

신용카드 거래 내역으로 보는 요식업 동향 분석

2 조 겨울 프로젝트

김휘람, 김예림, 박수진, 정다인

목차

1. 서론 (소논문의 총 요약본, 2 장~3 장 분량)

- 연구 배경 및 필요성

- 연구목표

- 데이터 소개

2. 본문 1

- 프로젝트 진행 과정

- a.분석 방법 설명

- b.분석 결과

3. 본문 2

- 프로젝트 분석을 통해 얻은 결과 그룹화

- a. 요식업 생태계 그림 확인

- b. 예상 결과와의 비교, 분석에 대한 개선점 제시

4. 결론

서론

우리나라의 고용경제는 꾸준한 문제거리로 대두되고 있다. 이런 상황 속에서도 창업률은 증가하고 있는 모순을 보인다. 이러한 배경 속에서 '평생직장'이라는 인식은 영향력이 약해져가고 실업이나 은퇴 후에 새로운 시작을 준비하는 4050 대와 취업이 되지 않은 2030 대가 각자 다른 이유로 창업에 도전하고 있다. 창업 중에서는 특히 요식업 시장의 소상공인 창업의 폐점율이 80%라는 높은 수치에도 불구하고 꾸준히 늘어나고 있는 강세를 보이고 있다. 요리는 일상생활에서도 쉽게 접할 수 있는 기술이기에 대부분이 '감'으로 업종별 주요 고객층과 매출을 가늠하고 요식업을 쉽게 느껴 도전하는 것으로 보인다. 하지만 폐점율이 80%라는 수치를 생각했을 때, 요식업 생태계에 '감' 하나만으로 뛰어드는 것은 위험한 일이다.

이제는 무분별한 데이터와 정보가 많은 '빅데이터 시대'에 맞춰 자신에게 필요한 정보만을 골라내는 능력이 요구된다. 요식업은 식사시간대인 12~1 시, 6 시~7 시에 강세를 보일 것이라는 선입견을 내려놓고 업종별로 어떤 매출액의 변동성을 가지고 있으며, 어떤 매출 건수의 변화를 보이는지 업종 별 데이터를 구체적으로 확인할 필요성이 생겼다. 이러한 필요성 하에 신용카드 거래내역 데이터를 이용, 우리나라 요식업 현황에 대한 전반적인 이해를 높이하고자 한다.

이러한 연구 목표 하에 2018 년부터의 신용카드 이용 데이터를 활용하여 요식업계에서 이용한 내역들만 추출, 실제 요식업계 시장은 어떤 양상을 띄고 거래가 이루어지고 있는지를 확인하고자 한다. 그리고 이러한 분석을 시각화시켜 요식업 창업을 준비하는 자가 쉽게 요식업계의 현황을 이해하고 니즈에 맞게 창업을 준비할 수 있게 도움을 줄 것으로 기대하고 있다.

본문 1)

a. 분석할 업종 선정

분석할 데이터는 상점 신용카드 매출 예측 경진대회에서 제공한 데이터로, 소상공인 매장의 업종, 결제할 때 사용된 카드의 카드사, 결제 금액, 결제 시각 등을 담고 있다. 해당 데이터의 가장 큰 장점은 수익이 발생한 매장의 업종이 구체적으로 분류되어 있어서 한 고객이 어떤 매장에서 얼마만큼의 금액을 썼는지 확인할 수 있다는 점이였다. 이러한 장점을 최대한 이용하고자 업종 별로 볼 수 있는 특징을 보기로 했다. 이러한 장점을 최대한 이용하고자 업종 별로 어떤 특징을 가지고 있는지 확인하는 것을 1 차적 목표로 세웠다.

	store_id	card_id	card_company	transacted_date	transacted_time	installment_term	region	type_of_business	amount
0	0	0	b	2016-06-01	13:13	0	NaN	기타 미용업	1857.142857
1	0	1	h	2016-06-01	18:12	0	NaN	기타 미용업	857.142857
2	0	2	c	2016-06-01	18:52	0	NaN	기타 미용업	2000.000000
3	0	3	a	2016-06-01	20:22	0	NaN	기타 미용업	7857.142857
4	0	4	c	2016-06-02	11:06	0	NaN	기타 미용업	2000.000000
...
6556608	2136	4663855	d	2019-02-28	23:20	0	제주 제주시	기타 주점업	-4500.000000
6556609	2136	4663855	d	2019-02-28	23:24	0	제주 제주시	기타 주점업	4142.857143
6556610	2136	4663489	a	2019-02-28	23:24	0	제주 제주시	기타 주점업	4500.000000
6556611	2136	4663856	d	2019-02-28	23:27	0	제주 제주시	기타 주점업	571.428571
6556612	2136	4658616	c	2019-02-28	23:54	0	제주 제주시	기타 주점업	5857.142857

6556613 rows × 10 columns

데이터를 분석해보니 확인해보니 의복 액세서리 및 모조 장신구 도매업, 한식 음식점업, 배관 및 냉·난방장치 도매업, 그 외 기타 스포츠시설 운영업, 간이음식 포장 판매 전문점, 안경 및 렌즈 소매업, 그 외 기타 분류 안된 상품 전문 소매업, 채소, 과일 및 뿌리작물 소매업, 서양식 음식점업 등 140 개 가량의 업종을 확인할 수 있었다. 업종이 너무 세분화 되어있어서 데이터 전체의 포괄적

특징은 살펴보기 어려웠기에, 특정 업종들만 골라낸 후 주제를 구체적으로 설정할 필요가 있었다. 연구 목표에 부합하도록 요식업과 관련된 음식을 다루고 있는 직종을 추출해냈다. 종류가 다양하면서 너무 많지도 않으며, 데이터의 양이 적절한 업종으로 요식업과 관련된 데이터가 적당함을 확인한 후, 우리나라 요식업계의 전반적인 특징을 분석하기로 분석 대상을 구체적으로 결정했다.

