

|  |  |  |  |
| --- | --- | --- | --- |
| **2020 혁신성장 빅데이터 분석 프로젝트 요약서** | | | |
| **팀 명** | **1조** | **팀 장** | **최재혁** |
| **팀 원** | **최재혁, 오현지, 최호경, 오인혜** | | |
| **주 제 명** | **치킨전문점 창업전략** | | |

|  |  |  |
| --- | --- | --- |
| 구분 | 프로젝트 상세내용 | |
| **프로젝트명** | 데이터 분석을 통한 치킨전문점 창업전략 | |
| **구현목표** | 치킨전문점을 창업할 점주들을 위한 맞춤 정보를 제공 | |
| **구현내용** | 1. PPT 표지 | 2. 데이터 분석 시각화 |
|  |  |
| 3. 데이터 분석 모델링 | 4. 결론 |
|  |  |
| **상세 설명** | 1. 저희 조가 주제로 선정한 ‘ 치킨 전문점 창업전략 ’ 의 PPT 자료의 표지 입니다.  치킨이라는 업종에 적합한 배경을 선택하였고, 치킨과 어울리는 노란색으로 포인트를  주었습니다.  2. R 프로그램을 이용하여 수집한 데이터를 분석한 후, 변수들의 내용을 시각화 하였습니다.  막대그래프를 이용하여 변수들 간의 수치를 더욱 효과적으로 비교할 수 있습니다. 그리고  서로 다른 색상을 활용하여 데이터의 가시성을 상승시켰습니다.    3. 데이터 분석 및 모델링 부분입니다. 선형 회귀분석을 통해 치킨전문점의 10대,20대 단골고객과 여성고객 간의 유의미한 관계 및 변수의 수치변화에 따른 예측 값을 도출해 내려고 했지만,  데이터의 성격이 비율이었기 때문에 한계가 있었습니다. 하지만 예측 정확성과 P-value 값을  통해 통계적으로 의미가 있다는 것을 확인하였습니다.  4. 마지막으로 PPT 자료의 결론 부분입니다. 모델링 부분에서 정확한 예측 값을 도출하지는 못했지만, 치킨전문점 창업과 밀접한 관련이 있는 변수들을 선정 및 분석하여 미래의 치킨전문점  창업주들에게 유용한 데이터를 제공할 수 있다고 생각합니다. | |

|  |  |  |
| --- | --- | --- |
| **각 팀원 역할** |  |  |
|  |  |
| **최재혁** : 조장역할 수행. 프로젝트 전반적인 방향 설정 및 세부 사항 관리. 5일 간의 프로젝트 동안 조원들에게 역할 분담 및 업무 할당. 조원들이 실시한 작업 검토 및 취합.  반복적인 피드백을 통해 프로젝트의 진행 상황 및 조원들의 업무 수행 확인.  매일 조별 작업일지 작성 후 강사님께 제출. 매일 아침 강사님과 조장들 간의 미팅 후 전파사항을 조원들에게 전달. 데이터 분석 모델링 및 최종 보고서 작성.  ( R 프로그램 활용 )  **오현지** : 데이터 수집 및 데이터 해석, 보고서 발표 역할 수행. 데이터 수집 및 EDA( 탐색적 데이터 분석 ), 분석 모델링 등 프로젝트 전 과정에 참여. 주로 시각화한 데이터의 해석을 전담( 데이터의 목적, 해석 및 결과 등 ). 완성된 PPT의 발표를 담당 (6/6(금). ( R 프로그램 활용 )  **최호경** : 데이터 수집 및 전반적인 EDA 과정 수행. 데이터 수집 및 EDA( 탐색적 데이터  분석 ), 분석 모델링 등 프로젝트 전 과정에 참여. 필요한 데이터의 추가 수집  혹은 프로젝트 진행 중 문제점에 봉착 시, 개선방안 적극 제시. PPT 제작 보조 역할 수행  ( R 프로그램 활용 )  **오인혜** : 데이터 분석 및 시각화, PPT 제작 역할수행. 데이터 수집 및 EDA( 탐색적 데이터  분석 ), 분석 모델링 등 프로젝트 전 과정에 참여. 주로 수집한 데이터의 분석 및 시각화 수행. 워드 클라우드 전담. PPT 제작을 주도적으로 진행.  ( R 프로그램 활용 ) | |

치킨전문점 창업전략

최재혁, 오현지, 최호경, 오인혜

2020년 6월 5일

### 1. 주제 선정 및 조사 개요

## 문제정의 : 치킨전문점 창업 전략

치킨전문점의 성공적인 창업을 위해 치킨전문점의 무분별한 진입을 방지하고 경영개선을 유도하기 위해 필요한 데이터 정보를 제공

## 조사 배경

우리나라 사람들의 치킨 사랑은 이루 말로 다 할 수 없을 정도입니다. 이에 따라 치킨집도 우후죽순으로 늘어나고 있습니다. 특히 퇴직자들에게 치킨집은 은퇴 후 한 번쯤 고려해보는 필수 사업 아이템입니다. 2019년 2월 기준 지방행정 인허가 자료를 보면 전국에서 영업 중인 치킨집은 8만 7,000곳으로 조사되었습니다. 맥도날드 가맹점이 2017년 기준 약 3만 4,000개임을 감안하면 많은 수입니다.

치킨집 창업의 가장 큰 장점은 적은 창업 비용입니다. 한 증권사에서 조사한 자영업 분석 보고서에 따르면 치킨집 창업 비용은 평균 약 5,700만 원으로 약 1억 1,000만 원인 커피숍의 절반 정도입니다. 평균비용이기 때문에 지역에 따라, 가게 크기에 따라, 카페형, 배달 전문형 같은 유형에 따라 비용이 천차만별로 달라질 수는 있습니다. 두 번째 장점은 치킨은 확실한 사업 아이템이라는 점입니다. 2018년 국내 외식 트렌드 조사에 따르면 응답자의 절반 이상이 치킨을 선호하는 것으로 나타났습니다.

하지만 치킨의 창업 매장보다 폐업 매장이 더 많은 것이 현실입니다. 자영업 분석 보고서에 따르면 치킨집의 창업 수와 폐업 수가 있는데 2018년 기준 창업 매장 수는 6,200개, 폐업 매장 수는 8,400개입니다. 2018년 치킨집 폐업율은 10.0 입니다. 한 해에 10곳 중 1곳이 치킨집 사업을 접고 있습니다. 경쟁이 심한 만큼 폐업을 하는 매장도 많은 상황입니다. 따라서 상기의 장단점의 이유를 바탕으로 변수를 지정하여 어떻게 하면 치킨전문점의 창업을 성공적으로 할 수 있는지에 대해 분석 하였습니다.

### 2. 데이터 전 처리

## 2-1. Excel 파일 읽기 및 데이터 확인

library( rJava ) # EDA에 필요한 패키지들 로드하기  
library( xlsx )  
library( mlbench )  
library( dplyr )

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library( ggplot2 )  
library( gridExtra )

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library( RColorBrewer )  
library( corrplot )

## corrplot 0.84 loaded

library( multilinguer )

## If you have any problem, please issue here.

## https://github.com/mrchypark/multilinguer/issues

library( wordcloud )  
library( wordcloud2 )  
library( brew )  
  
  
setwd("C:\\Workspace\\WorkspaceR") # 주 활용 data  
chicken.xlsx <- read.xlsx( file = "chicken.xlsx",   
 sheetIndex = 1,  
 encoding = "UTF-8" )  
  
setwd("C:\\Workspace\\workspaceR") # 성별원그래프 비율 data  
gender <- read.xlsx( file = "성별.xlsx",   
 sheetIndex = 1,  
 encoding = "UTF-8" )  
   
setwd("C:\\Workspace\\workspaceR") # 프랜차이즈 비율 data  
franchising <- read.xlsx( file = "franchise.xlsx",   
 sheetIndex = 1,  
 encoding = "UTF-8" )  
  
setwd("C:\\Workspace\\workspaceR") # 나이 비율 data  
age <- read.xlsx( file = "나이.xlsx",   
 sheetIndex = 1,  
 encoding = "UTF-8" )  
  
setwd("C:\\Workspace\\workspaceR") # 치킨전문점 업종전환이유 data   
reason <- read.xlsx( file = "업종전환이유.xlsx",   
 sheetIndex = 1,  
 encoding = "UTF-8" )  
  
setwd("C:\\Workspace\\workspaceR") # 서울치킨집수 data   
seoul.xlsx <- read.xlsx( file = "서울치킨집수.xlsx",   
 sheetIndex = 1,  
 encoding = "UTF-8" )  
  
setwd("C:\\Workspace\\workspaceR") #서울특별시 10대20대 여성고객 치킨통화량 수   
치킨통화량.xlsx <- read.xlsx( file = "서울6월치킨통화량.xlsx",   
 sheetIndex = 1,  
 encoding = "UTF-8" )  
  
class(chicken.xlsx);class(gender);class(franchising);class(age);class(reason);class(seoul.xlsx);class(치킨통화량.xlsx)

## [1] "data.frame"

## [1] "data.frame"

## [1] "data.frame"

## [1] "data.frame"

## [1] "data.frame"

## [1] "data.frame"

## [1] "data.frame"

dim(chicken.xlsx);dim(gender);dim(franchising);dim(age);dim(reason);dim(seoul.xlsx);dim(치킨통화량.xlsx)

