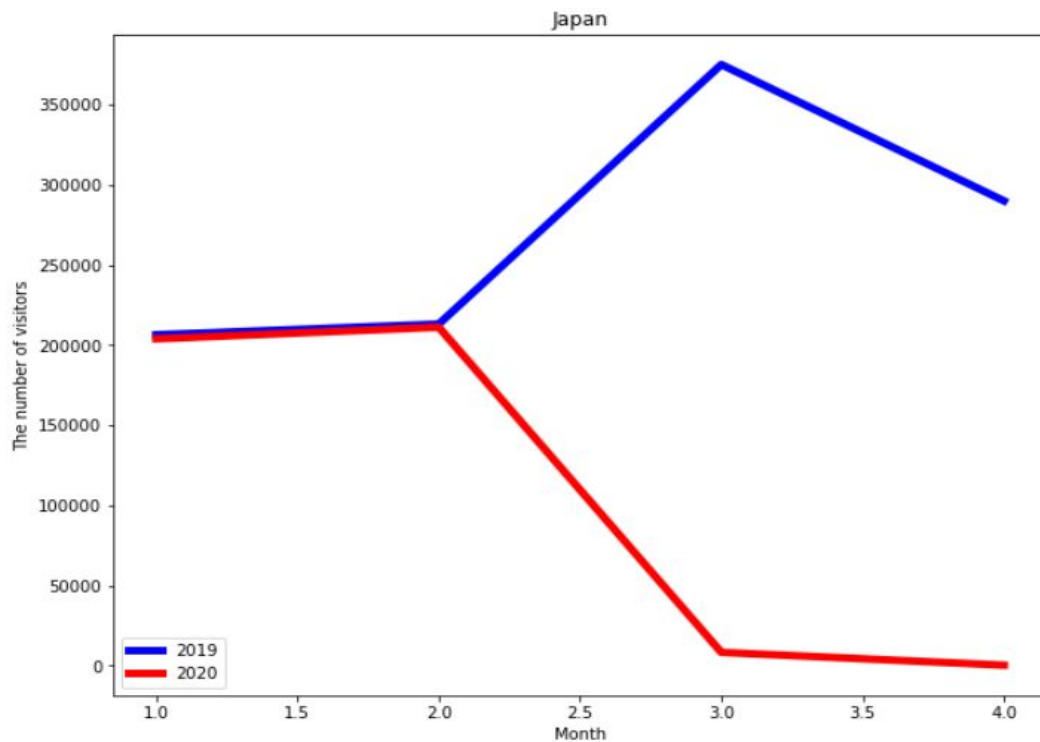
코로나의 영향





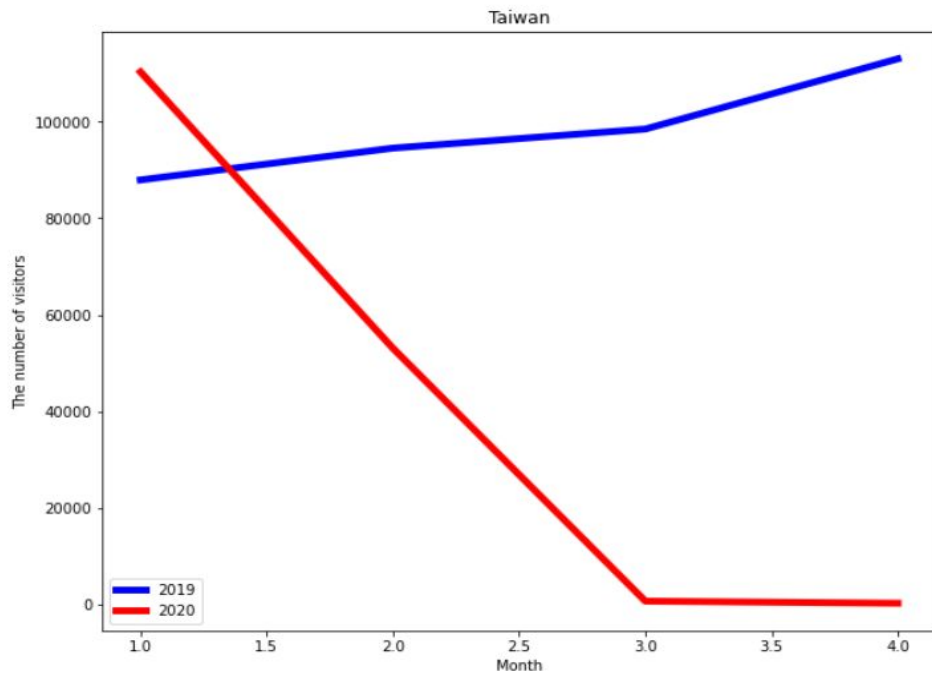
## 코로나가 미친 영향력

2020년 1월 말 이후, 감소: 코로나19 유행



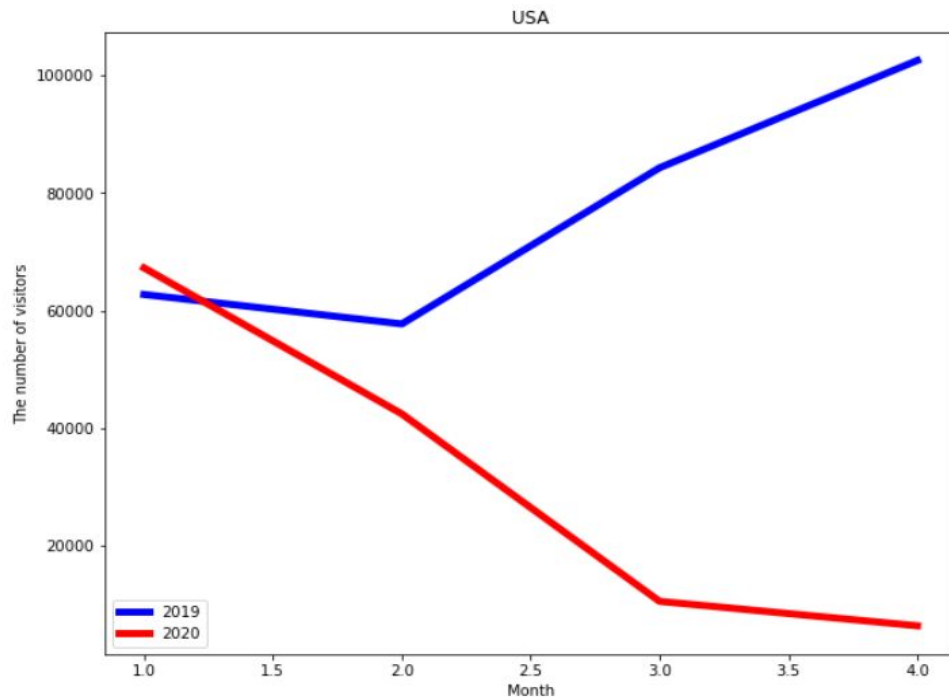
