

배달 음식 분석 및 예측

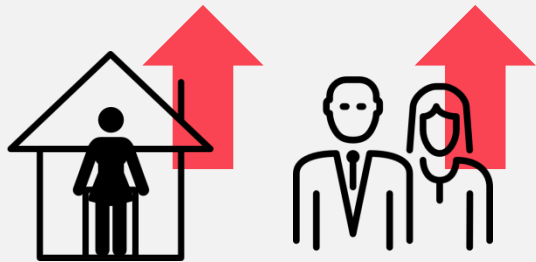
Food Delivery Services Analysis and Prediction



김형민, 윤유진, 안은정, 한미희

빅데이터

01. 개요



일인 가구 맞벌이 부부



배달앱 발달

| 편리한 배달 음식에 대한 수요 증가 |



배달 음식 이용에 큰 영향을 미칠 것으로
예상

01. 개요



대표적 음식 **치킨, 피자, 중국집** 의

이용 건수를 **날씨, 성별, 연령대**에 따라
분석

➡ 분석을 통해 **판매자**는 판매량을 예측하고, 이용자의 선호도에 따라
효율적인 운영에 이익을 줄 것.

02. 관련연구

인간의 감정은
외부적 요인에 의해 변화된다.
(Nowlis, 1956)



외부적 요인 중 **날씨**는
기분에 가장 큰 영향을 준다.
(Howarth, E. & Hoffman, M.S., 1984)



높은 기온, 강렬한 햇빛 > 부정적인 영향

적당한 햇빛 증가 > 행복 감정 촉진



춥고 비 오는 흐린 날 >

활동 제약

우울한 기분

03. 자료수집

기준일	요일	성별	연령대	시도	시군구	업종	통화건수
20161201	목	남	20대	서울특별시	강남구	중국집	5
20161201	목	여	30대	서울특별시	강남구	중국집	5
20161201	목	남	30대	서울특별시	강남구	중국집	43
20161201	목	여	30대	서울특별시	강남구	중국집	39
20161201	목	남	60대 이상	서울특별시	강남구	중국집	5
20161201	목	여	40대	서울특별시	강남구	중국집	41

배달 업종 이용 현황분석(2016년, 2017년)



서울 지역 배달 업종에 대한

기준일, 요일, 성별, 연령대, 지역, 업종, 통화량

이용현황 데이터

date	weather
20161201	cloud
20161202	sun
20161203	cloud
20161204	cloud
20161205	rain
20161206	sun
20161207	snow
20161208	sun

기상청 제공(2016년, 2017년 / 12.1.2월)



일별 날씨

맑음, 흐림, 비, 눈 4가지로 분류

04. 사용기술

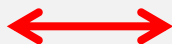
hadoop	대량의 자료를 처리할 수 있는 큰 컴퓨터 클러스터에서 동작하는 분산 응용 프로그램을 지원하는 프리웨어 자바 소프트웨어 프레임워크.
hive	하둡에서 동작하는 데이터 웨어하우스(Data Warehouse) 인프라 구조로서 데이터 요약, 질의 및 분석 기능을 제공.
SRG stat	R기반 시각화 도구
R	통계 계산과 그래픽을 위한 프로그래밍 언어이자 소프트웨어 환경

05. 분석

A그룹

2015년 12월, 2016년 1월, 2월

배달 현황, 날씨



비교, 예측

B그룹

2016년 12월, 2017년 1월, 2월

배달 현황, 날씨

1) 날씨와 연령별 배달음식 선호

```
create table A_food (  
    date string,  
    gender string,  
    age string,  
    food string)
```

```
create table 2016_food_weather (  
    date string,  
    gender string,  
    age string,  
    food string,
```

```
create table cloud(  
    date string,  
    gender string,  
    age string,  
    food string)
```

```
insert overwrite table sun  
select a.date , a.gender, a.age, a.food  
from dec_jan_food_weather a  
where weather = 'sun';
```

```
select count(*)  
from sun  
where age='teenager' and food = 'chicken'
```

날씨에 따른 각 연령대의 배달음식 선호도를 알기 위한 테이블 생성과 데이터 분류, 추출하기

05. 분석

2) 날씨별 연령대 배달음식 선호도

맑음	치킨	피자	중국집		흐림	치킨	피자	중국집
10대	41	27	32		10대	41	28	31
20대	33	30	27		20대	37	30	33
30대	37	27	36		30대	37	28	35
40대	35	28	37		40대	35	29	36
50대	35	25	40		50대	36	25	39
60대	33	21	46		60대	33	21	46
비	치킨	피자	중국집		눈	치킨	피자	중국집
10대	43	26	31		10대	42	27	31
20대	39	28	33		20대	36	31	33
30대	36	26	38		30대	37	28	36
40대	35	27	38		40대	36	29	36
50대	34	23	43		50대	35	26	39
60대	32	19	49		60대	34	21	45