## [1] 8 18

## [1] 2 2

## [1] 2 2

## [1] 3 2

## [1] 4 2

## [1] 25 2

## [1] 25 2

str(chicken.xlsx);str(gender);str(franchising);str(age);str(reason);str(seoul.xlsx);str(치킨통화량.xlsx)

## 'data.frame': 8 obs. of 18 variables:  
## $ 업종 : chr "한식" "중식" "일식" "서양식" ...  
## $ 평균방문고객수1일 : num 47.9 26.8 49 52.9 64.3 39.7 13.3 64.2  
## $ 평균배달수1일 : num 2.6 69.7 4.8 2.3 1.5 25.1 23.3 4.6  
## $ 평균테이크아웃수1일 : num 1.8 3.2 2.7 1.8 4.1 19.6 5 22.7  
## $ 매출액대비영업이익률2018: num 10.6 10.9 13.4 15.5 14.8 13.2 10.8 13.3  
## $ 월평균매출액 : num 1524 1848 2606 2094 2044 ...  
## $ 업종전환경험여부 : num 11.1 7.2 15.4 7.5 21 11.9 12.3 8.7  
## $ 전환이유영업이익감소 : num 47.4 52 31.3 58.2 59.4 50.1 34.3 64.3  
## $ 전환이유동종업종경쟁 : num 26.7 3.9 15.5 13.7 23.2 16.3 29.8 20.1  
## $ 배달앱이용비율 : num 6.9 37.2 15.7 12.2 8 55.2 57.4 7.3  
## $ 프랜차이즈가게비율 : num 9.1 5.8 11.5 26.5 38.9 74.9 68.8 22.5  
## $ 전국사업체수2018 : num 313562 24546 13436 12607 4547 ...  
## $ 남성고객 : num 54.4 54.3 57.4 37.6 39.2 42 54.1 41.3  
## $ 여성고객 : num 45.6 45.7 42.6 62.4 60.8 58 45.9 58.7  
## $ 고객10\_20대 : num 3 10.7 13.4 37.5 35.4 54.8 22.2 29  
## $ 고객30\_40대 : num 47 61.2 63.2 60.8 60.7 43.4 64 55.1  
## $ 고객50대이상 : num 50 28.1 23.4 1.7 3.9 1.8 13.8 15.9  
## $ 배달외식평균지출비용 : num 14556 16432 9739 18622 9606 ...

## 'data.frame': 2 obs. of 2 variables:  
## $ 성별: chr "여성" "남성"  
## $ 비율: num 54.1 45.9

## 'data.frame': 2 obs. of 2 variables:  
## $ 프랜차이즈여부: chr "프랜차이즈" "비프랜차이즈"  
## $ 비율 : num 68.8 31.2

## 'data.frame': 3 obs. of 2 variables:  
## $ 나이: chr "10대20대" "30대40대" "50대이상"  
## $ 비율: num 22.2 64 13.8

## 'data.frame': 4 obs. of 2 variables:  
## $ 이유: chr "영업이익감소" "동종업종경쟁" "구인의어려움" "기타"  
## $ 비율: num 34.3 29.8 11 24.9

## 'data.frame': 25 obs. of 2 variables:  
## $ 지역구 : chr "강서구" "강남구" "강동구" "강북구" ...  
## $ 치킨전문점수: num 224 357 211 138 245 308 214 116 255 186 ...

## 'data.frame': 25 obs. of 2 variables:  
## $ 지역구 : chr "강서구" "강남구" "강동구" "강북구" ...  
## $ 총통화량: num 250 347 173 138 98 72 175 101 140 42 ...

저희 조는 주제에 적합한 데이터를 활용하기 위해, 7개의 xlsx형식의 파일을 새롭게 제작하였습니다. 그리고 그 데이터는 ‘통계청’ 및 ‘2019년 외식산업경기전망지수 종합보고서.pdf’, ‘2019 국내 외식트렌드 조사 보고서.pdf’, ‘2019 식품외식산업 주요통계.pdf’, ‘SKT 데이터 허브’, ‘소상공인 상권정보시스템’ 등에서 수집하였습니다.

그리고 각각의 파일들의 기본 정보를 class, str, dim 함수를 이용하여 확인하였습니다. 확인 결과 7개 xlsx형식의 파일 (‘chicken.xlsx’, ‘gender’, ‘franchising’, ‘age’, ‘reason’, ‘seoul.xlsx’, ‘치킨통화량.xlsx’ 모두 data.frame 이었고, 행(관측치)과 열(변수)은 각각 ‘8행 18열’, ‘2행 2열’, ‘2행 2열’, ‘3행 2열’, ‘4행 2열’, ‘25행 2열’, ‘25행 2열’ 이었습니다.

## 2-2. 변수 지정

industry <- c("한식","중식","일식","서양식","기타외국식","피자햄버거","치킨전문점","김밥및간이음식점")  
visit<- chicken.xlsx %>% select(업종, 평균방문고객수1일)  
delivery <- chicken.xlsx %>% select(업종, 평균배달수1일)  
takeout <- chicken.xlsx %>% select(업종, 평균테이크아웃수1일)  
margin <- chicken.xlsx %>% select(업종, 매출액대비영업이익률2018)  
sales <- chicken.xlsx %>% select(업종, 월평균매출액)  
change <- chicken.xlsx %>% select(업종,업종전환경험여부)  
salesdown <- chicken.xlsx %>% select(업종,전환이유영업이익감소)  
compete <- chicken.xlsx %>% select(업종,전환이유동종업종경쟁)  
delivery\_app <- chicken.xlsx %>% select(업종,배달앱이용비율)  
all\_business <- chicken.xlsx %>% select(업종,전국사업체수2018)  
franchise<- chicken.xlsx %>% select(업종,프랜차이즈가게비율)  
male <- chicken.xlsx %>% select(업종,남성고객)  
female <- chicken.xlsx %>% select(업종,여성고객)  
young <- chicken.xlsx %>% select(업종,고객10\_20대)  
middle <- chicken.xlsx %>% select(업종,고객30\_40대)  
old <- chicken.xlsx %>% select(업종,고객50대이상)  
expenditure <- chicken.xlsx %>% select(업종,배달외식평균지출비용)  
expert <- seoul.xlsx %>% select(지역구,치킨전문점수)

저희 조가 선정한 18개의 변수(chicken.xlsx파일 중)는 치킨 전문점이 배달을 위주로 영업을 한다는 점에 착안하여 치킨 전문점과 경쟁 업종인 한식, 중식, 일식, 서양식, 기타외국식, 피자햄버거, 김밥및간이음식점과의 유의미한 비교를 목적으로 선택하였습니다. 주요 소비자들의 성별 및 연령대를 분석하여 치킨 전문점의 수익 증대에 도움을 줄 수 있는 정보를 제공하고자 노력하였습니다.

변수에 대해 설명하겠습니다. 먼저 chicken.xlsx 파일의 변수 중의 업종 중에서 일반 음식점업은 각종의 정식류를 조리하여 제공하는 산업활동으로서 한식과 일식, 중식, 서양식 및 기타 외국 음식점업을 운영하는 산업활동을 말합니다. 피자, 햄버거, 샌드위치 및 유사 음식점업은 피자, 햄버거, 샌드위치, 토스트 및 유사 음식을 직접 조리하여 일반소비자에게 판매하는 산업활동을 말합니다. 분식 및 김밥 전문점은 김밥, 만두, 찐빵, 라면, 떡볶이 등의 간이음식을 조리하여 제공하는 음식점을 운영하는 산업활동을 말합니다. 치킨전문점은 양념치킨, 프라이드치킨 등 치킨 전문점을 운영하는 산업활동을 말합니다.

1일 평균 방문 고객수는 각 업종에 하루동안 얼마나 많은 사람들이 방문한 수를 평균으로 나타낸 것입니다. 1일 평균 배달수는 각 업종에 하루동안 얼마나 많은 배달을 하였는지 평균으로 나타낸 것입니다. 1일 평균 테이크아웃 수는 각 업종에 하루 동안 얼마나 많은 테이크아웃을 하였는지 평균으로 나타낸 것입니다. 2018 매출액대비 영업이익률은 기업의 주된 영업활동에 의한 경영성과를 판단하기 위한 지표로서 제조 및 판매활동과 직접 관계가 없는 영업외손익을 제외한 순수한 영업이익만을 매출액과 대비한 것으로 판매 마진을 나타낸다고 볼 수 있습니다.

월평균매출액 기업의 주요 영업활동 또는 경상적 활동으로부터 얻는 수익으로서 상품 등의 판매 또는 용역의 제공으로 실현된 월평균 금액을 말합니다. 업종전환경험여부는 업종을 다른 업종으로 전환한 경험을 얼마나 하였는지를 표현한 것입니다. 전환이유동종업종경쟁은 업종 전환의 이유가 동종업종경쟁으로 인해 발생하였다는 것을 의미합니다. 전환이유영업이익감소는 업종 전환의 이유가 영업이익감소로로 인해 발생하였다는 것을 의미합니다. 배달앱 이용 비율은 각 업종에서 배달앱을 이용하는 비율입니다.

