

#### COVID19\_Small\_Business\_EDA

곽희원, 김아현, 박민춘, 이현지, 최디도



# STEP OF PRESENTATION

서론

거리두기 효과 분석

모빌리티 보고서 분석

코로나 전파로 인한 폐업 분석

배달 앱, 배달 대행 사용 분석

결론

#### 서론

전 세계적으로 유행하는 코로나19로 인한 펜데믹 상황

이 중 소상공인은 코로나 19 영향으로 폐업과 업종 축소를 하고 있는 이슈 등장

전국 소상공 업체 1392개사 중 99.5%는 6개월 지속될 경우 경영 상황이 악화될 것이며, 73%는 이후에도 지속될 경우 폐업 위기(소상공인연합회,20.4.9)



따라서 코로나19로 인한 확진자 현황, 정책, 이동성 자료, 배달 자료를 분석하여 코로나19가 실제 폐업에 영향을 주었는지, 그리고 앞으로 어떤 변화가 일어날지 예측

#### Library

```
# 필요 라이브러리 로드
import matplotlib.pyplot as plt
import matplotlib.dates as dt
import seaborn as sns
import pandas as pd
import numpy as np
import missingno as msno
from datetime import datetime
from matplotlib import rcParams
import plotly, io as pio
import plotly, express as px
import plotly.graph_objects as go
from datetime import datetime, timedelta, date
```

코로나19가 점차 유행함에 따라 정부는 국내에서의 관련 확진, 격리해제, 사망 수에 대해 촉각을 곤두세우고 있습니다.

따라서 매일 업데이트 되는 해당 데이터를 분석하고자 하였으며, 시각화를 통해 우리나라에 코로나19 대유행 시기가 언제였는지, 이후에 진행된 정부의 방역 정책에 의해 효과가 있었는지 확인하고자 합니다.

```
# Kaggle 'Real-time Covid 19 Data' 데이터 로드
time_series = pd.read_csv('/content/time-series-19-covid-KR.csv')
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 439 entries, 0 to 438
Data columns (total 5 columns):
   Column
                Non-Null Count Dtype
                                                 DAYCON의 Time.csv는 20.1.20 - 20.4.30, 약 3달간의
           345 non-null
   Date
                             object
                                                 데이터를 보유하고 있어, 좀 더 큰 범위의 최신화된 현황을 보고자
  Country/Region 345 non-null object
                                                 Kaggle의 Real-time Covid19 Data 활용했습니다.
2 Confirmed 345 non-null
                           float64
3 Recovered 345 non-null
                            float64
   Deaths
          345 non-null
                             float64
dtypes: float64(3), object(2)
memory usage: 17.3+ KB
# 전부 한국에 관한 데이터이므로, 국가 관련 열을 삭제
time_series = time_series.drop('Country/Region',axis = 1)
# 확진자를 기준으로 회복율 / 사망율 추가
time_series_ratio = time_series.assign(RatioOfRecovered = lambda x :(x['Recovered']/x['Confirmed'])*100,
                        RatioOfDeaths = lambda x :(x['Deaths']/x['Confirmed'])*100)
```

확진자 대비 회복 비율



확진자 대비 회복율에 대한 그래프입니다.

위의 그래프로 보아, 20년 2월 말, 20년 9월 초, 20년 12월 부분에서 큰 폭으로 감소했음을 확인할 수 있습니다.

이는 우리나라에서 발생한 대유행시기와 일치하고 있음을 알 수 있습니다.

확진자 대비 사망 비율



확진자 대비 사망율에 대한 그래프로, 위와 동일한 시기에서 큰 폭으로 변화하고 있음을 확인할 수 있습니다. 다만 이 경우 감소하고 있음을 보이고 있는데, 이는 확진자가 급격히 증가했기 때문에 일어난 현상이라고 분석할 수 있겠습니다.

```
# Policy data 호출
policy = pd.read_csv('/content/Policy.csv')
# Social Distancing Campaign 변수를 가지는 모든 행 출력
policy.loc[[28,29,30]]
     policy_id country type
                                                gov_policy detail start_date
28
             29
                    Korea Social Social Distancing Campaign
                                                               Strong
29
                    Korea Social Social Distancing Campaign
                                                               Strong
30
                    Korea Social Social Distancing Campaign
             31
 for i in range (28,31):
  if policy.loc[i,'detail'] =='Strong':
    policy.loc[i.'detail'] = 2
  else:
    policy.loc[i,'detail'] = 1
policy['Level'] = policy['detail']
                 gov_policy Level start_date
                                                    end date
   Social Distancing Campaign
                                       2020-03-22 2020-04-05
    Social Distancing Campaign
                                       2020-04-06
                                                  2020-04-19
```