A그룹

맑음	치킨	피자	중국집		흐림	치킨	피자	중국집
10대	41	27	32		10대	41	27	32
20대	37	30	33		20대	37	29	34
30대	37	27	36		30대	38	27	35
40대	36	27	37		40대	35	28	37
50대	35	25	40		50대	35	24	41
60대	33	20	47		60대	33	20	47
비	치킨	피자	중국집		눈	치킨	피자	중국집
10대	42	26	32		10대	42	28	30
20대	37	29	34		20대	37	30	33
30대	36	26	38		30대	36	29	35
40대	35	27	38		40대	34	28	38
50대	35	24	41		50대	35	25	40
60대	33	19	48		60대	34	21	45

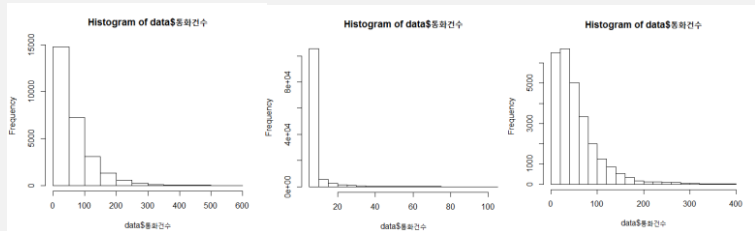
B그룹

▶ A그룹과 B그룹 모두 맑음, 흐림, 눈 오는 날 10, 20, 30대는 치킨집을 선호. 40, 50, 60대는 중국집을 선호

▶ 비 오는 날 10, 20대는 치킨집 선호. But 30대 중국집 선호. 40, 50, 60대 여전히 중국집 선호

05. 분석

3) 연령대별 통화 건수



치킨, 피자, 중국집 통화 건수에 대한 히스토그램

```
> 하 <- which(r<=quantile(r,0.33))
> 중 <- which(r<=quantile(r,0.66))
> 상 <- which(r<=quantile(r)[5])
> r[상] <- "상"
> r[중] <- "중"
> r[하] <- "하"
> table(r)
```

r

상	중	하
9112	8929	9104

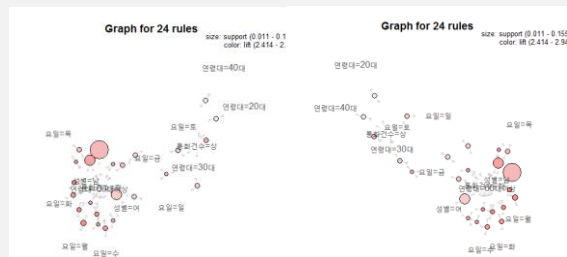
r

상	중	하
47410	6478	87079

연관성 분석을 위한 중국집 통화 건수 분할

	lhs	rhs	support	confidence	lift
[1]	{요일=월, 성별=남, 연령대=60대이상}	=> {통화건수=하}	0.01179703	0.9938272	2.939587
[2]	{요일=수, 성별=남, 연령대=60대이상}	=> {통화건수=하}	0.01179703	0.9907692	2.930542
[3]	{요일=화, 성별=남, 연령대=60대이상}	=> {통화건수=하}	0.01172376	0.9907121	2.930373
[4]	{요일=월, 연령대=60대이상}	=> {통화건수=하}	0.02348415	0.9876733	2.921385
[5]	{요일=월, 성별=여, 연령대=60대이상}	=> {통화건수=하}	0.01168712	0.9815385	2.903239
[6]	{요일=화, 연령대=60대이상}	=> {통화건수=하}	0.02315442	0.9753086	2.884812
[7]	{요일=수, 연령대=60대이상}	=> {통화건수=하}	0.02308115	0.9692308	2.866835