2018 전국 사업체 수는 2018년도 전국의 지역별 사업체의 규모 및 분포를 파악한 것입니다. 프랜차이즈가게 비율은 프랜차이즈 방식으로 벌이는 사업. 또는 그와 관련한 사업의 가맹점 비율입니다. 남성고객은 각 업종을 남성들이 이용한 비율입니다. 여성고객은 각 업종을 여성들이 이용한 비율입니다. 고객 10대~20대는 각 업종을 10대~20대 고객이 이용한 비율입니다. 고객 30대~40대는 각 업종을 30대~40대 고객이 이용한 비율입니다. 고객 50대 이상은 각 업종을 50대 이상의 고객이 이용한 비율입니다.

## 2-3. 결측치, 특이값 확인

sum(is.na(chicken.xlsx)) # 결측치 0

## [1] 0

sum(is.na(gender)) # 결측치 0

## [1] 0

sum(is.na(franchising)) # 결측치 0

## [1] 0

sum(is.na(age)) # 결측치 0

## [1] 0

sum(is.na(reason)) # 결측치 0

## [1] 0

sum(is.na(seoul.xlsx)) # 결측치 0

## [1] 0

sum(is.na(치킨통화량.xlsx)) # 결측치 0

## [1] 0

boxplot.stats(chicken.xlsx$평균방문고객수1일)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$평균배달수1일)$out # 특이값 1개( 69.7 )

## [1] 69.7

boxplot.stats(chicken.xlsx$평균테이크아웃수1일)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$매출액대비영업이익률2018)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$월평균매출액)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$업종전환경험여부)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$전환이유영업이익감소)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$전환이유동종업종경쟁)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$배달앱이용비율)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$배달대행이용비율)$out # 특이값 0

## NULL

boxplot.stats(chicken.xlsx$프랜차이즈가게비율)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$전국사업체수2018)$out # 특이값 1개( 313562 )

## [1] 313562

boxplot.stats(chicken.xlsx$남성고객)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$여성고객)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$고객10\_20대)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$고객30\_40대)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$고객50대이상)$out # 특이값 0

## numeric(0)

boxplot.stats(chicken.xlsx$배달외식평균지출비용)$out # 특이값 0

## numeric(0)

boxplot.stats(gender$비율)$out # 특이값 0

## numeric(0)

boxplot.stats(franchising$비율)$out # 특이값 0

## numeric(0)

boxplot.stats(age$비율)$out # 특이값 0

## numeric(0)

boxplot.stats(reason$비율)$out # 특이값 0

## numeric(0)

boxplot.stats(seoul.xlsx$치킨전문점수)$out # 특이값 1개( 357 )

## [1] 357

boxplot.stats(치킨통화량.xlsx$총통화량)$out # 특이값 0

## numeric(0)

변수별로 결측치나 특이값을 확인하기 위해 boxplo.stats함수를 이용하였습니다. 어떠한 파일의 변수에서도 결측치는 발견되지 않았고, chicken.xlsx 파일의 변수 평균배달수1일 과 전국사업체수2018에서 특이값이 각각 1개씩( 69.7, 131562 ) 발견되었습니다. 그리고 seoul.xlsx 파일의 변수 치킨전문점수 에서도 특이값이 1개( 357 ) 발견되었습니다.3개 특이값의 수치가 논리적으로 판단했을 때, 일반적인 범위를 넘어서지 않았기 때문에 EDA에 포함하기로 하였습니다.

bd <- ggplot( delivery, aes( y = 평균배달수1일 ) ) +   
 geom\_boxplot( fill = "#FF6666" )

결측치가 발견된 변수 평균배달수1일을 ggplot함수를 사용하여 상자수염그래프를 만들겠습니다.

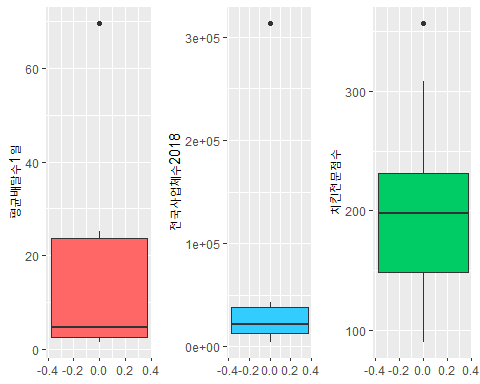
bb <- ggplot( all\_business, aes( y = 전국사업체수2018 ) ) +   
 geom\_boxplot( fill = "#33CCFF" )

결측치가 발견된 변수 전국사업체수2018을 ggplot함수를 사용하여 상자수염 그래프를 만들겠습니다.

bs <- ggplot( expert, aes( y = 치킨전문점수 ) ) +   
 geom\_boxplot( fill = "#00CC66" )

결측치가 발견된 변수 치킨전문점수를 ggplot함수를 사용하여 상자수염 그래프를 만들겠습니다.

grid.arrange(bd, bb, bs, ncol=3)



chicken.xlsx 파일의 평균 배달수1일과 전국사업체수2018, seoul.xlsx 파일의 치킨전문점수의 특이 값을 상자수염그래프로 한 화면에 표현하기 위해 grid.arrange함수를 사용하겠습니다. 3가지의 특이 값 모두 앞서 언급한 바와 같이 논리적으로 판단하였을 때, 충분히 존재할 수 있는 수치라고 생각했기 때문에 EDA 과정에 포함시켰습니다.

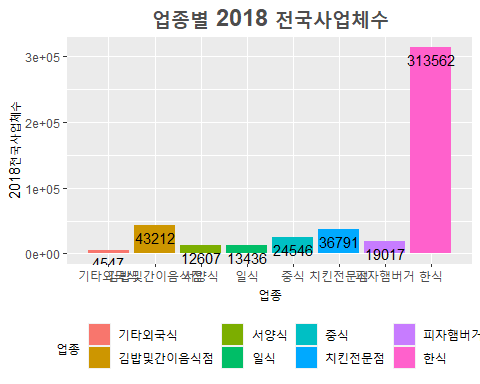
### 3. EDA( 탐색적 데이터분석 )

## 3-1.변수 시각화

저희 조는 치킨전문점의 창업전략 방법으로 선정한 변수들 각각의 데이터를 막대그래프와 원그래프를 이용하여 시각화 하였습니다.

# 전국사업체수2018 ( 막대그래프 )

k <- ggplot(chicken.xlsx , aes(x = 업종, y = 전국사업체수2018 )) +   
 geom\_bar(stat="identity", width = 0.9, aes(fill=업종)) +   
 labs(x = " 업종 ", y= " 2018전국사업체수 ", position = "dodge" )+  
 ggtitle( " 업종별 2018 전국사업체수 " ) +   
 geom\_text(aes(label= round(전국사업체수2018,2)),position = position\_dodge(width=1.8), vjust=1.5) +  
 theme(legend.position = "bottom", plot.title = element\_text(hjust=0.5, size=18, face="bold", colour = "gray30"))   
  
k



2018년 업종별 전국사업체 수를 한눈에 쉽게 파악하기 위해 ggplot함수를 써서 막대그래프로 표현했습니다. 그 결과 한식의 전국사업체수가 가장 높았고, 기타 외국식의 전국사업체수가 가장 낮았습니다. 그중 치킨 전문점은 한식과 김밥 및 간이 음식점 다음으로 높았습니다.

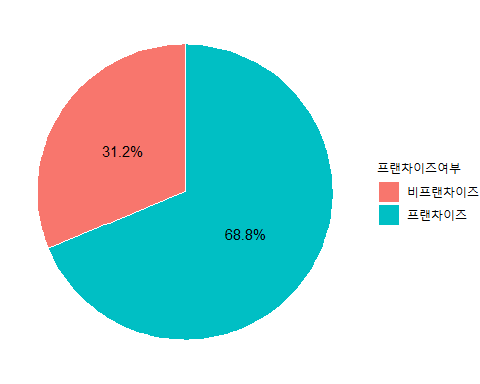
# 치킨집의 프랜차이즈가게비율 ( 원그래프 )

franchising

## 프랜차이즈여부 비율  
## 1 프랜차이즈 68.8  
## 2 비 프랜차이즈 31.2

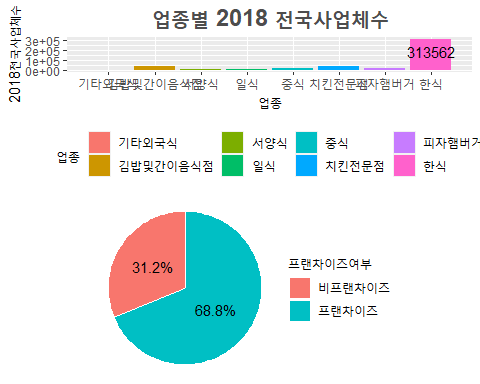
per <- c(68.8,31.2)  
  
fra <- ggplot(franchising, aes(x = "", y = 비율, fill = 프랜차이즈여부)) +   
 geom\_bar(width = 1, stat = "identity", color = "white") +  
 coord\_polar("y") +  
 geom\_text(aes(label = paste0(per,"%")),   
 position = position\_stack(vjust = 0.5)) +  
 theme\_void()  
  
fra

# 치킨집의 프랜차이즈가게비율 ( 원그래프 )



2018년 업종별 전국 사업체수의 3위였던 치킨 전문점을 프랜차이즈와 비 프랜차이즈로 나누었습니다. 그래프를 살펴보면 프랜차이즈 비율이 비 프랜차이즈 비율보다 크다는 것을 알 수 있습니다.