이를 위해 전체 14n 개의 (140 여개)업종 중 요식업과 관련된 업종을 추려내는 과정을 가졌다. 식당·주점·다방 등 접객시설을 갖추고 구내에서 직접 소비하도록 음식·음료 등을 조제하여 손님에게 판매하거나, 구입한 조제음식료품을 판매하는 식품접객업소로서의 영업활동이라는 요식업의 정의에 따라 적절하지 않은 업종들을 제거했고, 그 과정에서 건강보조식품 소매업, 육류 소매업 등은 제거되었다. 그 결과 **피자, 햄버거, 샌드위치 및 유사 음식점업, 제과점업, 치킨 전문점, 제과점업, 비알코올 음료점업, 음료 소매업, 기타 주점업, 중식 음식점업, 한식 음식점업, 일식 음식점업, 서양식 음식점업, 기타 외국식 음식점업, 김밥 및 기타 간이 음식점업, 간이음식 포장판매 전문점** 이렇게 13 개의 업종을 분석하기로 결정했다.

선택한 13 개의 업종 데이터의 'region' 변수에서 약 14 만 개의 결측치가 존재했다. 'region' 변수가 범주형 변수이며, 결측치가 전체 데이터의 약 10% 정도였기 때문에 결측치를 제거하는 방법을 선택하였다.

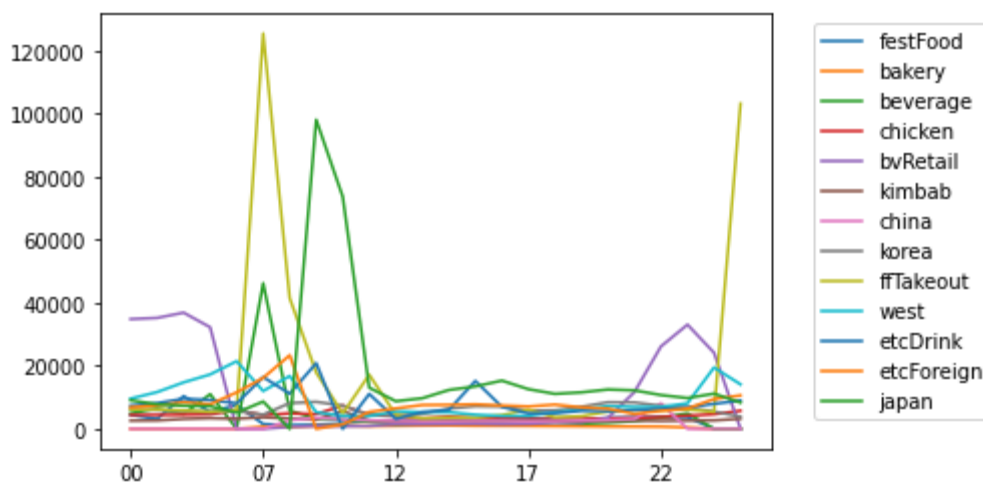
```
aToEtc.isna().sum()
```

```
store_id      0
card_id       0
card_company  0
transacted_date  0
transacted_time  0
installment_term  0
region        139782
type_of_business  0
amount        0
date          0
dtype: int64
```