## 코로나가 미친 영향력

2020년 2월 이후, 감소: 코로나19 유행



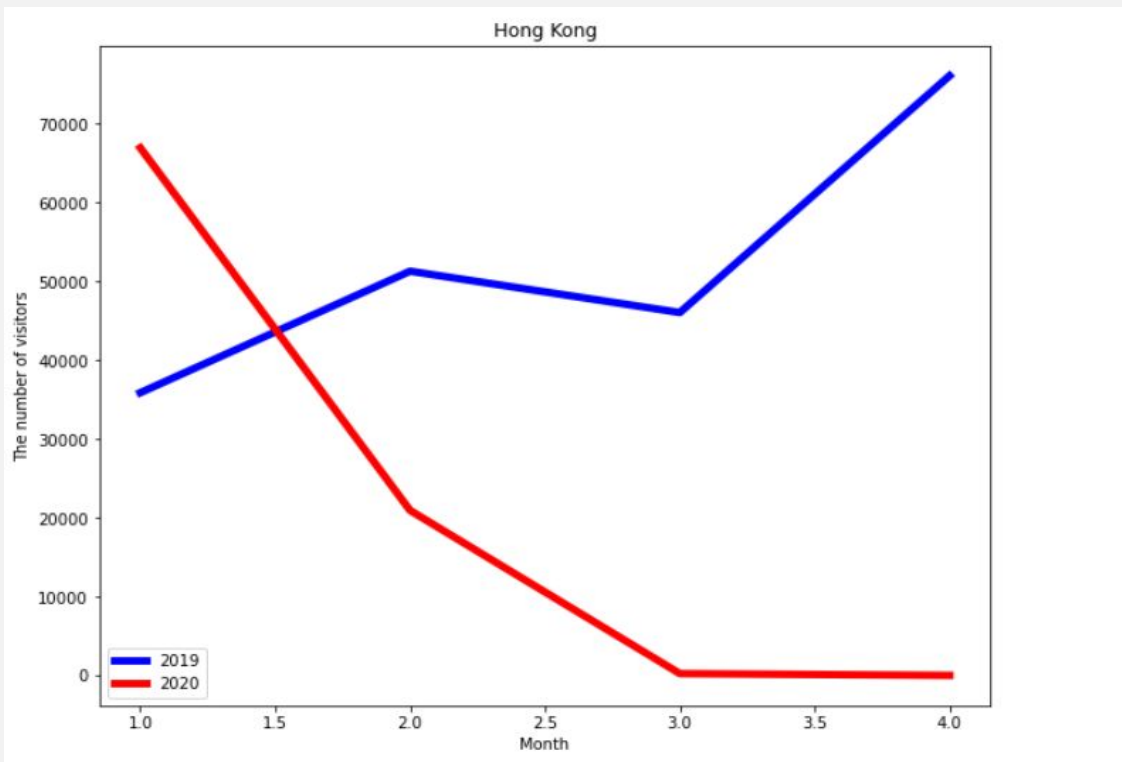
## 코로나가 미친 영향력

2020년 1월 이후, 감소: 코로나19 유행



## 코로나가 미친 영향력

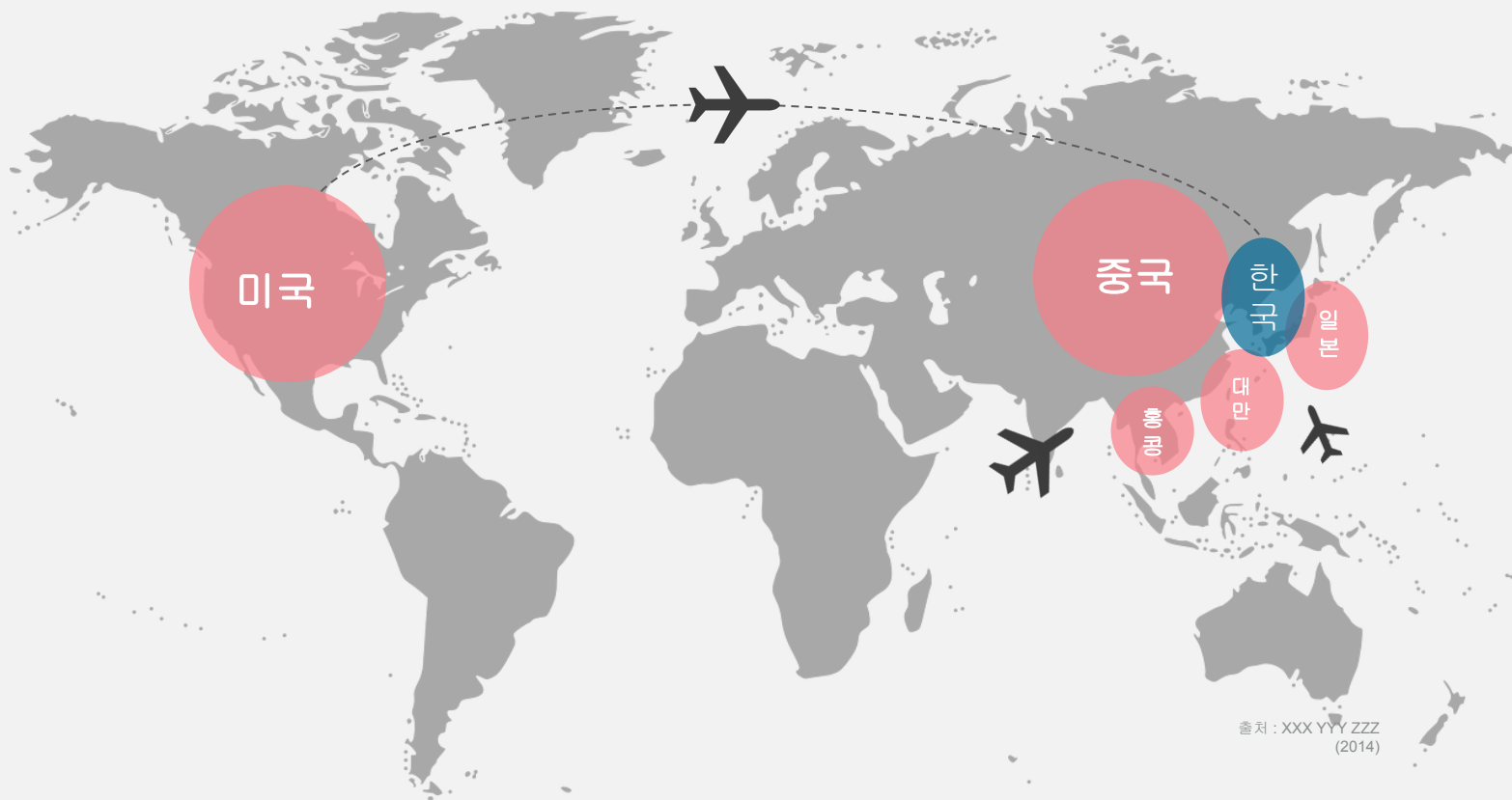
