



외국인 관광객 맞춤 랜드마크 가이드

LAND_MARK팀 (오경민, 이대한, 임성훈, 조혜진)

INDEX

01 팀 소개

02 프로젝트 개요

03 BUSINESS CHALLENGES

04 BIG DATA ANALYSIS

05 APPLICATION TO BUSINESS

06 기대효과

07 향후 과제

08 활용 기술 및 참고자료

01 팀 소개

- 팀명 : LAND_MARK
- 팀 구성 : 오경민, 이대한, 임성훈, 조혜진
- 구성원 역할

성명	역할	설명
오경민(팀장)	이미지 판별 CNN모델 구현	관련 데이터 수집, 가공, 모델 구현
임성훈	외국인 관광객들의 추이 파악	관련 데이터 수집, 시각화
이대한	외국인 관광객들의 국내 관광지 Top 3 파악	관련 데이터 수집, 시각화
조혜진	이미지 판별 CNN모델 구현	관련 데이터 가공, 모델 구현



02 프로젝트 개요

주제 및
분석방향 선정

시각화

예측결과 구현

01

데이터 수집 및 전처리

02

03

CNN 모델 (딥러닝)

04

결론 도출



03 BUSINESS CHALLENGES

01 주제 및 분석방향 제시

국내 체류 외국인, 15개월 만에 200만명대 회복했다

송고시간 | 2022-06-23 11:40



이상서 기자
기자 페이지

|"코로나, 엔데믹 전환하자 외국인 유입 늘어..."내년 250만명대 회복 예상"

(서울=연합뉴스) 이상서 기자 = 신종 코로나바이러스 감염증(코로나19)이 엔데믹(endemic-풍토병으로 굳어진 감염병)으로 전환하면서 국내 체류 외국인이 15개월 만에 200만 명대를 회복했다.

격리의무 해제...외국인 관광객이 온다

f t talk b blog | 0

격리의무 풀리면서 몰리는 외국인 관광객

코로나19가 잠잠해지면서 서울을 찾는 외국인 관광객도 부쩍 늘고 있다. 서울 종로구 경복궁과 청와대 일대는 이미 외국인 특수를 누리는 모습이다.

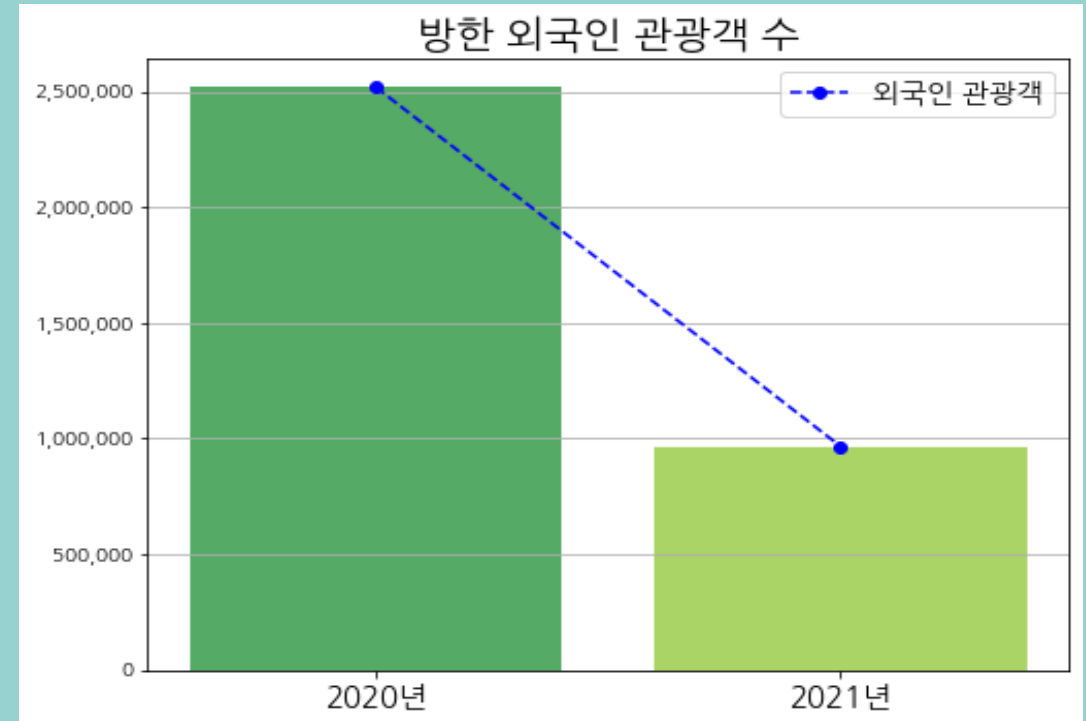
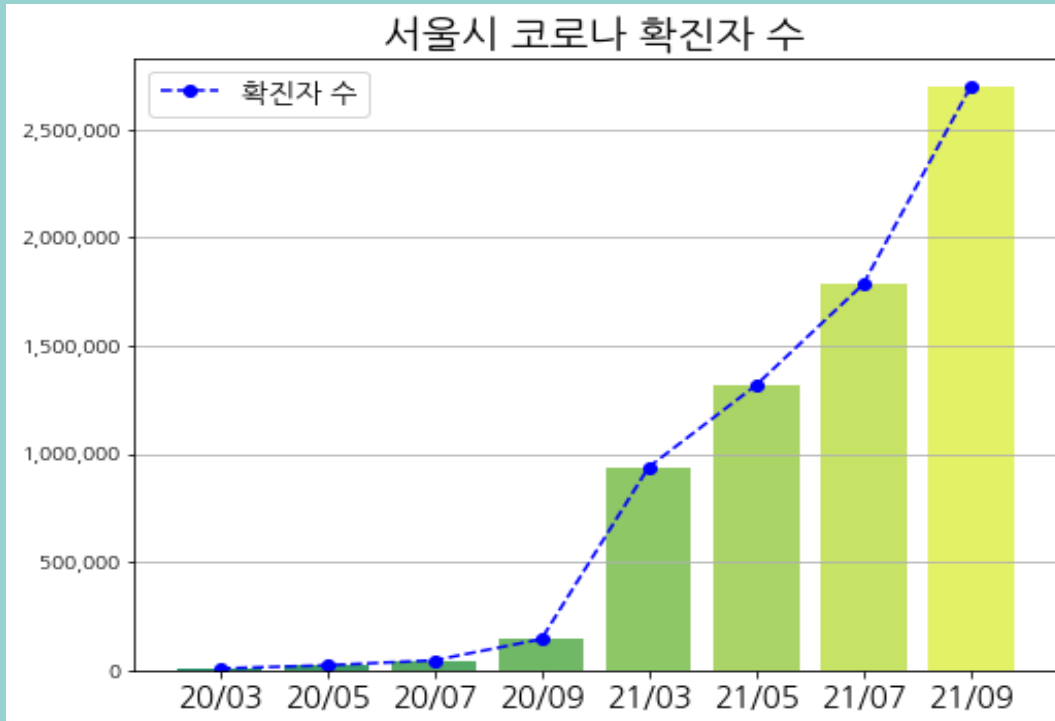
법무부에 따르면 지난 1~4월 관광 비자를 받고 한국을 찾은 외국인은 5만 4694명이다. 지난해 같은 기간 2만 1703명 대비 152.0% 증가한 규모다. 1월 7797명, 2월 7855명, 3월 1만 421명, 4월 2만 8621명으로 매달 관광객 수도 증가 추세다.

- 위드코로나 전환으로 외국인 관광객의 방한 증가
- 외국인 관광객들이 랜드마크 사진을 찍으면
랜드마크의 설명, 주변 관광지, 맛집이 나오게 하여
한국을 관광하는데에 도움을 줄 수 있도록 한다.



03 BUSINESS CHALLENGES

02 코로나에 따른 방한 외국인 관광객 수



코로나 확진자 수가 늘어날 수록
방한 외국인 관광객의 수가 낮아지는 반비례 관계가 존재함.

