Food delivery scene in London 2021

Francesco Mantovani

3/9/2021

Check for presence of necessary packages and install missing ones. Open necessary packages. Make sure system language is set to english

```r
if (!require("pacman")) install.packages("pacman")
```

```
## Loading required package: pacman
```

```r
if (!require("tidyverse")) install.packages("tidyverse")
```

```
## Loading required package: tidyverse

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
if (!require("data.table")) install.packages("data.table")
```

```
## Loading required package: data.table

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose
```

```r
if (!require("formattable")) install.packages("formattable")
```

```
## Loading required package: formattable
```

```r
if (!require("fastDummies")) install.packages("fastDummies")
```

```
## Loading required package: fastDummies
```

```r
if (!require("caTools")) install.packages("caTools")
```

```
## Loading required package: caTools
```

```r
if (!require("caret")) install.packages("caret")
```

```
## Loading required package: caret

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift
```

```r
if (!require("car")) install.packages("car")
```

```
## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'car'

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

## The following object is masked from 'package:purrr':
##
##     some
```

```r
if (!require("gridExtra")) install.packages("gridExtra")
```

```
## Loading required package: gridExtra

##
## Attaching package: 'gridExtra'

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

```r
if (!require("corrplot")) install.packages("corrplot")
```

```
## Loading required package: corrplot

## corrplot 0.84 loaded
```

```r
library(tidyverse)
library(formattable)
library(data.table)
library(fastDummies)
library(caTools)
library(caret)
library(car)
library(gridExtra)
library(corrplot)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'

## The following objects are masked from 'package:data.table':
##
##     dcast, melt

## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
library(tinytex)
library(knitr)

Sys.setenv(LANG = "en")
```

Read databases & visualize data

```r
res <- readRDS("resturants-mibe.rds")
del <- readRDS("delivery-mibe.rds")
```

```r
res[1:5,]
```

```
## # A tibble: 5 x 7
##   restaurant_id rest_name  rest_brand rest_postcode rest_neighborho~ rest_rating
##           <dbl> <chr>      <chr>      <chr>         <chr>                  <dbl>
## 1        191295 Baba Wali~ <NA>       NW97DY        Hendon                  NA
## 2         54515 Burger & ~ Burger & ~ W1W7JE        Fitzrovia                4.7
## 3        113653 Afta Eats  <NA>       HA90TG        Wembley                 NA
## 4        184167 Europa 2 ~ Europa 2 ~ SE255QF       Croydon                  3.8
## 5         84922 Julia Dom~ <NA>       SW151JP       Putney                   4.3
## # ... with 1 more variable: rest_menu_item_price <list>
```