Social Distancing Campaign

2020년 6월 28일부터 정부는 사회적 거리두기 체계를 3단계로 정비하고, 단계별 전환 기준 및 실행 방안을 제시하였습니다. 이 방안에 맞춰 detail 인덱스의 Strong과 Weak 변수를 각각 거리두기 단계 2와 1로 변경했습니다.

또한 detail 인덱스의 정보를 새롭게 만든 Level 인덱스에 저장했습니다.

end\_date

2020-04-19

2020-03-22 2020-04-05

2020-04-20 2020-05-05

2020-04-06

Weak

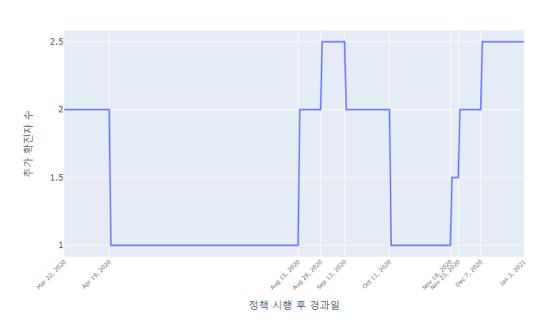
2020-04-20 2020-05-05

```
policy.loc[3] = ['Distancing in daily life', 1 , '2020-05-06', '2020-06-27'] # 생활속 거리두기 20.5.6-20.6.27 policy.loc[4] = ['Social Distancing Campaign', 1, '2020-06-28', '2020-08-15'] # 단계별 적용 - 1단계 거리두기 20.6.25 policy.loc[5] = ['Social Distancing Campaign', 2, '2020-08-16', '2020-08-29'] # 2단계 거리두기 20.8.16-20.8.29/ 8월 policy.loc[6] = ['Social Distancing Campaign', 2.5, '2020-08-30', '2020-09-13'] # 강화 2단계 거리두기 20.8.30-20.9. policy.loc[7] = ['Social Distancing Campaign', 2, '2020-09-14', '2020-10-11'] # 2단계 거리두기 20.9.14-20.10.11 policy.loc[8] = ['Social Distancing Campaign', 1, '2020-10-12', '2020-11-18'] # 1단계 거리두기 20.10.12-20.11.23 policy.loc[9] = ['Social Distancing Campaign', 1.5, '2020-11-19', '2020-11-23'] policy.loc[10] = ['Social Distancing Campaign', 2, '2020-11-24', '2020-12-07'] # 2단계 거리두기 20.11.24-20.12.07 policy.loc[11] = ['Social Distancing Campaign', 2.5, '2020-12-08', '2020-12-23'] # 2단계 거리두기 20.12.8-20.12.23 policy.loc[12] = ['Social Distancing Campaign', 2.5, '2020-12-24', '2021-01-03'] # 5인이상 집합금지 시작
```

#### gov\_policy Level start\_date end\_date 0 Social Distancing Campaign 2 2020-03-22 2020-04-05 1 Social Distancing Campaign 2 2020-04-06 2020-04-19 2 Social Distancing Campaign 1 2020-04-20 2020-05-05 3 Distancing in daily life 1 2020-05-06 2020-06-27 4 Social Distancing Campaign 1 2020-06-28 2020-08-15 5 Social Distancing Campaign 2 2020-08-16 2020-08-29 6 Social Distancing Campaign 2.5 2020-08-30 2020-09-13 7 Social Distancing Campaign 2 2020-09-14 2020-10-11 8 Social Distancing Campaign 1 2020-10-12 2020-11-18 9 Social Distancing Campaign 1.5 2020-11-19 2020-11-23 10 Social Distancing Campaign 2 2020-11-24 2020-12-07 11 Social Distancing Campaign 2.5 2020-12-08 2020-12-23 12 Social Distancing Campaign 2.5 2020-12-24 2021-01-03

경기도 공식 블로그를 참고하여 DACON 데이터에 없는 2020년 거리두기 데이터를 인덱스에 맞게 추가했습니다.