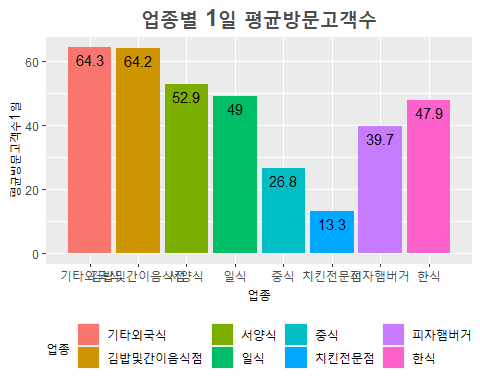
grid.arrange(k, fra, nrow=2)



두 변수를 나타내어 치킨 전문점은 전체 사업체수가 3위를 기록하고 보통 프랜차이즈를 선호한다는 것을 알 수 있습니다.

# 평균방문고객수1일 ( 막대그래프 )

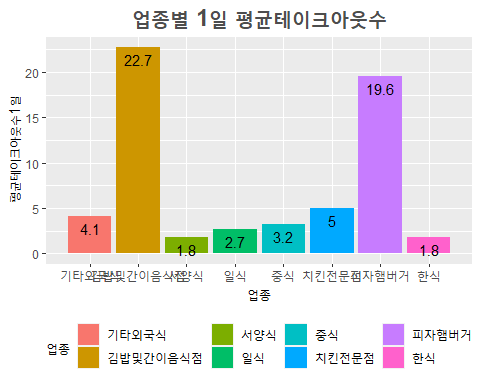
a <- ggplot(chicken.xlsx , aes(x = 업종, y = 평균방문고객수1일 )) +   
 geom\_bar(stat="identity", width = 0.9, aes(fill=업종)) +   
 labs(x = " 업종 ", y= " 평균방문고객수1일 ", position = "dodge" )+  
 ggtitle( " 업종별 1일 평균방문고객수 " ) +   
 geom\_text(aes(label= round(평균방문고객수1일,2)),position = position\_dodge(width=1.8), vjust=1.5) +  
 theme(legend.position = "bottom", plot.title = element\_text(hjust=0.5, size=18, face="bold", colour = "gray30"))   
  
a



각 업종별로 하루 동안의 평균방문 고객 수를 쉽게 보기 위해 ggplot함수를 이용해서 막대그래프를 그렸다. 그 결과, 고객들이 평균적으로 기타외국식 음식점에 가장 많이 방문하고 치킨전문점은 유달리 적게 방문한다는 사실을 발견할 수 있습니다. 따라서 치킨전문점의 매장을 넓히는 것은 회피해야 합니다.

# 평균테이크아웃수1일 ( 막대그래프 )

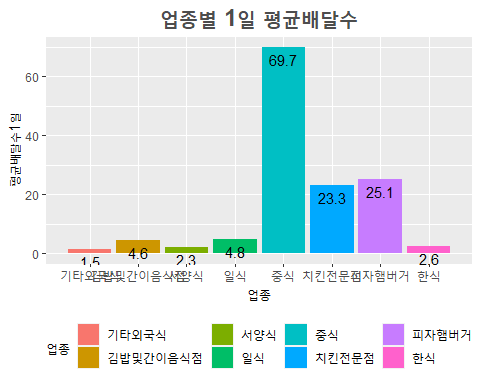
c <- ggplot(chicken.xlsx , aes(x = 업종, y = 평균테이크아웃수1일 )) +   
 geom\_bar(stat="identity", width = 0.9, aes(fill=업종)) +   
 labs(x = " 업종 ", y= " 평균테이크아웃수1일 ", position = "dodge" )+  
 ggtitle( " 업종별 1일 평균테이크아웃수 " ) +   
 geom\_text(aes(label= round(평균테이크아웃수1일,2)),position = position\_dodge(width=1.8), vjust=1.5) +  
 theme(legend.position = "bottom", plot.title = element\_text(hjust=0.5, size=18, face="bold", colour = "gray30"))   
  
c



각 업종별로 하루 동안의 평균 테이크아웃 수를 쉽게 보기위해 ggplot함수를 이용해서 막대그래프를 그렸습니다. 그 결과, 한식을 주로 이용하고 간편하게 먹을 수 있는 패스트푸드를 선호 하며, 치킨은 그 외의 업종 중에 선호하는 것을 알 수 있습니다.

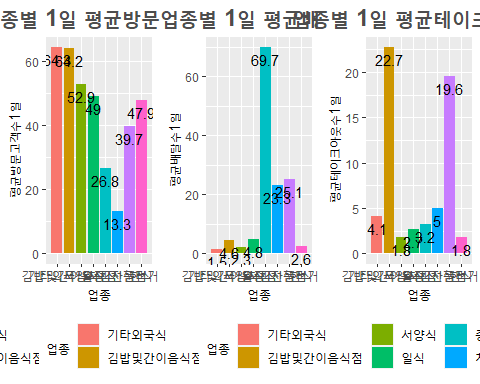
# 평균배달수1일 ( 막대그래프 )

b <- ggplot(chicken.xlsx , aes(x = 업종, y = 평균배달수1일 )) +   
 geom\_bar(stat="identity", width = 0.9, aes(fill=업종)) +   
 labs(x = " 업종 ", y= " 평균배달수1일 ", position = "dodge" )+  
 ggtitle( " 업종별 1일 평균배달수 " ) +   
 geom\_text(aes(label= round(평균배달수1일,2)),position = position\_dodge(width=1.8), vjust=1.5) +  
 theme(legend.position = "bottom", plot.title = element\_text(hjust=0.5, size=18, face="bold", colour = "gray30"))   
  
b



각 업종별로 하루 동안의 평균배달 수를 쉽게 보기 위해 ggplot함수를 이용해서 막대그래프를 그렸습니다. 그 결과, 평균적으로 중식의 배달수가 유달리 높았고 기타외국식음식점의 배달수가 가장 낮은 것을 볼 수 있습니다. 치킨 전문점의 경우 중식보단 비율이 다소 낮지만 방문과 포장에 비하면 월등이 높다는 사실을 알 수 있습니다.

grid.arrange(a, b, c, ncol=3) # ggplot 분할 화면

평균방문고객수1일, 평균배달수1일, 평균테이크아웃수1일을 한 화면에서 막대그래프로 비교해보기 위해 gird.arragne함수를 썼습니다. 종합하여 나타내 보면 치킨 전문점은 평균 방문고객과 평균 테이크아웃수가 평균 배달수 보다 낮기 때문에 굳이 목 좋은 장소에 점포를 얻을 필요가 없으며, 인테리어에 큰 투자없이 주방만 잘 갖추면 창업하는데 도움이 될 수 있습니다.

# 배달앱이용비율 ( 원그래프 )

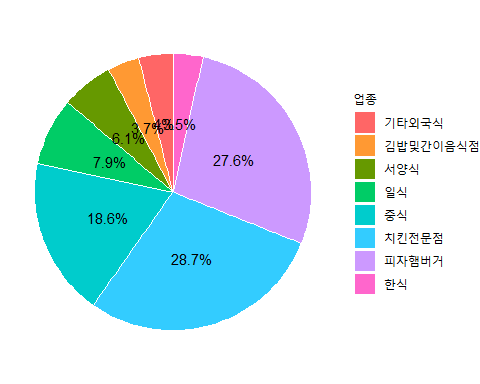
delivery\_app$배달앱이용비율 <- round(delivery\_app$배달앱이용비율/199.9,3)  
delivery\_app$배달앱이용비율

## [1] 0.035 0.186 0.079 0.061 0.040 0.276 0.287 0.037

pie(delivery\_app$배달앱이용비율,  
 labels=delivery\_app$배달앱이용비율)

i <- ggplot(delivery\_app,aes(x="",y=배달앱이용비율,fill=업종))+  
 geom\_bar(width=1,stat="identity",color = "white" )+  
 coord\_polar("y")+  
 geom\_text(aes(label=paste0(round(배달앱이용비율\*100,1),"%")),  
 position=position\_stack(vjust=0.5)) +   
 scale\_fill\_manual(values = c("#FF6666","#FF9933","#669900","#00CC66","#00CCCC","#33CCFF","#CC99FF","#FF66CC")) +  
 theme\_void()  
  
i

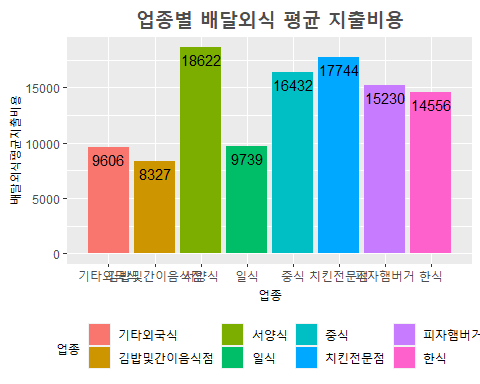
# 배달앱이용비율 ( 원그래프 )



업종별로 배달앱 이용 비율을 표현하기 위해 ggplot함수를 이용하여 원그래프를 그려보았습니다. 그 결과 치킨전문점이 가장 많이 배달앱을 이용한다는 것을 알 수 있었고 한식전문점이 가장 적게 배달앱을 이용한다는 것을 알 수 있었습니다.