```r
del[1:5,]
```

```
## # A tibble: 5 x 3
##   rest_key neighborhood_name rest_delivery_time_min
##      <dbl> <chr>                              <dbl>
## 1    98636 the-city                              10
## 2   167932 the-city                              10
## 3      902 the-city                              15
## 4    22555 the-city                              15
## 5    29850 the-city                              10
```

Part 1 - Restaurants Information

1) Top 10 neighborhoods by number of restaurants

Check for missing values:

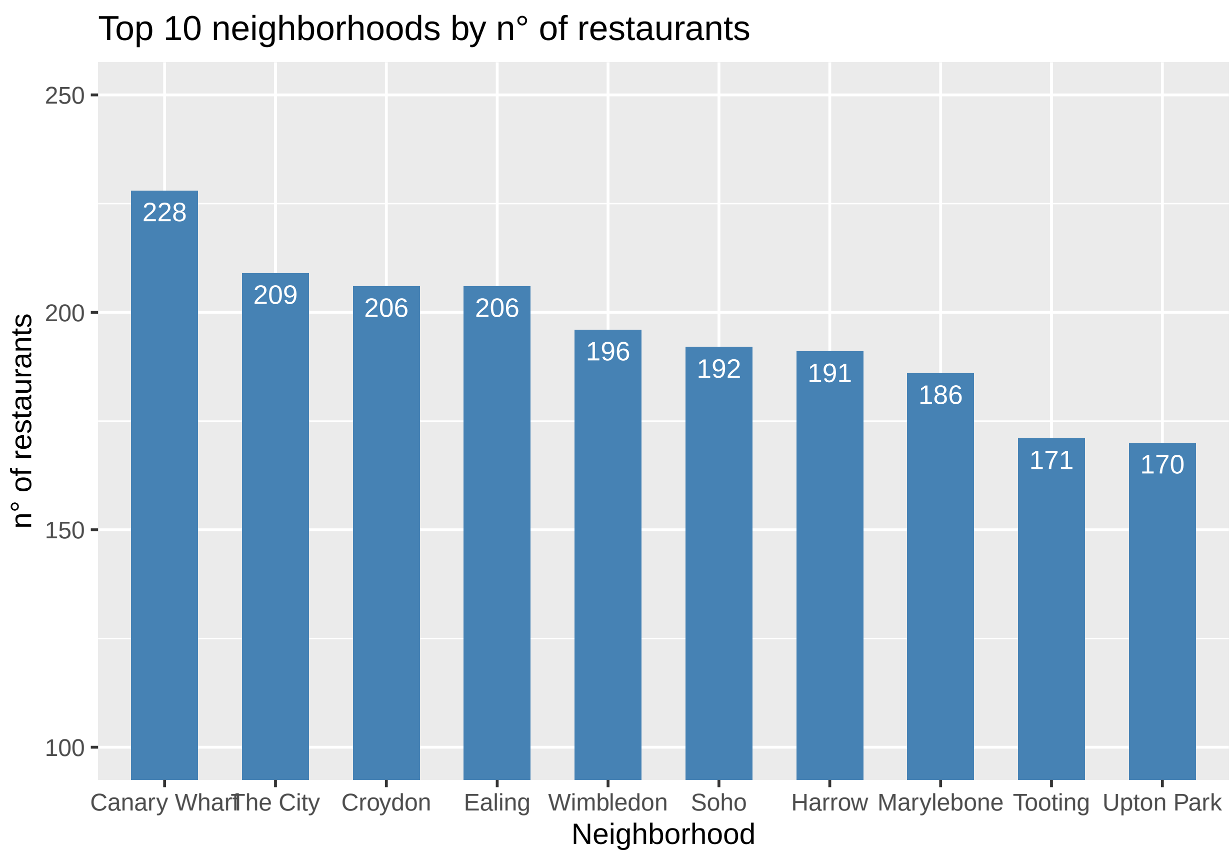
```r
sapply(res, function(x) sum(is.na (x)))
```

```
##        restaurant_id            rest_name           rest_brand
##                    0                    0                 3009
##        rest_postcode    rest_neighborhood          rest_rating
##                    0                    0                  850
## rest_menu_item_price
##                   14
```

No missing values in Neighborhoods column (all restaurant have indicated their neighborhood)

Column Chart

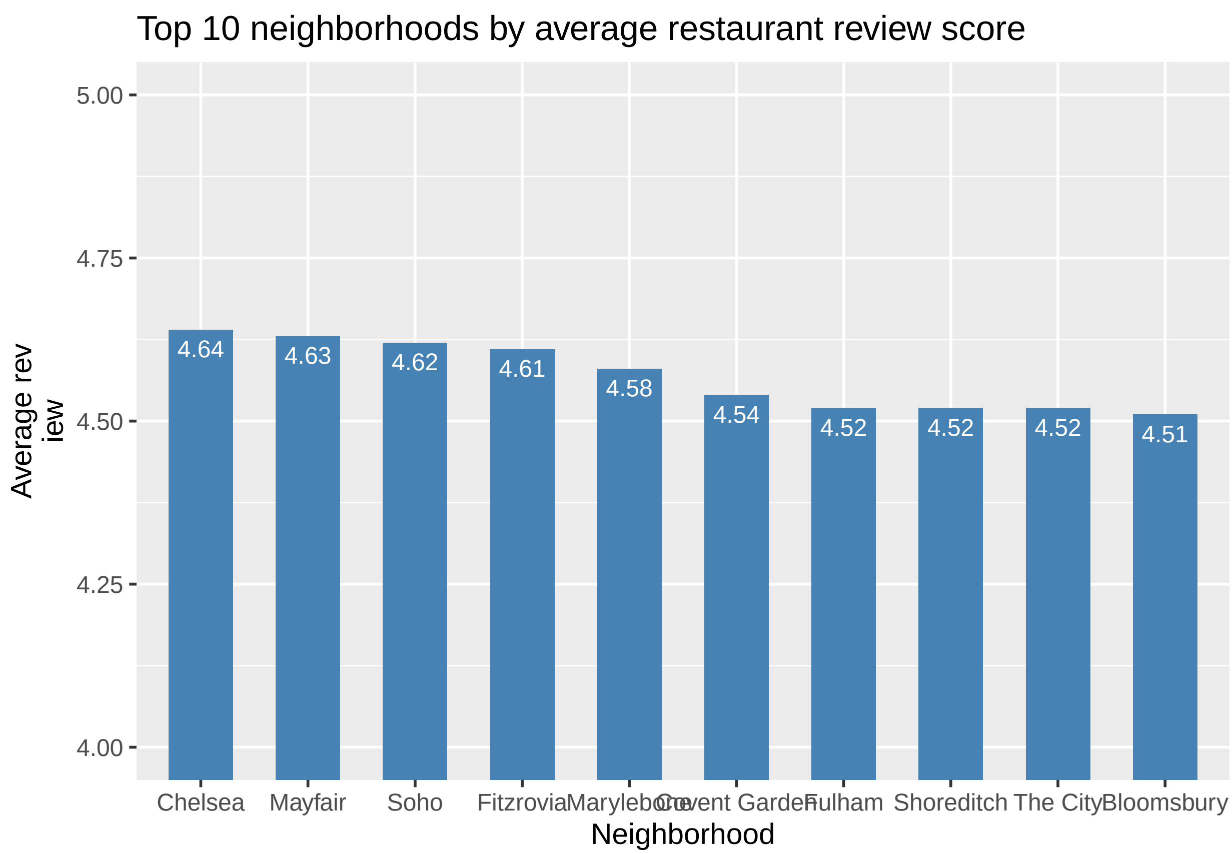
```r
x1 <- sort(table(res$rest_neighborhood),decreasing=T)
x1 <- x1[1:10]
x1 <- as.data.frame(x1)