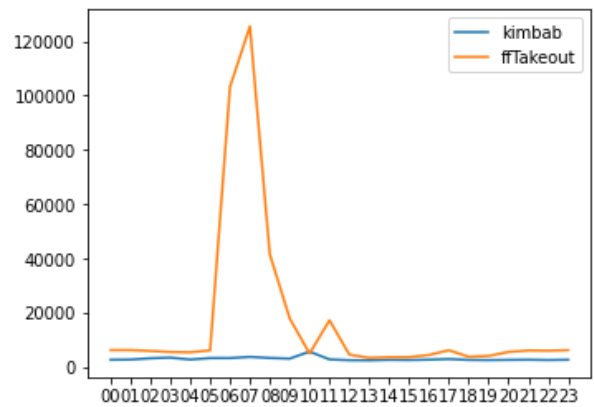
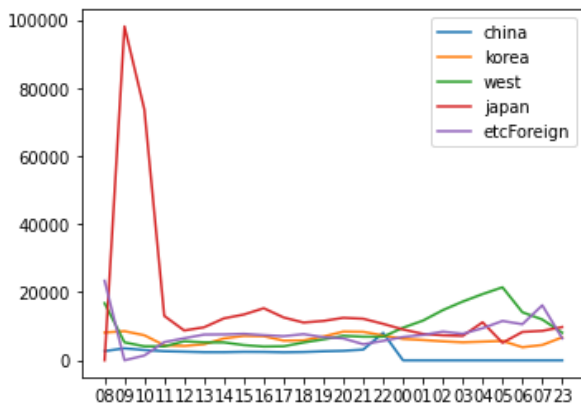
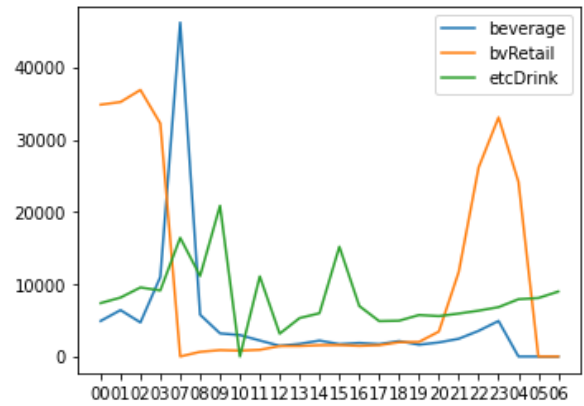
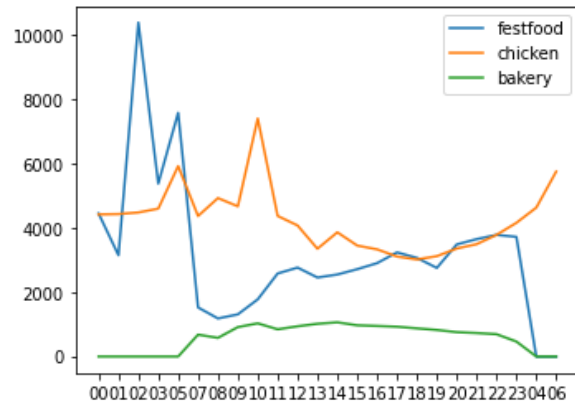
또, 선택한 13 개의 업종 데이터 중 거래가 일어난 지역을 기입하는 region 변수에서 약 14 만 개의 결측치가 존재함을 확인했다. 결측치를 대체하기에는 region 은 범주형 변수이며, 결측치가 전체 데이터의 약 10% 정도였기에 대체값을 넣기보다는 결측치를 제거하는 방법으로 선택하여 데이터를 전처리했다.

b. 업종별 통계수치

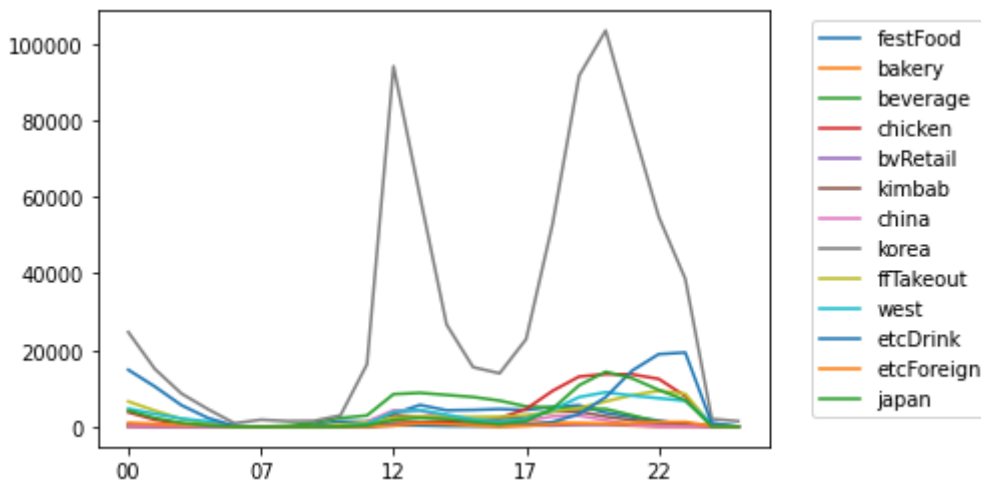
요식업의 전체적인 매출의 특징을 확인하는 것이 목적이었기에 각 업종별로 무엇을 분석 할 지 정했다. 시간대 별 매출액 합계의 변화, 시간대 별 건수의 변화를 통해 해당 업종의 이용률이 시간대 별로 다른지, 혹은 시간대 별로 찾는 물품의 금액이 상이한지 확인하기로 했다. 또, 업종의 평균 매출금액, 업종별 매출금액의 분산, 업종별 최소 매출과 최대 매출을 확인하기로 했다. 그 결과는 다음과 같았다.



(전체 그래프 - 평균)



업종 통계에서 가장 눈에 띄는 것은 한식 음식점업이었다. 한식 음식점업의 거래 데이터 수부터 다른 업종의 데이터 수보다 많았고, 시간대 별 매출과 시간대 별 거래 건수를 그래프로 표시했을 때 단위가 다른 업종에 비해 컸다. 요식업계에 한식 음식점업이 차지하는 비중이 크고, 이가 끼치는 영향이 상당할 것으로 예측된다.



c. 비슷한 양상을 가진 업종끼리의 그룹화

위의 통계수치를 통해 시간대별 이용 건수, 시간대 별 매출액 합계의 양상이 비슷한 업종들의 그룹화가 가능했다. 그룹은 총 5 개의 그룹이 나왔으며, 그 다섯 그룹에 포함되지 않는 특징을 가지는 etc의 특징을 가지는 업종도 존재했다.