코로나 기간 동안의 거리두기 단계 변화



```
# DACON patientInfo ElOIE
patient_info = pd.read_csv('/content/PatientInfo.csv')
# 기존 DACON의 policy 데이터에 있던 거리두기 정책 3개에 대한 확진자 dataframe 생성
df campaign1 = extract by date(patient info, [policy.loc[0]['start date'], policy.loc[0]['end date']], 'confirmed date')
df_campaign2 = extract_by_date(patient_info, [policy.loc[1]['start_date'], policy.loc[1]['end_date']], 'confirmed_date')
df_campaign3 = extract_by_date(patient_info, [policy.loc[2]['start_date'], policy.loc[2]['end_date']], 'confirmed_date')
# 정책 기간 경과에 따른 확진자 수를 계산한 dataframe 생성
df_campaign1_count = df_campaign1,groupby(['confirmed_date']).size().reset_index(name='campaign1')
df_campaign2_count = df_campaign2,groupby(['confirmed_date']).size().reset_index(name='campaign2')
df_campaign3 count = df_campaign3.groupby(['confirmed_date']).size().reset_index(name='campaign3')
# 시각화를 위해 하나의 dataframe으로 합함
df_campaign = pd.concat([df_campaign1_count['campaign1'], df_campaign2_count['campaign2'], df_campaign3_count['campaign3']], axis=1)
```

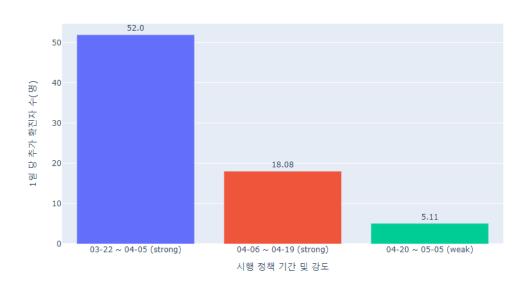
2월부터 4월까지 시행된 거리 두기 정책에 따른 확진자 수



첫 2주간 실행된 2단계 거리두기에서는 초반 9일간은 확진자 수 상승세를 보이다 10일부터 급격히 줄어듦을 볼 수 있습니다.

이후 연장된 2단계 거리두기 정책에서는 꾸준히 확진자가 줄어듦을 볼 수 있으며 약화된 1단계 거리두기 정책시기에도 확진자 수가 줄어듦을 볼 수 있습니다.

2월부터 4월까지 시행된 거리 두기 정책에 따른 1일 당 추가 확진자 수



거리두기 정책 기간이 모두 같지는 않아 1일 당 확진자 수 평균을 확인해본 결과 거리두기 시행에 따라 확진자 수가 줄어듦을 확인할 수 있습니다.

따라서, 4월까지의 거리두기 정책은 효과적이였다 볼 수 있습니다.

```
# 2020년 서울시 코로나19 확진자 데이터
patient_seoul = pd.read_csv('/content/seoul_covid19.csv', encoding='utf-8')
list_campaigns = list()
list names = list()
list_days = list()
# policy dataframe의 기간을 이용해 campaign 별 확진자 dataframe 생성
for index, campaign in enumerate(policy, iterrows()):
 name = '/'.join(campaign[1]['start_date'].split('-')[-2:]) + '~' + '/'.join(campaign[1]['end_date'].split('-')[-2:])
 list names.append(name + '(' + str(campaign[1]['Level']) + ')')
 campaign = extract by date(patient seoul. [campaign[1]['start date']. campaign[1]['end date']]. '확진일')
 campaign = campaign.groupby(['확진일']).size().reset_index(name='count'+str(index))
 list campaigns.append(campaign)
  list_days.append(count_days(campaign, '확진일'))
df campaign = pd.concat([campaign['count'+str(index)] for index. campaign in enumerate(list campaigns)], axis=1)
```

2020년 하반기에는 코로나 확산 정도가 심해지며 정부의 거리두기 정책이 세분화되며 전국적 유행 단계인 2.5단계를 시행하기에 이릅니다. DACON의 데이터만으로는 하반기의 거리두기 단계에 따른 정책의 효율성을 볼 수 없어 서울 열린 데이터 광장의 **서울시 코로나 19 확진자** 현황 데이터와 kaggle의 **Real-time Covid 19 Data**를 추가 확보하여 정책의 효율성을 파악해보았습니다.

