

2019-1
데이터사이언스입문 프로젝트

잇츠 잇츠(ITS EATs):
데이터를 이용한 배달 수요 예측 및 서비스 제안

이은지
김재인
김문정



팀원 소개



김문정

프로젝트 소개
문제 정의
데이터 소개
데이터 정리 과정



김재민

언어 및 도구
모델링 소개
프로젝트 해결방향
진행 일정 소개



이은지

예상 최종산출물
분석 기대효과
홍보 방안 소개



- 푸드테크란?

식품(Food)과 기술(Technology)를 결합한 신조어

- 푸드테크 활용 사례



- 푸드테크란?

식품(Food)과 기술(Technology)를 결합한 신조어

- 푸드테크 활용 사례





2 문제 정의

가상의 신규 배달앱 A 업체



- 서울 지역을 중심으로 배달앱 시장 진출 예정
- 비용 대비 효율적인 앱 홍보 방향과 서비스에 대한 솔루션 필요

3 데이터 소개

SK telecom Big Data Hub

데이터 카탈로그 인포그래피 이용가이드 공지사항

공공/교통 과학/기술 생활 경제/사회

18년 12월 서울시 중국음식 판매업종 이용 통화량

서울지역 중국집에 대한 2018년 12월 한달 통화량을 기반으로 지역별 이용 현황 데이터를 제공합니다.

다운로드 수: 17

CSV XLSX XLS TSV

▶서울지역 중국집에 대한 한달 통화량을 기반으로 지역별 이용 데이터를 제공합니다.
 ▶SK 텔레콤의 통화량 데이터와 업종 데이터를 기반으로 추출한 데이터로 전체 중국집 서비스 이용 현황이 반영되어 있지 않음을 감안하여 데이터를 활용하시기 바랍니다.
 ▶본데이터는 월 1회 정기적으로 업데이트 됩니다.