x1p <- ggplot(x1, aes(x = Var1, y = Freq)) +
 geom_bar(fill = "steelblue", stat = "identity", width=0.6) +
 geom_text(aes(label=Freq), vjust=1.6, color="white", size=3.5) +
 ggtitle("Top 10 neighborhoods by n° of restaurants") +
 xlab("Neighborhood") + ylab("n° of restaurants") + coord_cartesian(ylim=c(100,250))
x1p
```



2) Top 10 neighborhoods by number restaurant review score

Let's check for missing values first and then compute the average review per neighborhood

```
x2 <- res[!is.na(res$rest_rating),]
x2 <- aggregate(x2[, 6], list(x2$rest_neighborhood), mean)
x2$rest_rating <- round(x2$rest_rating, digits = 2)
x2 <- x2[order(x2$rest_rating, decreasing = TRUE),]
x2 <- x2[1:10,]

x2p <- ggplot(x2, aes(x=reorder(Group.1, -rest_rating), y = rest_rating)) +
 geom_bar(fill = "steelblue", stat = "identity", width=0.6) +
 geom_text(aes(label=rest_rating), vjust=1.6, color="white", size=3.2) + coord_cartesian(ylim=c(4,5)) +
 ggtitle("Top 10 neighborhoods by average restaurant review score") + xlab("Neighborhood") + ylab("Avera
iew")
x2p
```



3) Top 10 Chains by number of restaurants

For this task, two fundamentals corrections to the data have been made: 1. The first result is "Gets drinks delivered". This is not a brand, but simply refers to the fact that that restaurant delivers alcholic beverages to their costumers. It shouldn't be in that column. The first result is excluded. 2. The big brand Co-Operative was labeled under two names: "Co-op" and "Co-operative". It is therefore necessary to modify one, otherwise they would appear separately in the ranking.

```
x3 <- res[!is.na(res$rest_brand),]
x3$rest_brand[x3$rest_brand == "Co-operative"] <- "Co-op"
x3 <- sort(table(x3$rest_brand),decreasing=T)
x3 <- x3[2:11]
x3 <- as.data.frame(x3)
names(x3)[names(x3) == "Var1"] <- "Chain"
names(x3)[names(x3) == "Freq"] <- "Number of restaurants"
x3
```

```
##              Chain Number of restaurants
## 1              KFC                    42
## 2     PizzaExpress                    42
## 3   Pret A Manger                     33
## 4            Co-op                    24
## 5      Burger King                    22
## 6             itsu                    22
```

```
## 7          Pure                     21
## 8        Wasabi                     20
## 9          LEON                     19
## 10   Papa John's                    18
```

The final table excludes "Gets drink delivered" and gathers "Co-op" in one.

4) Average menu price and number of items

```
avg <- res$rest_menu_item_price %>% map(mean)
count <- res$rest_menu_item_price %>% map(length)
res$avg <- avg
res$count <- count
x4 <- select(res,-c(rest_menu_item_price))
x4 <- x4[1:10,]
x4
```

