



## <목차>

<a href="#"><u>1. 주제 선정 과정</u></a>	1
<a href="#"><u>2. 치킨집 매출 데이터 설명</u></a>	1
<a href="#"><u>3. Research Question 선정</u></a>	2
<a href="#"><u>4. 관련 데이터 모색과 수집 outline</u></a>	2
<a href="#"><u>5. 잠재적 이슈</u></a>	4
<a href="#"><u>6. 앞으로의 계획</u></a>	4

## 1. 주제 선정 과정

서로의 관심사를 공유하며 토의한 결과 많은 주제들이 제시되었다. 그 예시를 몇 개 보자면, 인구 구조 변화와 관련하여 한국 기대수명의 분석, 혼인율이 낮아지는 원인 분석, SNS를 바탕으로 새로 나온 영화의 관객 수와 매출 예측, 소셜 미디어 및 인터넷 기사들의 감정분석을 통하여 대통령 수행의 긍정/부정 비율·정당별 선호도의 조사 등 많은 의견이 있었다. 그러나

- 혼인율처럼 월별로 데이터가 존재하는 경우 데이터의 양이 매우 적어진다.
- 기사 감정분석처럼 분석이 너무 어려운 경우 해석이 힘들고 감당이 어렵다.
- 기대수명 분석처럼 원인이 너무 명확해 당연한 결과만을 도출하는 경우 분석의 의미가 없다.

등의 이유로 위의 주제들을 기각하였다. 긴 논의 중 마침 팀 내에 부모님께서 치킨 배달 전문점을 운영하는 팀원이 있었는데, 이 치킨집의 매출 데이터를 가져와 코로나와 다양한 변수들과 연관 지어서 의미 있는 분석을 해보는 것은 어떨까라는 의견이 나왔다. 해당 데이터로 분석을 진행할 경우 우리의 분석 결과가 의미 있게 쓰일 수 있고, 코로나라는 현재 상황과도 큰 관련이 있어보여 이 주제를 선택하게 되었다.

## 2. 치킨집 매출 데이터 설명

분석의 핵심이 되는 것은 치킨집 매출 데이터이다. 개인정보 보호를 위해 정확한 매출 액수가 아닌 판매 건수가 표시된 데이터를 사용하였다.(그림1 참고) 이 데이터는 개업일인 2015년 10월 15일부터 시작하여 현재까지의 일별 판매 건수를 담고 있다. 매장의 정식 명칭은 [REDACTED]에 위치하고 있다. 매장의 배달 범위는 대전시 도마동/변동이다.

	A	B	C
1	날짜	요일	판매갯수
2	2015-10-15	목	1
3	2015-10-16	금	4
4	2015-10-17	토	17
5	2015-10-18	일	7
6	2015-10-19	월	17
7	2015-10-20	화	20
8	2015-10-21	수	32
9	2015-10-22	목	24
10	2015-10-23	금	22
11	2015-10-24	토	43
12	2015-10-25	일	23
13	2015-10-26	월	26

\*\*\* 정보 보호를 위해 이미지를 삭제하였습니다. \*\*\*

▲ (그림1) 일별 판매건수 데이터

### 3. Research Question 선정

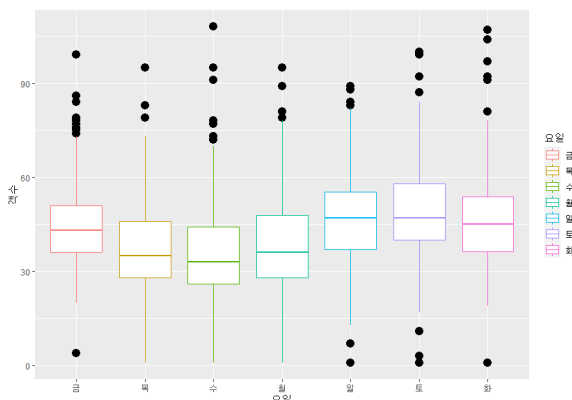
분석 결과가 현실에서 유용하게 쓰였으면 좋겠다는 생각으로, research question은 코로나가 해당 치킨집의 매출에 미치는 영향은 어느 정도인지, 코로나 외에도 매출에 영향을 미치는 다른 유의한 변수는 없는지 살펴보고 매출 극대화 방안을 탐구하는 것으로 선정하였다. 조원별 research question은 해당 주제로 통일하였으며, 각각 다른 모델을 이용하여 분석을 시도할 계획이다.