통화량에 대한 분석 결과

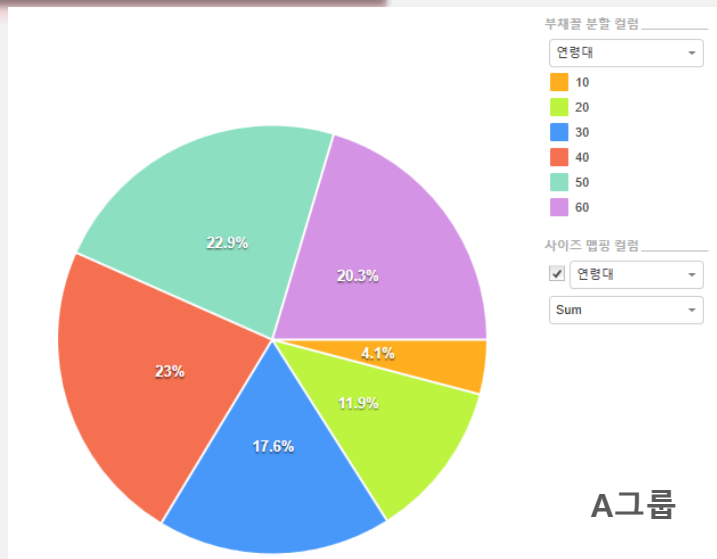


통화 건수에 대한 분석 그래프

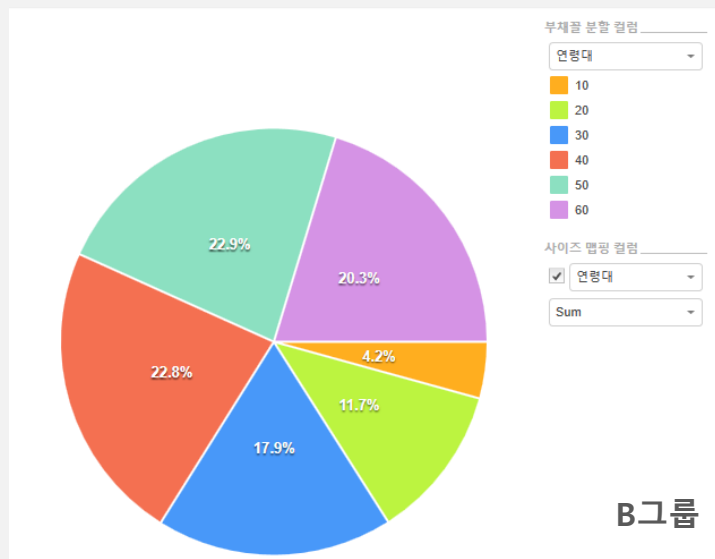
연령대별 통화 건수를 알아보기 위한 분석

05. 분석

3) 연령대별 통화 건수



40대 (23%)가 가장 높았고 50대, 60대
30대, 20대 순이며, 10대 (4.1%)가 가장 낮음.

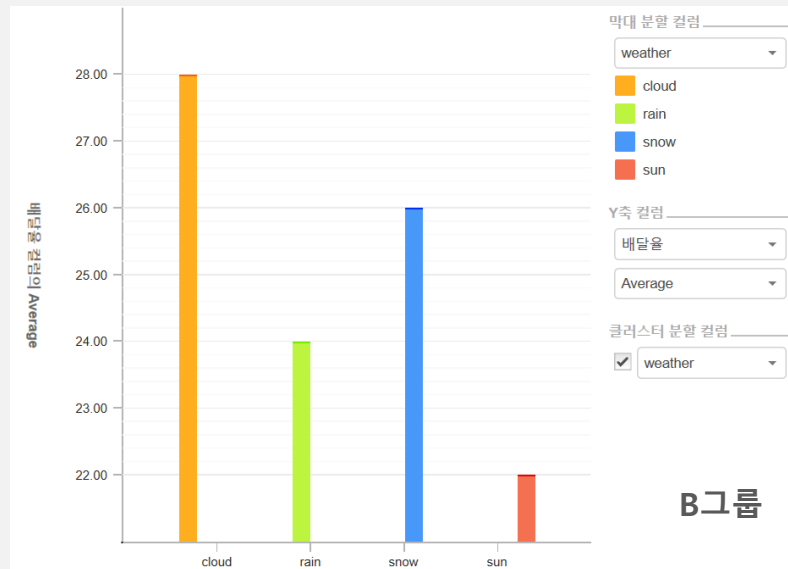
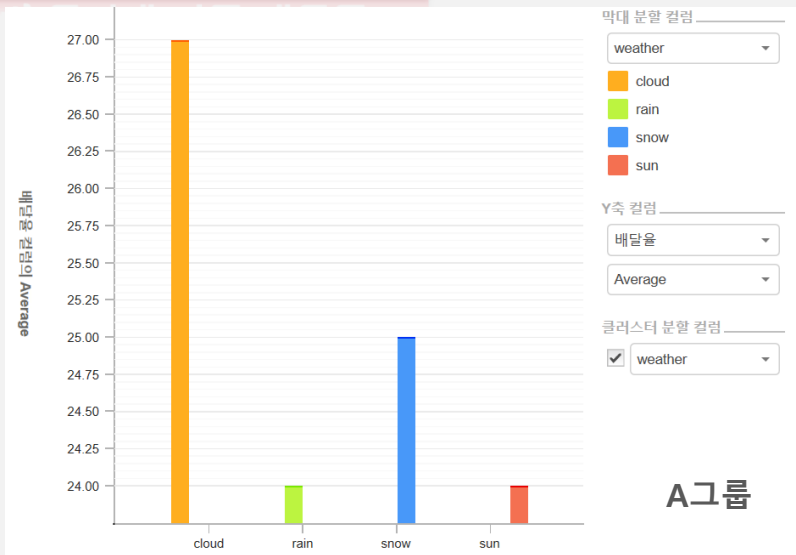


50대 (22.9%)가 가장 높았고 40대, 60대
30대, 20대 순이며, 10대 (4.2%)가 가장 낮음.

▶ 배달 업체를 이용하는 주 고객층은 40대, 50대, 60대라는 것을 알 수 있으며
A와 B그룹의 수치가 거의 동일

05. 분석

4) 날씨에 따른 배달율



▶ A, B그룹 모두 **흐림** > 눈 > 비 > 맑은 날 순으로 배달율이 높았다.

Q&A



Thank you