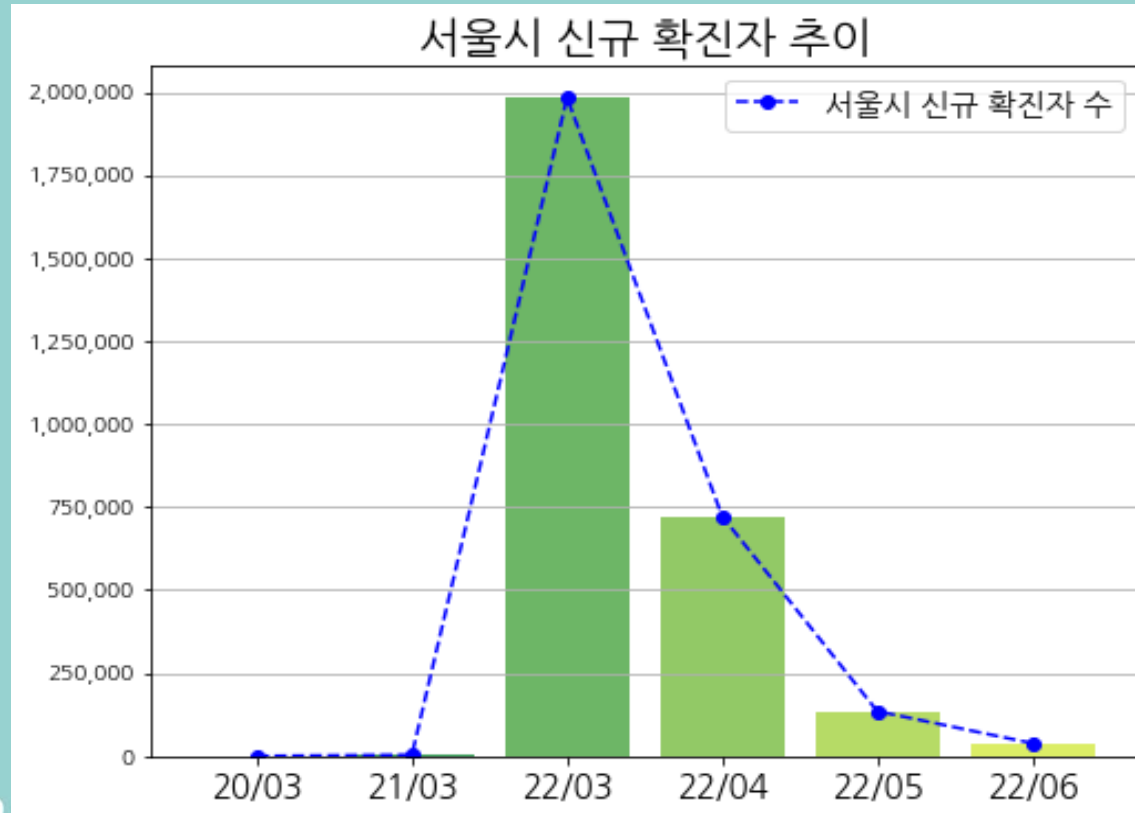
따라서 코로나 유행이 방한 외국인 관광객의 수에 영향을 주고 있음을 확인.

코로나 유행이 잦아들면 방한 외국인 관광객들의 수가 증가할 것이라고 예측 가능.



03 BUSINESS CHALLENGES

03 외국인의 한국관광 방향성



- 22/06부분은 22/06/24일까지의 기준

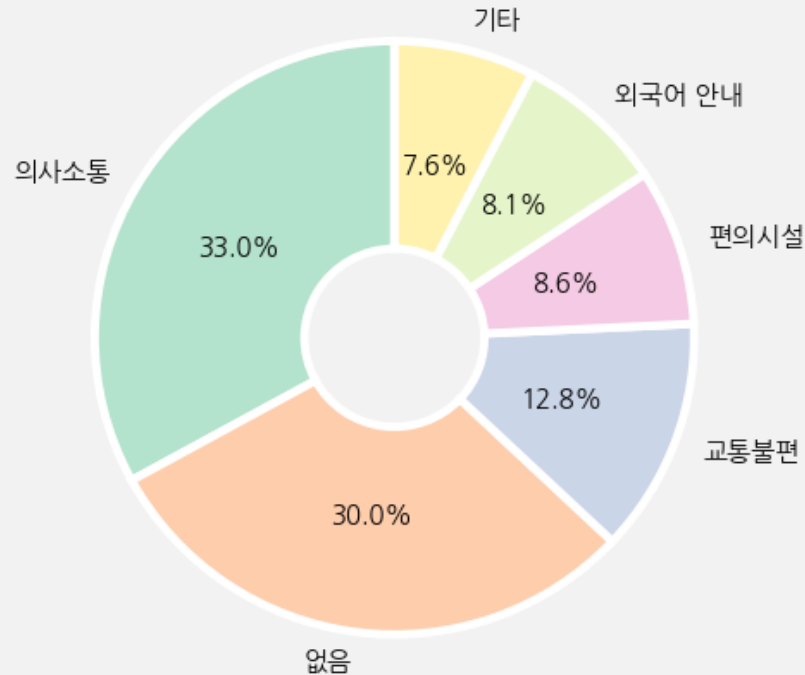
- 서울 신규 확진자 수가 2021년 5월 이후 폭발적으로 증가.
- 2022년 6월 현재, 2021년 5월 이전의 상태로 회복중.
- 코로나 유행이 끝나가는 것으로 보이고, 이에 외국인의 한국 관광 활성화를 점진적으로 희망해 볼 수 있음.



03 BUSINESS CHALLENGES

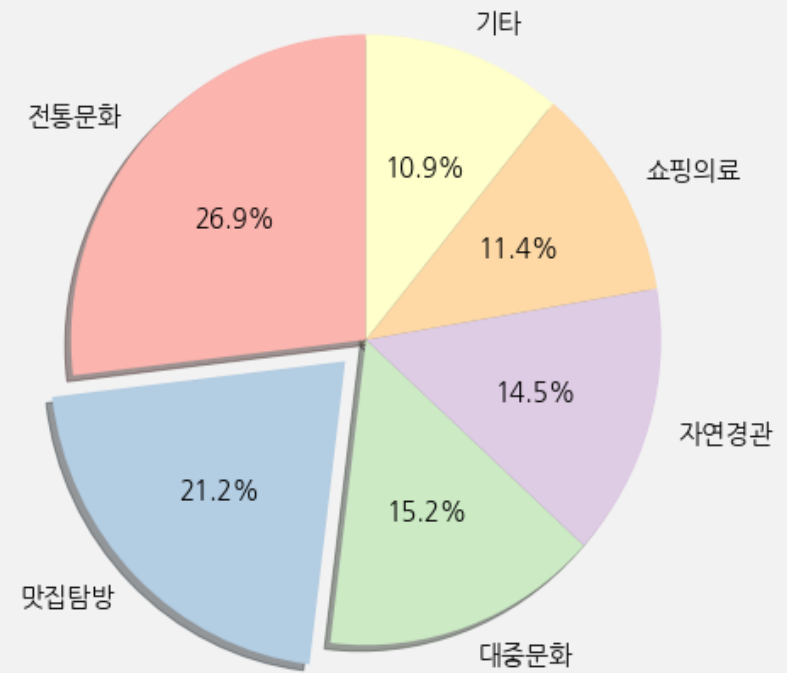
04 외국인 관광객들의 생각

한국 여행 중 가장 불편했던 점



외국 관광객들이 한국여행에서 언어적 어려움을 느낀다(의사소통, 외국어 안내)
→ 번역을 통해 언어적 어려움에 도움을 주어 더욱 원활한 여행을 기대

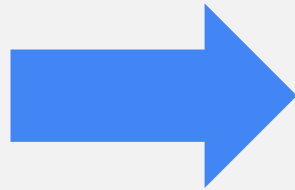
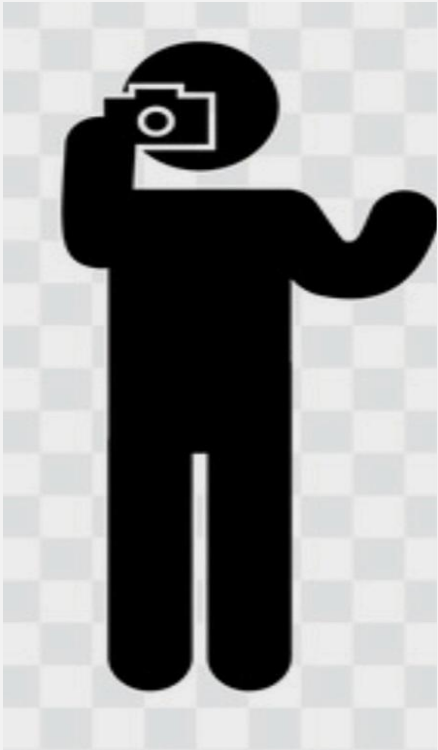
한국을 관광 목적으로 선택한 이유