1048576

전체 데이터 개수

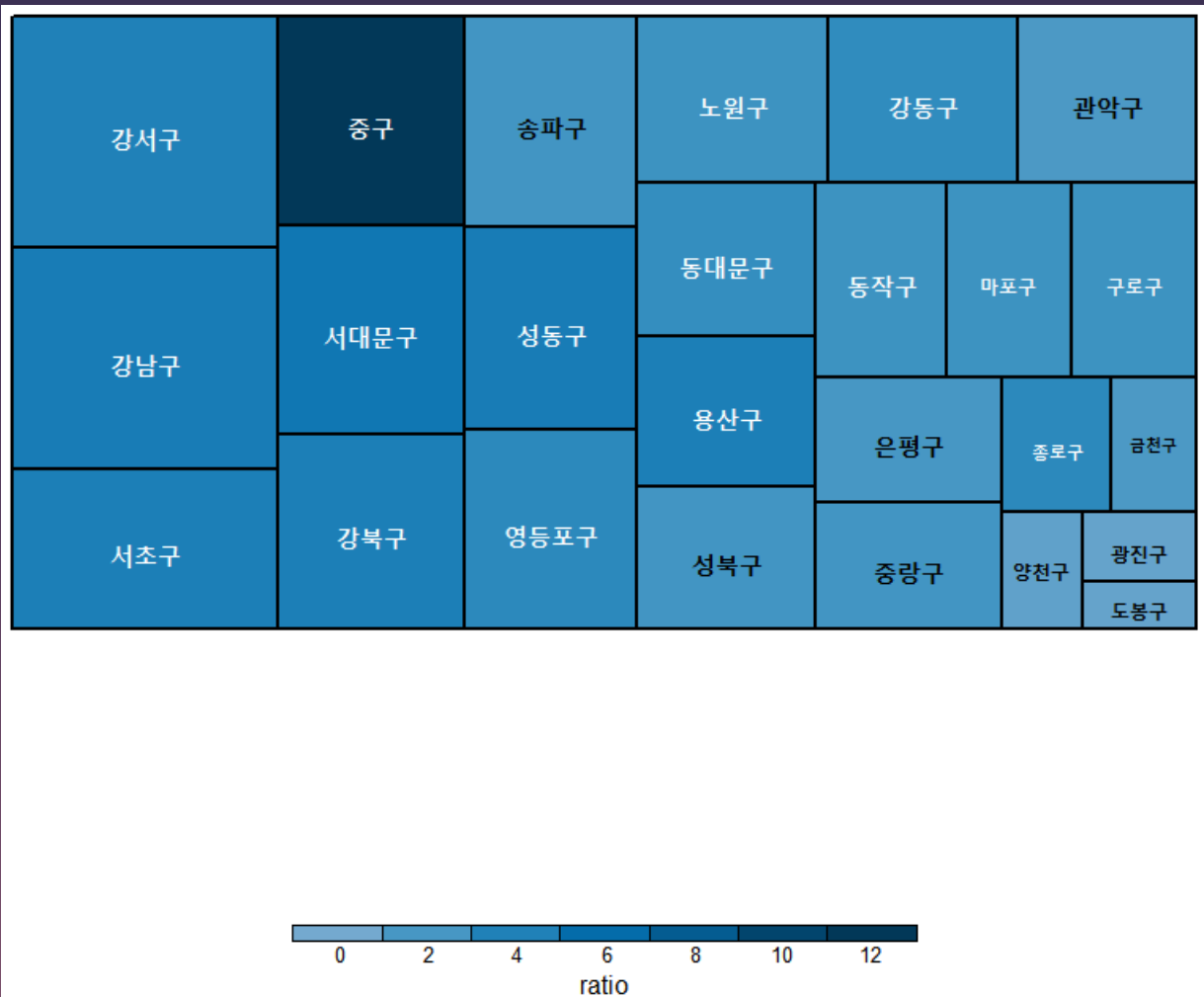
9개

해당 데이터에 주어진 변수

Response data			
컬럼 정보	컬럼명	컬럼설명	컬럼타입
	기준일	기준일	Number
	요일	요일	Text
	성별	성별	Text
	연령대	연령대	Text
	시도	시도	Text
	시군구	시군구	Text
	읍면동	읍면동	Text
	업종	업종	Text
	통화건수	통화건수	Number



데이터 소개



1048576

전체 데이터 개수

9개

해당 데이터에 주어진 변수



데이터 소개



25개구
서울시 행정구

365일
2018년 1년간 자료

서울시/기상청
두 개의 관측시스템

4

데이터 정리

```
#2. pizza
```{r}
pizza <- read.csv("pizza.csv",header=TRUE, sep=",")
```

# 기본키 일치시키기
```{r}
rain$date = rain$date %>% str_remove_all(pattern = '-')
pizza$date = as.character(pizza$date)
rain$date = as.character(rain$date)

pizza$region = as.character(pizza$region)
rain$region = as.character(rain$region)
```


# RAIN과 pizza 데이터를 date와 region에 따라 병합
```{r}
pizza_final= left_join(pizza,rain,by=c("date","region"))
View(pizza_final)

write.csv(x = pizza_final, file = 'c:\\users\\김문정\\Desktop\\2019\\데이터사이언스입문\\weather\\pizza_final.csv')
```
```

- 기상 자료 중 강수량과 평균 기온에 관한 정보 추출
- left_join함수를 이용해 배달량 데이터에 rain 변수와 temp 변수 추가




R Studio

 [rstudio::conf](#) [Products](#) [Resources](#) [Pricing](#) [About Us](#) [Blogs](#) [Search](#)

RStudio

Open source and enterprise-ready professional software for R


[Download RStudio](#)[Discover Shiny](#)[shinyapps.io Login](#)[Discover RStudio Connect](#)



RStudio

RStudio makes R easier to use. It includes a code editor, debugging & visualization tools.






[Download](#) [Learn More](#)



Shiny

Shiny helps you make interactive web applications for visualizing data. Bring R data analysis to life.

[Learn More](#)



R Packages

Our developers create popular packages to expand the features of R. Includes ggplot2, dplyr, R Markdown & more.

