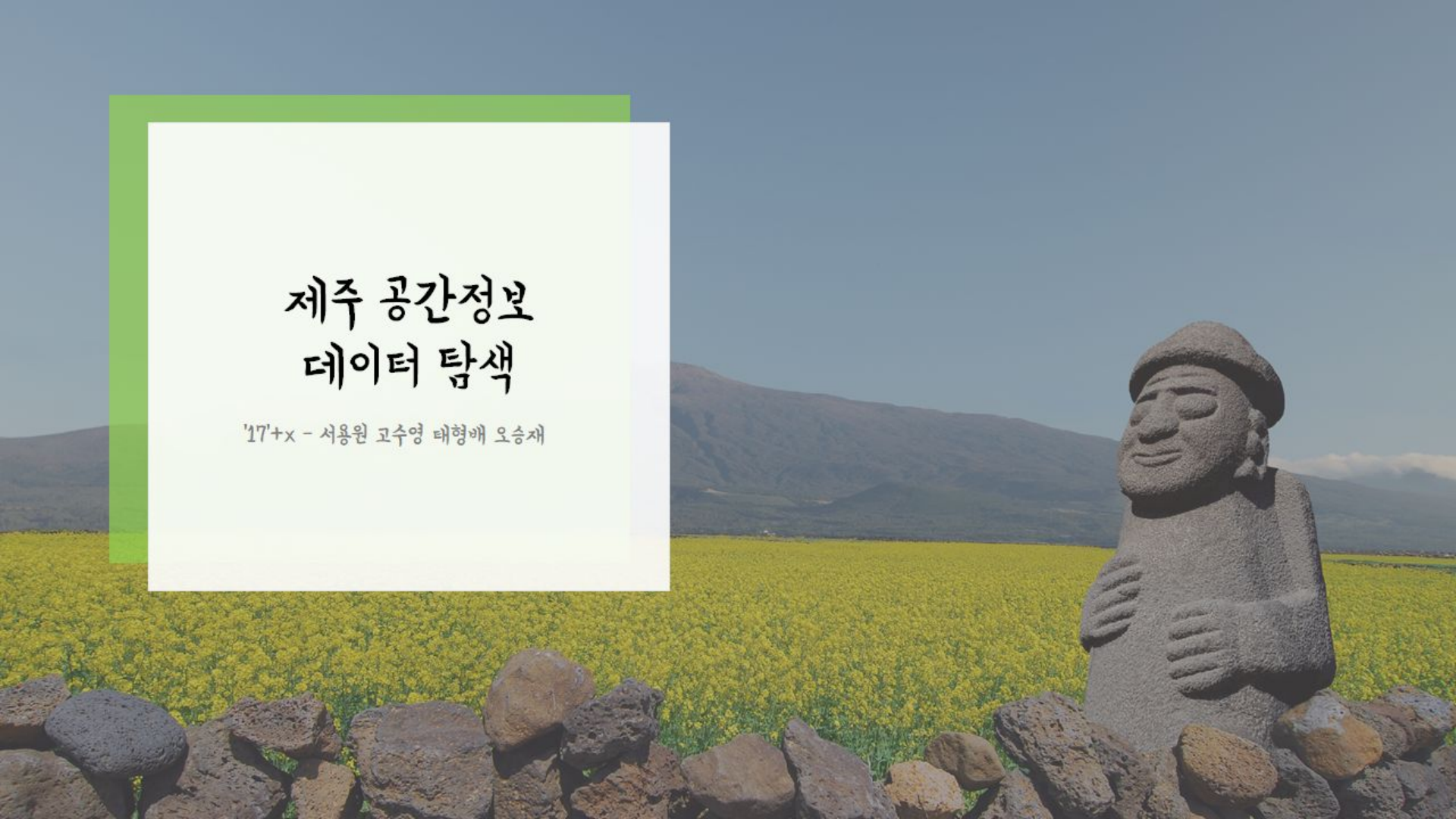


# 제주 공간정보 데이터 탐색

'17'+x - 서용원 고수영 태형배 오승재



# 데이터 설명

# 프로젝트 소개



**공간정보 탐색적 데이터 분석 경진대회**  
국토도시 빅데이터 원터스쿨 | 국토연구원 | 시각화 | 유저평가

🏆 상금 : 100만원

🕒 2020.12.14 ~ 2020.12.27 17:59 [+ Google Calendar](#)

👥 455팀 🏠 마감

[🔗](#) [참여중](#)

## 1. 주제

- 공간 정보를 활용한 탐색적 데이터 분석 경진대회

## 2. 배경

- 공간 정보를 활용하여 올 5~8월까지의 제주 지역 데이터를 분석하여 다양한 인사이트 발굴

## 3. 목표

- 공간 정보를 활용한 탐색적 데이터 분석 및 시각화
- 공간 정보에 대한 일반인의 관심을 제고할 수 있는 인사이트 발굴

## 4. 주최 / 주관

- 주최 : 국토연구원
- 주관 : DACON

# 데이터 설명

data.zip(27.8MB)

- KRI-DAC\_Jeju\_data5.txt(37MB)
- KRI-DAC\_Jeju\_data6.txt(38MB)
- KRI-DAC\_Jeju\_data7.txt(41.6MB)
- KRI-DAC\_Jeju\_data8.txt(37.4MB)

## 데이터 정의

- YM : 기준년월
- SIDO : 지역대분류명
- SIGUNGU : 지역중분류명
- FranClass : 소상공인구분
- Type : 업종명
- Time : 시간대
- TotalSpent : 총사용금액
- DisSpent : 재난지원금 사용금액
- NumOfSpent : 총 이용건수
- NumOfDisSpent : 총 재난지원금 이용건수
- POINT\_X, POINT\_Y : X,Y 좌표

YM	SIDO	SIGUNGU	FranClass	Type	Time	TotalSpent	DisSpent	NumOfSpent	NumOfDisSpent	POINT_X	POINT_Y
202005	제주특별자치도	제주시	영세	일반한식	00시	363000	66500	10	2	877006	1.47977e+06
202005	제주특별자치도	제주시	영세	단란주점	00시	1180000	0	3	0	877006	1.47982e+06
202005	제주특별자치도	제주시	중소1	편의점	00시	157670	6850	20	2	877057	1.47962e+06
202005	제주특별자치도	제주시	영세	편의점	00시	46600	0	2	0	877056	1.47977e+06
202005	제주특별자치도	제주시	영세	주점	00시	66000	0	2	0	877055	1.47987e+06
202005	제주특별자치도	제주시	영세	스낵	00시	159000	0	3	0	877205	1.47987e+06
202005	제주특별자치도	제주시	영세	서양음식	00시	26000	26000	1	1	877377	1.48577e+06
202005	제주특별자치도	제주시	중소1	편의점	00시	215910	34250	27	4	877431	1.48502e+06
202005	제주특별자치도	제주시	영세	스낵	00시	53000	0	2	0	877430	1.48507e+06
202005	제주특별자치도	서귀포시	중소	편의점	00시	21420	0	1	0	878884	1.47378e+06

# 데이터 업종 분류작업

- 업종별 분석을 하기 위해서 데이터 업종 분류작업을 하였습니다.
- 기존 데이터의 업종이 206개 정도였고 겹치는 부분이 많았기 때문입니다. (\*데이콘에서 참고했습니다.)

205개  20개

## < 분류 코드 >

`all_data.groupby("Type").size()` # 205개 정도의 업종이 존재하는 것을 확인 할 수 있습니다.

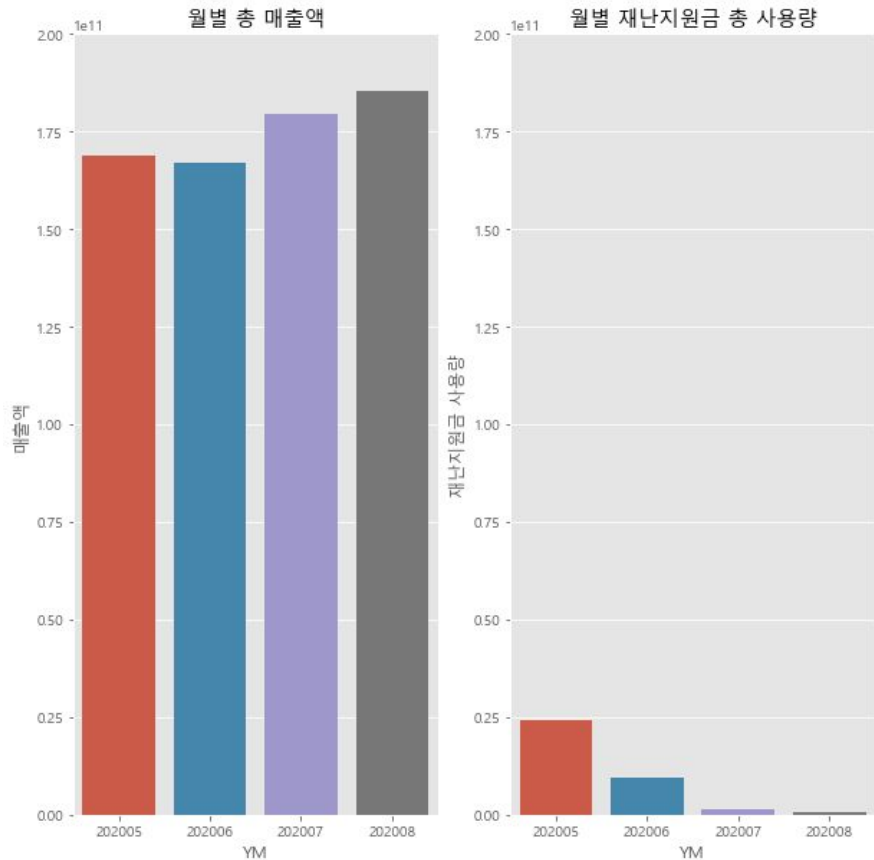
```
def trans_type(df):
    data=df.Type.copy()
    data.replace(dict.fromkeys(['택시', '기타교통수단', '통신기기'], '교통/통신'))
    data.replace(dict.fromkeys(['대형할인점', '농축협직영매장', '농협하나로클럽', '편의점', '슈퍼마켓', '정육점', '기타유통업',
                                '연쇄점', '상품권', '인삼제품', '홍삼제품', '농축수산물', '주류판매점', '기타건강식', '건강식품(회원제형태)',
                                '마트/편의점(유통)'], '마트/편의점(유통)', inplace=True)
    data.replace(dict.fromkeys(['기타농업관련', '비료농약사료종자', '농기계', '미곡상'], '농업'), inplace=True)
    data.replace(dict.fromkeys(['사무서비스', '기타대인서비스', '화물운송', '보관창고업', '종합용역', '조세서비스', '가례서비스',
                                '공공요금', '위탁급식업', '기타용역서비스', '기타수리서비스', '기타운송', '가정용품수리', '견인서비스', '부동산중개임대',
                                '부동산산양', '신변잡화수리', '사무통신기기수리', '법률회계서비스', '사무서비스(회원제형태)', '정보서비스', '기타업종',
                                '기타비영리유통', '서비스/기타'], '서비스/기타'), inplace=True)

    ...

    df["MainType"]= data
    return df
```