거리두기 정책에 따른 확진자 수 변동



sum(파란색)의 경우 해당 정책 기간 동안 추가 확진된 확진자 수의 총합을 의미하며, a day(주황색)은 해당 정책 기간 동안 발생한 확진자 수를 정책 기간으로 나누어 1일 당 추가 확진된 확진자 수를 의미합니다. a day를 기준으로 보았을 때 거리두기 2단계에서 1단계로 바꾸면 확진자수가 늘어남을 볼 수 있습니다. (2020-08-15 이후 확진자 증가, 2020-11-18 이후 확진자 증가로 거리 두기 정책 단계 강화 시행)

확진자 대비 회복 비율 + 거리두기 시행 단계



앞서 보았던 확진자 대비 회복율에 대한 그래프에 2020년에 시행되었던 사회적 거리두기 시점을 더하여 분석해보고자 합니다. 2020년 3월 말에 거리두기 2단계가 시행된 후, 점차 회복세를 보이고 있음을 확인할 수 있으며, 이후 4월까지 유지한 결과 큰 폭의 변화없이 현상유지하고 있음을 확인할 수 있습니다.

8월말에 다시 유행함에 따라, 거리두기 1단계를 시행했으나 큰 효과를 보지 못한채 확진자가 더욱 증가함에 따라 회복율은 감소하고 있음을 확인할 수 있습니다. 따라서 더욱 강화된 2단계, 2.5 단계를 시행하게 되었으며, 이는 다시 회복세로 들어서게 된 계기가 되었음을 확인할 수 있습니다. 이로써 정부의 거리두기 관련 방역 정책이 효과를 보이고 있음을 확인할 수 있습니다.

#### 모빌리티 보고서 분석

```
google_2020_df = pd.read_csv("/content/2020_KR_Region_Mobility_Report.csv")
google_2021_df = pd.read_csv("/content/2021_KR_Region_Mobility_Report.csv")
```

앞에서 살펴본 정책이 사람들에게 영향을 미쳤는지 코로나19 지역사회 이동성 데이터를 통해서 확인해보고자 합니다. 2020~2021년 이동성 데이터를 통해 분석했으며 기준값 0%는 코로나 이전 2020년 1월 3일부터 2020년 2월 6일의 중앙 값을 기준으로 나온 데이터입니다.

또한 해당 데이터는 서울 기준(Seoul Metropolitan)으로 작성하였습니다.

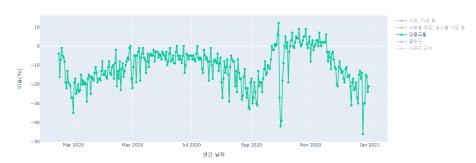
```
google_2020_df1 = google_2020_df.loc[seoul_start_index:, ['date','retail_and_recreation_percent_change_from_baseline']]
google_2020_df2 = google_2020_df.loc[seoul_start_index:, ['date','grocery_and_pharmacy_percent_change_from_baseline']]
google_2020_df3 = google_2020_df.loc[seoul_start_index:, ['date','transit_stations_percent_change_from_baseline']]
google_2020_df4 = google_2020_df.loc[seoul_start_index:, ['date','workplaces_percent_change_from_baseline']]
google_2020_df5 = google_2020_df.loc[seoul_start_index:, ['date','residential_percent_change_from_baseline']]
```

- retail\_and\_recreation : 식당, 카페 데이터
- grocery\_and\_pharamacy : 식료품 매장, 식자재 창고, 농산물 시장 데이터
- transit\_stations: 지하철, 버스, 기차역 데이터
- workplaces : 직장 데이터
- residential: 거주지 근처 데이터

# 모빌리티 보고서 분석







2020년 이동 추이 데이터



→ 식당, 카페 등 → 식료용 매장, 농산물 시장 등 → 대중교통 → 합퇴근



2020년 이동 추이 데이터

# 모빌리티 보고서 분석





연간 날짜

2021년 이동 추이 데이터





# 코로나 전파로 인한 폐업 분석

2020년 02월의 대구 신천지 사태, 2020년 05월의 이태원 클럽 집단 확진 사태, 2020년 08월의 사랑 제일 교회 사태로 사회적 거리두기 정책이 강화되었으며, 모빌리티 보고서에 따르면 유동 인구 또한 적어짐을 확인할 수 있었습니다.

해당 기간에는 집단 확진 사건들이 있었던 대구와 서울을 중심으로 유동 인구 감소로 인한 소상공인들의 경영 악화 문제가 수면 위로 드러났습니다.