```
## # A tibble: 10 x 8
##    restaurant_id rest_name rest_brand rest_postcode rest_neighborho~ rest_rating
##            <dbl> <chr>     <chr>      <chr>         <chr>                  <dbl>
##  1        191295 Baba Wal~ <NA>       NW97DY        Hendon                  NA
##  2         54515 Burger &~ Burger & ~ W1W7JE        Fitzrovia                4.7
##  3        113653 Afta Eats <NA>       HA90TG        Wembley                 NA
##  4        184167 Europa 2~ Europa 2 ~ SE255QF       Croydon                  3.8
##  5         84922 Julia Do~ <NA>       SW151JP       Putney                   4.3
##  6        194571 Kin + De~ <NA>       E146AB        Canary Wharf             4.4
##  7        136833 Costa     Costa      E16GQ         Brick Lane               4.6
##  8        200652 Bingo Wi~ <NA>       E29DT         Bethnal Green            4.2
##  9         28019 Crussh    Crussh     E145NY        Canary Wharf             4.8
## 10        141517 Simple H~ Simple He~ EC4N7BE       The City                 4.6
## # ... with 2 more variables: avg <list>, count <list>
```

5) Number of menu items for 5 most expensive and cheapest restaurants

```
res3 <- res
res3$avg <- as.numeric(res$avg)
res3$count <- as.numeric(res$count)
res3 <- select(res3,-c(rest_menu_item_price))
res3 <- as.data.frame(res3)
res3 <- res3[order(res3[, 7], decreasing = T), ]
res3 <- res3[!is.na(res3$avg),]
res3 <- res3[!grepl("Catering", res3$rest_name),]
top5 <- res3[2:6,]
top5[1, 2] = "Attilus\nCaviar"
top5[2, 2] = "Radio\nAlice"
top5[3, 2] = "Organic\nPress"
top5[4, 2] = "Kenza"
```

```
top5[5, 2] = "Levant"
bot5 <- top_n(res3, -5, wt=avg)
bot5 <- bot5[1:5,]
bot5[1, 2] = "Little\nMoons Soho"
bot5[2, 2] = "LittleMoons\nMayfair"
bot5[3, 2] = "Metro Coffe\nExpress"
bot5[4, 2] = "Snow\nBallz"
bot5[5, 2] = "Bambusa\nFitzrovia"

p <- ggplot(top5, aes(x =reorder(rest_name, -count), y = count)) +
 geom_bar(fill = "red", stat = "identity", width=0.6) +
 geom_text(aes(label=count), vjust=-0.3, color="black", size=3.2)+ ylab("Items on Menu") + xlab("most ex
nsive")
q <- ggplot(bot5, aes(x =reorder(rest_name, -count), y = count)) +
 geom_bar(fill = "steelblue", stat = "identity", width=0.6) +
 geom_text(aes(label=count), vjust=-0.3, color="black", size=3.2) + ylab("") + xlab("cheapest")
test <- grid.arrange(p, q, nrow =1,
top="Number of Items in menu for Most expensive and Cheapest restaurants in London",
bottom= "Restaurant Name")
```

Part 2 - Restaurants Delivery Times

The second dataset contains a large amount of data with three features: restaurant Id, Neighborhood name and Delivery time.

No extra information is provided.

Since restaurants Id appear more then once, we assume that each record represent a delivery that restaurant has made.

We drop rows with null values in "rest_delivery_time_min" and merge the two data sets.

```
del <- del[!is.na(del$rest_delivery_time_min),]
names(del)[names(del) == "rest_key"] <- "restaurant_id"
merged <- merge(res, del, by= "restaurant_id")
```

6) Count number of neighborhoods where each restaurant deliver.

```
x6 <- data.table(merged)
x6 <- x6[, .(res_neigh_count = uniqueN(neighborhood_name)),
 by = restaurant_id]
x6t <- x6[1:5,]