한국에 관광을 온 이유 큰 목적은 한국에만 있는 전통문화, 음식이다.
→ 한국의 랜드마크 소개와, 음식 소개, 주변 관광지 추천을 통해 풍부한 여행을 기대

03 BUSINESS CHALLENGES

05 서비스 구현

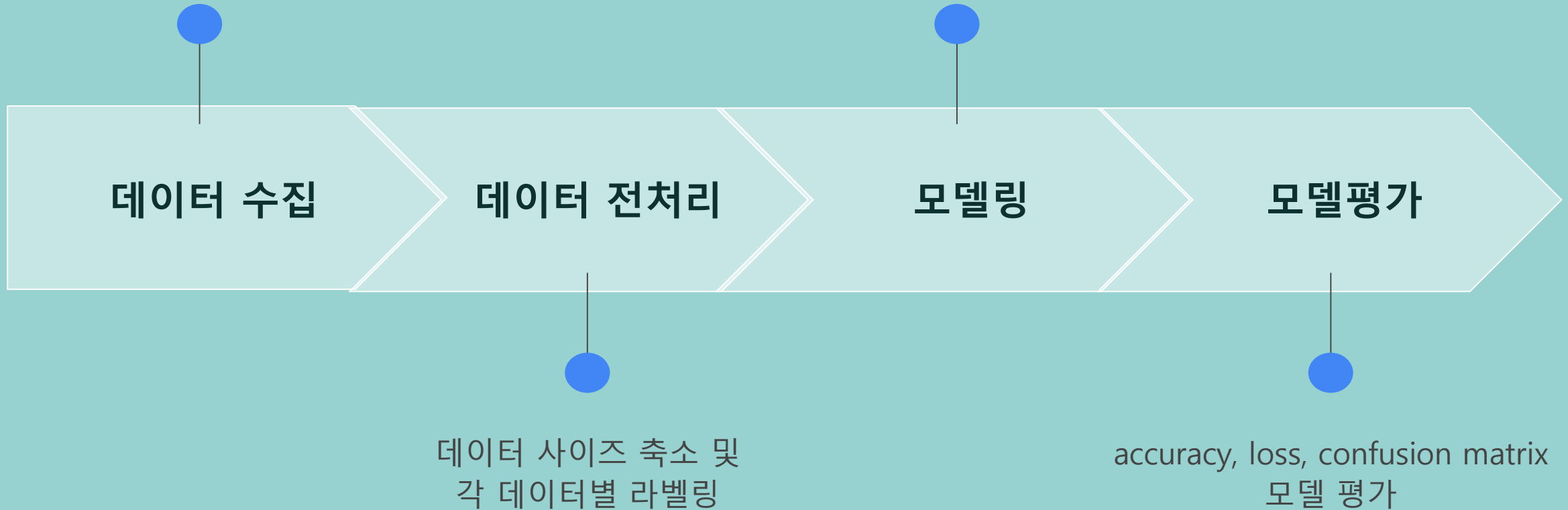


각 랜드마크의 번역 된 간단한 설명과 주변 관광지, 맛집을 소개해주는 서비스를 제공하여 더욱 풍부한 관광을 기대

04 BIG DATA ANALYSIS

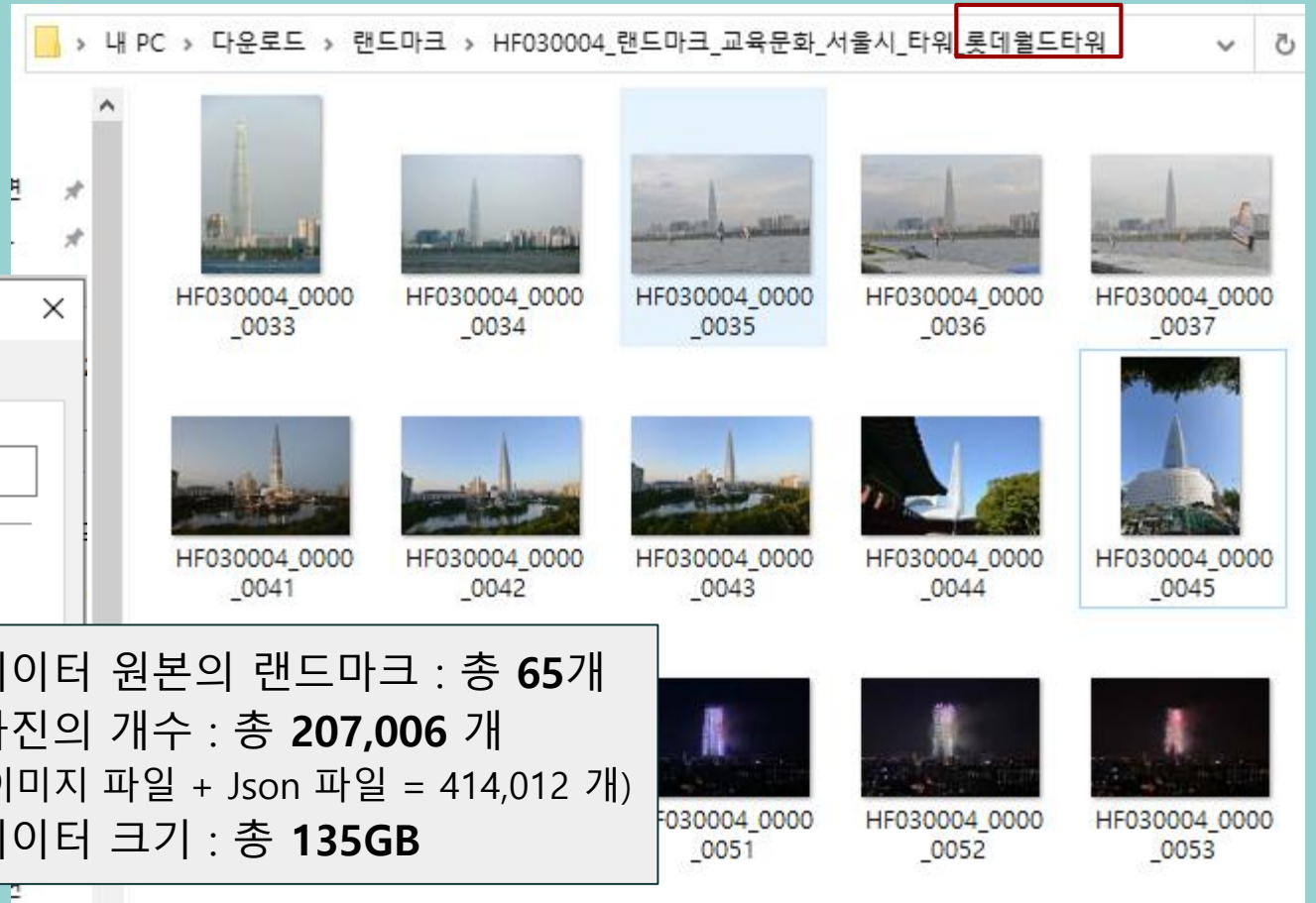
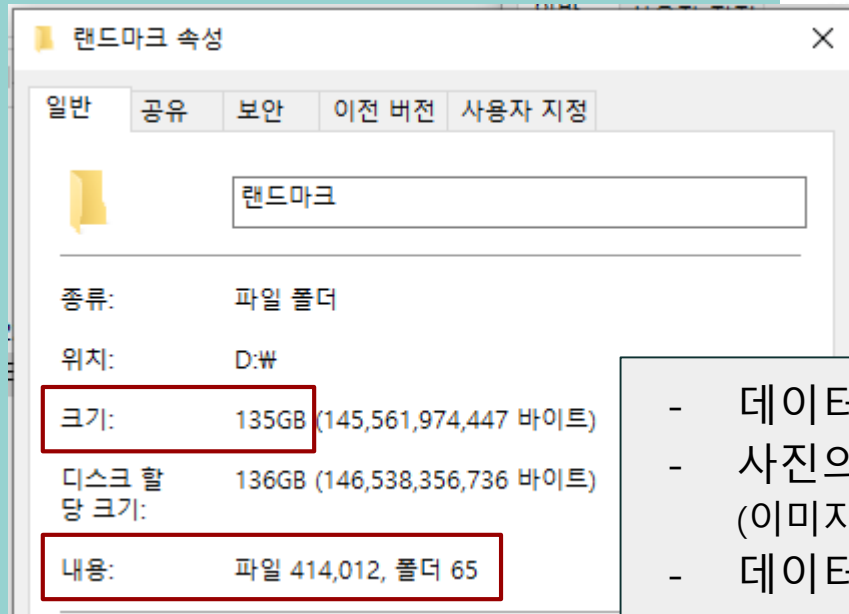
국내 여행지 3곳 선정
서울, 부산, 제주 랜드마크 이미지 수집

CNN모델을 이용하여
랜드마크 분류 모델 생성



04 BIG DATA ANALYSIS

01 국내 여행지 선정 이유?