# 배달외식평균지출비용 ( 막대그래프 )

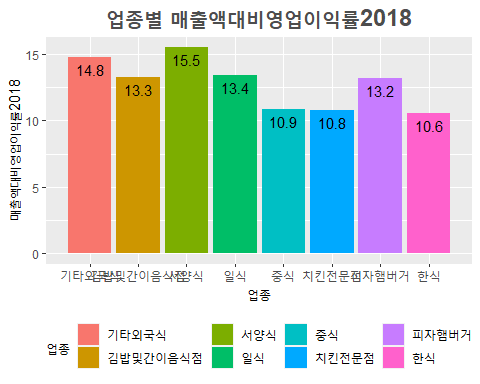
r <- ggplot(chicken.xlsx , aes(x = 업종, y = 배달외식평균지출비용 )) +   
 geom\_bar(stat="identity", width = 0.9, aes(fill=업종)) +   
 labs(x = " 업종 ", y= " 배달외식평균지출비용 ", position = "dodge" )+  
 ggtitle( " 업종별 배달외식 평균 지출비용 " ) +   
 geom\_text(aes(label= round(배달외식평균지출비용,2)),position = position\_dodge(width=1.8), vjust=1.5) +  
 theme(legend.position = "bottom", plot.title = element\_text(hjust=0.5, size=18, face="bold", colour = "gray30"))   
  
r



2018년 업종별 배달 외식 평균 지출 비용의 비율을 한눈에 쉽게 파악하기 위해 ggplot함수를 써서 막대그래프로 표현했습니다. 그 결과 업종별 배달 외식 평균 지출 비용의 비율은 서양식이 가장 높았고, 업종별 배달 외식 평균 지출 비용의 비율 중 김밥 및 간이음식점이 가장 낮았습니다. 치킨전문업이 평균 배달 외식 지출비용에서 1위인 서양식과 차이가 나지 않는 모습을 볼 수 있습니다.

# 매출액대비영업이익률2018 ( 막대그래프 )

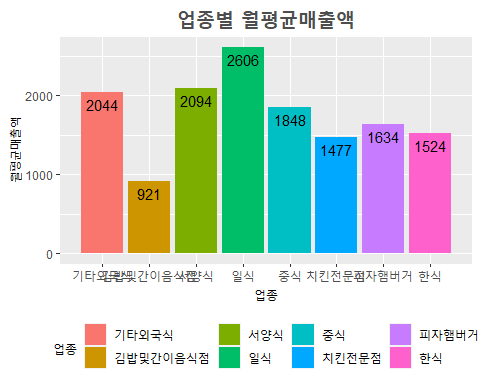
d <- ggplot(chicken.xlsx , aes(x = 업종, y = 매출액대비영업이익률2018 )) +   
 geom\_bar(stat="identity", width = 0.9, aes(fill=업종)) +   
 labs(x = " 업종 ", y= " 매출액대비영업이익률2018 ", position = "dodge" )+  
 ggtitle( " 업종별 매출액대비영업이익률2018 " ) +   
 geom\_text(aes(label= round(매출액대비영업이익률2018,2)),position = position\_dodge(width=1.8), vjust=1.5) +  
 theme(legend.position = "bottom", plot.title = element\_text(hjust=0.5, size=18, face="bold", colour = "gray30"))   
  
d



2018년 업종별 매출액 대비 영업이익률를 한눈에 쉽게 파악하기 위해 ggplot함수를 써서 막대그래프로 표현했습니다. 그 결과 서양식음식점의 영업이익률이 가장 높았고, 한식의 영업이익률이 가장 낮았습니다.

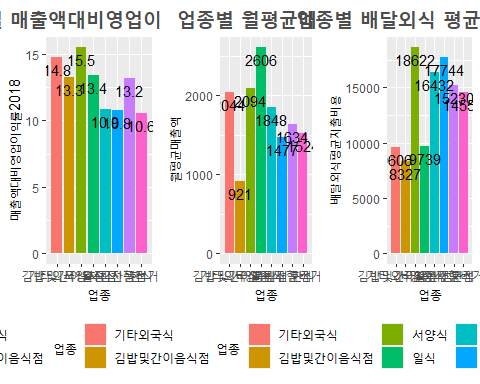
# 월평균매출액 ( 막대그래프 )

e <- ggplot(chicken.xlsx , aes(x = 업종, y = 월평균매출액 )) +   
 geom\_bar(stat="identity", width = 0.9, aes(fill=업종)) +   
 labs(x = " 업종 ", y= " 월평균매출액 ", position = "dodge" )+  
 ggtitle( " 업종별 월평균매출액 " ) +   
 geom\_text(aes(label= round(월평균매출액,2)),position = position\_dodge(width=1.8), vjust=1.5) +  
 theme(legend.position = "bottom", plot.title = element\_text(hjust=0.5, size=18, face="bold", colour = "gray30"))   
  
e



업종별 월 평균 매출액을 한눈에 쉽게 파악하기 위해 ggplot함수를 써서 막대그래프로 표현했습니다. 그 결과 일식의 월 평균 매출액이 가장 높았고, 김밥및 간이음식점의 월 평균 매출액이 가장 낮았습니다. 치킨전문점은 다른 업종에 비해 월평균 매출액은 낮으나 보통 수준은 유지하고 있습니다.

grid.arrange( d, e, r, ncol=3)



2018년 업종별 매출액 대비 영업이익률과 업종별 월 평균 매출액의 막대그래프를 한 화면에 표현하기 위해 grid.arrange 함수를 이용했습니다. 치킨전문점의 매출액 대비 영업이익률이 낮기 때문에 이익률을 높이도록 사이드 메뉴를 준비해야합니다.

# 업종전환경험여부 ( 원그래프 )

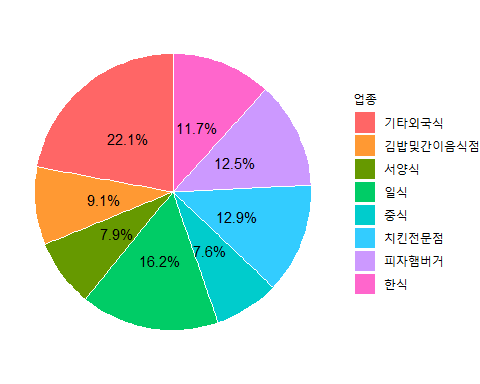
change$업종전환경험여부 <- round(change$업종전환경험여부/95.1,3)  
change$업종전환경험여부

## [1] 0.117 0.076 0.162 0.079 0.221 0.125 0.129 0.091

pie(change$업종전환경험여부,  
 labels=change$업종전환경험여부)

f <- ggplot(change,aes(x="",y=업종전환경험여부,fill=업종))+  
 geom\_bar(width=1,stat="identity",color = "white" )+  
 coord\_polar("y")+  
 geom\_text(aes(label=paste0(round(업종전환경험여부\*100,1),"%")),  
 position=position\_stack(vjust=0.5)) +   
 scale\_fill\_manual(values = c("#FF6666","#FF9933","#669900","#00CC66","#00CCCC","#33CCFF","#CC99FF","#FF66CC")) +  
 theme\_void()  
  
f

# 업종전환경험여부 ( 원그래프 )

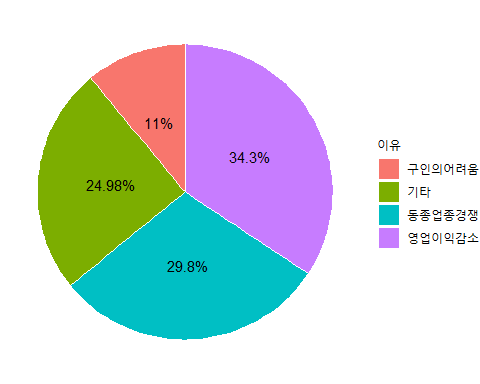


각 업종별로 업종전환경험여부를 알아보기 위해 pie함수를 이용해서 원그래프를 그려보았습니다. 그 결과 기타외국식 음식점을 운영하는 사장님들이 가장 많이 업종전환을 경험하였고 중식음식점을 운영하는 사장님들이 가장 적게 업종전환경험을 한 것을 볼 수 있었습니다. 치킨 전문점은 다소 상위에 위치하나 낮은 비율과 비슷한 양상을 보입니다.