x6t
```

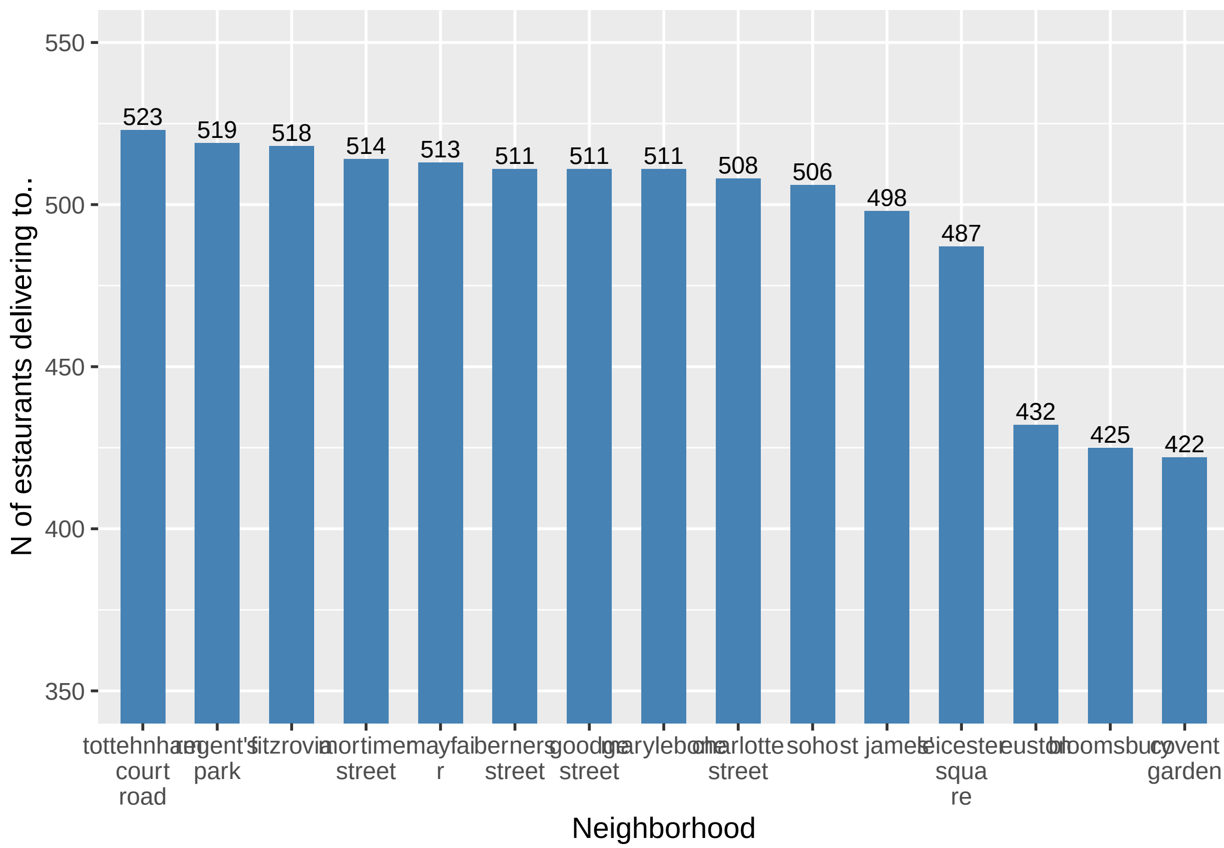
```
##    restaurant_id res_neigh_count
## 1:             3               9
## 2:             5              10
## 3:             8              10
## 4:            10               9
## 5:            15               4
```

Object x6 contains a table where the number of different neighborhoods to which each restaurant delivers.

Object x6t is a table where only 5 restaurants are reported to give a visual example.

7) Top 15 neighborhoods by restaurant delivery time

```
x7 <- data.table(merged)
x7 <- x7[, .(res_count = uniqueN(restaurant_id)), by = neighborhood_name]
x7 <- x7 [order(x7$res_count, decreasing = TRUE),]
x7 <- x7[1:15,]

#Put \n in between names for better visualization
x7$neighborhood_name <- c("tottehnham\ncourt\nroad", "regent's\npark", "fitzrovia", "mortimer\nstreet", 
r", "marylebone", "berners\nstreet", "goodge\nstreet", "charlotte\nstreet", "soho", "st james'", "leices
re", "euston", "bloomsbury","covent\ngarden")

x7p <- ggplot(x7, aes(x = reorder(neighborhood_name, -res_count), y = res_count)) +
geom_bar(fill = "steelblue", stat = "identity", width=0.6) +
geom_text(aes(label=res_count), vjust=-0.3, color="black", size=3.2) +
ylab("N of estaurants delivering to..") + xlab("Neighborhood") +
coord_cartesian(ylim=c(350,550))

x7p
```



8) Compute average Delivery time per restaurant

```
x8 <- aggregate(rest_delivery_time_min~restaurant_id, merged, mean)
x8$rest_delivery_time_min <- round(x8$rest_delivery_time_min, digits = 2)
names(x8)[names(x8) == "rest_delivery_time_min"] <- "avg_del_time"
x8t <- x8[1:5,]

x8t
```

```
##   restaurant_id avg_del_time
## 1             3        19.61
## 2             5        17.12
## 3             8        18.28
## 4            10        18.62
## 5            15        21.79
```

x8 is a table comprehensive of all restaurants and their respective delivery times. x8t is a short table reporting an example of first five restaurants

9) Report a table with top 20 restaurants by average delivery time.