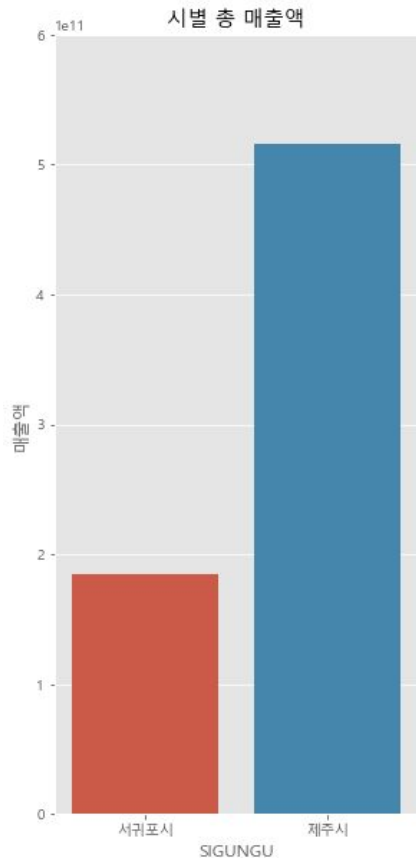
# 데이터 탐색 / 시각화

# 월별 매출액 비교



- 월별 매출액을 비교해 보았을때 점점 증가하는 양상
- 월별 재난지원금 사용금액에서 **5, 6, 7, 8**월로 갈수록 금액이 줄어들었다.
- 사람들이 코로나 사회에 적응해 감에 따라 코로나 발생 직후 많이 떨어져 있던 소비가 점점 올라오는 것이라는 것을 알 수 있습니다.
- **20년 5월**에 1차 재난지원금을 지급받고 사람들이 비교적 바로바로 사용했다는 것을 알 수 있습니다.

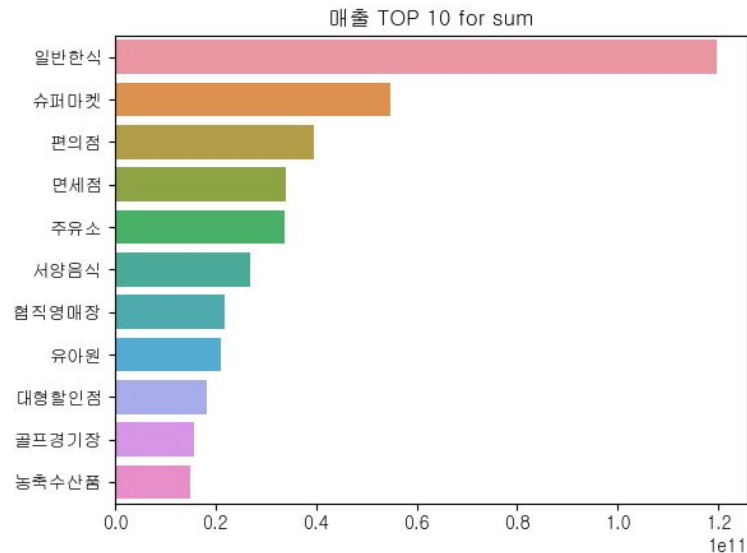
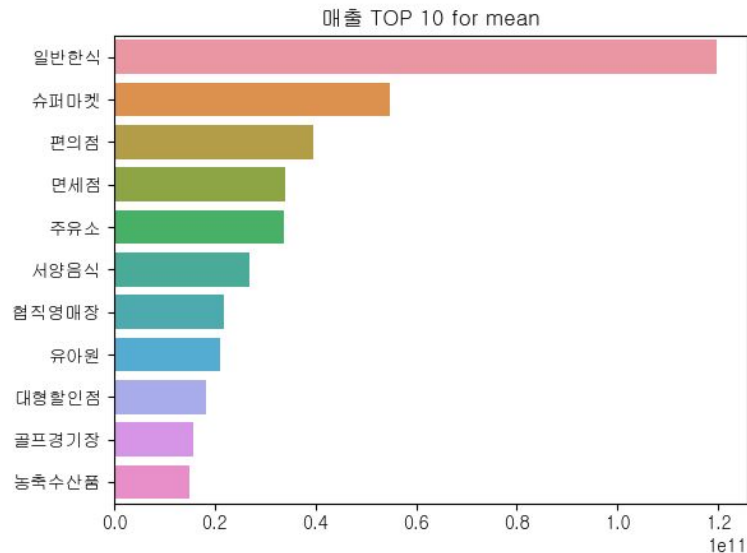
# 지역구에 따른 매출비교 총매출액/재난지원금 사용금액



- 제주도는 제주시와 서귀포시로 구성
- 제주시가 서귀포시보다 약 **2.5배** 정도 매출액이 크게 나타남  
→ 이유 : 제주시 인구가 서귀포시 인구의 **2.5배** 정도 되며 제주시에 공항 등의 편의 시설이 많이 분포 매출액의 차이가 존재
- 총 매출액과 총 재난지원금 사용금액의 규모를 비교해 보았을때 재난지원금의 규모가 매출액에서 큰 비중을 차지하는 수준은 아니라고 판단됨

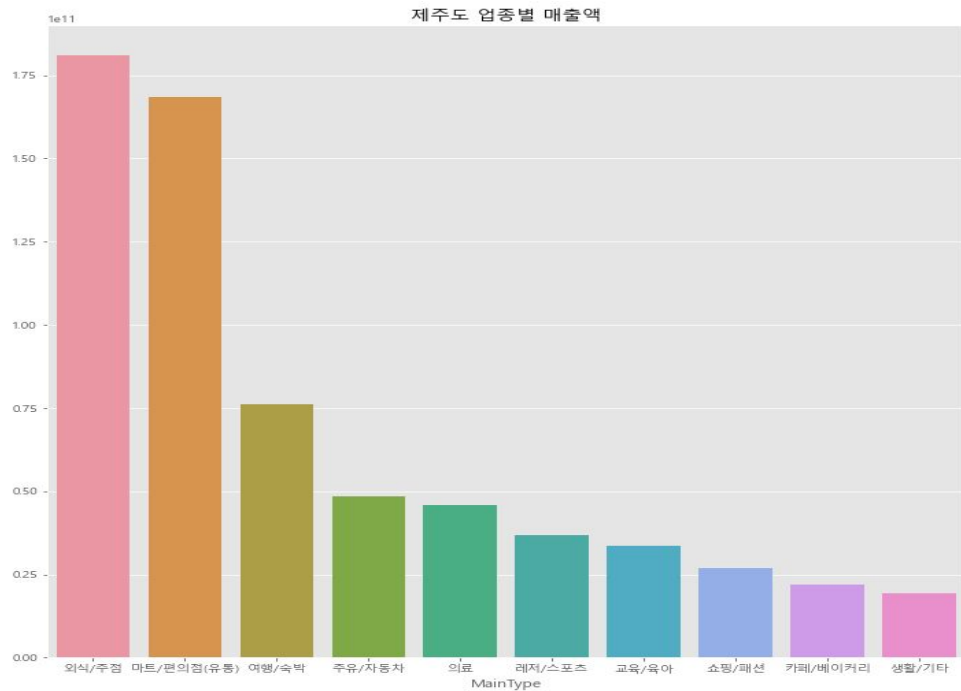


# 총매출액 TOP 10 분석 소분류



- 가장 많이 나온 매출 TOP 10은 일반한식, 슈퍼마켓, 편의점, 면세점, 주유소, 서양음식, 협직영매장, 유아원, 대형할인점, 골프경기장, 농축산물 순이다.
- 외식이나 장보기, 쇼핑 등 을 많이 했음을 알수 있다.
- 평균(mean)으로 분석한 것과 합계(sum)으로 분석한 것의 매출 TOP 10의 차이는 없다.

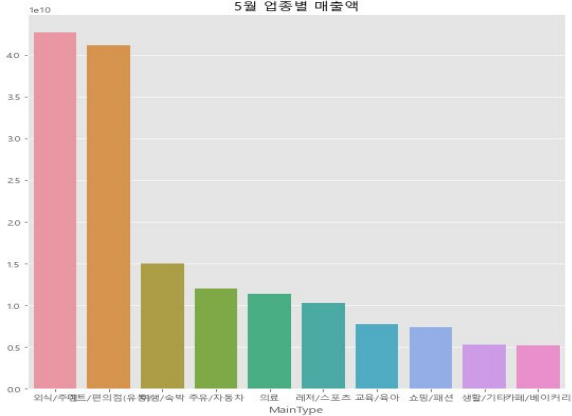
# 총매출액 TOP 10 분석 대분류



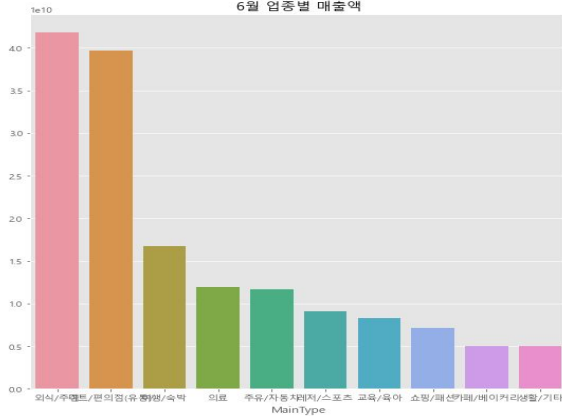
- 업종별 매출액 상위 10개 다른 지역들과 마찬가지로 외식/주점, 마트/편의점에서 매출이 높은 것을 확인
- 특이점은 여행/숙박이 업종별 매출의 3위를 기록하고 있다는 점입니다. 이것은 관광업이 발달한 제주도의 특징으로 볼 수 있습니다.