# 치킨전문점 사장님의 업종전환 이유

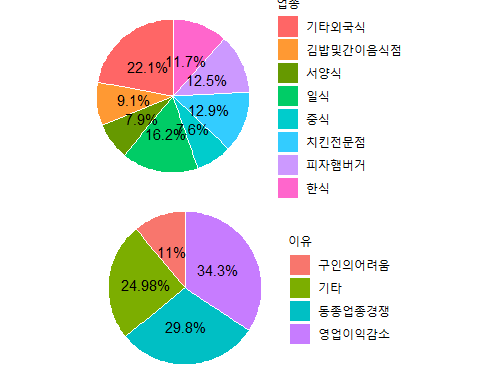
per5 <- c(34.3,29.8,11,24.98)  
  
cha <- ggplot(reason, aes(x = "", y = 비율, fill = 이유)) +   
 geom\_bar(width = 1, stat = "identity", color = "white") +  
 coord\_polar("y") +  
 geom\_text(aes(label = paste0(per5,"%")),   
 position = position\_stack(vjust = 0.5)) +  
 theme\_void()  
  
cha

# 치킨전문점 사장님의 업종전환 이유



치킨전문점의 업종전환 이유는 닭의 단가가 점점 올라가고 배달비가 생기는 등 영업이익의 감소가 일어났고, 같은 업계와의 경쟁 또한 심화되었기 때문입니다.

grid.arrange( f, cha, nrow=2)

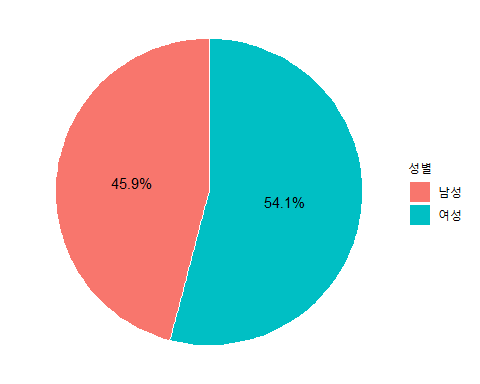


업종변환 경험 여부와 치킨전문점 사장님의 업종전환 이유를 원그래프를 한 화면에 표현하기 위해 grid.arrange 함수를 이용했습니다. 치킨전문점의 경우 업종전환의 경험이 3번째로 높지만 기타 외국식을 제외하면 비슷한 양상을 보입니다. 치킨전문점의 업종전환 이유는 여러가지가 있는데 영업이익 감소, 동종업종의 경쟁, 구인의 어려움이 있었습니다. 따라서 치킨전문점 창업을 준비하는 창업주들은 창업할 때, 영업이익감소와 동종업종경쟁의 부분을 고려해야 할 것입니다.

# 치킨집을 이용하는 남,여 성별비율 ( 원그래프 )

per2 <- c(54.1,45.9)  
gen <- ggplot(gender, aes(x = "", y = 비율, fill = 성별)) +   
 geom\_bar(width = 1, stat = "identity", color = "white") +  
 coord\_polar("y") +  
 geom\_text(aes(label = paste0(per2,"%")),   
 position = position\_stack(vjust = 0.5)) +  
 theme\_void()  
  
gen

# 치킨집을 이용하는 남, 여 성별비율 ( 원그래프 )

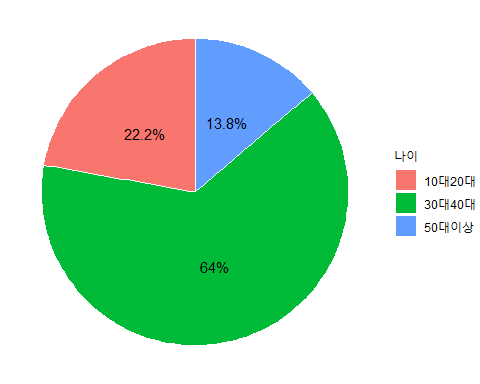


치킨전문점을 이용하는 성별 비율은 여성이 남성보다 다소 높습니다. 그러나 유의미한 차이는 아니므로 여성과 남성 모두를 고려해야 할 것입니다.

# 치킨집을 이용하는 사람들의 나이 비율( 원그래프 )

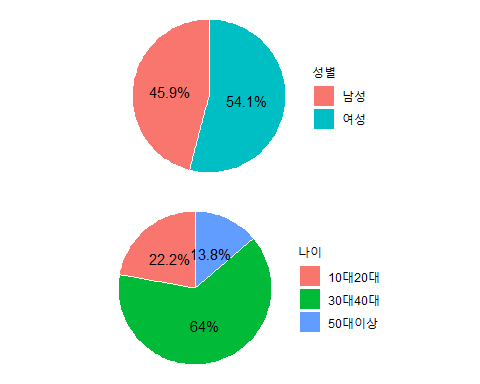
per3 <- c(22.2,64,13.8)  
sas <- ggplot(age, aes(x = "", y = 비율, fill =나이)) +   
 geom\_bar(width = 1, stat = "identity", color = "white") +  
 coord\_polar("y") +  
 geom\_text(aes(label = paste0(per3,"%")),   
 position = position\_stack(vjust = 0.5)) +  
 theme\_void()  
  
sas

# 치킨집을 이용하는 사람들의 나이 비율( 원그래프 )



치킨 전문점을 이용하는 사람들의 나이비율은 30-40대가 가장 높았고 그 다음이 10-20대, 그리고 50대 이상입니다. 이 데이터를 통해 치킨 소비자들의 특성을 살펴보면, 50대 이상의 중장년층 보다는 40대 이하의 연령대를 타겟으로 삼아 창업을 준비해야 함을 알 수 있습니다.

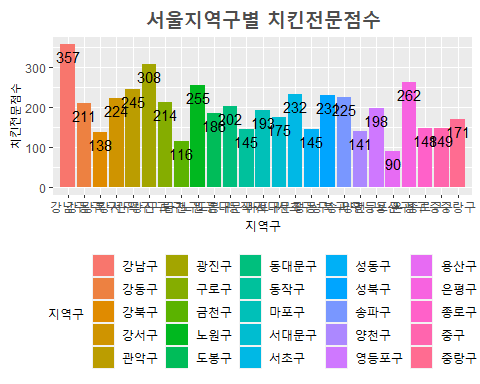
grid.arrange( gen,sas, nrow=2)



치킨전문점을 이용하는 성별 비율과 치킨전문점을 이용하는 나이비율을 원그래프로 한 화면에 표현하기 위해 grid.arrange 함수를 이용했습니다. 치킨전문점을 이용하는 성별은 비슷하며, 나이대는 30-40대가 가장 많이 이용했습니다.

# 서울 지역구별 치킨전문점 수

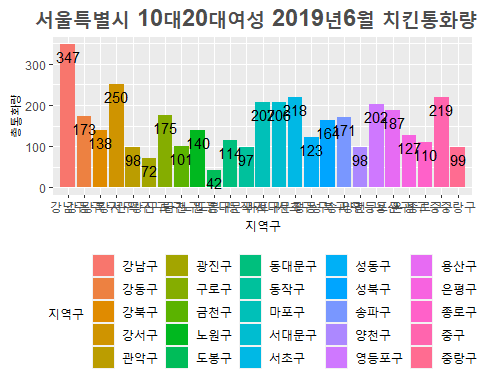
gu <- ggplot( seoul.xlsx, aes(x = 지역구, y = 치킨전문점수 )) +   
 geom\_bar(stat="identity", width = 0.9, aes(fill=지역구)) +   
 labs(x = " 지역구 ", y= " 치킨전문점수 ", position = "dodge" )+  
 ggtitle( " 서울지역구별 치킨전문점수 " ) +   
 geom\_text(aes(label= round(치킨전문점수,2)),position = position\_dodge(width=1.8), vjust=1.5) +  
 theme(legend.position = "bottom", plot.title = element\_text(hjust=0.5, size=18, face="bold", colour = "gray30"))   
  
gu



서울 지역구별 치킨 전문점 수를 막대 그래프로 표현하였습니다. 2019년 6월 기준, 서울특별시 지역구에서 가장 많은 치킨 전문점이 있는 곳은 강남구(357개)이며, 가장 적은 곳은 용산구(90개)입니다. 가장 인구가 많은 서울을 기준으로 보았을 때, 서울에서 창업을 하기 위해서는 치킨전문점의 수가 많은 지역구 보다는 적은 곳에서 하는 곳이 상대적으로 경쟁이 수월할 것입니다.

# 서울특별시 10대,20대 여성 2019년 6월 치킨 통화량

call <- ggplot(치킨통화량.xlsx , aes(x = 지역구, y = 총통화량 )) +   
 geom\_bar(stat="identity", width = 0.9, aes(fill=지역구)) +   
 labs(x = " 지역구 ", y= " 총통화량 ", position = "dodge" )+  
 ggtitle( "서울특별시 10대20대여성 2019년6월 치킨통화량 " ) +   
 geom\_text(aes(label= round(총통화량,2)),position = position\_dodge(width=1.8), vjust=1.5) +  
 theme(legend.position = "bottom", plot.title = element\_text(hjust=0.5, size=18, face="bold", colour = "gray30"))   
  
call



2019년 6월 기준 서울특별시 지역구 10대-20대 여성고객의 치킨 배달 통화량의 수를 막대 그래프로 표현했습니다. 치킨 배달 통화량의 건수가 가장 많은 지역구는 강남구(347건)이며, 가장 적은 치킨 배달 통화량 건수를 기록한 지역구는 도봉구(42건)입니다. 따라서 치킨전문점 창업을 서울 기준 10대,20대 여성을 타겟으로 한다면 배달 통화량이 많은 지역구를 선택하는 것이 창업성공에 도움을 줄 수 있을 것입니다.