우선 시간대별 이용률이 특정 시간대에 급증하여 두 개의 봉우리 형태에 물려있는 업종이 있다. 주로 12시~14시, 17시~20시에 이용률이 증가하고 매출액 합계가 증가하는 업종으로 중식 음식점업, 한식 음식점업, 서양식 음식점업, 김밥 및 기타 간이 음식점업이 있었다. 해당 음식점업들은 점심시간, 저녁시간에 이용률이 올라가는, 식사와 관련된 A 그룹으로 분류할 수 있다.

특이하게도 일식 음식점업은 이와 달리 양상을 보였다. 일식 음식점업 역시 12시~14시 사이에 이용률이 살짝 증가하긴 하나, 그 증가폭이 A 그룹과 달리 미세할 정도이고, 오히려 17시에서 24시까지 저녁시간에 커다란 봉오리를 가진다. 즉, 저녁 시간과 새벽 시간에 대부분의 이용자들이 있음을 알 수 있다. 일식 음식점업처럼 시간대별 이용률이 저녁과 새벽에 걸쳐 커다란 봉우리 하나를 가지고 있는 업종은 음료 소매업, 기타 주점업이 있었다. 이를 통해 이 업종들은 통상적인 점심 식사시간, 저녁 식사시간보다는 저녁에 이용률이 높아지는 주류를

다루는 요식업이라 유추 가능했다. 따라서 이 그룹은 식사 외의 주류와 관련된 B 그룹으로 분류했다.

또 특정 봉오리를 형성하기보다는 오후 시간 이후로 꾸준한 이용률을 가지고 있었던 업종들도 있었다. 이러한 부류에 해당하는 두 업종은 피자, 햄버거, 샌드위치 및 유사 음식점업과 제과점업이었다. 제대로 된 식사보다는 간단한 끼니를 해결할 수 있는 그룹으로 보이며, 이를 C 그룹이라 분류하기로 한다.

시간대별 이용 건수에서는 특이한 사항이 보이지 않으나 시간대별 매출액 평균에서 이상한 특징을 보이는 업종은 두 개가 있었다. 비알코올 음료점업과 간이음식 포장 판매 전문업종이 그 두 업종이었다. 이 업종 둘은 아침 시간, 특히 오전 7시에 시간대별 매출액 평균이 유의할 정도로 높은 값을 보이고 그 이후는 거의 평평했다. 즉, 오전 7시에 구입하는 물건의 값이 다른 시간대의 물건의 값보다 유의할 정도로 높았음을 보여준다. 아침에 해당 요식업에서 판매하는 음식이 다른 시간의 음식에 비해 특징을 가진다고 유추할 수 있었으며, 이 그룹을 D 그룹으로 분류했다. 그리고 해당 A, B, C, D 그룹 중 어디에도 해당되지 않는 업종인 기타 외국식 음식점업을 etc 로 두었다.

<표 1> 데이터 환경의 특징

구분(그룹)	업종 종류	특징
A	중식 음식점업 한식 음식점업 서양식 음식점업	12 시~14 시, 17 시~20 시에 시간별 매출액 합계가 증가한다. 즉, 식사시간에 매출액이 유의미한 변화를 보인다.

	김밥 및 기타 간이 음식점업	
B	음료 소매업 일본 음식점업 기타 주점업	17 시~14 시에 시간별 매출액 합계가 증가한다. 저녁시간부터 새벽까지 주류가 팔리는 시간에 매출액이 유의미한 변화를 보인다.
C	피자, 햄버거, 샌드위치 및 유사 음식점업제과점업	눈에 띄는 매출 변화는 없지만 오후부터 꾸준한 매출액 합계를 유지한다. 시간과 관계없이 늘 구입할 수 있는 것으로 예측된다.
D	비알코올 음료점업과 간이음식 포장 판매 전문업종	시간대별 매출액 평균이 오전 7 시에 유의미하게 높다. 즉 아침시간에 특정 메뉴를 찾는 사람들이 많은 것을 유추할 수 있다.

이런 그룹화를 통해 우리는 요식업계에 있는 13 개의 업종을 매출 시간에 따라 4 가지의 경향성으로 요약할 수 있었다. 또, 이 그룹들이 요식업에 어떤 영향을 끼치고 있는지 파악함으로써 우리나라 전반적인 요식업계의 근황을 요약할 수 있으리라 기대한다.

이런 그룹화를 통해 우리는 요식업계에 있는 13 개의 업종을 매출 시간에 따라 4 가지의 경향성으로 요약할 수 있었다. 이는 '내가 시작할 요식업계 사업이 이런

식품을 판매하고 있으니 무엇을 중심으로 둘 것이다'와 같은 막연한 계획을 가지고 창업을 시작하려고 하는 사람들에게 보다 현실적인 수치로 계획을 수립하게 하는 참고자료가 되어줄 것으로 예상된다. 또, 13 개의 업종을 4 가지로 간소화하면서도 그 특징은 생략하지 않음으로써 전체적인 요식업의 경향성을 해당 4 가지 업종으로 표현할 수 있기에 앞으로 추가적인 분석을 할 때 설명이 용이하다.