2020년 1월 이후, 감소: 코로나19 유행



## 코로나가 미친 영향력

2020년 1월 이후, 감소: 코로나19 유행





출처 : XXX YYY ZZZ  
(2014)

# Insight



## 결론

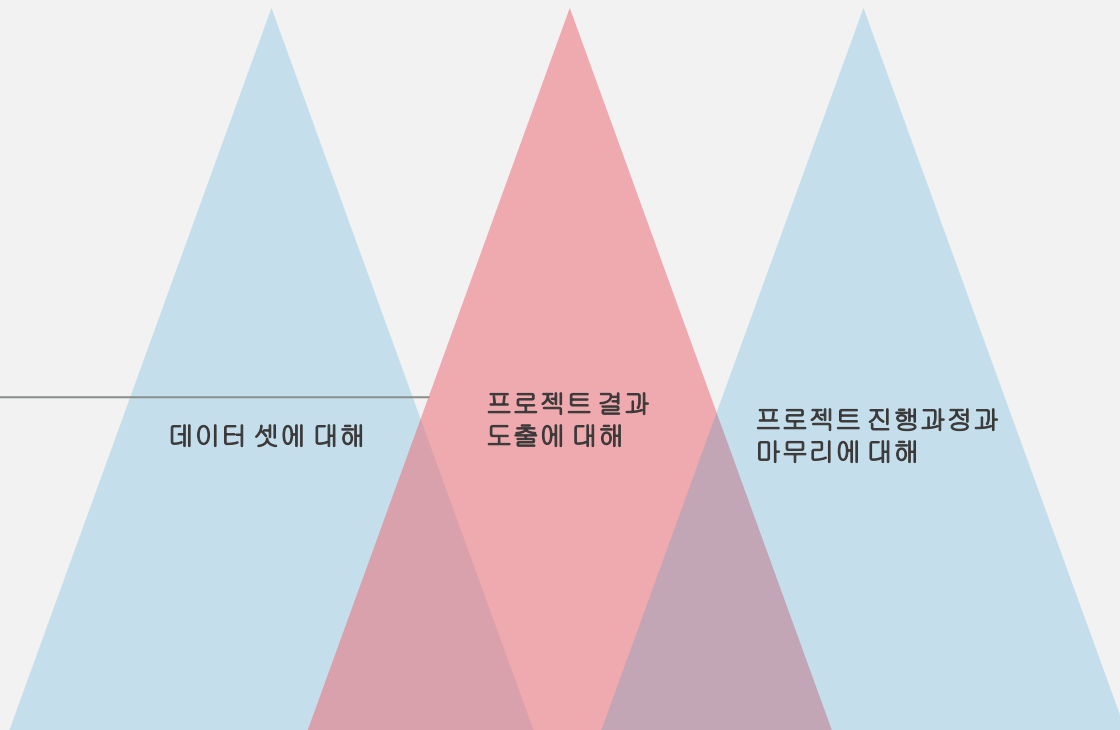
데이터셋이 원하는 마케팅 전략의 우선순위를 도출했습니다.

마케팅 전략

1. 공휴일 + 이슈
2. 공휴일
3. 이슈

# 004

## 회고



# 데이터 셋에 대해

## ·자료 구성

이 데이터 셋은 **date, nation, visitor, growth, share**를 모든 파일의 공통 항목으로, **age, gender, purpose**를 각 파일 별 세부 항목으로 함.