# 총매출액 TOP 10 분석 월별 대분류

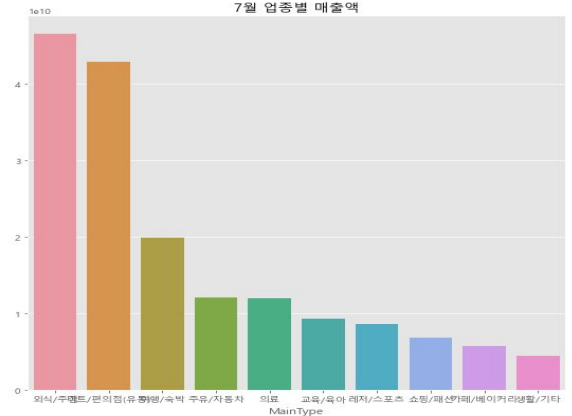
5월 업종별 매출액



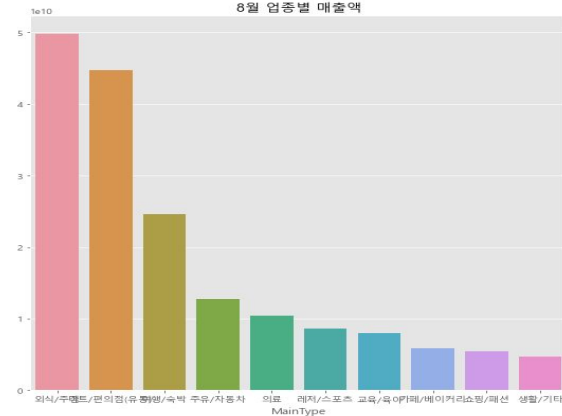
6월 업종별 매출액



7월 업종별 매출액



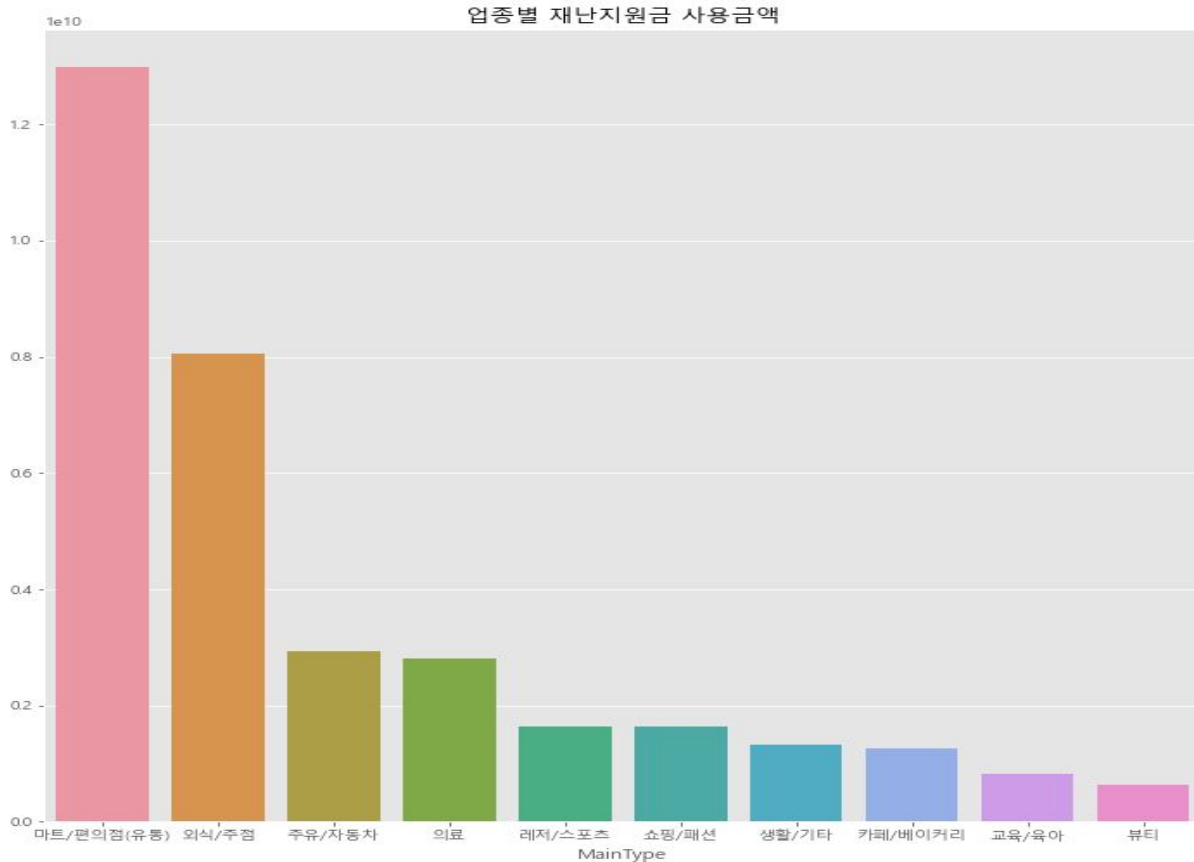
8월 업종별 매출액



● 월별 업종별 매출액의 순위에는 큰 변화가 없는것을 확인 할 수 있습니다.

● 4개월 정도의 데이터 밖에 있지 않기 때문에 특정 월과 매출과의 관계를 정확하게 파악하기는 어렵다는 판단

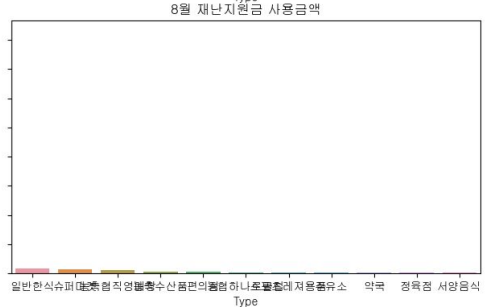
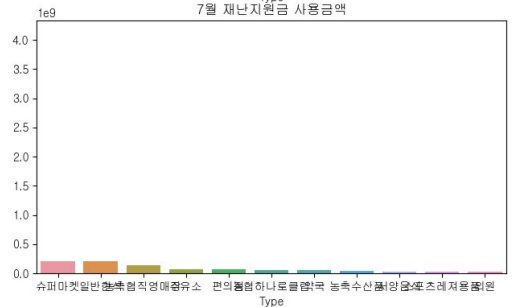
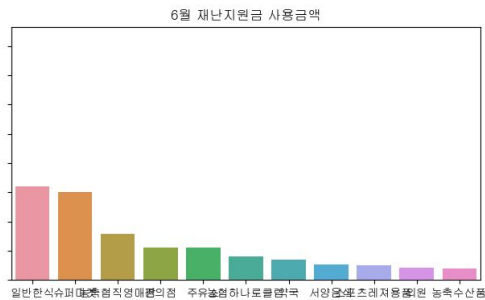
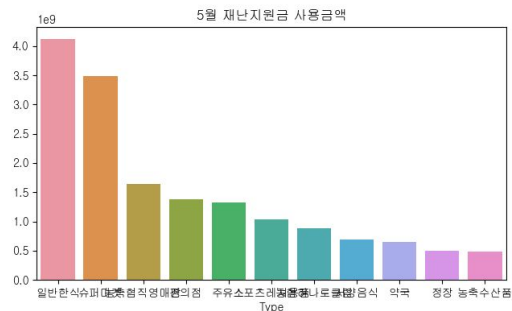
# 재난지원금 사용금액 TOP 10 분석 대분류



- 매출액 순위와 비교했을 때 단순 매출 상위 10개에서는 여행/숙박이 3위를 기록했지만 재난지원금 사용금액의 상위 10개에는 여행/숙박이 없는 것을 확인 할 수 있습니다.

- 이것으로 보아 재난지원금이 여행등과 같은 여가 비용이 아닌 생활비용으로 많이 사용되었다는 것을 알 수 있습니다.

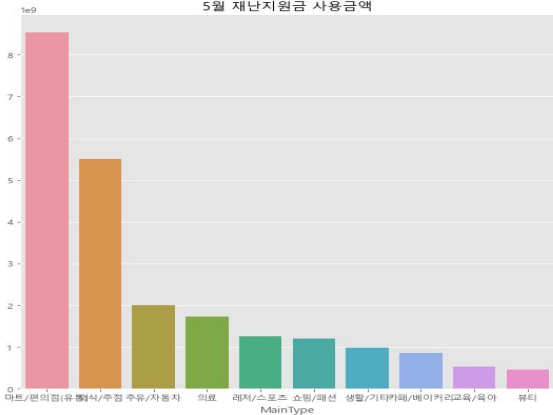
# 재난지원금 사용금액 TOP 10 분석 월별 소분류



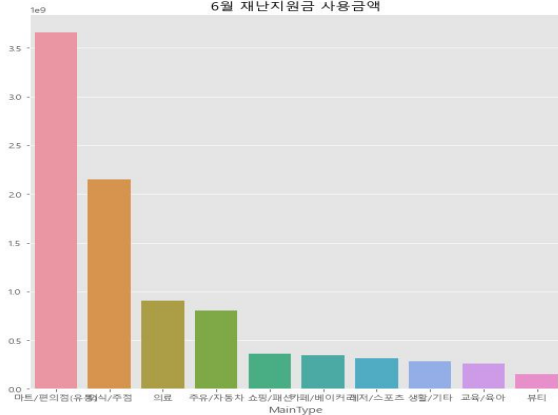
5월달에는 재난지원금 사용을 많이 했지만 6, 7, 8월로 갈수록 재난지원금 사용금액이 확연히 줄은 것을 볼 수 있다.