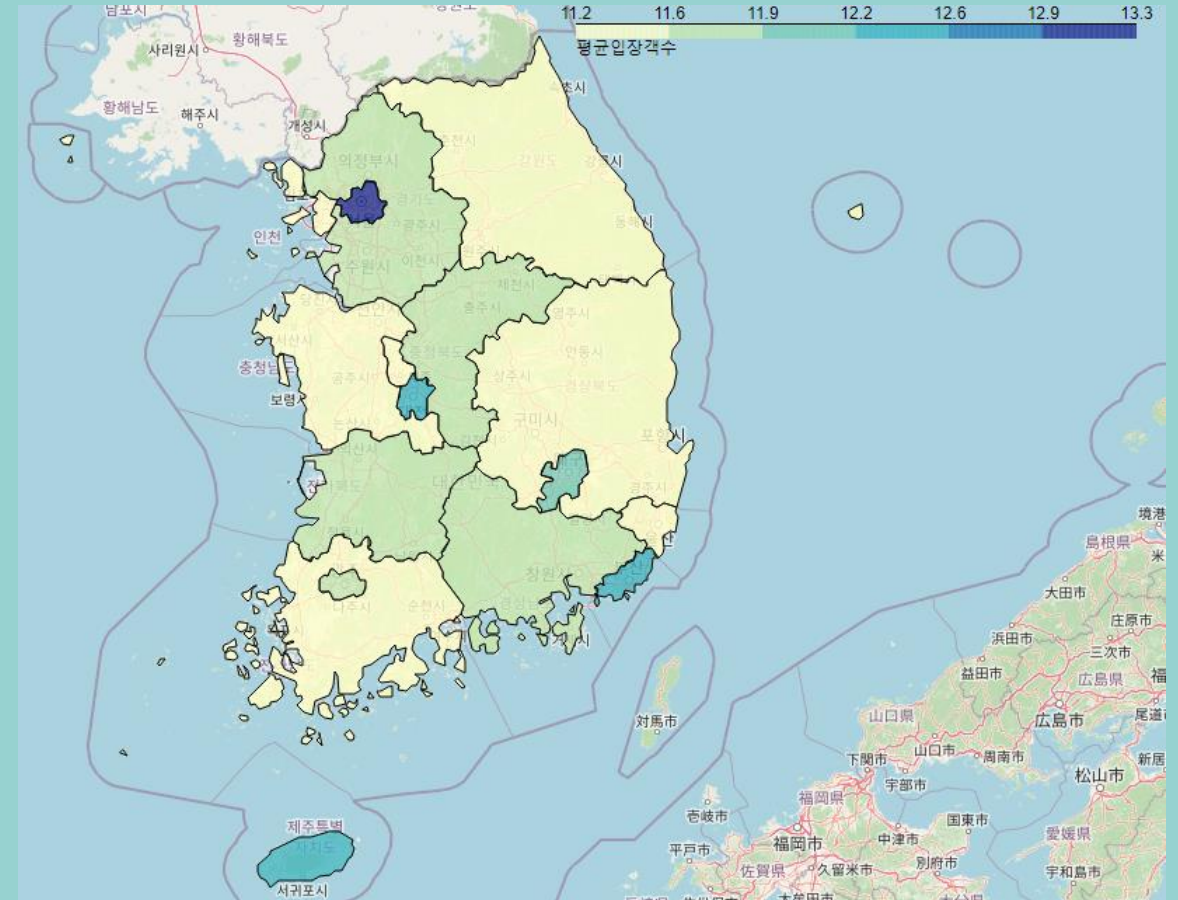
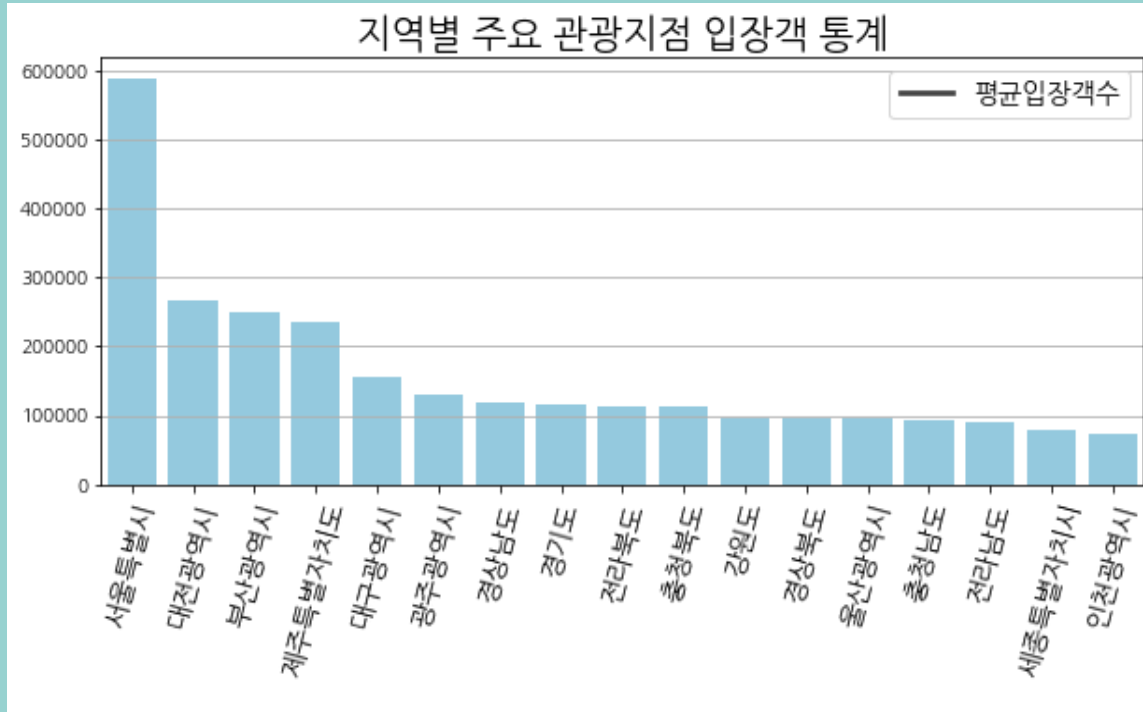
- 데이터 원본의 랜드마크 : 총 **65**개
- 사진의 개수 : 총 **207,006** 개
(이미지 파일 + Json 파일 = 414,012 개)
- 데이터 크기 : 총 **135GB**

원본 데이터 크기가 너무 크고, 파일의 수도 너무 많아서
외국인들이 많이 찾는 국내 여행지 3곳을 선정하여
그 여행지의 랜드마크를 가지고 모델링 학습 데이터로 활용.

여행 관광객들에게 사랑받는 국내 여행지 3곳을 알아볼 예정.

04 BIG DATA ANALYSIS

02 국내 여행지 선정



2021년도 전국 시도 총 17개의 지역 중에서
관광객들이 많이 찾는 지역은 상위 4개의 지역은 **서울, 대전, 부산, 제주도**로 확인.
대전의 경우에는 랜드마크에 아파트가 포함되어 있어서 제외.