연구를 통해 유의미한 분석 결과가 도출된다면, 해당 매장의 매출 향상에 도움이 될 수 있을 뿐만 아니라 판매량 예측을 통한 재고 관리, 매출 감소 예상에 따른 재정 대비 등 다양한 긍정적 효과가 발생할 것으로 기대된다. 또한 추후에 수업이 끝나고 나서도 매출 데이터는 계속 업데이트될 것이고, 이에 따라 데이터를 붙여나가면서 후속 연구를 계속 진행할 수 있을 것이다.

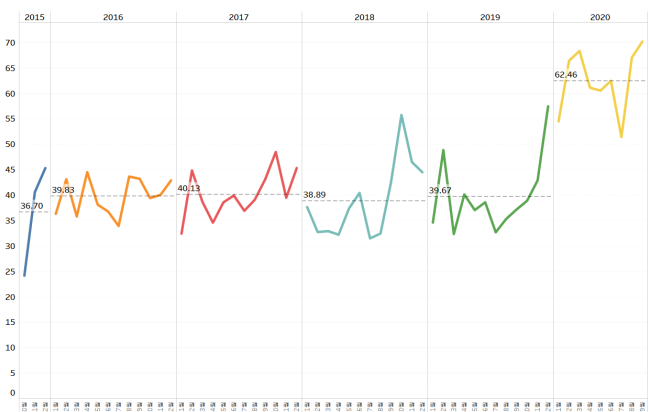
### 4. 관련 데이터 모색과 수집 outline

매출 극대화 방안 모색이라는 주제에 따라, 종속변수인 매출에 영향을 미치는 변수를 고려하여 데이터를 수집하고 통합할 계획이다. 해당 치킨집의 특성을 조사하고 이를 토대로 매출에 영향을 미칠 것으로 예상되는 변수들에 대해 브레인스토밍을 진행한 결과는 다음과 같다.

- ◆ 복날, 공휴일, 연휴, 주말 등은 매출이 높다.(그림3 참조)
- ◆ 코로나 시기에 매출은 증가하였다.(그림4 참조)



▲(그림3) 요일별 매출 상자그림



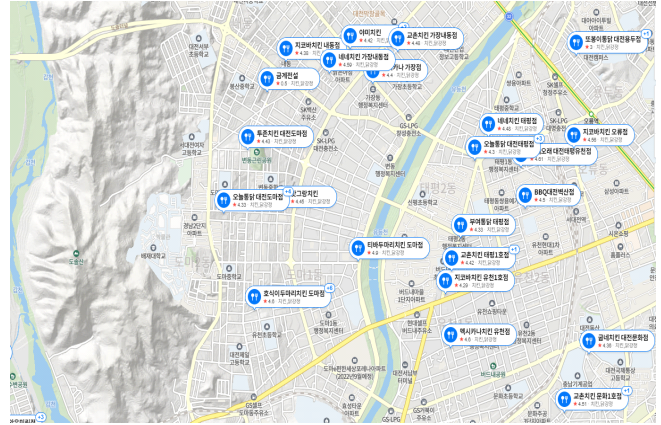
▲ (그림4) 월별 일 매출평균, 코로나 시기에 오히려 매출이 증가

- ◆ 프랜차이즈 업체(BHC)이기 때문에 해당 업체의 평판이 높아지면 매출 또한 증가할 것이다. 특히 브랜드 평판 지수는 본사의 광고와 홍보 등의 마케팅이 매출에 미치는 영향 또한 간접적으로 보여줄 수 있을 것이다.
- ◆ 신메뉴가 출시했을 때 일시적으로 매출이 올라가는 경향이 있다.
- ◆ 해당 지역 인구분포 변화와 온도, 강수량 등의 변수도 매출과 상관관계가 있을 것이다.

- ◆ 매장이 배재대학교와 약 1km 거리로 인근에 위치하고 있어(그림5 참조) 대학생이 주 고객층을 이루므로 대학교의 학사일정을 변수로 추가할 필요가 있다.
- ◆ 주변에 다른 브랜드의 동종업체가 많이 분포하고 있고(그림6 참조), 개업과 폐업이 활발하다. 주변 경쟁업체의 개수 또한 매출에 직접적인 영향을 끼치는 변수 중 하나일 것이다.



▲(그림5) 배달 구역(변동, 도마1동, 도마2동)과 배재대 위치 (붉은 핀 = 매장)



▲ (그림6) 매장 근처 동종 업체(치킨집) 분포

위 특성들을 반영할 수 있는 데이터셋 후보와 수집 계획은 다음과 같다.