# 재난지원금 사용금액 TOP 10 분석 월별 대분류

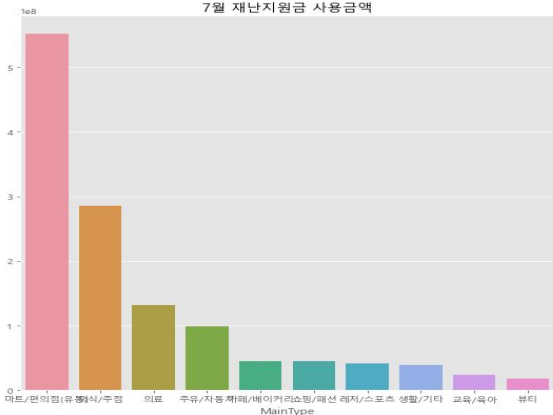
5월 재난지원금 사용금액



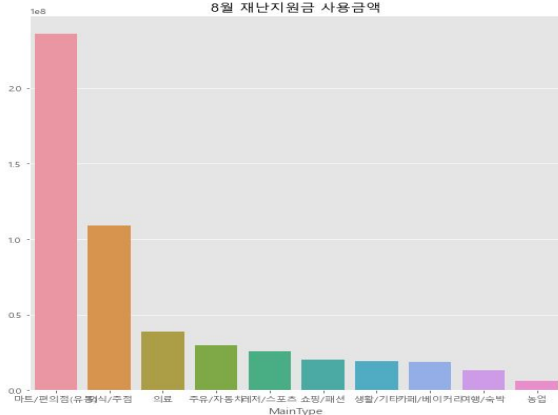
6월 재난지원금 사용금액



7월 재난지원금 사용금액



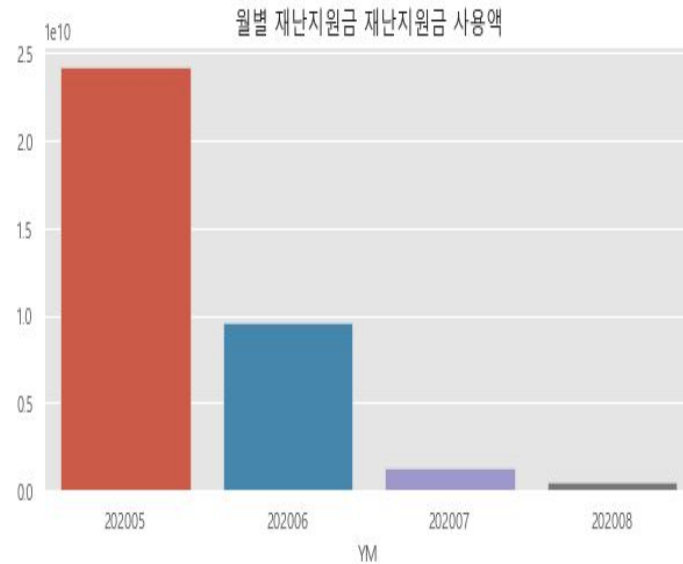
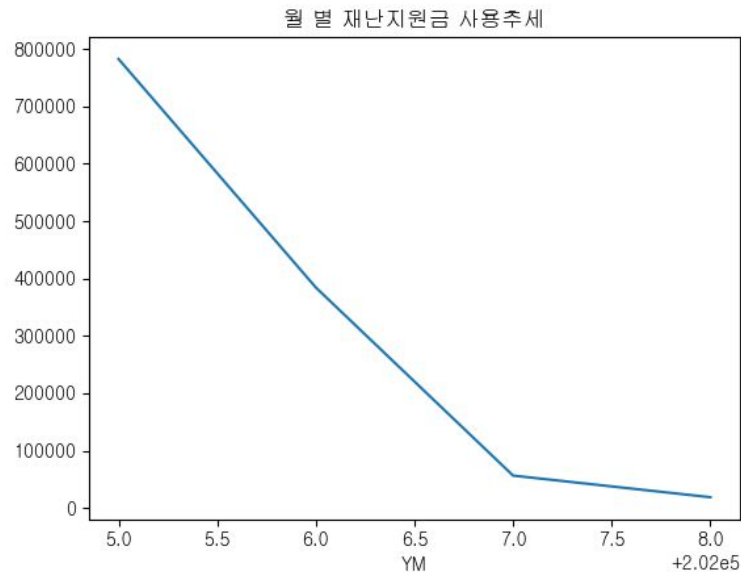
8월 재난지원금 사용금액



● 업종별 매출액 비교 그래프와 마찬가지로 월별로 큰 차이가 없는 것을 확인 할 수 있었습니다.

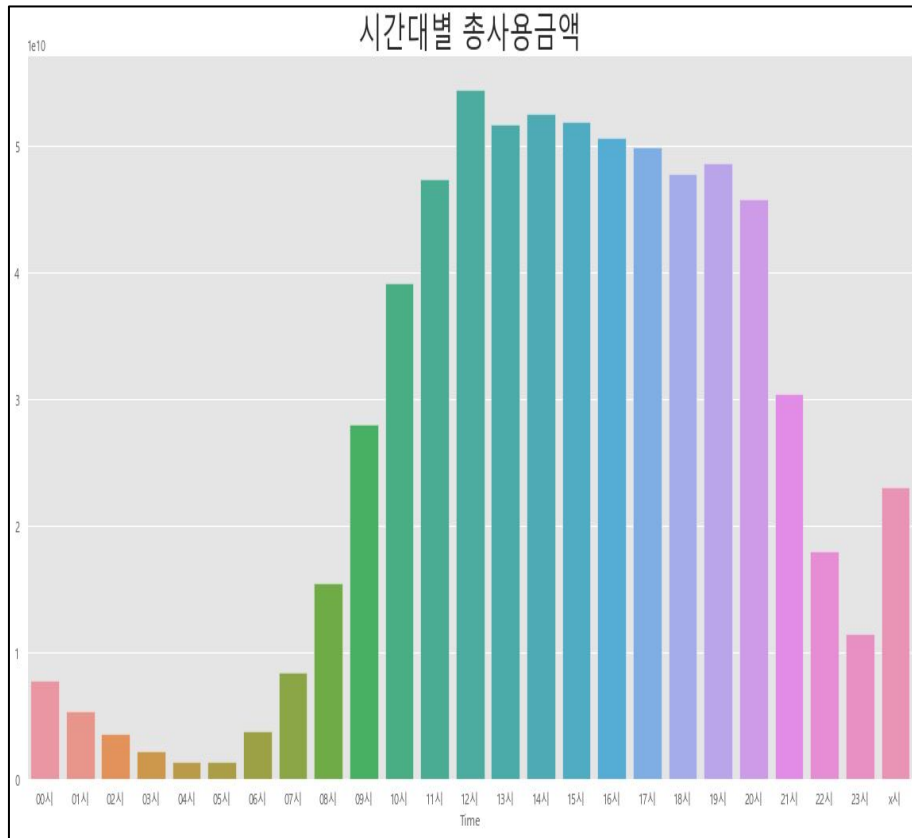
● 월별 매출액 그래프와 같이 특정 달과 재난지원금 사용금액 간의 정확한 관계를 파악하기는 어렵습니다.

# 월별 재난지원금 사용추세



5월에서 8월로 갈수록 재난지원금 사용 추세가 하락세 인것을 볼 수 있다.

# 시간대별 총사용금액

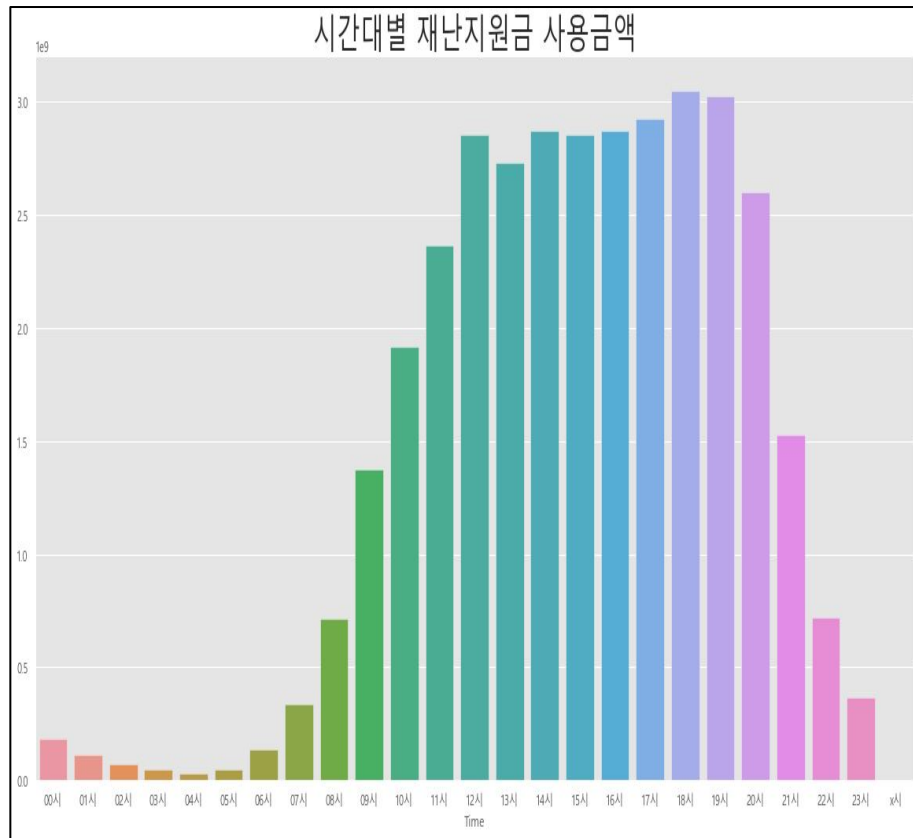


```
plt.figure(figsize=(20,10))
plt.title('시간대별 총사용금액', fontsize=30)
data=pd.concat([x,y,z,w], axis=0)
x=data.groupby('Time')['TotalSpent'].sum()
sns.barplot(y=x.values,x=x.index)
```

- 시간대별 총사용 금액(5월~8월 합한 값)
- 수치가 가장 높은 구간: 12시
- 수치가 가장 낮은 구간: 4시~5시
- 05시부터 12시까지 상승세를 보이다가 14시부터 사용금액이 감소 그러다 다시 23시부터 24시까지 증가함을 보입니다.
- 유동인구가 많을 것이라고 기대되는 시간에 소비가 증가하는 것을 알 수 있습니다.