04 BIG DATA ANALYSIS

03 CNN 모델링 데이터 전처리

폴더명	랜드마크 장소	원본 데이터 수	학습데이터 수
HF030001_랜드마크_주거생활_서울시_빌딩_63빌딩	63빌딩	3,622	250
HF030002_랜드마크_인물사건_서울시_동상_이순신장군동상	이순신장군동상	3,240	250
HF030003_랜드마크_교육문화_서울시_타워_남산타워	남산타워	2,206	250
HF030004_랜드마크_교육문화_서울시_타워_롯데월드타워	롯데월드타워	3,754	250
HF030005_랜드마크_교육문화_부산시_타워_부산타워	부산타워	3,884	200
HF030058_랜드마크_교육문화_제주시_센터_43기념관	43기념관	3,923	200
HF030059_랜드마크_주거생활_제주시_빌딩_롯데시티호텔제주	롯데시티호텔제주	3,803	200
HF030068_랜드마크_교통통신_부산시_다리_광안대교	광안대교	3,465	200
HF030069_랜드마크_주거생활_부산시_빌딩_해운대아이파크	해운대아이파크	2,564	200
HF030070_랜드마크_교통통신_부산시_다리_영도대교	영도대교	3,604	200
		34,065	2200

서울(4), 부산(4), 제주(2)의 랜드마크 총 10 곳으로 데이터 선정.

각 랜드마크의 사진의 수가 2,500 ~ 3,900개 정도로 **총 34,065개 확인.** (파일크기 총 **21GB**)

파일 수가 너무 방대하여 CNN 모델링을 할 수 없어서

서울의 랜드마크는 250개, 나머지 지역은 200개로 파일의 수를 축소. (**총 2200개**)

04 BIG DATA ANALYSIS

03 CNN 모델링 데이터 전처리

```
1 # x 데이터 생성
2 import cv2
3 import os
4 import numpy as np
5
6 def make_train_data(landmark_type, DIR):
7     for img in os.listdir(DIR):
8         label = landmark_type # 라벨링
9         path = os.path.join(DIR, img) # 이미지 파일 경로 주소
10
11         img = cv2.imread(path, cv2.IMREAD_COLOR) # 컬러 이미지 가져오기
12         img = cv2.resize(img, (200, 150)) # 이미지 사이즈 재지정
13
14         X.append(img)
15         Z.append(label) # y 데이터
16
17
18 1 X=[]
19 2 Z=[]
20 3
21 4 path = '/content/drive/MyDrive/프로젝트/data/랜드마크/' # 이미지 파일 폴더 경로
22 5
23 6 DIRS = os.listdir(path)
24 7 DIRS.sort() # 정렬
25 8
26 9 for dir in DIRS:
27 10     make_train_data(dir, path + dir) # 함수 사용해서 데이터 가져오기
28 11     print(len(X))
```

사진 크기	1920 x 1280
너비	1920픽셀
높이	1280픽셀

6 DIRS

```
['HF030001_랜드마크_주거생활_서울시_빌딩_63빌딩',
 'HF030002_랜드마크_인물사건_서울시_동상_이순신장군동상',
 'HF030003_랜드마크_교육문화_서울시_타워_남산타워',
 'HF030004_랜드마크_교육문화_서울시_타워_롯데월드타워',
 'HF030005_랜드마크_교육문화_부산시_타워_부산타워',
 'HF030058_랜드마크_교육문화_제주시_센터_43기념관',
 'HF030059_랜드마크_주거생활_제주시_빌딩_롯데시티호텔제주',
 'HF030068_랜드마크_교통통신_부산시_다리_광안대교',
 'HF030069_랜드마크_주거생활_부산시_빌딩_해운대아이파크',
 'HF030070_랜드마크_교통통신_부산시_다리_영도대교']
```

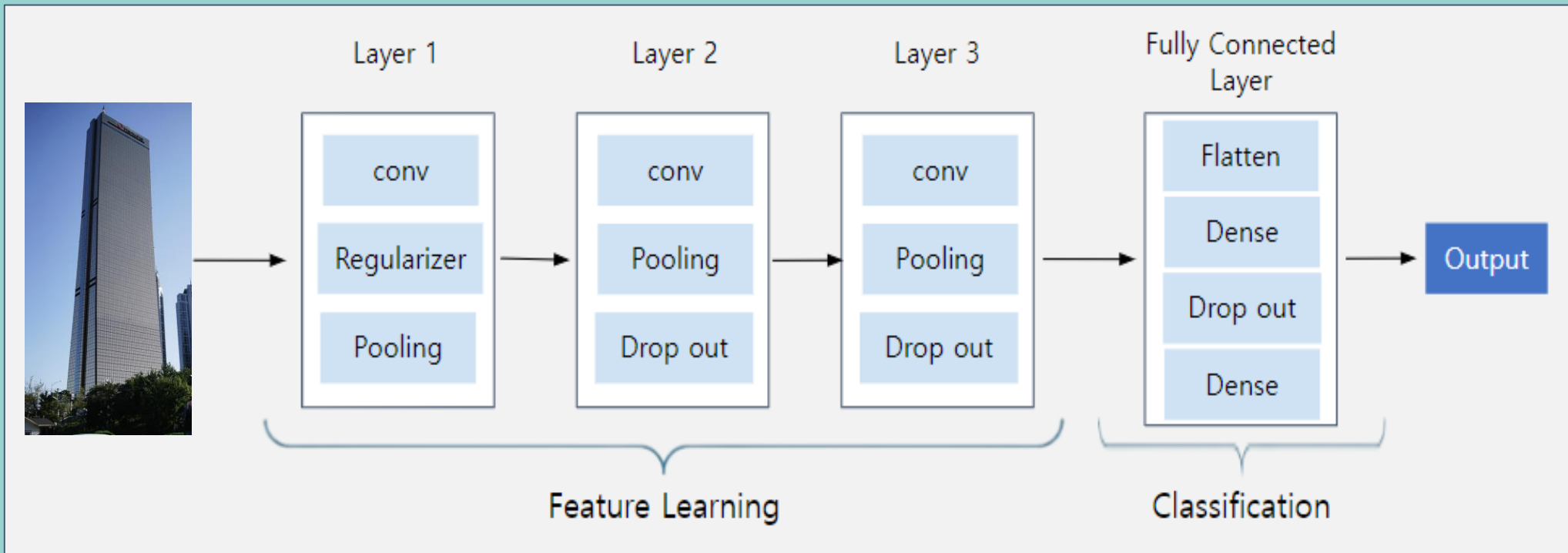
원본의 사진 크기가 **1920 * 1280**으로 커서
이미지 파일을 가져올 때 사이즈를
200 * 150으로 축소하여 가져옴.

이미지 파일을 가져올 때,
os.listdir를 통해 이미지 폴더의 이름들을
리스트에 담고,
그 폴더에 있는 이미지들을 하나씩 가지고 옴.

04 BIG DATA ANALYSIS

04 CNN 모델링

- **CNN모델**(Convolutional Neural Network)이란, 합성곱 신경망 모델로 합성곱 계층을 통해 데이터의 특징을 추출 후, 그 특징들을 기반으로 분류하는 딥러닝 모델.
- 다수의 Layer층을 보유한 네트워크 구조로 Convolution + ReLU + Pooling 방식으로 작동.
- 이미지 분류시, 이미지의 공간 정보를 유지하면서 인접 이미지와의 특징을 효과적으로 인식.