본문 2)

대략적인 요식업계의 4 가지 경향성을 시간에 따라 어떤 형상을 띄는지를 기준으로 분류했으나, 이것만으로 요식업계의 현황을 대략적으로 확인했다고 볼 수 있는지에 의문이 있다. 원데이터에 있던 변수들 중 요식업계의 현황을 보다 자세히 볼 수 있게 해주는 요소는 없나 검수해봤다. 결제할 때 사용했던 카드 회사의 경우 요식업계 현황을 확인할 때에 큰 도움이 되지 않았을 것 같았기에, 남은 변수 중 지역이라는 변수에 집중하여 보다 면밀하게 우리나라 전체의 요식업 현황을 설명하기로 했다.

우선 전체 데이터를 그룹별로 나누어 각각 a, b, c, d, etc 변수에 저장하였다. 지역에 따라 각 그룹을 비교하고자 하였는데 'region' 변수가 시/군/구 단위까지 나와있어 지역별로 비교하기엔 너무 많았다. 따라서 'region' 변수에서 앞 두 글자를 추출하여 도/광역시 단위로 나타내었다.

```
# 지역에서 도/광역시 단위로 뽑아내기
a['Do'] = a['region'].str[0:2]
b['Do'] = b['region'].str[0:2]
c['Do'] = c['region'].str[0:2]
d['Do'] = d['region'].str[0:2]
etc['Do'] = etc['region'].str[0:2]
```

각 그룹의 지역별 거래 건수를 지도에 시각화하기 위해 각 그룹의 거래 건수를 지역별로 구해 데이터프레임으로 나타내었다. 아래는 그 코드와 결과(데이터 프레임)이다.

```
aCount = aGroupByDo['group']
aCount = pd.DataFrame(aCount)
aCount.rename(columns = {'group' : 'A'}, inplace = True)

bCount = bGroupByDo['group']
bCount = pd.DataFrame(bCount)
bCount.rename(columns = {'group' : 'B'}, inplace = True)

cCount = cGroupByDo['group']
cCount = pd.DataFrame(cCount)
cCount.rename(columns = {'group' : 'C'}, inplace = True)

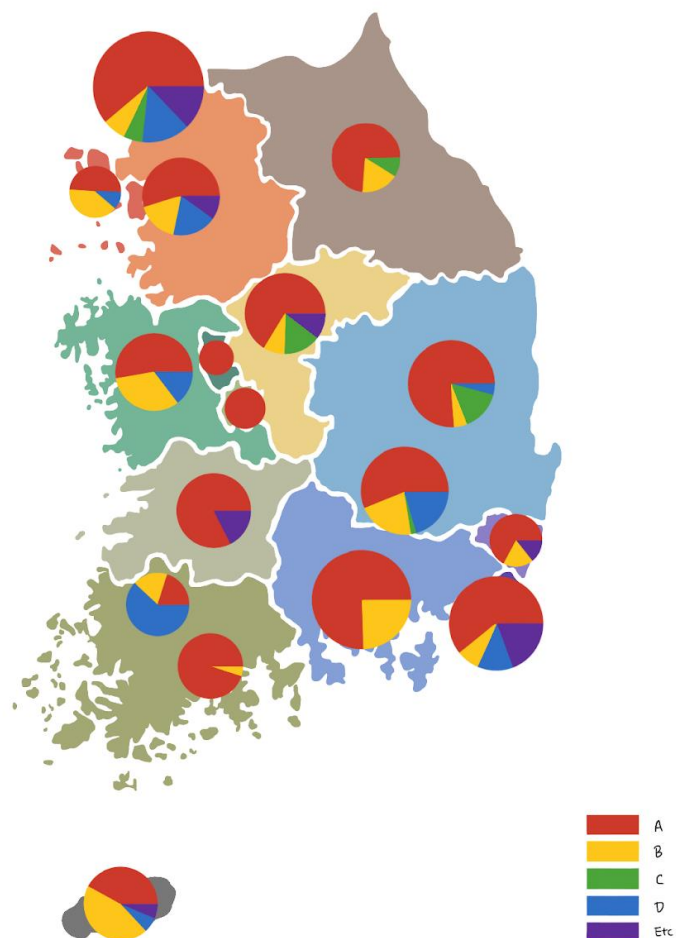
dCount = dGroupByDo['group']
dCount = pd.DataFrame(dCount)
dCount.rename(columns = {'group' : 'D'}, inplace = True)

etcCount = etcGroupByDo['group']
etcCount = pd.DataFrame(etcCount)
etcCount.rename(columns = {'group' : 'Etc'}, inplace = True)
```