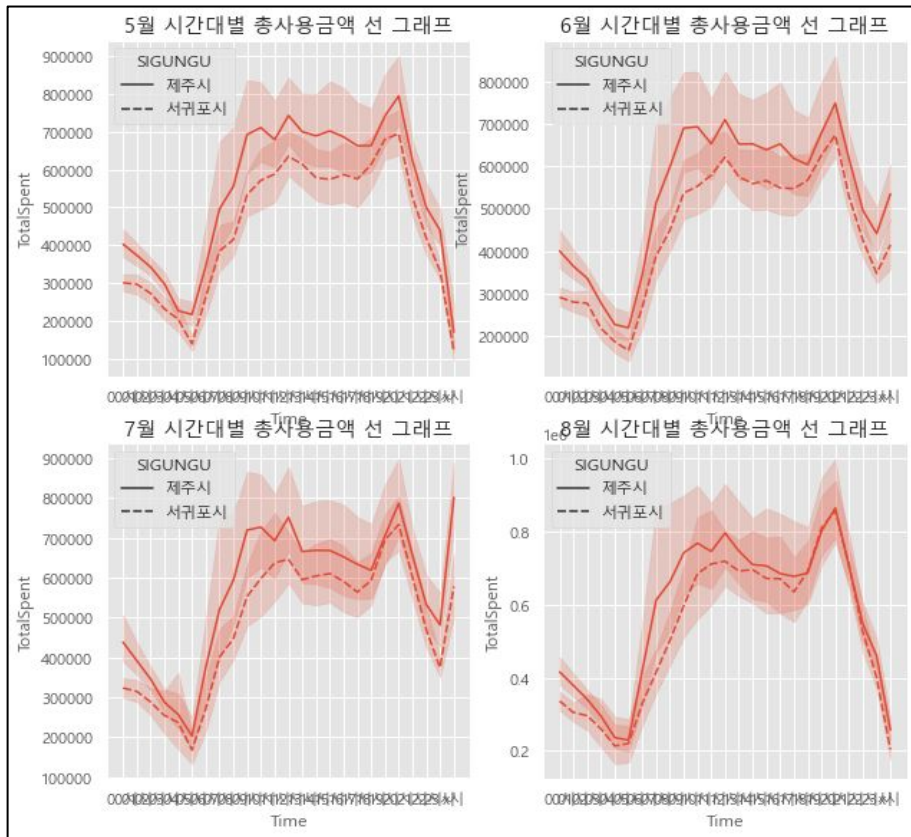
# 시간대별 재난지원금 사용금액



```
plt.figure(figsize=(20,10))
plt.title('시간대별 총사용금액', fontsize=30)
data=pd.concat([x,y,z,w], axis=0)
x=data.groupby('Time')['TotalSpent'].sum()
sns.barplot(y=x.values,x=x.index)
```

- 시간대별 재난지원금 사용금액 (5월~8월 합한 값)
- 수치가 가장 높은 구간: 18시
- 수치가 가장 낮은 구간: 4시
- 4시부터 18시까지 상승세가 보입니다.
- 19시부터 재난지원금 사용이 하강세를 보입니다.

# 시간대별 총사용금액 꺾은선 그래프

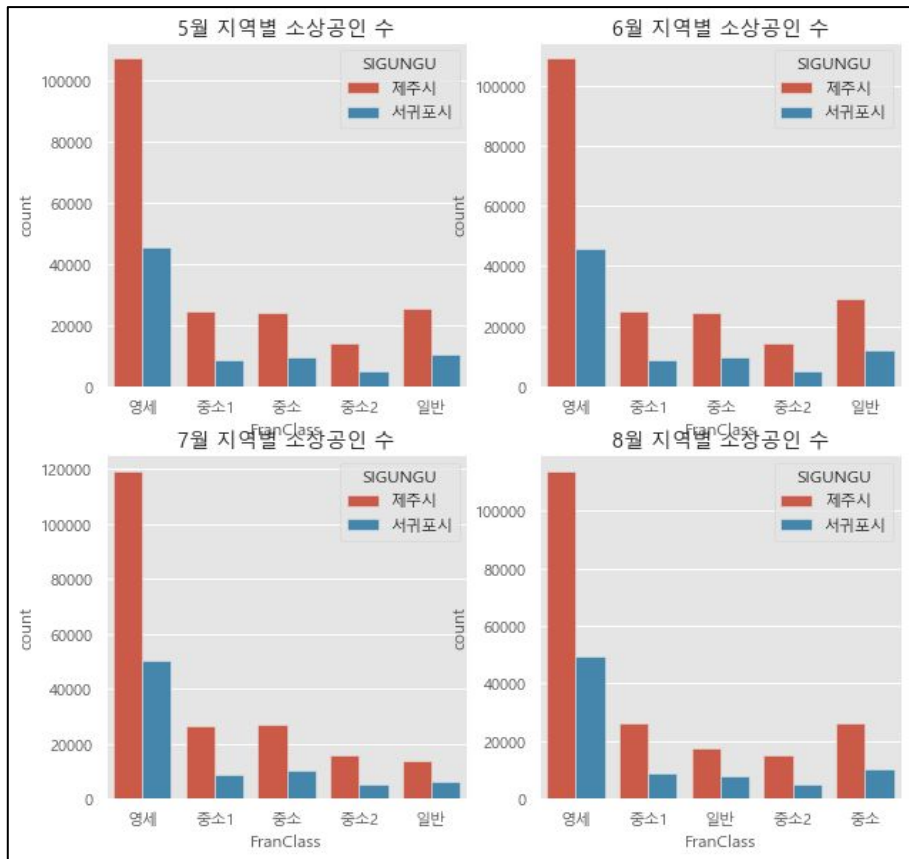


```
plt.figure(figsize=(10,10))
plt.subplot(221)
plt.title('5월 시간대별 총사용금액 선 그래프')
sns.lineplot(x='Time', y='TotalSpent', data=x, style='SIGUNGU')
plt.subplot(222)
plt.title('6월 시간대별 총사용금액 선 그래프')
sns.lineplot(x='Time', y='TotalSpent', data=y, style='SIGUNGU')
plt.subplot(223)
plt.title('7월 시간대별 총사용금액 선 그래프')
sns.lineplot(x='Time', y='TotalSpent', data=z, style='SIGUNGU')
plt.subplot(224)
plt.title('8월 시간대별 총사용금액 선 그래프')
sns.lineplot(x='Time', y='TotalSpent', data=w, style='SIGUNGU')
```

- 수치가 가장 낮은 구간: 4시~5시
- 수치가 가장 높은 구간: 20시~22시
- 제주시가 서귀포시에 비해 비교적 시간대별 총 사용량이 많다
- 시별 사용금액의 분포 또한 제주시가 넓게 나타남.

# 소상공인구분에 따른 탐색

지역구에 따른 월별 분포



```
plt.figure(figsize=(10,10))
plt.subplot(221)
plt.title('5월 지역별 소상공인 수')
sns.countplot(x='FranClass', hue='SIGUNGU', data = x)
```

```
plt.subplot(222)
plt.title('6월 지역별 소상공인 수')
sns.countplot(x='FranClass', hue='SIGUNGU', data = y)
```

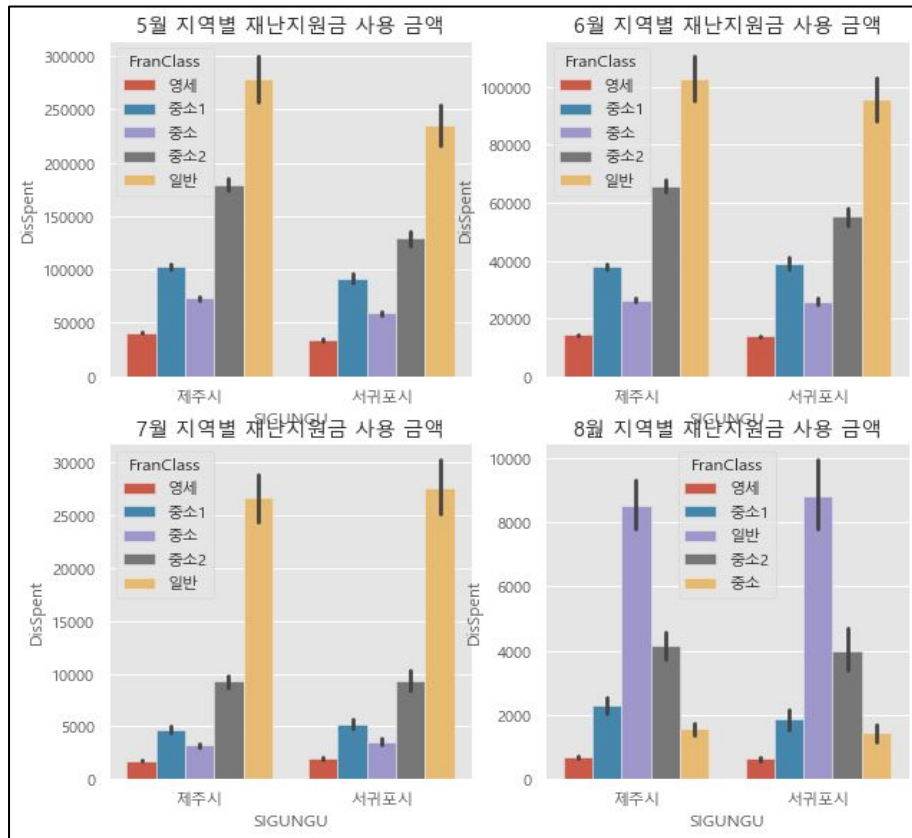
```
plt.subplot(223)
plt.title('7월 지역별 소상공인 수')
sns.countplot(x='FranClass', hue='SIGUNGU', data = z)
```

```
plt.subplot(224)
plt.title('8월 지역별 소상공인 수')
sns.countplot(x='FranClass', hue='SIGUNGU', data = w)
```

지역별 소상공인들의 수를 비교해보면 제주시가 대체적으로 모든 소상공인에서 높다는 것을 알 수 있습니다.  
>> 제주시에 더 많은 소상공인이 있다는 것을 알 수 있습니다.

# 소상공인구분에 따른 탐색

## 지역구에 따른 월별 재난지원금 사용금액



```
plt.figure(figsize=(10,10))
plt.subplot(221)
plt.title('5월 지역별 재난지원금 사용 금액')
sns.barplot(y='DisSpent',x='SIGUNGU',data=x,hue='FranClass')
```

```
plt.subplot(222)
plt.title('6월 지역별 재난지원금 사용 금액')
sns.barplot(y='DisSpent',x='SIGUNGU',data=y,hue='FranClass')
plt.subplot(223)
```

```
plt.title('7월 지역별 재난지원금 사용 금액')
sns.barplot(y='DisSpent',x='SIGUNGU',data=z,hue='FranClass')
```

```
plt.subplot(224)
plt.title('8월 지역별 재난지원금 사용 금액')
sns.barplot(y='DisSpent',x='SIGUNGU',data=w,hue='FranClass')
)
```