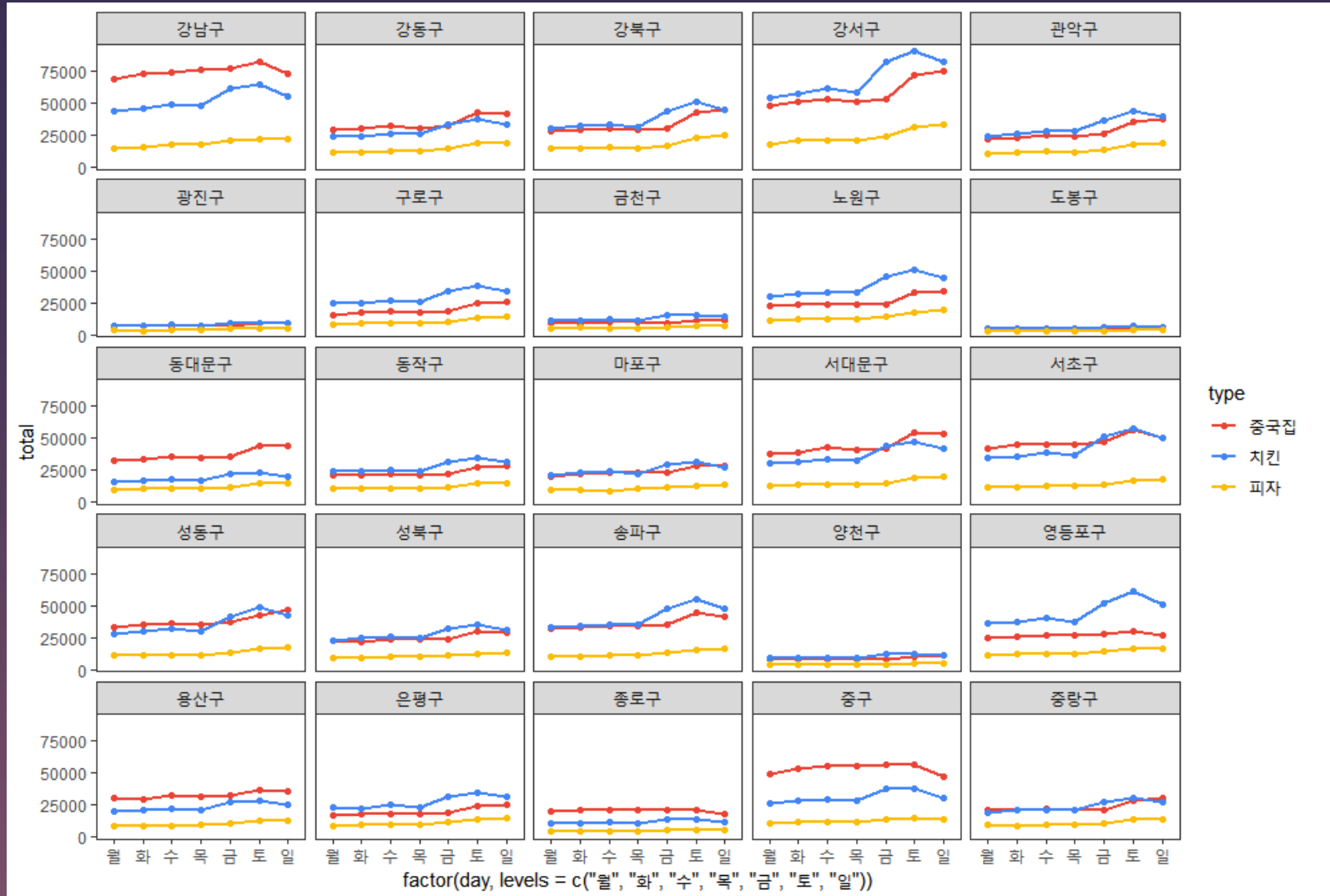
[Learn More](#)

Shiny(샤이니)

The screenshot displays the ShinyApps.io admin dashboard. The left sidebar contains navigation links: Dashboard, Applications (with sub-links for All, Running, Sleeping, and Archived), Settings, and Account. The main content area features a summary card for 'ONLINE APPLICATIONS' showing 1 application. Below this, a table lists application statuses: Running (0), Sleeping (1), and Archived (0). To the right, a 'RECENT APPLICATIONS' table shows a single application with ID 23668, Name 'demo', and Status 'Sleeping'. A user menu for 'Test User' is open in the top right, showing options for Profile, Tokens, and Log out. The footer includes copyright information for RStudio Inc. and a link to the Terms of Use.