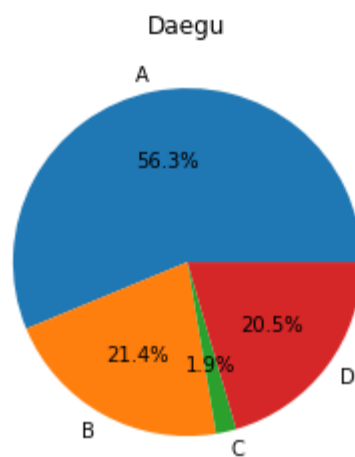
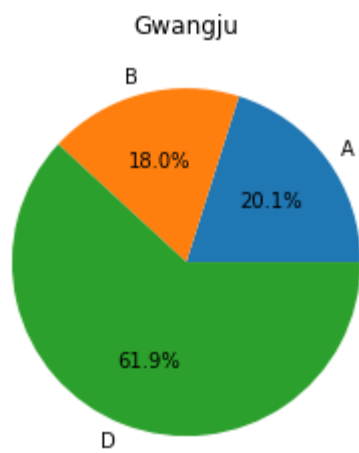
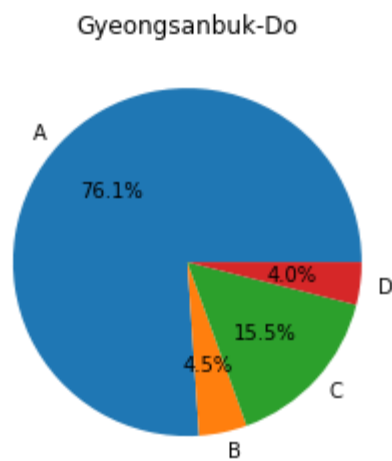
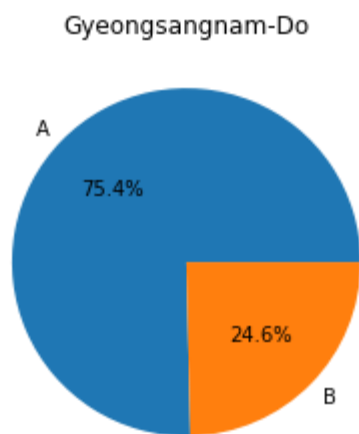
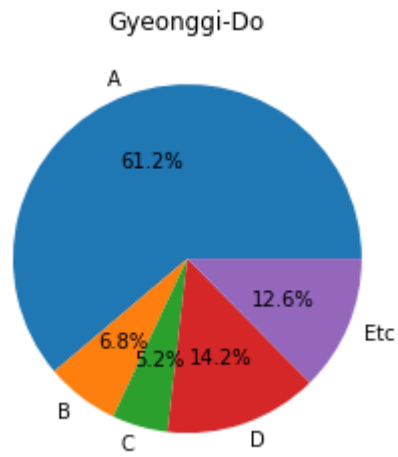
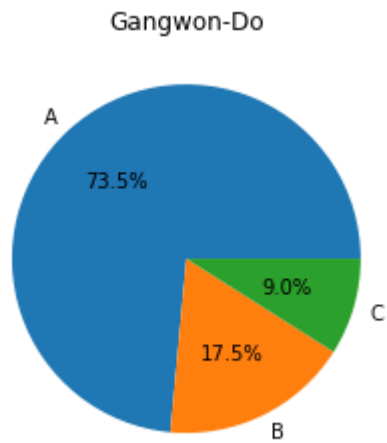
```
allCount = pd.concat([aCount, bCount, cCount, dCount, etcCount], axis = 1)
allCount
```

	강원	경기	경남	경북	광주	대구	대전	부산	서울	세종	울산	인천	전남	전북	제주	충남	충북
A	28779.0	181790.0	95887.0	79455.0	4809.0	43953.0	9492.0	77768.0	38100.0	7046.0	12103.0	8017.0	40164.0	38965.0	26447.0	29401.0	54082.0
B	6845.0	20307.0	31347.0	4660.0	4308.0	16713.0	NaN	10013.0	12308.0	NaN	3374.0	6489.0	2097.0	NaN	28190.0	18355.0	7000.0
C	3513.0	15330.0	NaN	16204.0	NaN	1468.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	12297.0
D	NaN	42312.0	NaN	4130.0	14790.0	15998.0	NaN	15625.0	12968.0	NaN	NaN	1863.0	NaN	NaN	4005.0	8182.0	NaN
Etc	NaN	37403.0	NaN	NaN	NaN	NaN	NaN	24539.0	6706.0	NaN	2549.0	NaN	NaN	8284.0	4156.0	NaN	8376.0

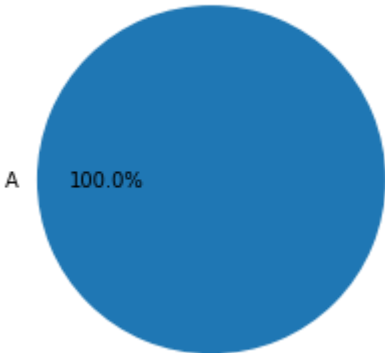
구해진 거래 건수를 지도 위에 파이 차트로 나타내었다. 파이 차트의 크기가 클수록 거래 건수가 많음을, 작을 수록 거래 건수가 적음을 의미한다.



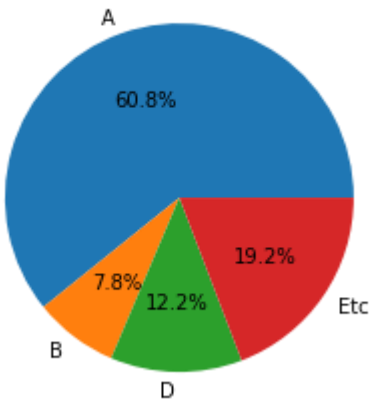
자세한 비율은 아래 파이차트와 같다.