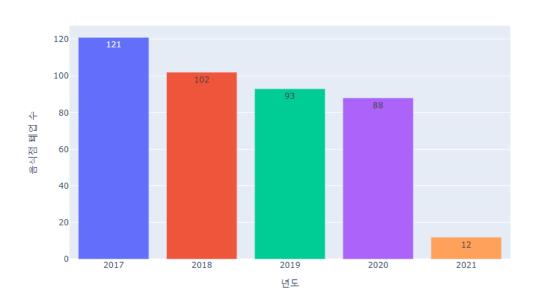
2020, 2021년 대구와 서울의 음식점 폐업을 이전 년도들과 비교해 코로나로 인한 음식점 폐업 정도를 살펴보겠습니다.

# 코로나 전파로 인한 폐업 분석 - 대구

```
# 대구 내 음식점 데이터
closed_daegu = pd.read_csv('/content/res_dague.csv', encoding='utf-8')
# 폐업일자의 format을 datetime으로 변경
closed_daegu['폐업일자'] = pd.to_datetime(closed_daegu['폐업일자'], format='%Y%m%d')
# 5개년도 비교를 위해 2017 이후 폐업 음식점들 추출
closed_daegu = closed_daegu[(closed_daegu['폐업일자'] >= '2017-01-01')]
# resample을 통해 각개년도 1월 1일 기준으로 년도 별 폐업으로 묶음
closed_year = closed_daegu,resample('YS', on='폐업일자').count()
```

# 코로나 전파로 인한 폐업 분석 - 대구

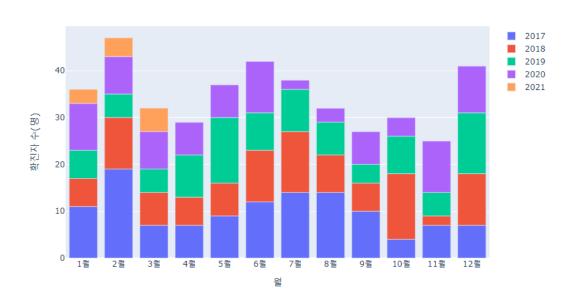
대구 음식점 폐업 수



년도 별 대구 음식점 폐업 수는 예상과는 다르게 오히려 줄고 있습니다. 2021년은 아직 3월까지 데이터만 있는 것을 고려해 4배를 곱해주어도 48건 밖에 되지 않는 것을 확인할 수 있습니다.

# 코로나 전파로 인한 폐업 분석 - 대구

대구 음식점 폐업 수



월 별 데이터에서도 2021(주황색), 2020(보라색)년도의 폐업 수보다 2017(파란색), 2018(빨간색)년의 폐업 수가 더 많은 것을 확인할 수 있습니다.

# 코로나 전파로 인한 폐업 분석 - 서울

df\_서울 = pd.read\_csv("/content/res\_seoul.csv", encoding='utf-8') # 서울데이터 읽어오기

```
      df_서울['인허가일자'] = pd.to_datetime(df_서울['인허가일자'], format='%Yxmxd', errors = 'coerce') #날짜형식변환

      df_서울['페업일자'] = pd.to_datetime(df_서울['페업일자'], format='%Yxmxd', errors = 'coerce') #날짜형식변환

      df_서울['데업일자'] = pd.to_datetime(df_서울['페업일자'], dt.year > 2021)].index, inplace=True)
      # 페업일자 오타로 추정됨 2021년 이후 지무기

      df_서울['df_서울['데업일자'].dt.year < 2017)].index, inplace=True)</td>
      # 四업일자 오타로 추정됨 2021년 이후 지무기

      # 2017년 이전 data drop
      # 2017년 이전 data drop

      df_서울['연도'], df_서울['월'] = df_서울['페업일자'].dt.year, df_서울['페업일자'].dt.month #연도와 월 분리

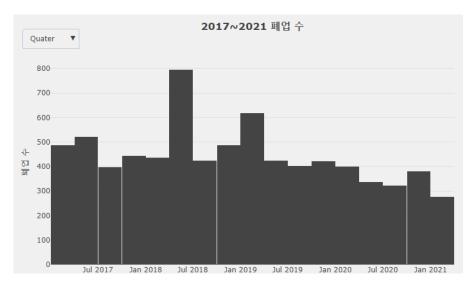
      df_서울['시도','시군구']] = df_서울['소재지전체주소'].str.split(' ', expand=True).iloc[:,[0,1]] #시도 시군구 분리

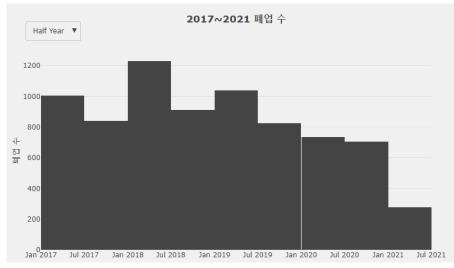
      df_서울['영업기간'] = df_서울['페업일자'] - df_서울['인허가일자'] #

      df_서울 = df_서울['시도','시도구','위생업태명','페업일자','연도','월','인허가일자','사업장명','소재지전체주소','영업기간','좌표정보(x)','좌표정보(y)']]

      df_서울.drop(df_서울['서울['시도'].isnull() == True].index, inplace=True)
```