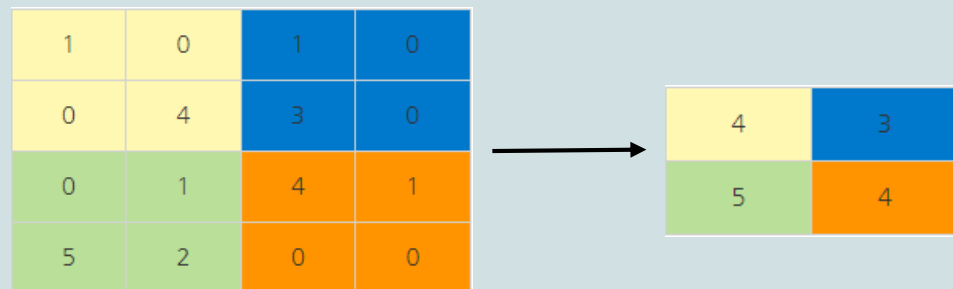
04 BIG DATA ANALYSIS

04 CNN 모델링

```
4 model = Sequential()
5 model.add(Conv2D(32, (3, 3), input_shape=(150,200,3), activation = 'relu', kernel_regularizer= l2(0.001)))
6 model.add(MaxPooling2D(pool_size=(2, 2)))
7
8 model.add(Conv2D(32, (3, 3),))
9 model.add(Activation('relu')) # 따로 add 취할 수 있음.
10 model.add(MaxPooling2D(pool_size=(2, 2)))
11 model.add(Dropout(0.5)) # 50%만 사용
12
13 model.add(Conv2D(64, (3, 3)))
14 model.add(Activation('relu'))
15 model.add(MaxPooling2D(pool_size=(2, 2)))
16 model.add(Dropout(0.5)) # 50%만 사용
17
18 model.add(Flatten())
19
20 # 분류모델
21 model.add(Dense(64, kernel_regularizer= l2(0.001)))
22 model.add(Activation('relu'))
23
24 model.add(Dropout(0.5)) # 분류층에서 50%만 사용
25
26 model.add(Dense(10))
27 model.add(Activation('softmax'))
```

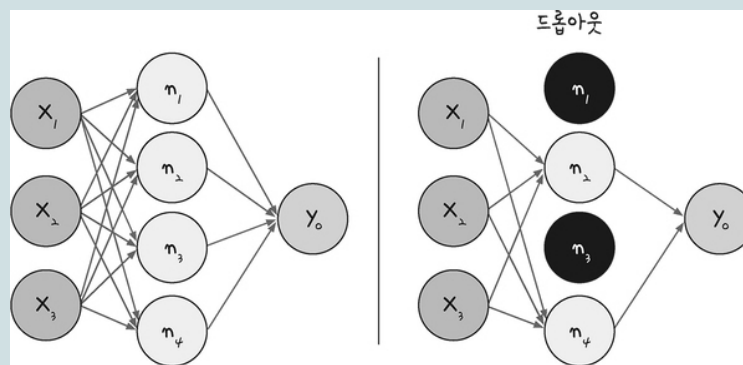
MaxPooling이란, 이미지 차원을 축소하는 기법 중 하나로 정해진 구역 안에서 최댓값을 뽑아내는 기법이다.

예시)



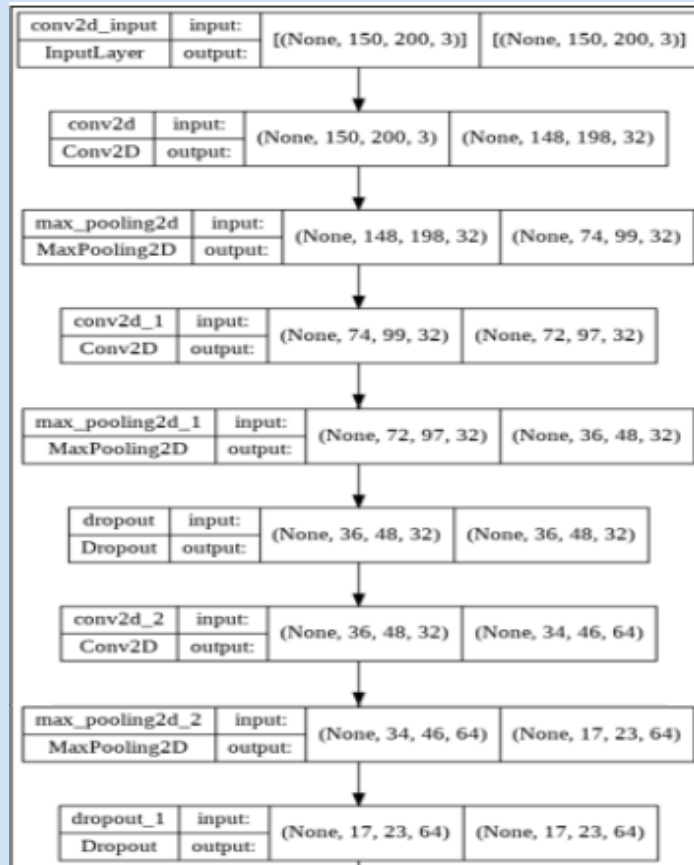
Dropout이란, 과적합을 방지하기 위한 기법으로, 은닉층에 배치된 노드 중 일부를 정지시키는 기법이다.

예시)

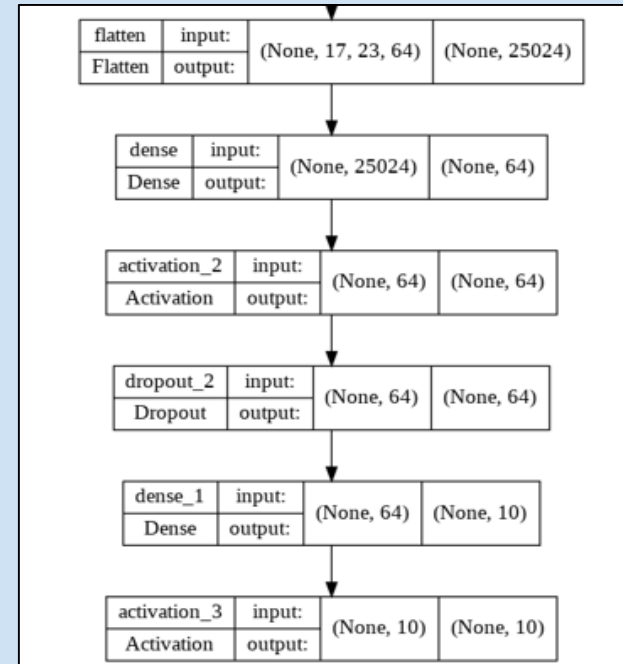


04 BIG DATA ANALYSIS

04 CNN 모델링



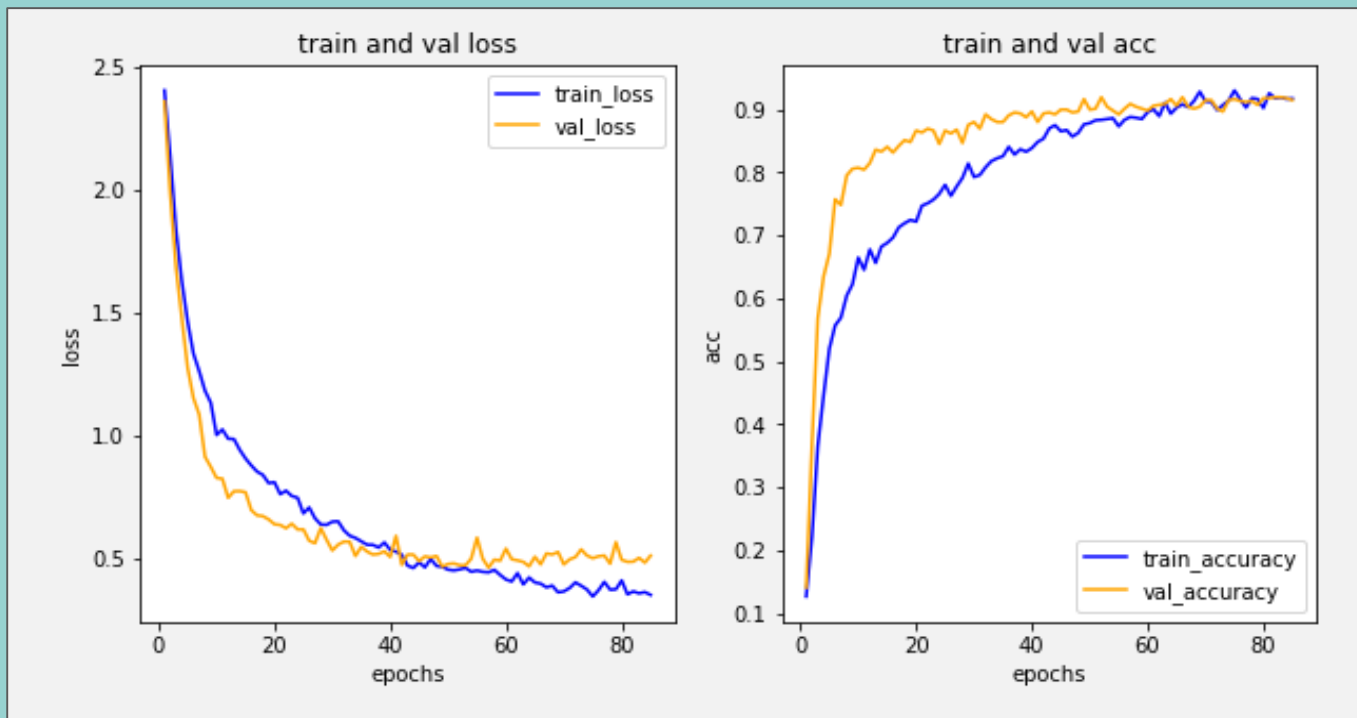
Feature Learning



Classification

04 BIG DATA ANALYSIS

04 CNN 모델링 평가



l2 규제를 은닉층에서 한번,
분류층에서 한번 총 2번을 주고,

dropout를 세 번 준 결과,

epoch = 85에서
train 데이터의 정확도와
test 데이터의 정확도가 동일해짐.