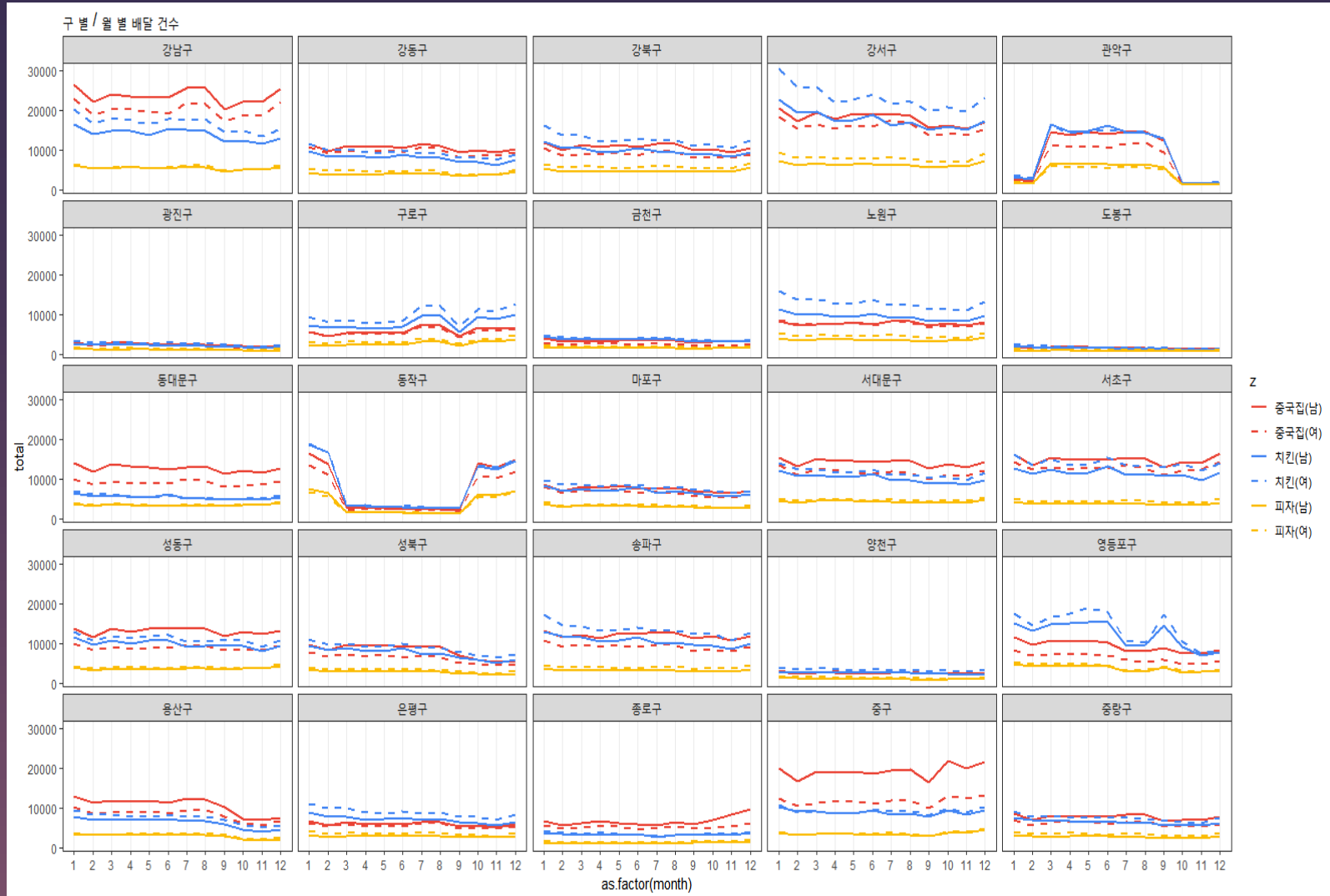
| Id | Name | Status |
|-------|------|----------|
| 23668 | demo | Sleeping |