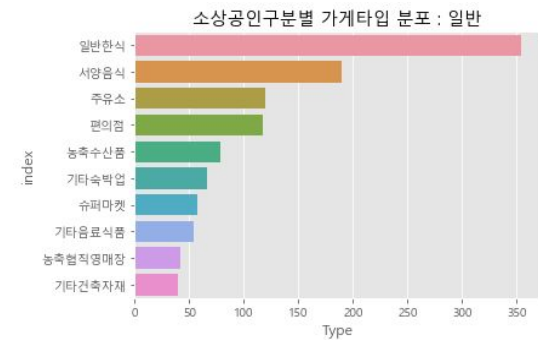
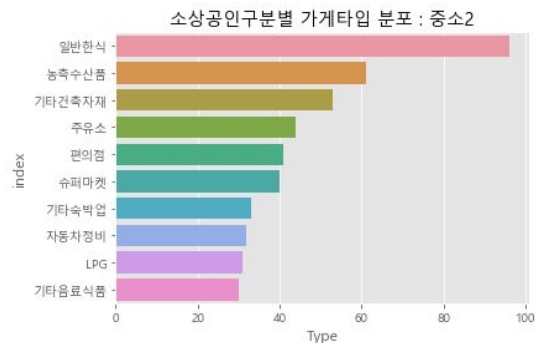
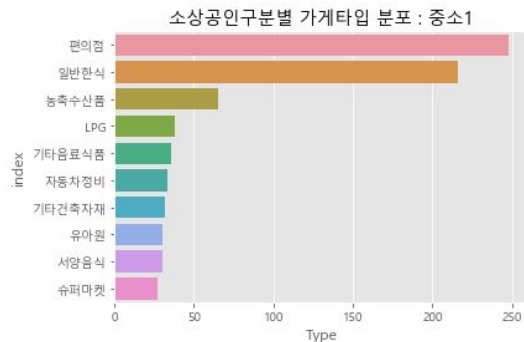
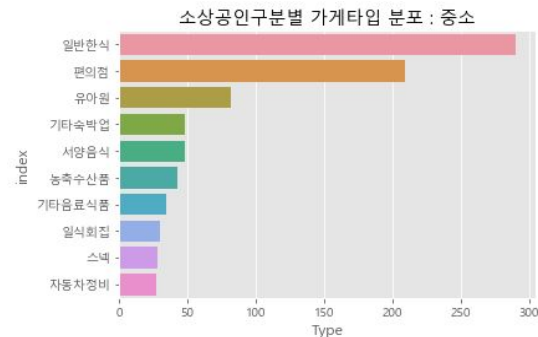
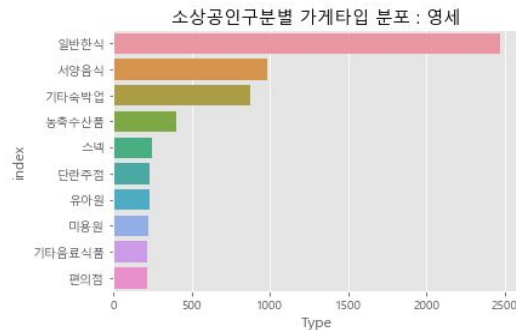
모든 달에 소상공인중에서 일반 부분에서 재난지원금이 가장 많이 사용되었고 가장 적은 부분은 영세 부분임을 알수 있습니다.

# 좌표데이터 활용

## : 맛집 탐색

# 소상공인구분별 가게타입 분포

■ 소상공인구분 (20년 ...	
구분	정의
영세	연매출 3억 이하
중소	5억이하
중소1	10억이하
중소2	30억이하
일반(대형)	30억초과



# 좌표 변환

- 데이터에서 주어진 X,Y 좌표가 우리가 흔히 사용하는 좌표가 아니기 때문에 변환이 필요합니다.(\*데이콘 참고)

```
from pyproj import Proj, transform
```

```
def lation(df):
```

```
    proj_ITRF2000 = Proj(init='epsg:5179')
```

```
    proj_WGS84 = Proj(init='epsg:4326')
```

```
    x,y=list(df['POINT_X']), list(df['POINT_Y'])
```

```
    X,Y=transform(proj_ITRF2000, proj_WGS84, x, y)
```

```
    df['TRANS_X']=X
```

```
    df['TRANS_Y']=Y
```

```
    df.head(3)
```

```
    return df
```

- 위의 작업을 마치면 우리가 흔히 사용하는 위도, 경도 좌표를 데이터에 추가할 수 있습니다.

# 분석방향

별첨 : 사용불가업종 세부

No	업종명	No	업종명
1	주류판매(유통)	21	농협(상물권)
2	커피, 스탠드	22	귀금속, 금, 은, 보석
3	단란주점	23	세금공과금
4	유흥주점	24	면세점
5	실내골프장	25	백화점
6	실외골프장	26	CATV 상품판매
7	노래방	27	기타4
8	총포류 판매	28	전자상거래(다품목취급)
9	카지노	29	결제대행(PG)
10	성인용품판매점	30	생명보험
11	안마업	31	손해보험
12	스포츠마사지	32	손해보험특종
13	룸싸롱	33	손해보험장기, 기존
14	국장식당	34	손해보험자동차
15	나이트클럽	35	손해보험장기, 신규
16	카바레	36	합인점/슈퍼마켓 *일부 대형업체만 제한 - 이마트/롯데마트/홈플러스(SSM 매장 포함)
17	비디오방/전화방		
18	어린이집/유치원	37	가전, 가전용품 *일부 대형업체만 제한
19	상품권 판매		하이마트/전자랜드/LG 전자베스트샵/삼성디지털프라자
20	복권 판매		

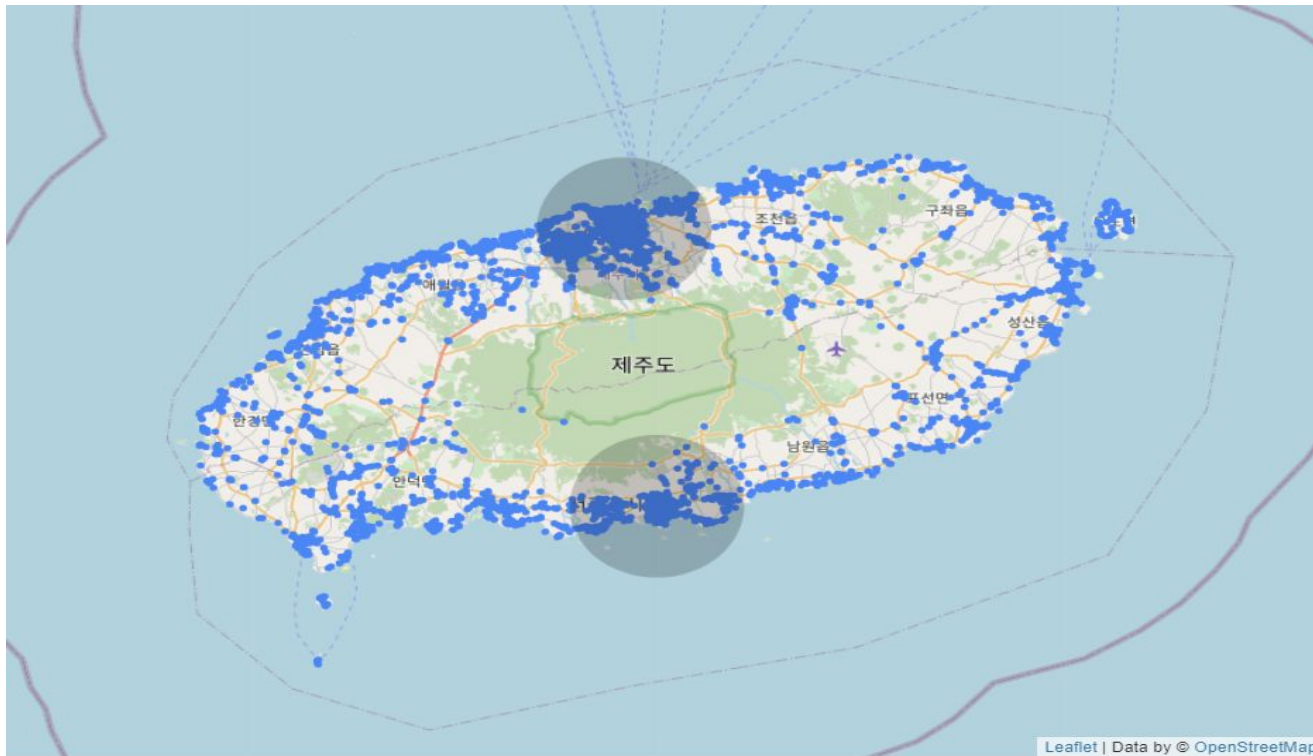
'서양음식', '일반한식',  
'일식회집', '중국음식'에 대해서만 탐색



```
aaa.replace(dict.fromkeys(['서양음식',  
'일반한식', '일식회집', '중국음식'],  
'외식'), inplace=True)  
data['Maintype']=aaa
```



# 제주도 음식점 분포도

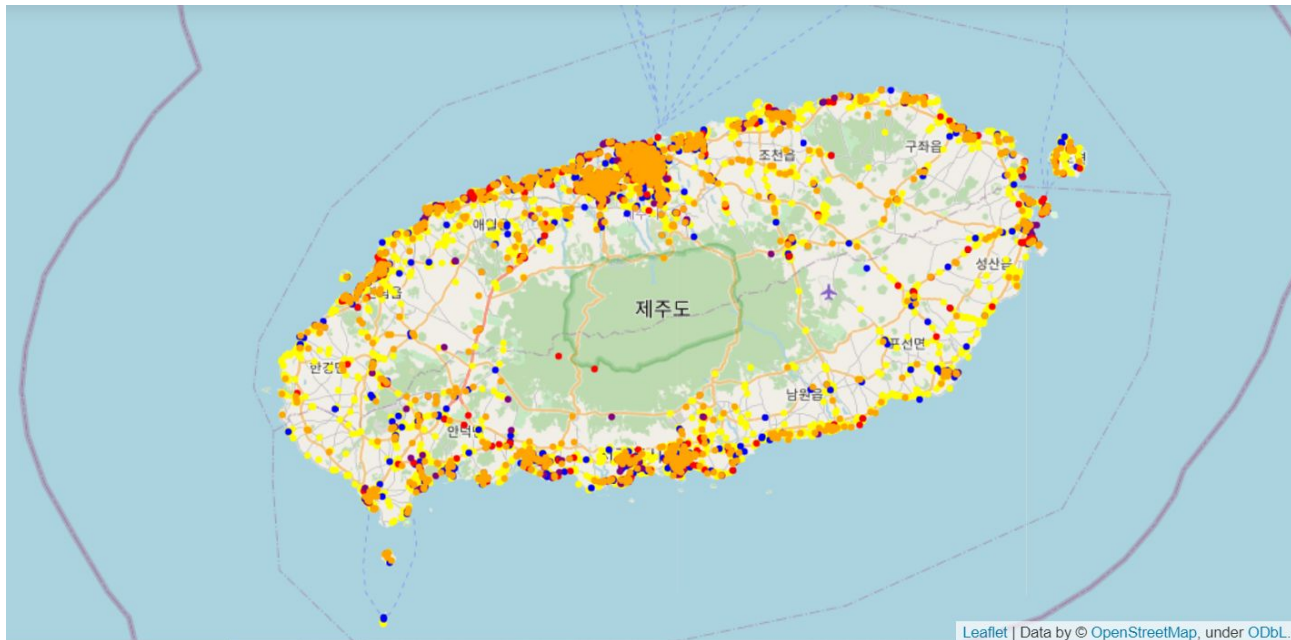


- 제주도의 음식점 대부분은 해변과 시청, 공항 주변에 분포하고 있습니다.