최적의 모델은 epoch = 52에서 생성.

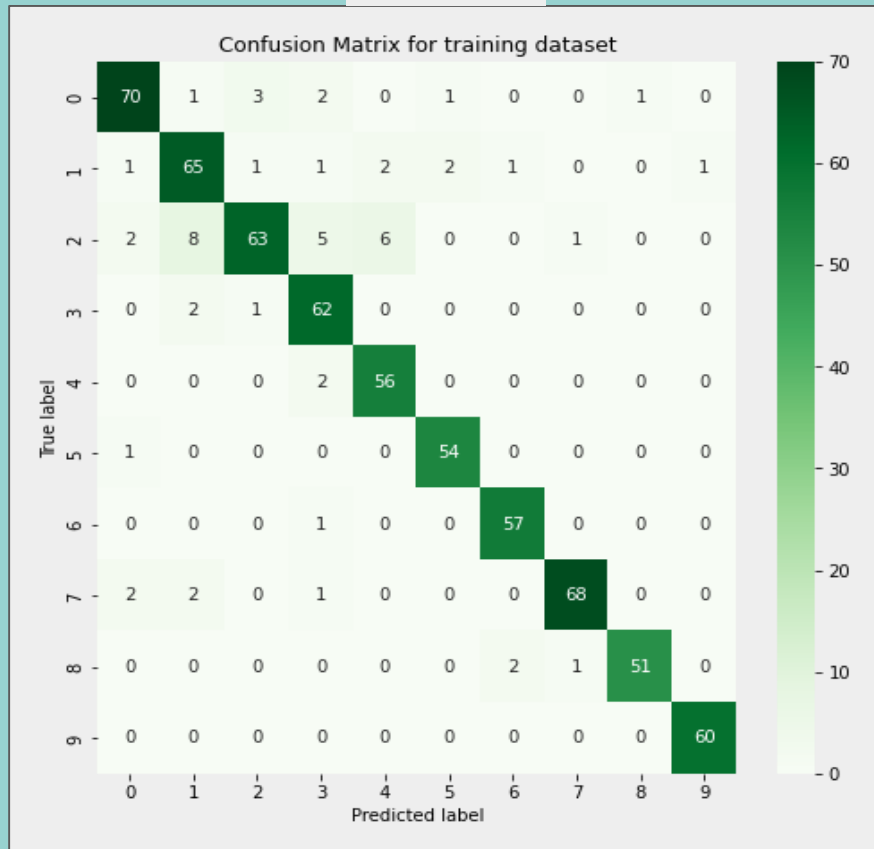
```
Epoch 52/100
152/154 [=====>.] - ETA: 0s - loss: 0.4527 - accuracy: 0.8842
Epoch 52: val_accuracy improved from 0.91667 to 0.91970, saving model to /content/drive/MyDrive/프로젝트/data/best_model.h5
154/154 [=====] - 3s 17ms/step - loss: 0.4530 - accuracy: 0.8838 - val_loss: 0.4699 - val_accuracy: 0.9197
```

```
Epoch 85/100
152/154 [=====>.] - ETA: 0s - loss: 0.3480 - accuracy: 0.9184
Epoch 85: val_accuracy did not improve from 0.91970
154/154 [=====] - 2s 15ms/step - loss: 0.3488 - accuracy: 0.9175 - val_loss: 0.5091 - val_accuracy: 0.9152
Epoch 85: early stopping
```

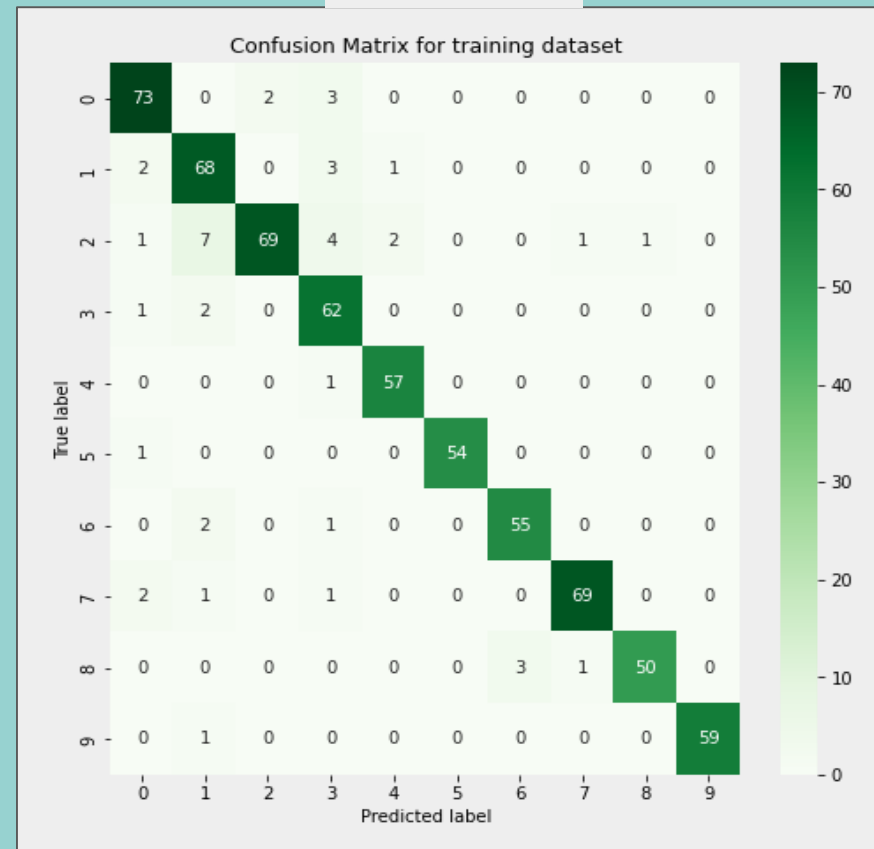
04 BIG DATA ANALYSIS

04 CNN 모델링 평가

기존 모델



최적의 모델



test데이터에 대한 model의 예측값을 혼동행렬도 표시한 결과,
최적의 모델의 성능이 더 좋기 때문에 **best_model**를 예측 모델로 확정.

04 BIG DATA ANALYSIS

04 CNN 모델링 평가 - classification report

	precision	recall	f1-score	support
HF030001_랜드마크_주거생활_서울시_빌딩_63빌딩	0.91	0.94	0.92	78
HF030002_랜드마크_인물사건_서울시_동상_이순신장군동상	0.84	0.92	0.88	74
HF030003_랜드마크_교육문화_서울시_타워_남산타워	0.97	0.81	0.88	85
HF030004_랜드마크_교육문화_서울시_타워_롯데월드타워	0.83	0.95	0.89	65
HF030005_랜드마크_교육문화_부산시_타워_부산타워	0.95	0.98	0.97	58
HF030058_랜드마크_교육문화_제주시_센터_43기념관	1.00	0.98	0.99	55
HF030059_랜드마크_주거생활_제주시_빌딩_롯데시티호텔제주	0.95	0.95	0.95	58
HF030068_랜드마크_교통통신_부산시_다리_광안대교	0.97	0.95	0.96	73
HF030069_랜드마크_주거생활_부산시_빌딩_해운대아이파크	0.98	0.93	0.95	54
HF030070_랜드마크_교통통신_부산시_다리_영도대교	1.00	0.98	0.99	60
accuracy			0.93	660
macro avg	0.94	0.94	0.94	660
weighted avg	0.94	0.93	0.93	660

05 APPLICATION TO BUSINESS

01 예측 결과 전처리

- NAVER 개발자 센터에서 **파파고 API**를 신청 후 받아서
랜드마크의 설명과 주변 관광지를 영어로 번역.
- 맛집은 영어로 번역하면 찾아가기 더 힘들 것 같아서 번역에서 제외.

```
1 # 파파고 API로 번역하기 (하루에 무료 10,000자)
2 import requests
3
4 def get_translate(text):
5     client_id = "" # <-- client_id 기입
6     client_secret = "" # <-- client_secret 기입
7
8     data = {'text' : text,
9             'source' : 'ko',
10            'target' : 'en'} # 한글을 영어로 번역해달라.
11
12     url = "https://openapi.naver.com/v1/papago/n2mt"
```

남산타워

N서울타워(영어: N Seoul Tower)는 대한민국 서울특별시 용산구 용산동 2가 남산 공원 정상 부근에 위치한 전파 송출 및 관광용 타워이다. 1969년에 착공하여 1975년 7월 30일 완공되었다. 높이는 236.7m, 해발 479.7m이다.