- 변수의 통계적 특성 파악
- 변수 간 상관관계 분석



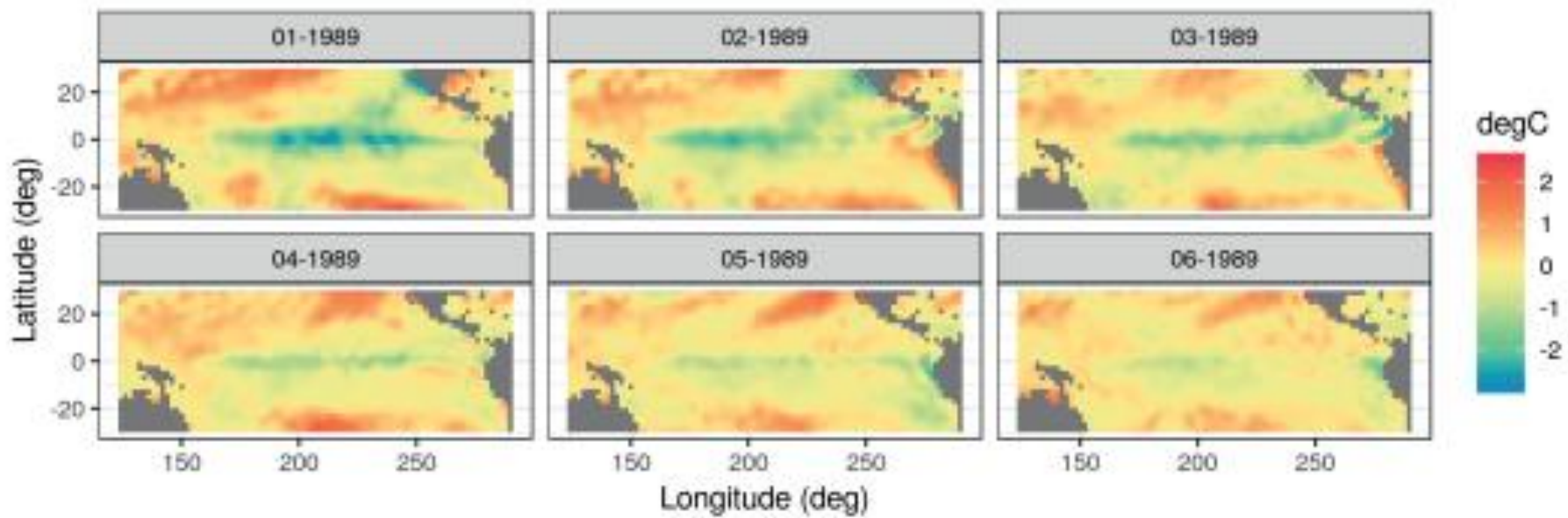
행정구별 배달 건수 비교

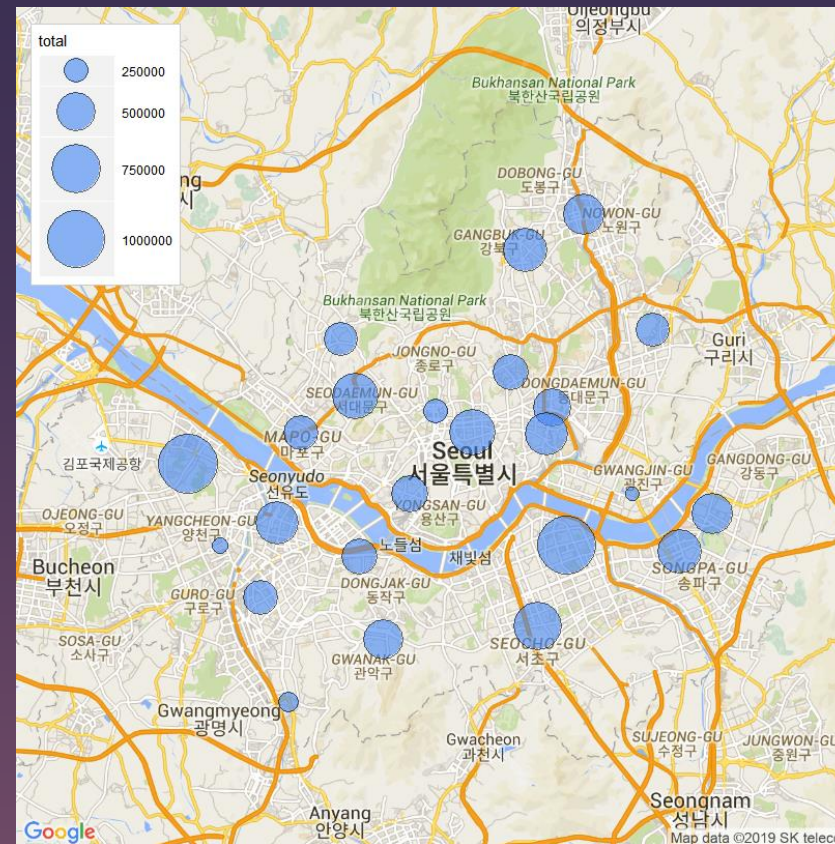
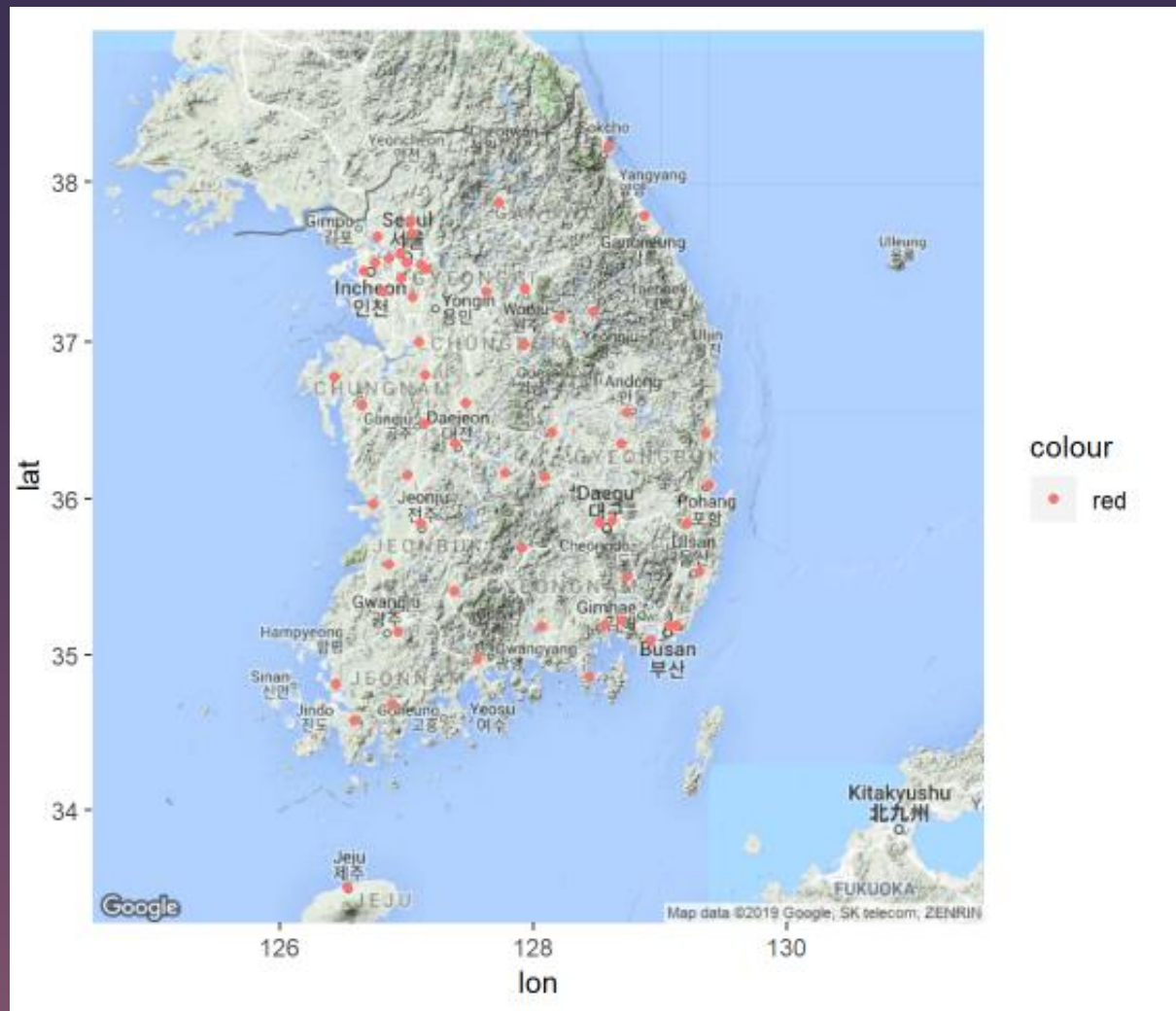
- 변수의 통계적 특성 파악
- 변수 간 상관관계 분석



성별에 따른 배달 건수 비교

| | | |
|-----------------------|----------------------|------------------------------------|
| Supervised Learning | Classification Model | Binary Classification |
| | | Multinomial Classification |
| | | Support Vector machine |
| | | Decision Tree |
| | Regression Model | Linear regression |
| | | Locally weighted linear regression |
| | | Ridge |
| Unsupervised Learning | Clustering Model | Lasso |
| | | K means |
| | | Density estimation |
| | | Expectation maximization |
| | | Pazen window |
| | | DBSCAN |