## ·좋았던 점

파일 별 공통 항목을 가지고 있어 이용 목적에 따라 세부 항목과 연계하여 분석이 가능.

공통 항목 자체 분석: **date**에 따른 추이, **nation**별 추이 등.

공통 항목끼리 연계 분석: **date-nation, date-visitor, nation-growth** 등.

세부 항목 자체 분석: **age, gender** 등 항목에 대한 분석.

공통 항목-세부 항목 연계 분석: **date-age, nation-gender** 등.



데이터 셋에 대해

## ·아쉬웠던 점

데이터의 구축 목적이 외국인 방문자 대상 관광·마케팅 정책 수립이라는 측면에서 볼 때,

나이, 성별, 방문 목적 항목들을 서로 연계하여 분석, 도출 작업을 할 수가 없었음.

세부 지역별 데이터가 없어서 방문객들이 구체적으로 어느 지역을 선호해서 가는지 알 수 없음. 연령, 성별, 방문 목적 등의 데이터가 세부 지역별로 연계될 수 있었다면 지역별 관광, 마케팅 정책 수립에 더욱 도움이 되었을 것.

1개월 단위의 시계열 데이터라 상대적으로 더 짧은 기간에 집중된 데이터가 표현되지 않는 한계가 있음. **ex)** 주말에 집중되었던 **BTS** 콘서트로 인한 방문객 등.



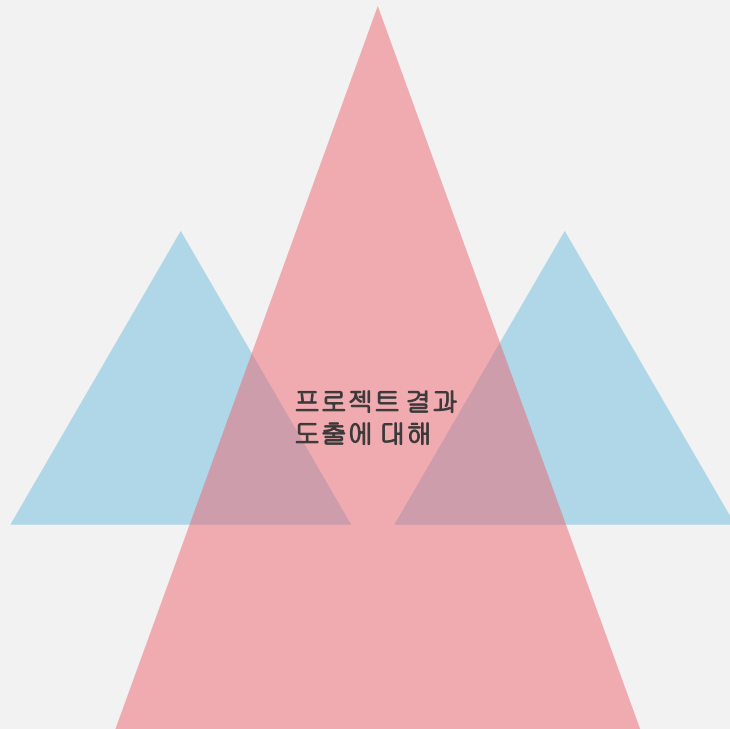
데이터 셋에 대해

# 프로젝트 결과 도출에 대해

·탐색적 데이터 분석(**Exploratory Data Analysis**)을 잘 수행 하였나?