주변 맛집 : 더백푸드트럭, 더플레이스다이닝, 도깨비코티지

주변 관광지 : 리움미술관, 경리단길, 남산 케이블카



남산타워

N Seoul Tower is a tower located near the summit of Namsan Park in Yongsan-dong 2-ga, Yongsan-gu, Seoul, South Korea. Construction began in 1969 and was completed on July 30, 1975. It is 236.7m high and 479.7m above sea level.

주변 맛집 : 더백푸드트럭, 더플레이스다이닝, 도깨비코티지

주변 관광지 : Leeum Museum of Art, Gyeongridan-gil, Namsan Cable Car

05 APPLICATION TO BUSINESS

02 화면 구현



롯데월드타워

Lotte World Tower is a skyscraper located in the Lotte World Mall complex in Sincheon-dong, Songpa-gu, South Korea. Construction began in 2010 with 123 stories above ground and a height of 555m and was completed on December 22, 2015, and exterior construction was completed by completing the spire construction around March 2016, and was completed on December 22, 2016. It opened on April 3, 2017.

주변 맛집 : 도쿄등심, 비채나, 봉땅

주변 관광지 : Seoul Sky (Observation Deck), Lotte World Aquarium, Jamsil River Park



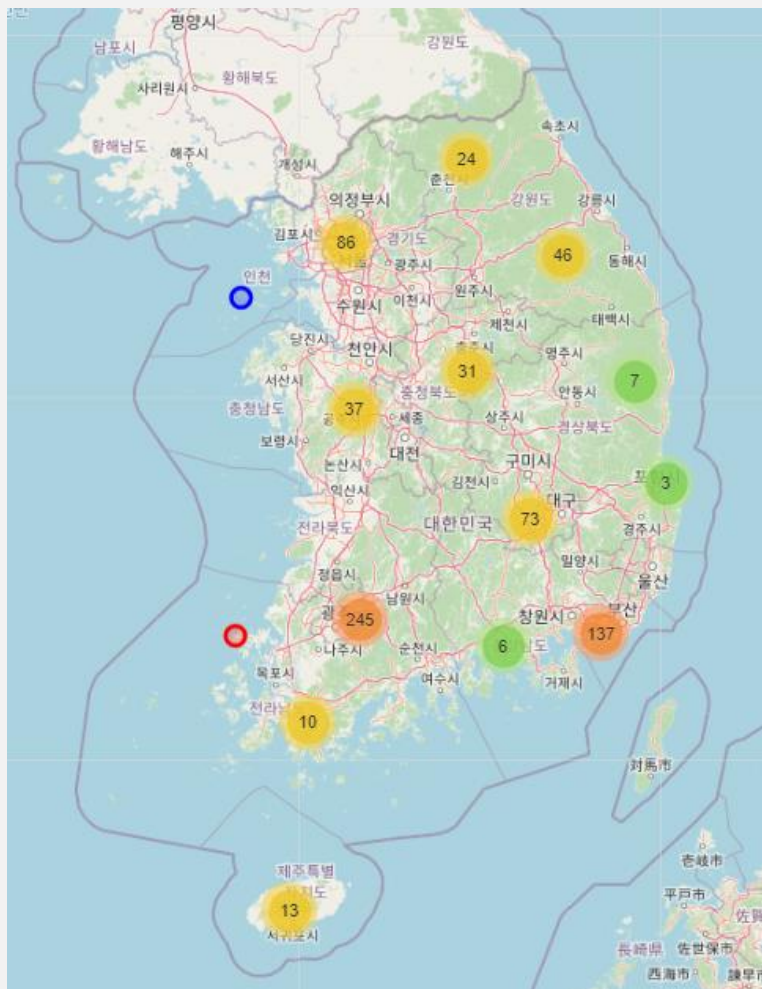
광안대교

The Gwangan Bridge or Diamond Bridge is a bridge located in Busan and is part of Busan Metropolitan City Line 66. This bridge is Korea's largest double-decker bridge connecting Square 49 in Namcheon-dong, Suyeong-gu, and Centum City in Udong, Haeundae-gu.

주변 맛집 : 오유미당, 아텔라베일리, 삼굽살

주변 관광지 : Gwangalli Beach, Gwangalli Marine Sports Center, and Minrak Waterfront Park

06 기대효과



외국인들의 방한 이유는 한국만의 전통문화, 음식 때문이 크다. 그렇기에 각 지역특색에 따라 혹은 외국인들의 이동코스에 따라 서비스를 제공한다면 더욱 풍부하고 원활한 여행을 기대할 수 있다.

A 동해일미

강원도 강릉시 경강로2024번길 29

B 강릉짬뽕순두부 동화가든 본점

강원도 강릉시 초당순두부길77번길 15

C 순두부젤라또 1호점

강원도 강릉시 초당순두부길 95-5

D 카페 틈마루

강원도 강릉시 난설헌로 232 카페 틈마루

E 보헤미안박이추커피

강원도 강릉시 사천면 해안로 1107

A 카페기와

강원도 강릉시 초당원길54번길 15

B 하슬라아트월드

강원도 강릉시 강동면 율곡로 1441

C 하조대전망대

강원도 양양군 현북면 하륜길 56

D 너울집

강원도 속초시 신흥2길 38 너울집

E 봉브레드

강원도 속초시 동해대로 4344-1

07 향후 과제

코로나로 인해 방한 외국인들의 한국 여행에 대한 생각에 대해 최신자료가 없음.
나중에 최신 설문조사 자료를 분석하여 서비스 개선방향 생각.

컴퓨터의 성능 때문에 데이터의 수와 이미지의 사이즈를 축소하여 모델을 학습한 결과,
남산타워, 이순신 동상, 63빌딩 처럼 높게 솟은 랜드마크들의 예측 정확도가 떨어짐.
데이터를 추가하여 모델 학습 예정.

서비스를 기획을 했지만, 구현 및 제공까지 시간이 부족했음.
후에 어플로서 외국인 관광객들에게 제공될 수 있도록 프로그램 개발 예정.

08 활용 기술 및 참고자료

- 사용언어 : Python
- 데이터 전처리 : Pandas, Numpy, OS
- 시각화 라이브러리 : Seaborn, Matplotlib, Folium
- CNN 라이브러리 : Keras, Tensorflow

- DATA 목록

출처	파일명	활용
서울열린데이터광장	(전국기준)국적별 외국인 방문객 통계	막대 그래프
	서울시 코로나19 확진자 발생동향	막대 그래프
https://www.seoul.go.kr/coronaV/coronaStatus.d 으	서울시 신규 코로나19 확진 수	막대 그래프
한국여행업협회-KATA	KATA 관광통역안내소 2차 설문조사결과 통계자료	한국 여행 중 가장 불편했던 점 (파이) 한국을 관광 목적지로 선택한 이유 (파이)
관광지식정보 시스템 https://know.tour.go.kr/	주요관광지점입장객	막대그래프, 지도 시각화
AI 허브 (aihub.or.kr)	한국형 사물 이미지 - 랜드마크	CNN 모델링

- 참고문헌 : 모두의 딥러닝 개정 3판