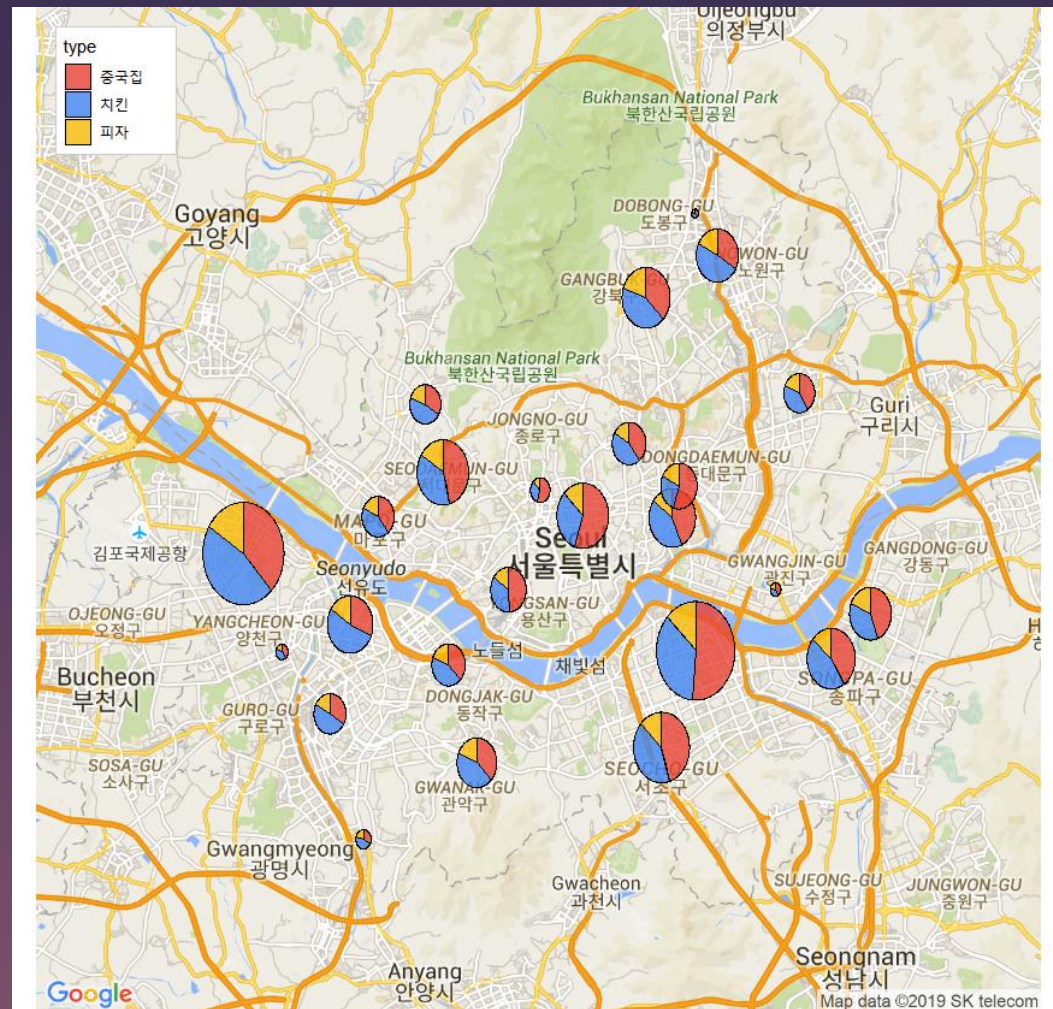


각 행정구별 총 배달 건수

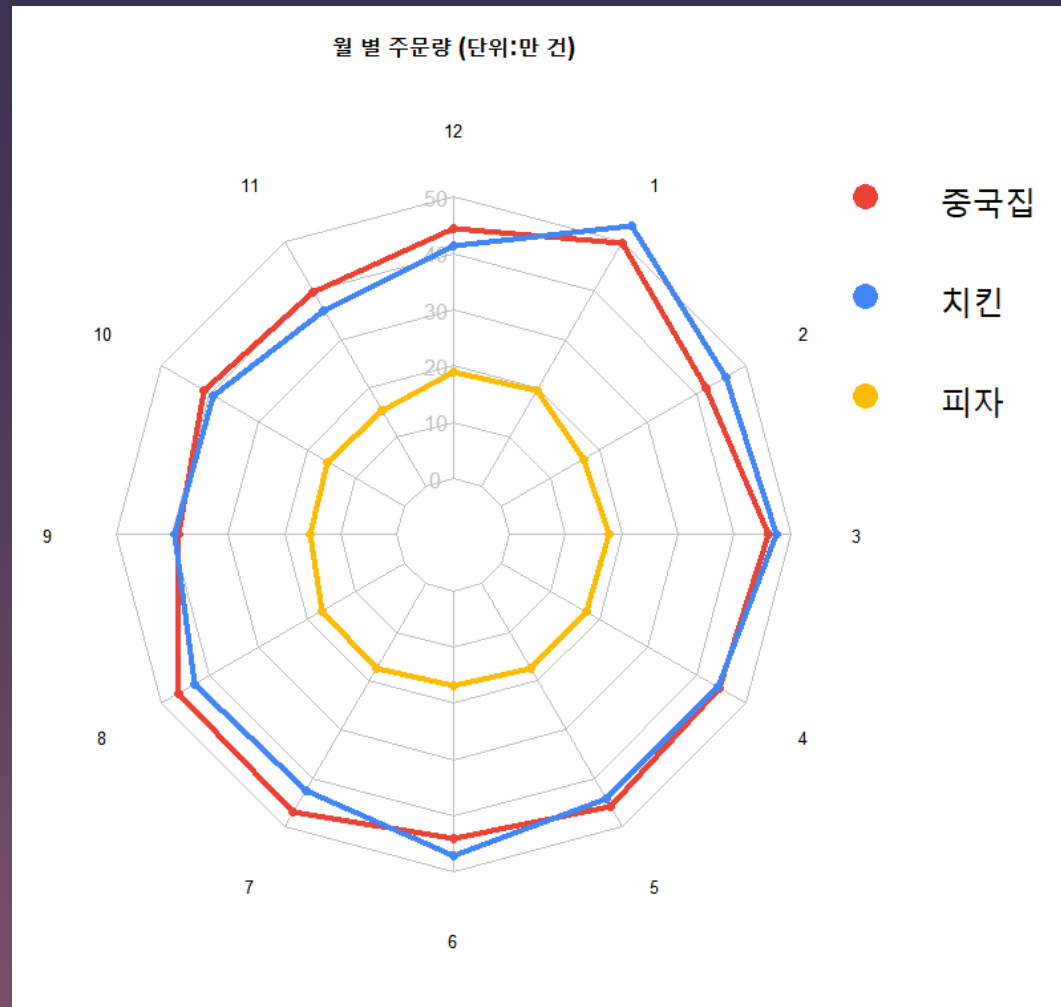
| TO DO | 4.25-
5.1 | 5.2-
5.8 | 5.9-
5.15 | 5.16-
5.22 | 5.23-
5.29 | 5.30-
6.6 | 6.7-
6.13 |
|--------|--------------|-------------|--------------|---------------|---------------|--------------|--------------|
| EDA | | | | | | | |
| 모델링 | | | | | | | |
| 시각화 | | | | | | | |
| 결과 분석 | | | | | | | |
| 서비스 제안 | | | | | | | |

- EDA 를 통한 시각화
- 모델링의 결과
- 결과 해석 및 기대 효과
 - 서비스 제안

- EDA 를 통한 시각화
- 모델링의 결과
- 결과 해석 및 기대 효과
 - 서비스 제안



- EDA 를 통한 시각화
- 모델링의 결과
- 결과 해석 및 기대 효과
 - 서비스 제안



기대효과 (서비스 제안)

1. 위치 기반 서비스 이용

“강남구 도곡동, 금요일에 치킨 주문량이 많다”면?



기대효과 (서비스 제안)

1. 위치 기반 서비스 이용

“강남구 도곡동, 금요일에 치킨 주문량이 많다”면?

2. 기상 정보 이용

“비가 오면 중국집 주문량이 많다”면?

기대효과 (서비스 제안)

1. 위치 기반 서비스 이용

“강남구 도곡동, 금요일에 치킨 주문량이 많다”면?

2. 기상 정보 이용

“비가 오면 중국집 주문량이 많다”면?





THANK YOU