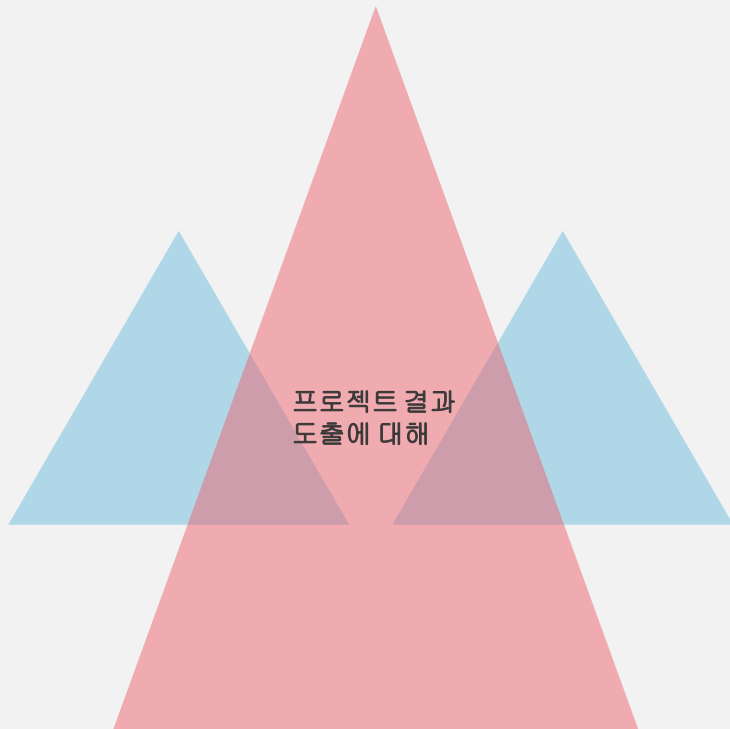
충분한 데이터 탐색 과정을 거쳤나?

- 1) 사전에 각 팀원 개개인이 데이터 탐색을 통해 데이터의 구성과 장단점을 어느 정도 파악한 상태에서 협업을 진행함.
- 2) 모임에서 의견을 나누고 데이터에 대해 각자 잘 이해하지 못한 부분들을 서로 도와 보충하는 과정을 거침. ex) share 항목의 의미 등.



데이터의 **전처리**가 잘 이루어졌나? - 데이터의 전처리 작업은 학습 모델에 적용할 때 뿐만 아니라 분석을 위해서도 필요.

- 1) **결측치와 중복 데이터**의 확인은 불필요 했나? - 데이터가 복잡하지 않고 직관적인 편이며, 결측치나 중복 데이터가 육안으로는 관측된 적이 없었고, 있다 하더라도 결과 도출에 미치는 영향이 미미할 거라 판단하고 별도의 확인 과정을 거치지 않음. 그러나 오류 예방과 정밀한 결과 도출을 위해서는 간과하지 말았어야 함.
- 2) **이상치** 검토가 필요한 경우였나? - 데이터의 성격과 프로젝트 목적 상 임의로 이상치를 규정하면 오히려 데이터에 왜곡을 초래할 가능성이 큼.
- 3) **정규화**가 필요한 부분이 없었나? - 전체 60개국에 대한 그래프와 달리 개별 국가에 대한 그래프는 각각 방문객 규모의 차이가 커서 세로축의 스케일이 다르므로 증감 폭에 대한 오해를 유발할 가능성이 있음. 그러나 이것은 정규화에 의한 통일보다는 발표 과정에서 적절한 설명을 통해 해결하는 것이 시각적인 효과를 유지할 수 있는데 도움이 됨.



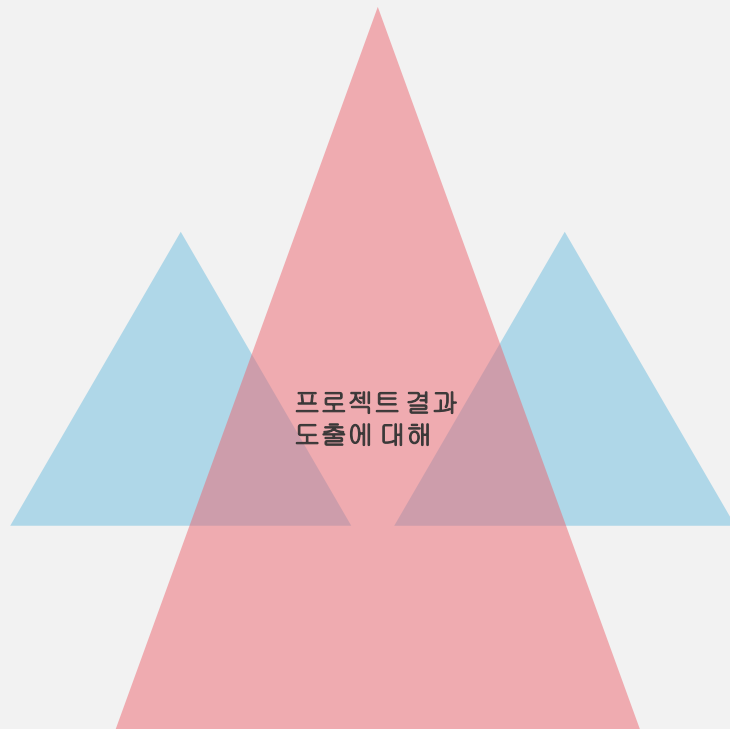
## ·결과물을 잘 이끌어냈는가?