# [음식점] 소상공인 분포도

## ◆소상공인구분

구분	정의
영세	연매출 3억 이하
중소	5억 이하
중소1	10억 이하
중소2	30억 이하
일반	30억 초과

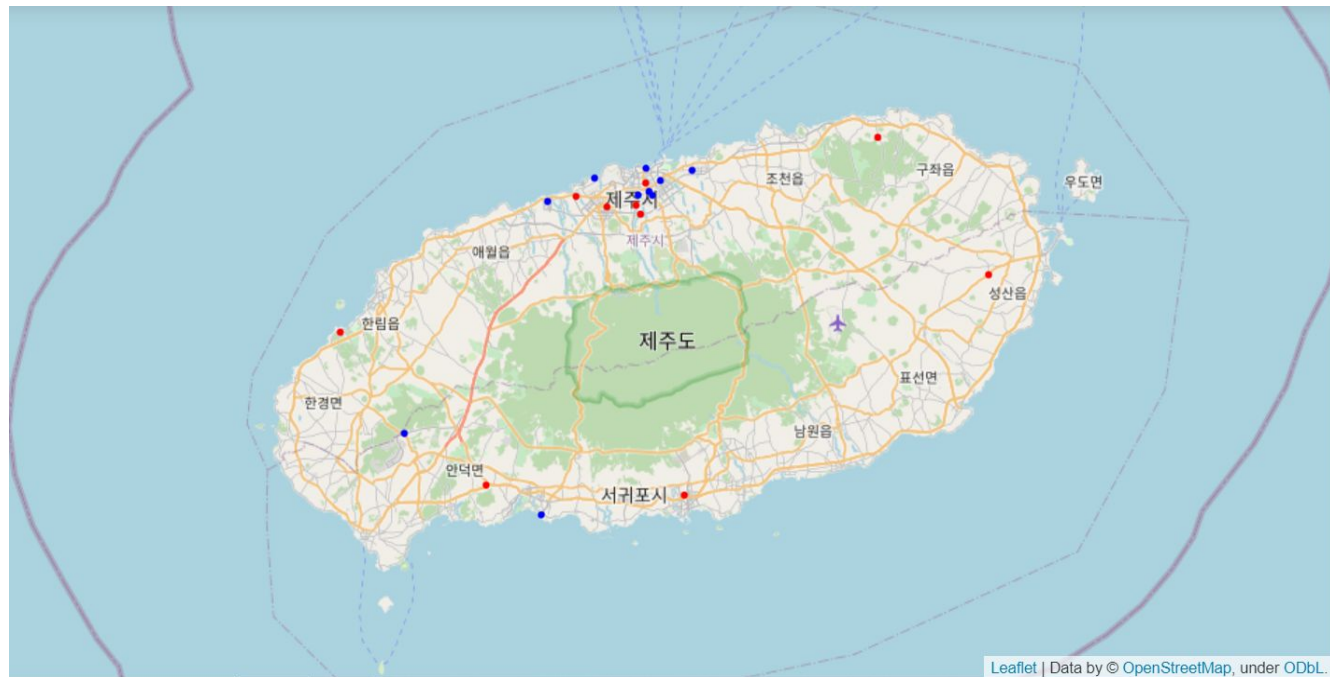


# 총매출/거래빈도에 따른 '외식' 상위 Top 10

◆상위 10개

● Totalspent(총매출)

● NumofSpent(거래빈도)



# 총매출 음식점 Top 10 좌표

TotalSpent	Maintype	TRANS_X	TRANS_Y
10790000	외식	126.522	33.4804
10600000	외식	126.754	33.5416
7023000	외식	126.37	33.2644
6100000	외식	126.489	33.4863
5016000	외식	126.863	33.4322
5000000	외식	126.527	33.5053
4530000	외식	126.565	33.2564
4487000	외식	126.227	33.3864
4400000	외식	126.458	33.4945
4287800	외식	126.517	33.4876

## <TotalSpent Top 10>

1. 부민장례식장
2. 김녕농협장례문화센터
3. 춘심이네 본점
4. 거부
5. 제주해룡
6. 제주칼호텔 주변
7. 한빛장례식장 입구
8. 싱싱잇
9. 이호해수사우나 부근
10. 정확한 위치x, 스시 호시카이 부근

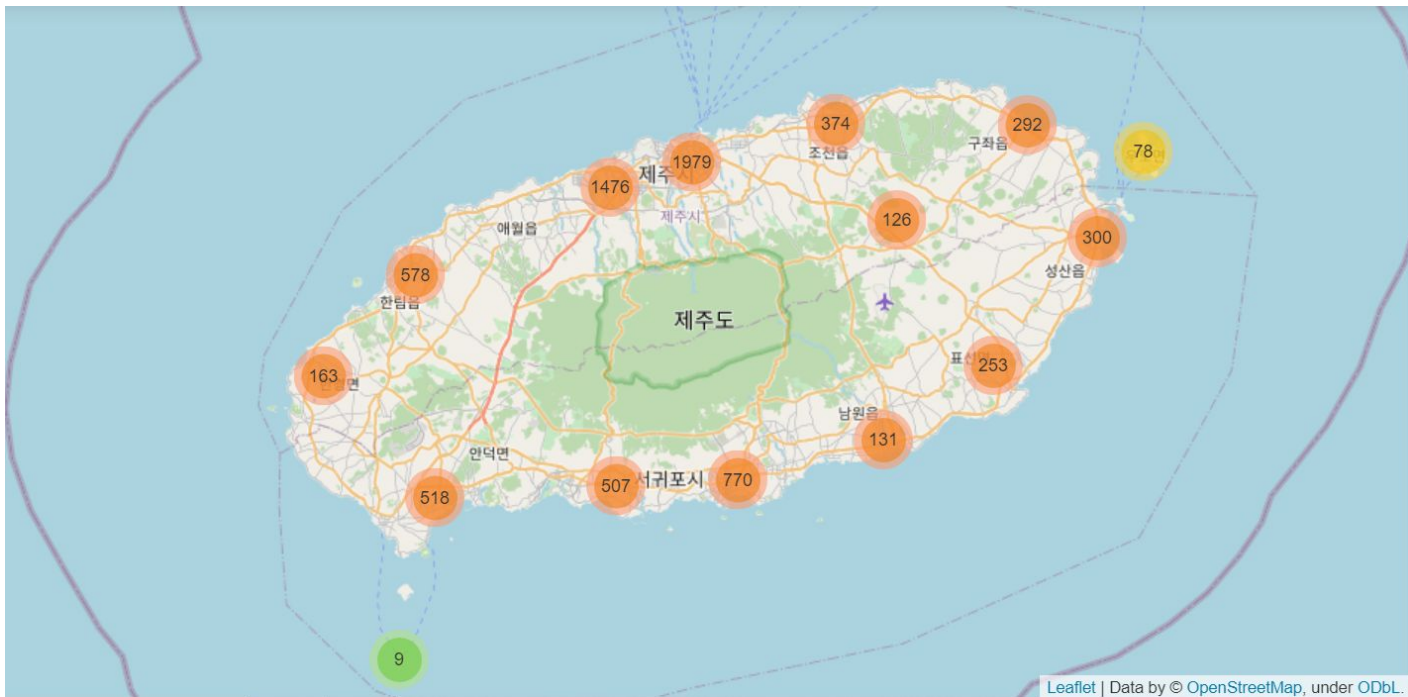
# 거래빈도 음식점 Top 10 좌표

NumofSpent	Maintype	TRANS_X	TRANS_Y
220	외식	126.533	33.4954
189	외식	126.476	33.5092
180	외식	126.541	33.5072
179	외식	126.519	33.4957
171	외식	126.431	33.4906
168	외식	126.572	33.5153
138	외식	126.424	33.2408
135	외식	126.29	33.3057
132	외식	126.53	33.4985
110	외식	126.527	33.517

## <NumofSpent Top 10>

1. 정확한 위치x 맥도날드 부근
2. 좌표 위치 근처 음식점x
3. 정확한 위치x 스타벅스와 해와달 삼계탕 사이
4. 정확한 위치x 제주시종합운동장 부근
5. 고수라김밥 근처
6. 에어바우트 커피 근처
7. 제주관광공사면세점 (안의 음식점으로 추정)
8. 오설록
9. 라라쿡푸마라탕 (제주시청점)과 신마포갈매기 사이
10. 제주월드21(안의 음식점으로 추정)

# Folium Cluster

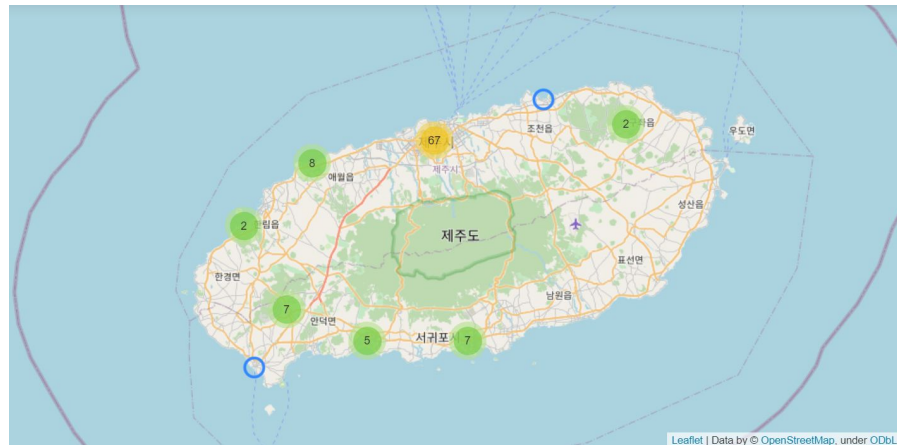


좌표변환으로 인해 위치가 정확하지 않아 좌표로 맛집을 찾는 것이 어렵다고 판단  
->folium cluster를 이용해 제주도 어느 지역에 맛집이 많은지 탐색