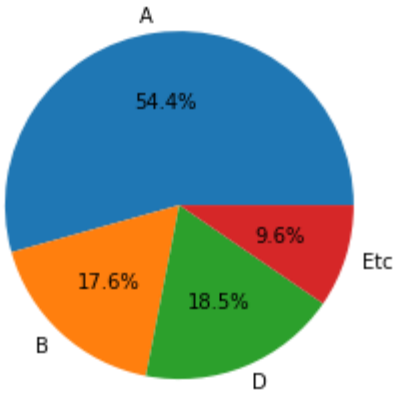
Daejeon



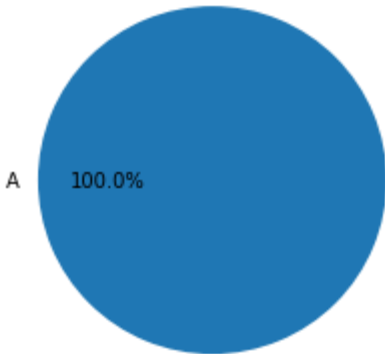
Busan



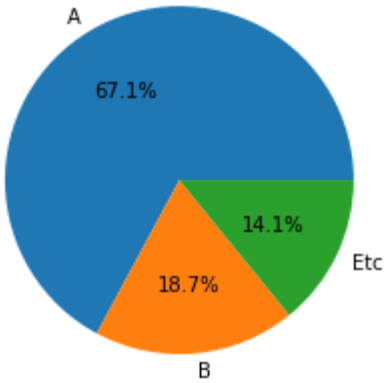
Seoul



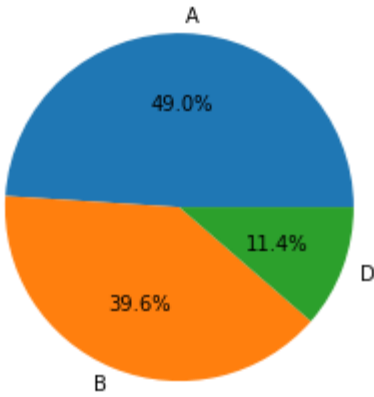
Sejong

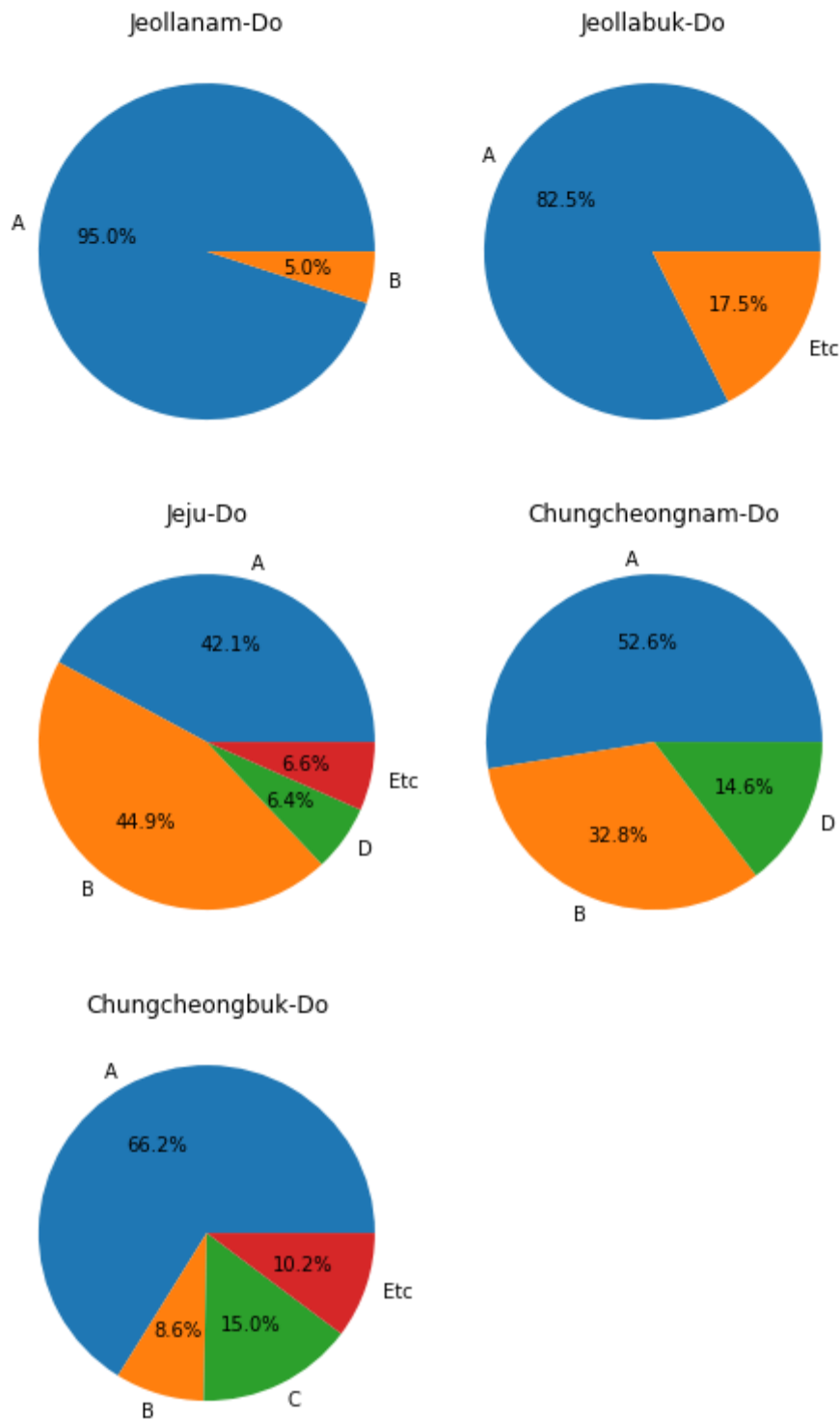


Ulsan



Incheon





해당 시각화를 통해 지역 간에도 요식업의 특징이 다르다는 사실을 알 수 있었다. 파이차트가 차지하는 비중을 통해 지자체 별 요식업 특징을 파악할 수 있다. 주식을 담당하고 있는 A의 비중은 모든 지역에서 가장 높은 비율을

차지하고 있었다. 즉, 대부분의 요식업계에 가장 큰 영향을 미치고 있는 그룹은 점심과 저녁 식사시간에 그 비중이 높아지는 한식, 중식, 서양식과 김밥 등의 기타 간이 음식점이라는 사실을 확인할 수 있다. 이와 달리, 주식 외에 다른 상품을 판매하는 업종이 성행하는 특이한 경우의 지역도 있었는데 광주와 제주도가 그 경우다. 광주의 경우는 A가 가장 높은 비율을 차지하지 않고 D인 비알코올 음료점업이 60% 이상을 차지하는 특징을 보였다. 또, 제주도의 경우 A가 약 42%, B가 약 44%로, 점심과 저녁에 판매 비중이 높아지는 A와 저녁엔 높아지는 B가 비슷한 비율을 차지하고 있었다는 점을 유의할 필요가 있다.

파이차트에 나타나는 그룹의 다양성 역시 눈여겨 볼 필요가 있다. 대부분의 지역은 적어도 3개 이상의 그룹의 조합으로 구성된 파이차트를 보였으나, 경남, 대전, 세종, 전남, 전북의 경우는 2가지 이하의 조합만으로 구성되어 있었다. 이러한 지역의 특징은 지자체의 인구밀도가 낮다는 특징을 보였다. 세종과 대전의 경우는 인구밀도는 높았으나 거래 수 자체가 9492 건, 7049 건으로 가장 낮은 건수를 가지고 있었기 때문에 A로만 구성된 파이차트가 나온 것으로 예상된다. 또 지역의 도시발전도가 높을수록 분류의 조합이 더 다양해지고 있음을 서울과 부산과 같이 인구밀도가 높은 도시에서 보이는 파이차트를 통해 유추할 수 있다.