시각화는 데이터 셋 자체에서 이끌어낸 부분에 외부 자료들을 보충하는 방식으로 구성.

데이터 분석을 통한 인사이트와 결론 도출은 아래와 같이 2가지 반대 방향으로 접근하는 프로세스를 거쳐, 예측한 원인과 실제 조사된 내용의 일치 여부를 확인하는 틀을 구성함.

- 1) 외국인들의 한국 방문에 영향을 미칠 수 있는 국내외 이슈 조사 ⇒ 실제 데이터에 영향을 미쳤는지 분석
- 2) 실제 데이터에서 방문자 수 증감 부분 조사 ⇒ 이것의 원인(국내외 이슈 등) 조사, 분석

예측이 많이 빗나간 결과가 나오기는 했지만, 이와 같이 예측과 현실이 불일치하는 부분을 확인할 수 있다는 것이 이 데이터 셋의 존재 의미이며, 이를 통해 관광·마케팅 정책에 반영할 수 있는 유의미한 결과물을 얻을 수 있었음.





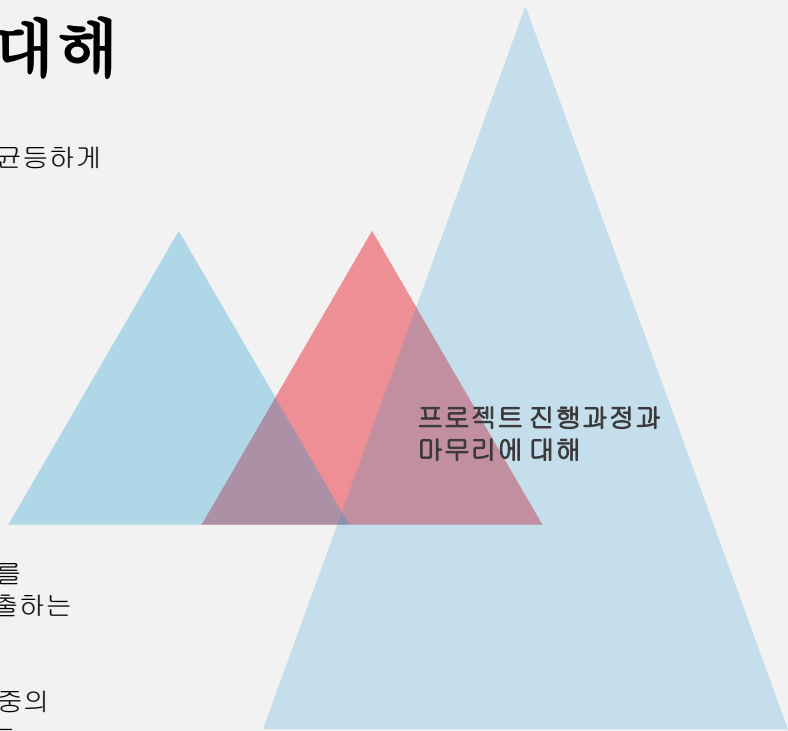
# 프로젝트 진행과정과 마무리에 대해

팀원들의 협업은 다음과 같은 방식으로 진행하였고 효율적이면서도 비교적 균등하게 작업이 배분된 것으로 생각됨.

- 003 상위 5개국 분석 부분은 같이 실시간으로 협업으로 진행.
- 001 데이터 소개, 002 방문 국가, 004 회고 부분은 각자 배분받아 주도적으로 작업.
- 최종적으로 001~004 전체를 실시간 협업으로 마무리.

## 팀원들의 마무리 소감

- 프로젝트를 진행하면서 데이터 분석을 어떻게 해야하는지에 대한 이해도를 얻었지만 시계열 이해도가 부족해서 시계열을 이용하여 데이터 자료를 도출하는 것을 이해하지 못했습니다.
- 이번 프로젝트에서 예측한 가설이 많이 빗나갔는데, 이번에 크게 느낀 점 중의 한 가지는 가설을 세워 적중률 높게 예측하는 것이 그리 쉬운 것이 아니라는 것입니다.



Questions