# 음식점 총매출/거래빈도 Top100



총매출 (TotalSpent)  
상위 Top 100

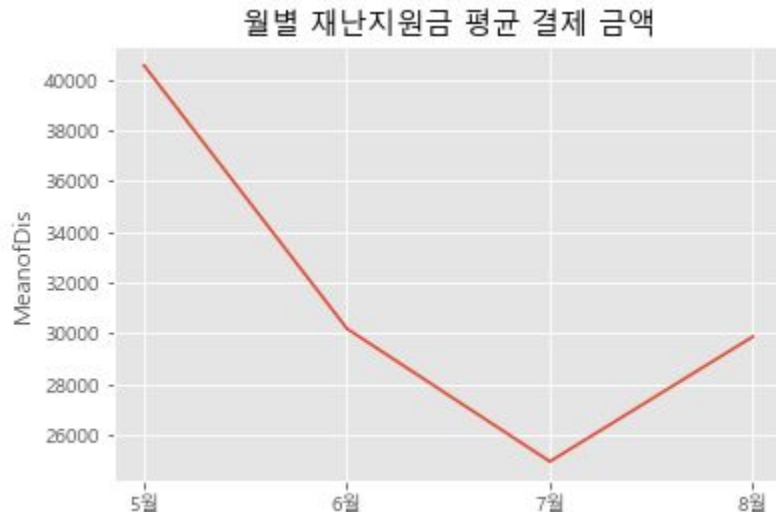
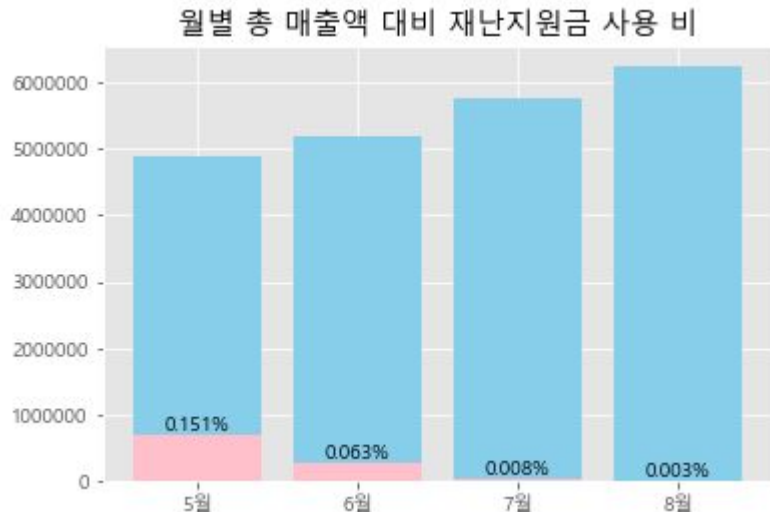


거래빈도 (NumofSpent)  
상위 Top 100

# 재난지원금 소비 패턴 분석



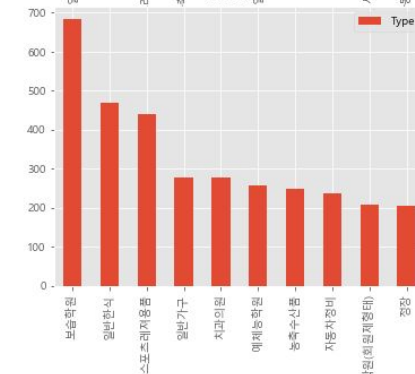
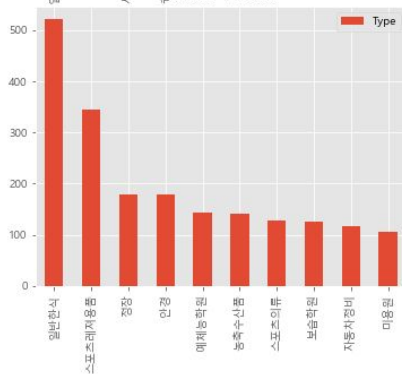
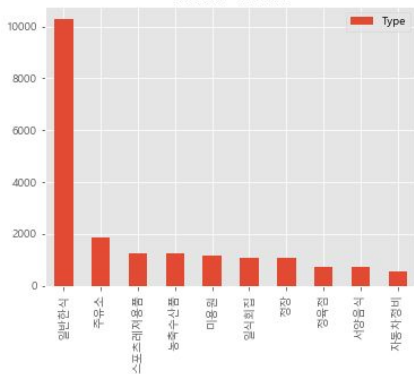
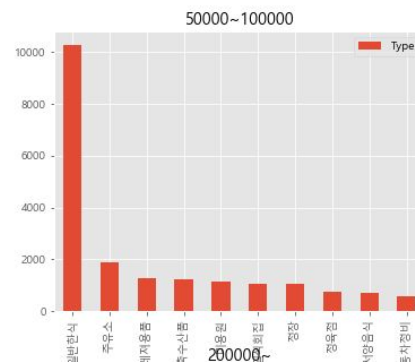
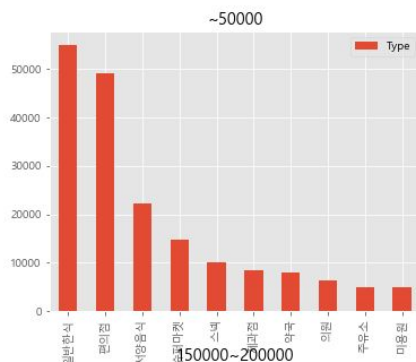
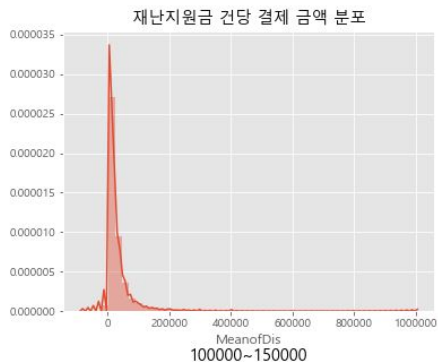
# 월별 재난지원금 소비 추이



- 총매출액에 대한 재난지원금 소비금액의 비는 을 살펴보면 5월에 15%로 가장 많고 달이 지날수록 감소한다.
- 반면, 월별 재난지원금 1회 결제 금액 평균을 살펴보면 5월에 가장 큰 금액으로 나타나며 7월까지 감소하다가 재난지원금 사용 마지막 달인 8월에 다시 1회 결제 금액 평균이 증가하였다.
- 전체적인 총 재난지원금 소비금액은 줄어들이지만 마지막 달에 남은 가격을 소진하기 위해 1회 소비 금액이 증가한 것으로 추측된다.

# 재난지원금 1회 사용금액별 업종분포

재난지원금 건당 가격 범주별 업종분포 top10



- 전체적으로 일반한식 및 음식점에 대한 소비가 많고 1회 소비금액의 분포도 넓게 나타난다
- 1회 소비금액이 15만원 이상일 경우 일반 한식과 다른 부류의 상점과의 도수 차이가 크게 나지 않다가 20만 이상의 경우 1위가 보습학원으로 대체되었다.
- 15만원 이하의 소액결제에의 경우 일반한식 및 음식점, 편의점 등의 부류의 상점의 도수가 크게 나타나지만 15만원 이상부터 그 분포가 크게 줄어드는 패턴을 보임

# 결론

# 결론

- 제주도의 총 월별 매출액은 증가하였지만 재난지원금 사용금액은 월별로 줄어들고 있는 추세입니다.
- 제주시가 서귀포시에 비해 전체적인 규모가 크다는 것을 알 수 있습니다. (경제규모, 자영업자, 인구 등)
- 업종별 매출액(+재난지원금 사용금액)은 월별로 큰 변화없이 순위가 유지되었습니다.
- 제주도는 낮 12시에 가장 소비가 활발하고, 04~05시에 가장 소비가 소극적 입니다.
- 재난지원금은 18시에 가장 소비가 활발하고, 04시에 가장 소비가 소극적 입니다.
- 재난지원금은 소상공인구분에서 일반에서 가장 많이 쓰였고, 영세에서 가장 적습니다.
- 일반한식 사업을 하시는 소상공인 수가 가장 많습니다.
- 전체적으로 음식점에 대한 소비가 많습니다.
- 재난지원금의 사용제한기간이 재난지원금 소비 유형에 영향이 있다는 것을 알 수 있습니다.

# 아쉬운 점

compare\_data\_20\_4

	년월	시도명	지역구분	읍면동명	업종코드	업종명	이용자구분	관광구분	연령대	성별	이용자수	이용건수	이용금액	매장수	업종명대분류	데이터기준일자	매장별금액평균
1571533	2020-04	서귀포시	도심	강정동	156111	한식음식점	내국인	NaN	30.0	여성	1	1	38000	2953	음식점업	2020-12-15	12.868270
1571534	2020-04	서귀포시	도심	강정동	156111	한식음식점	내국인	NaN	30.0	여성	1	1	46000	2953	음식점업	2020-12-15	15.577379
1571535	2020-04	서귀포시	도심	강정동	156111	한식음식점	내국인	NaN	50.0	여성	1	1	11000	2953	음식점업	2020-12-15	3.725025
1571536	2020-04	서귀포시	도심	강정동	156111	한식음식점	내국인	NaN	50.0	여성	1	1	25000	2953	음식점업	2020-12-15	8.465967
1571537	2020-04	서귀포시	도심	강정동	156111	한식음식점	내국인	개별	20.0	여성	98	102	2735300	2953	음식점업	2020-12-15	926.278361
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1596586	2020-04	제주시	읍면	한림읍	156114	서양식점	제주도민	도민	50.0	여성	188	207	3778850	1588	음식점업	2020-12-15	2379.628463
1596587	2020-04	제주시	읍면	한림읍	156114	서양식점	제주도민	도민	60.0	여성	72	78	1364400	1588	음식점업	2020-12-15	859.193955
1596588	2020-04	제주시	읍면	한림읍	156114	서양식점	제주도민	도민	60.0	여성	62	66	1269480	1588	음식점업	2020-12-15	799.420655
1596589	2020-04	제주시	읍면	한림읍	156114	서양식점	제주도민	도민	70.0	여성	11	12	307700	1588	음식점업	2020-12-15	193.765743
1596590	2020-04	제주시	읍면	한림읍	156114	서양식점	제주도민	도민	70.0	여성	33	39	577600	1588	음식점업	2020-12-15	363.727960

8794 rows x 17 columns

외부데이터를 구해 평균 비교검정을 시도  
 목표: 재난지원금 제도에 따른 소비증진 효과에 대해 탐색  
 -> 기존 데이터와 다른 구성  
 -> 외부데이터에 존재하지 않는 금액의 단위  
 의 이유로 실시하지 못하였다.

*END*