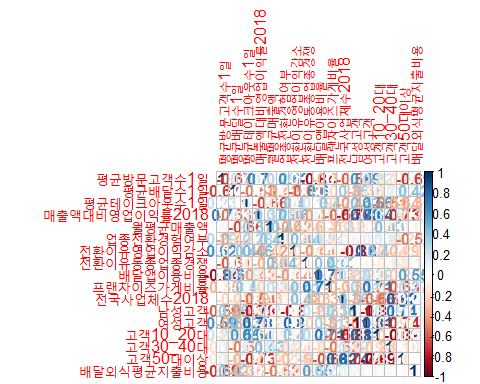
## 3-2. 변수들 간의 상관도 분석

상관계수가 0.8/-0.8 이상인 변수들을 살펴보면, 배달앱이용비율-평균방문고객수1일(-0.84), 남성고객-전환이유영업이익감소(-0.8), 여성고객-전환이유영업이익감소(0.8), 고객10\_20대-남성고객(-0.81), 고객10\_20대-여성고객(0.81), 고객50대이상-전국사업체수2018(0.82), 고객50대이상-고객10\_20대(-0.89)으로 나왔습니다.

group <- chicken.xlsx[ ,c( "평균방문고객수1일", "평균배달수1일", "평균테이크아웃수1일", "매출액대비영업이익률2018",  
 "월평균매출액", "업종전환경험여부", "전환이유영업이익감소", "전환이유동종업종경쟁",  
 "배달앱이용비율", "프랜차이즈가게비율", "전국사업체수2018", "남성고객","여성고객", "고객10\_20대", "고객30\_40대", "고객50대이상", "배달외식평균지출비용" ) ]  
  
group.cor <- cor( group )  
group.cor

## 평균방문고객수1일 평균배달수1일 평균테이크아웃수1일  
## 평균방문고객수1일 1.00000000 -0.67012857 0.22797888  
## 평균배달수1일 -0.67012857 1.00000000 -0.03991757  
## 평균테이크아웃수1일 0.22797888 -0.03991757 1.00000000  
## 매출액대비영업이익률2018 0.69653243 -0.52747397 0.11785960  
## 월평균매출액 0.04436944 -0.04820290 -0.65564326  
## 업종전환경험여부 0.28829779 -0.41546251 -0.17465809  
## 전환이유영업이익감소 0.61750504 -0.08893562 0.39568090  
## 전환이유동종업종경쟁 0.01453658 -0.60959320 0.02294519  
## 배달앱이용비율 -0.84447443 0.60046957 0.18795573  
## 프랜차이즈가게비율 -0.34995247 -0.02950186 0.39927702  
## 전국사업체수2018 0.02933250 -0.21519765 -0.21158183  
## 남성고객 -0.59210321 0.36127424 -0.42947381  
## 여성고객 0.59210321 -0.36127424 0.42947381  
## 고객10\_20대 0.21931277 -0.16765573 0.56345406  
## 고객30\_40대 -0.17230019 0.10505219 -0.51293156  
## 고객50대이상 -0.14352405 0.12227905 -0.33564125  
## 배달외식평균지출비용 -0.69015708 0.41929625 -0.35836070  
## 매출액대비영업이익률2018 월평균매출액 업종전환경험여부  
## 평균방문고객수1일 0.69653243 0.04436944 0.28829779  
## 평균배달수1일 -0.52747397 -0.04820290 -0.41546251  
## 평균테이크아웃수1일 0.11785960 -0.65564326 -0.17465809  
## 매출액대비영업이익률2018 1.00000000 0.36806026 0.27789491  
## 월평균매출액 0.36806026 1.00000000 0.39740290  
## 업종전환경험여부 0.27789491 0.39740290 1.00000000  
## 전환이유영업이익감소 0.45669312 -0.41502759 -0.19969267  
## 전환이유동종업종경쟁 -0.17112118 -0.35724403 0.43644538  
## 배달앱이용비율 -0.41925855 -0.12771369 -0.14952307  
## 프랜차이즈가게비율 0.07316411 -0.23383944 0.23515477  
## 전국사업체수2018 -0.54061056 -0.29536583 -0.13510538  
## 남성고객 -0.78176552 0.19758599 -0.04671979  
## 여성고객 0.78176552 -0.19758599 0.04671979  
## 고객10\_20대 0.64494122 -0.08616169 0.12903492  
## 고객30\_40대 0.14771250 0.42546453 0.14344423  
## 고객50대이상 -0.73033477 -0.11214903 -0.19977031  
## 배달외식평균지출비용 -0.27666384 0.02366174 -0.52404226  
## 전환이유영업이익감소 전환이유동종업종경쟁  
## 평균방문고객수1일 0.61750504 0.01453658  
## 평균배달수1일 -0.08893562 -0.60959320  
## 평균테이크아웃수1일 0.39568090 0.02294519  
## 매출액대비영업이익률2018 0.45669312 -0.17112118  
## 월평균매출액 -0.41502759 -0.35724403  
## 업종전환경험여부 -0.19969267 0.43644538  
## 전환이유영업이익감소 1.00000000 -0.21369841  
## 전환이유동종업종경쟁 -0.21369841 1.00000000  
## 배달앱이용비율 -0.43648075 -0.03678819  
## 프랜차이즈가게비율 -0.13583650 0.40562629  
## 전국사업체수2018 -0.07653687 0.42092328  
## 남성고객 -0.80265496 0.02727454  
## 여성고객 0.80265496 -0.02727454  
## 고객10\_20대 0.39778584 -0.04096279  
## 고객30\_40대 -0.24168515 -0.12743044  
## 고객50대이상 -0.29368733 0.10199548  
## 배달외식평균지출비용 -0.18435429 -0.13155339  
## 배달앱이용비율 프랜차이즈가게비율 전국사업체수2018  
## 평균방문고객수1일 -0.84447443 -0.34995247 0.02933250  
## 평균배달수1일 0.60046957 -0.02950186 -0.21519765  
## 평균테이크아웃수1일 0.18795573 0.39927702 -0.21158183  
## 매출액대비영업이익률2018 -0.41925855 0.07316411 -0.54061056  
## 월평균매출액 -0.12771369 -0.23383944 -0.29536583  
## 업종전환경험여부 -0.14952307 0.23515477 -0.13510538  
## 전환이유영업이익감소 -0.43648075 -0.13583650 -0.07653687  
## 전환이유동종업종경쟁 -0.03678819 0.40562629 0.42092328  
## 배달앱이용비율 1.00000000 0.71408201 -0.30018034  
## 프랜차이즈가게비율 0.71408201 1.00000000 -0.33445183  
## 전국사업체수2018 -0.30018034 -0.33445183 1.00000000  
## 남성고객 0.23398117 -0.31050557 0.36111618  
## 여성고객 -0.23398117 0.31050557 -0.36111618  
## 고객10\_20대 0.28621547 0.71453739 -0.56335031  
## 고객30\_40대 -0.07406750 -0.21441420 -0.51963700  
## 고객50대이상 -0.25834121 -0.63105058 0.82194561  
## 배달외식평균지출비용 0.54857531 0.26703036 0.07998930  
## 남성고객 여성고객 고객10\_20대 고객30\_40대  
## 평균방문고객수1일 -0.59210321 0.59210321 0.21931277 -0.172300192  
## 평균배달수1일 0.36127424 -0.36127424 -0.16765573 0.105052193  
## 평균테이크아웃수1일 -0.42947381 0.42947381 0.56345406 -0.512931562  
## 매출액대비영업이익률2018 -0.78176552 0.78176552 0.64494122 0.147712496  
## 월평균매출액 0.19758599 -0.19758599 -0.08616169 0.425464529  
## 업종전환경험여부 -0.04671979 0.04671979 0.12903492 0.143444232  
## 전환이유영업이익감소 -0.80265496 0.80265496 0.39778584 -0.241685153  
## 전환이유동종업종경쟁 0.02727454 -0.02727454 -0.04096279 -0.127430442  
## 배달앱이용비율 0.23398117 -0.23398117 0.28621547 -0.074067496  
## 프랜차이즈가게비율 -0.31050557 0.31050557 0.71453739 -0.214414200  
## 전국사업체수2018 0.36111618 -0.36111618 -0.56335031 -0.519637004  
## 남성고객 1.00000000 -1.00000000 -0.80609176 0.191642459  
## 여성고객 -1.00000000 1.00000000 0.80609176 -0.191642459  
## 고객10\_20대 -0.80609176 0.80609176 1.00000000 -0.281317239  
## 고객30\_40대 0.19164246 -0.19164246 -0.28131724 1.000000000  
## 고객50대이상 0.73557647 -0.73557647 -0.89199542 -0.182854845  
## 배달외식평균지출비용 0.09599632 -0.09599632 0.04072414 -0.001516265  
## 고객50대이상 배달외식평균지출비용  
## 평균방문고객수1일 -0.14352405 -0.690157082  
## 평균배달수1일 0.12227905 0.419296254  
## 평균테이크아웃수1일 -0.33564125 -0.358360698  
## 매출액대비영업이익률2018 -0.73033477 -0.276663838  
## 월평균매출액 -0.11214903 0.023661737  
## 업종전환경험여부 -0.19977031 -0.524042263  
## 전환이유영업이익감소 -0.29368733 -0.184354293  
## 전환이유동종업종경쟁 0.10199548 -0.131553385  
## 배달앱이용비율 -0.25834121 0.548575310  
## 프랜차이즈가게비율 -0.63105058 0.267030360  
## 전국사업체수2018 0.82194561 0.079989299  
## 남성고객 0.73557647 0.095996318  
## 여성고객 -0.73557647 -0.095996318  
## 고객10\_20대 -0.89199542 0.040724138  
## 고객30\_40대 -0.18285484 -0.001516265  
## 고객50대이상 1.00000000 -0.041008232  
## 배달외식평균지출비용 -0.04100823 1.000000000

corrplot( group.cor, method = "number" )