#### 코로나 전파로 인한 폐업 분석 - 서울





분기별 추세를 살펴보면 소폭 감소하는 것을 알 수 있습니다. 반기별 추세를 보면 아직 1/4밖에 지나지 않아 수치를 가산하더라도 앞 3년 대비 근 2년간은 음식점의 폐업수가 오히려 준 것을 볼 수 있습니다.

#### 코로나 전파로 인한 폐업 분석 - 서울

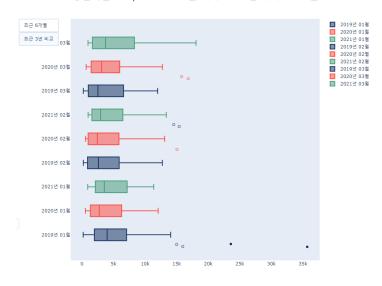




10k

15k

영업기간 Boxplot - 2019년 대비 2020년 대비 2021년



최근 6개월: 중앙값의 변화는 크게 일어나지 않았지만, 조금씩 영업기간이 긴 업체의 폐업이 늘어났다는 것을 알 수 있습니다.

최근 3년 비교: 2019-2020년에는 폐업률이 감소했지만, 21년에는 소폭 증가했고, 특히 21년 3월이 전년의 같은 시점과 대비하여 늘었다는 것을 알 수 있습니다.

폐업에 대한 데이터를 분석해본 결과, 많은 소상공인들이 폐업할 것이라는 예상과는 다르게 폐업하지 않았습니다. 그렇다면 왜 이러한 결과가 발생했을까요?

저희는 모빌리티 데이터에서 다시 한번 그 이유를 추론할 수 있었습니다. 재택이나 거주지의 움직임이 제한되었고, 소상공인과 소비자들은 이에 적응할 필요가 있었습니다. 따라서 저희는 배달앱과 배달대행을 통해 소상공인들이 어려움을 극복해나가지 않았을까 유추했고 이를 분석하기 위해 국가 통계 포털의 배달 앱, 배달 대행 사용량 데이터를 분석했습니다.

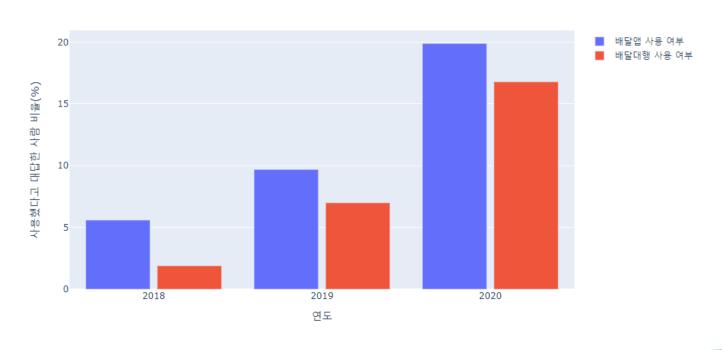
```
# 배달 앱, 배달 대행 사용량 데이터, 한글 사용을 위해 CP949 사용
app_agent_use_df = pd.read_csv("<u>/content/2018-2020_delivery.csv</u>", encoding='CP94<mark>9</mark>')
```

	Region	Year	Type	Usable(Y)	Usable(N)	Cost (Month)
0	Seoul	2018	Арр	5.6	94.4	256369.6
1	Seoul	2018	Agent	1.9	94.4	430653.5
2	Seoul	2019	App	9.7	90.3	338523.9
3	Seoul	2019	Agent	7.0	93.0	729198.3
4	Seoul	2020	Арр	19.9	80.1	729198.3