- ① 공휴일 및 복날 정보 : 공공데이터포털에서 오픈API를 통해 공휴일 정보를 얻을 수 있고, 기존 달력을 참고하여 복날의 날짜를 얻을 수 있다.
- ② 배재대학교의 종강, 개강 일시 데이터 : 배재대학교 홈페이지의 학사일정에서 찾을 수 있다.
- ③ 날짜별 주변 동종업체의 개수 데이터 : 매장 주변 특정 반경을 정하여 인허가 데이터에서 해당 반경 내에 주소를 둔 치킨집들의 개/폐업 날짜를 참고하면, 일별 반경 내 주변 치킨집의 개수를 파악할 수 있을 것이다.
- ④ BHC 브랜드 파워 및 신메뉴의 출시 일자 데이터 : 매월마다 한국기업평판연구소 등에서 진행한 프랜차이즈 업계별 브랜드 평판 분석결과를 참고하고, 본사 홈페이지에서 메뉴 출시날짜 등을 인용하면 될 것이다.
- ⑤ 기온, 날씨, 인구 데이터 : 해당 치킨집의 배달 범위에 맞는 데이터를 조사해야 할 것이다.
- ⑥ 코로나 검사/확진자 데이터 : [https://github.com/jooeungen/coronaboard\\_kr](https://github.com/jooeungen/coronaboard_kr) 에서 얻을 수 있다.

위의 데이터들을 찾고 난 이후에도 지속적인 탐색 과정을 통해 데이터를 추가/삭제/수정해나갈 계획이다. 그러나 이와 동시에 데이터를 통합하는 과정에서 고려해 봐야 할 잠재적 이슈들도 있을 것이다.

## 5. 잠재적 이슈

### ① 데이터의 population이 다를 수 있다.

i) 전국 단위로 수집된 데이터들을 대전 지역에 국한된 데이터와 어떻게 통합할지 결정해야 한다.

2020년 7월		
순위	치킨전문점(전국)	브랜드지수
1	교촌치킨	2,265,497
2	BBQ	1,821,750
3	BHC	1,759,388
4	멕시카나	1,610,367
5	굽네치킨	1,529,626

▲ 2020년 7월 치킨전문점 브랜드 평판 순위  
'전국'을 기준으로 조사한 데이터이다.

날짜	지점	평균기온
2015-01-01	133(대전)	-5.6
2015-01-02	133(대전)	-4.7
2015-01-03	133(대전)	-2.9
2015-01-04	133(대전)	2
2015-01-05	133(대전)	3.2

▲ 일별 평균기온  
'대전' 지역 기준의 평균기온이다.

ii) 월별과 일별 데이터를 합칠 때, 다른 시간대의 값들을 어떻게 통합할지 결정해야한다.

1	날짜	동이름	세대수	총인구수
2	2015년 10월	도마1동	7911	17771
3	2015년 10월	도마2동	8474	20936
4	2015년 10월	변동	7486	18034
5	2015년 11월	도마1동	7893	17711
6	2015년 11월	도마2동	8451	20850

▲ '월별' 도마1,2동/변동 인구수

	A	B	C
1	날짜	요일	판매갯수
2	2015-10-15	목	1
3	2015-10-16	금	4
4	2015-10-17	토	17
5	2015-10-18	일	7
6	2015-10-19	월	17

▲ '일별' 판매건수 데이터

### ② 동일한 feature에 수치가 다르게 나올 수 있다.

만일 동일한 feature에 대해 조사기관이나 측정기준에 따라 수치가 다르게 나올 경우, 이것들을 어떻게 통합할 것인지 고민해봐야 할 것이다.

### ③ 결측치가 존재할 수 있다.

브랜드 평판지수는 월별로 꾸준히 조사하는 데이터가 아니어서, 중간중간 빠져있는 경우가 있다. 이런 경우 어떻게 결측치를 채워넣을 것인지 결정해야한다.

## 6. 앞으로의 계획

[4. 관련 데이터 모색과 수집 outline]에서 명시한 바와 같이, 매출에 큰 영향을 끼칠 것이라고 예상되는 변수를 최우선으로 수집할 계획이다.

이외에도 어떤 유의한 변수가 더 있을지 고민해보고, 그것들을 어떻게 측정하며 모델에 이용할 수 있을지 서로 상의해 가면서 필요에 따라 데이터를 추가/수정/삭제하려고 한다. 그 밖에 수업에서 배우는 방법론을 활용할 수 있도록 데이터를 변형하며 지속적으로 데이터를 업데이트함과 동시에 그에 적절한 모델을 생각해볼 계획이다.