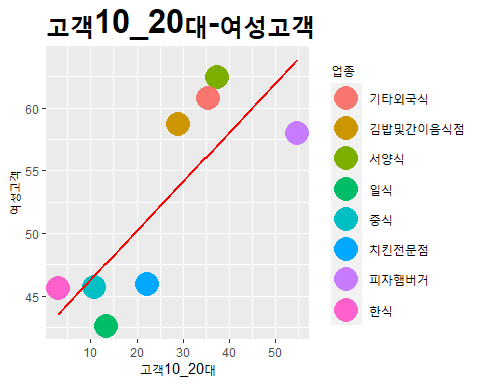


각 변수들 간의 상관계수를 구한 후, cor 함수를 활용하여 숫자로 시각화 하였습니다. 상관계수가 높은 변수들 중에서 논리적으로 판단하였을 때, 그 수치가 의미하는 관계가 유의미하다고 판단되는 관계에 대해서 산점도로 시각화 하였습니다.

# 고객 10\_20대이상-여성고객에 대한 산점도

z <- ggplot( data = chicken.xlsx, aes( x = 고객10\_20대,  
 y = 여성고객,  
 color = 업종 ) ) +   
 geom\_point( size = 8 ) +  
 stat\_smooth(method = 'lm', se=F, color='red') +  
 ggtitle( "고객10\_20대-여성고객" )+  
 theme( plot.title = element\_text( size = 25,  
 face = "bold",  
 colour = "black" ) )  
z

## `geom\_smooth()` using formula 'y ~ x'



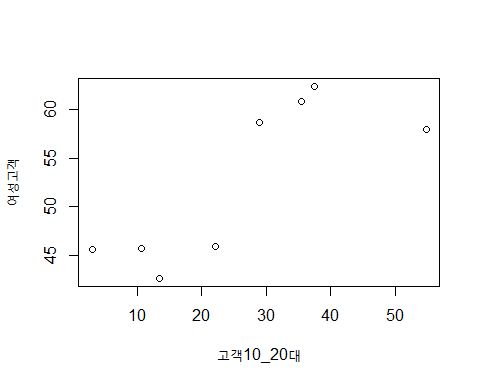
고객 10\_20대-여성고객에 대한 산점도가 0.81로 양의 상관관계가 높습니다. 이 데이터가 의미하는 것은 10대,20대 고객이 증가할 수록, 여성고객도 함께 증가한다는 것이므로, 치킨전문점 창업을 할 때 10대,20대 여성을 메인 타켓으로 설정하는 것이 좋은 창업전략이라고 생각합니다.

## 3-3. 단순선형 회귀분석

상관계수가 높았던 변수들 중에서 고객10\_20대와 여성고객을 회귀분석의 대상으로 선정하였습니다.

# 산점도를 통한 선형 관계 확인

plot( 여성고객~고객10\_20대, data = chicken.xlsx )



회귀모델 구하기 종속변수(여성고객) ~ 독립변수(고객10\_20대) 순으로 지정하였습니다.

model2 <- lm( 여성고객~고객10\_20대, chicken.xlsx )  
model2

##   
## Call:  
## lm(formula = 여성고객 ~ 고객10\_20대, data = chicken.xlsx)  
##   
## Coefficients:  
## (Intercept) 고객10\_20대   
## 42.3765 0.3917

summary(model2)

##   
## Call:  
## lm(formula = 여성고객 ~ 고객10\_20대, data = chicken.xlsx)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.8410 -5.0619 0.5904 4.6594 5.3352   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 42.3765 3.5461 11.950 2.08e-05 \*\*\*  
## 고객10\_20대 0.3917 0.1174 3.337 0.0157 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.243 on 6 degrees of freedom  
## Multiple R-squared: 0.6498, Adjusted R-squared: 0.5914   
## F-statistic: 11.13 on 1 and 6 DF, p-value: 0.01568

**예측 정확성 : 59.14%**  
**P-value : 0.01568** ( P-value가 0.05보다 작으면 통계적으로 의미가 있는데, 저희가 구한 값은 0.01568이므로 통계적 의미가 있다고 말할 수 있습니다. )

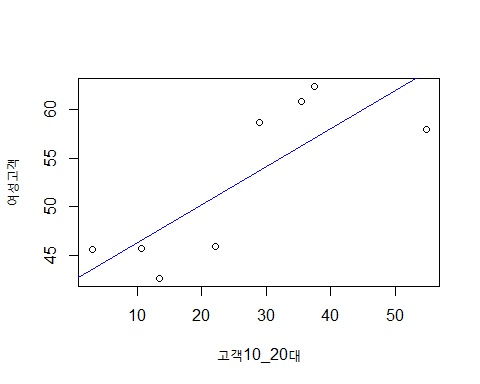
**Multiple R-squared** : 전체 변동 량의 **약 64%** 정도를 이 모형을 통해 설명할 수 있습니다.

# 여성고객 = 0.3917(고객10\_20대) X 고객10\_20대 + 42.3765

plot( 여성고객~고객10\_20대, data = chicken.xlsx )  
model2

##   
## Call:  
## lm(formula = 여성고객 ~ 고객10\_20대, data = chicken.xlsx)  
##   
## Coefficients:  
## (Intercept) 고객10\_20대   
## 42.3765 0.3917

abline( model2, col = "blue" )



coef( model2 ) # coef : 회귀계수 출력함수

## (Intercept) 고객10\_20대   
## 42.3765326 0.3916881

coef( model2 )[ 1 ] # b값 (42.37653)

## (Intercept)   
## 42.37653

coef( model2 )[ 2 ] # W값 (0.3916881)

## 고객10\_20대   
## 0.3916881

b1 <- coef( model2 )[ 1 ]  
W1 <- coef( model2 )[ 2 ]  
  
  
고객10\_20대 <- 21.5  
여성고객 <- W1 \* 고객10\_20대 + b1  
여성고객 # (50.79783)

## 고객10\_20대   
## 50.79783

회귀분석한 모델을 토대로 실제 변수의 수치를 입력하여 예측하려 하였으나, 데이터 수집 과정 중에 해당 변수의 데이터가 수치가 아닌 비율로 나타나고 있었기 때문에 예측하지 못하였습니다.

### 4. 결론

치킨 전문점을 창업을 염두에 두고 계신 창업주분들을 대상으로 현재 영업하고 있는 치킨전문점의 정보를 분석하고 소비자들의 동향을 예측 하였습니다. 데이터를 분석한 결과, 2018년 기준 치킨전문점의 사업체 수는 한식, 분식 다음으로 많은 점포가 있었고, 치킨전문점들의 프랜차이즈 비율은 70%를 육박하였습니다. 이것은 치킨전문점이 다른 요식업계와 비교하여 상대적으로 동종업종 간의 경쟁이 심하고 대부분의 소비자들이 프랜차이즈 점포를 선호한다는 것입니다.

치킨전문점의 판매형태는 소비자들이 직접 매장을 방문하는 것보다는 대부분 배달이나 테이크아웃 위주입니다. 또한 배달앱의 사용률도 치킨전문점이 월등이 높아 따라서 치킨전문점을 창업할 창업주들은 소비자들의 매장 방문을 고려하여 매장의 위치와 인테리어가 따로 필요 없으며, 단지 잘 갖춰진 주방만 있으면 됩니다. 그리고 치킨전문점의 매출액 대비 영업이익률이 낮기 때문에 이익률을 높이도록 사이드 메뉴 또한 준비해야합니다.

하지만 치킨전문점의 업종전환 이유는 여러가지 원인이 있는데 영업이익 감소, 동종업종의 경쟁, 구인의 어려움이 있었습니다. 따라서 치킨전문점 창업을 준비하는 창업주들은 창업할 때, 영업이익감소와 동종업종경쟁의 부분을 고려해야 할 것입니다. 치킨전문점을 이용하는 성별은 비슷하며, 나이대는 30-40대가 가장 많이 이용했습니다. 따라서 40대이하의 연령층을 주 타겟으로 창업 전략을 세워야 합니다.

더 구체적으로 지역을 선정하여 가장 인구가 많은 서울을 기준으로 보았을 때, 서울에서 창업을 하기 위해서는 치킨전문점의 수가 많은 지역구 보다는 적은 곳에서 하는 곳이 상대적으로 경쟁이 수월할 것입니다. 치킨 배달 통화량의 건수로 보자면 가장 많은 지역구는 강남구(347건)이며, 가장 적은 치킨 배달 통화량 건수를 기록한 지역구는 도봉구(42건)입니다. 따라서 치킨전문점 창업을 서울 기준 10대,20대 여성을 타겟으로 한다면 배달 통화량이 많은 지역구를 선택하는 것이 창업성공에 도움을 줄 수 있을 것입니다.