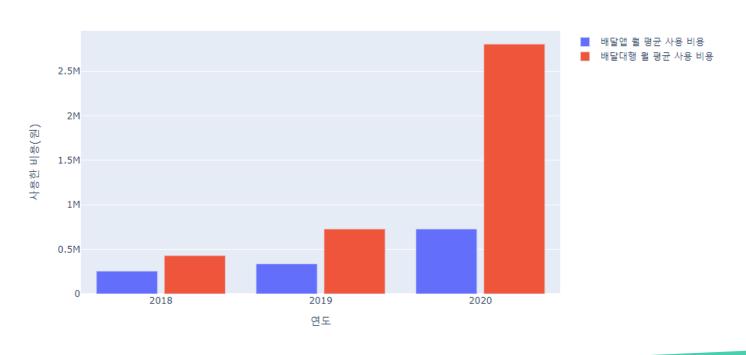
연도별 각 데이터를 확인하기 위해 다음과 같은 전처리를 진행했습니다.

```
app_use_usable = app_agent_use_df.loc[0:6:2, ['Year', 'Usable(Y)']]
app_use_cost = app_agent_use_df.loc[0:6:2, ['Year', 'Cost(Month)']]
agent_use_usable = app_agent_use_df.loc[1:6:2, ['Year', 'Usable(Y)']]
agent_use_cost = app_agent_use_df.loc[1:6:2, ['Year', 'Cost(Month)']]
```

2018~2020년 배달앱/배달대행 사용량



2018~2020년 배달앱/배달대행 월 평균 사용비용



#### 결론

코로나19과 국내 소상공인의 폐업 상황에 얼마나 영향을 미쳤을까?



폐업 점포수의 감소..왜?



배달 앱 / 배달 대행을 통해서 위기 극복을 하고 있음을 추론 가능 + 정부의 지원 정책

폐업 비용의 부담으로 인해 폐업 결정 지연

(http://newslabit.hankyung.com/article/202005071929G)

# 결론

지금은 2021년이고 5인 이상 거리두기 정책이 시행된 지 5개월이 되어가는 시점이며 백신 보급이 진행되고 있습니다. 하지만 여전히 집단 감염과 코로나 변종 바이러스 등장으로 코로나 확진자 수는 줄어들지 않고 있습니다.

분서한 바와 같이 오랫동안 진행되고 있는 코로나 사태를 버티기 위해 사람들의 배달 앱 사용량은 늘고 있으며, 이러한 사람들의 요구를 빠르게 받아 들여 배달 서비스를 함께 제공하는 상인들이 있습니다.

하지만 여전히 여건과 비용의 문제, 시스템 사용에 익숙하지 못한 문제들로 인해 많은 소상공인들은 코로나 19의 사각지대에 놓여있습니다. 이러한 소상공인들을 위해 정책적으로 배달 시스템을 이용할 수 있는 교육과 지원이 필요합니다.

더 나아가, 저희의 분석이 단순 배달앱/배달대행 사용량으로 그치는 것이 아니라 올해, 의학적으로 코로나 19를 극복하기 위해 백신이 필요한 것처럼 소상공인들에게서 코로나 19를 극복하기 위한 직접적인 정책이 펼쳐졌으면 좋겠습니다.

# 출처

#### 도입부 기사 통계 뉴스

https://biz.chosun.com/site/data/html\_dir/2020/04/10/2020041003722.html?utm\_source=urlcopy&utm\_medium=share&utm\_campaign=biz

배달앱 및 배달대행 이용현황: https://kosis.kr/statHtml/statHtml.do?orgId=114&tblId=DT\_114054\_016

구글 모빌리티 이동 보고서: https://www.google.com/covid19/mobility/

서울시 코로나19 확진자 현황: <a href="http://data.seoul.go.kr/dataList/OA-20279/S/1/datasetView.do">http://data.seoul.go.kr/dataList/OA-20279/S/1/datasetView.do</a>

확진자 정보 데이터(DACON): https://dacon.io/competitions/official/235590/data/

Real-time Covid 19 Data(Kaggle): <a href="https://www.kaggle.com/gauravduttakiit/covid-19">https://www.kaggle.com/gauravduttakiit/covid-19</a>

# Q&A