지자체 별 요식업 특징을 분석한 결과에 유의해야하는 점은 해당 데이터에서 거래 지역에 결측치가 많았다는 점이다. 거래 지역 결측치의 경우 대체할 수 있는 값을 넣기 힘들었으므로 거래 지역을 가지고 있는 데이터로만 분석했다. 그 결과 위의 차트의 세종과 대전처럼 원래 거래 데이터가 적은 지역은 파이 차트를 구성하는 그룹의 다양성이 현실의 거래보다 적게 나타나는 한계가 있었다.

이를 통해 우리는 구체적으로 요식업이 어떤 시간대에 강세를 가지는지, 또 어느 지역에서 어떤 종류의 요식업이 인기를 끌고 있는지 대략적으로 파악할 수 있었다.

;

결론

총 13 개의 업종을 살펴본 결과, 대한민국 요식업 강자는 한식 업종이었다. 또한, 시간대 별로 분석해보았을 때, 일반 식사의 경우 점심, 저녁 시간대의 매출이 높았으며 테이크 아웃이나 음료의 경우 아침 시간대에 매출이 높은것으로 보아 끼니는 간단하게 해결하려는 현대인들의 바쁜 일상을 엿볼 수 있었다.

저녁 시간 이후에는 주류 업계가 강세를 보였다, 특히 일식업이 큰 매출액을 기록한 것으로 보아 일식업에 이자카야와 같은 주점이 포함되어 있는 걸로 유추할 수 있다.

시간대별로 분석한 것을 보니, 아침 시간대에는 간이 포장 음식으로 끼니를 해결하고 점심 저녁 시간에 맞춰 하루 세끼를 챙겨먹고 저녁에는 사람들과 음주문화를 즐기는 현대인의 일상이 녹아 들어있음을 알 수 있다. 따라서 요식업계에 뛰어드는 자영업자의 경우, 이러한 시간대별 매출 흐름을 알면 업종 선택에 있어서 큰 도움을 얻을 수 있을 것으로 보인다. 시간대별로 비슷한 매출 양상을 보이는 업종끼리 그룹화하여 지역별로도 분석을 해보았는데, 수도권과 비수도권에서 특이한 동향을 파악할 수 있었다. 비수도권의 경우 강세를 보이는 업종이 단조로웠지만 인구밀도가 높은 서울, 경기, 부산의 경우 다양한 그룹으로

분포했다. 이를 통해 인구밀도와 업종 분포의 다양성이 비례할 것으로 유추할 수 있었다.