```r
x9 <- x8[order(x8$avg_del_time),]
x9 <- merge(res, x9, by = 'restaurant_id')
res2 <- x9
myvars <- c("restaurant_id", "rest_name", "rest_postcode", "avg_del_time", "rest_rating")
x9 <- x9[myvars]
x9 <- x9[1:20,]
row.names(x9) <- NULL
names(x9)[names(x9) == "restaurant_id"] <- "Rest ID"
names(x9)[names(x9) == "rest_name"] <- "Rest Name"
names(x9)[names(x9) == "rest_postcode"] <- "Rest Postcode"
names(x9)[names(x9) == "rest_rating"] <- "Rest Rating"

x9
```

```
##    Rest ID                                                           Rest Name
## 1        3                                                      Busaba Chelsea
## 2        5                                                       Rossopomodoro
## 3        8                                              New Culture Revolution
## 4       10                                                           Mandaloun
## 5       15                                       Busaba St Christopher’s Place
## 6       16                                                   Busaba Bloomsbury
## 7       18 <U+0001F1EF><U+0001F1F5><U+0001F1E7><U+0001F1F7> YOOBI <U+0001F363>
## 8       19                                                               Noura
## 9       20                                                          Dozo Sushi
## 10      21                                                              Levant
## 11      23                                                 Princi - Wardour St
## 12      24                                                                Goat
## 13      25                                                               Jak’s
## 14      26                                                             Sartori
## 15      34                                                      Koshari Street
## 16      41                                                       Rossopomodoro
## 17      42                                                             Cây Tre
## 18      43                                                           Ping Pong
## 19      46                                                           Ping Pong
## 20      48                                                  Beiteddine Express
##    Rest Postcode avg_del_time Rest Rating
## 1         SW35UZ        19.61         4.6
## 2        SW109NB        17.12         4.5
## 3         SW35EP        18.28         4.7
## 4        SW109TW        18.62         4.8
## 5         W1U1BU        21.79         4.7
## 6        WC1E7DF        22.40         4.7
## 7         W1F0LL        20.62         4.8
## 8         W1J5HP        23.65         4.7
## 9         W1D4TP        62.16         4.8
## 10        W1U2SJ        21.77         4.8
## 11        W1F0UT        21.56         4.5
## 12       SW109QL        17.69         4.6
## 13        SW32HT        13.27         4.8
## 14       WC2H7JE        25.55         4.8
```

```
## 15       WC2N4EA        15.58         4.5
## 16       WC2H9EP        23.11         4.4
## 17        W1D4PZ        20.93         4.8
## 18        W1U1DZ        23.75         4.8
## 19        W1F7JL        22.50         4.8
## 20        W1J8AG        22.93         4.5
```

It's interesting to notice how most restaurants among the faster deliverers are part of chains.

Part 3 - Open analysis

Business Context:

Food delivery has been a growing trend for the past several years and faced a steep acceleration since the pandemic hit. We see delivery apps booming, restaurants change their business model and the birth of innovative food-service solutions like "Dark Kitchen".

Food delivery growth might not stop anytime soon, but the market has already been flooded with new players. Is there potential for new entrants? What can current players (both food delivery services and restaurants) do to keep growing, foster innovation and defend from competition?

Through this analysis, we investigate what factors are the most relevant for businesses in the field and what they should work on to improve their service.

According to the available data, 3 major business questions were identified.

1) Does delivery time impacts ratings?

Restaurants ratings, provided through apps such as Yelp, TripAdvisor or Google Reviews, are extremely important in nowadays digital work. A recent study found that a 1/2 star increase resulted in a 30-40 percent increase in 7pm bookings, while a 1 star increase could result in up to 9% in revenue.

Understanding what factors impact ratings can help businesses to improve their services and ultimately increase revenues. Since nowadays most deliveries are operated by third-party services (mostly delivery apps), such insights might be useful also for them.

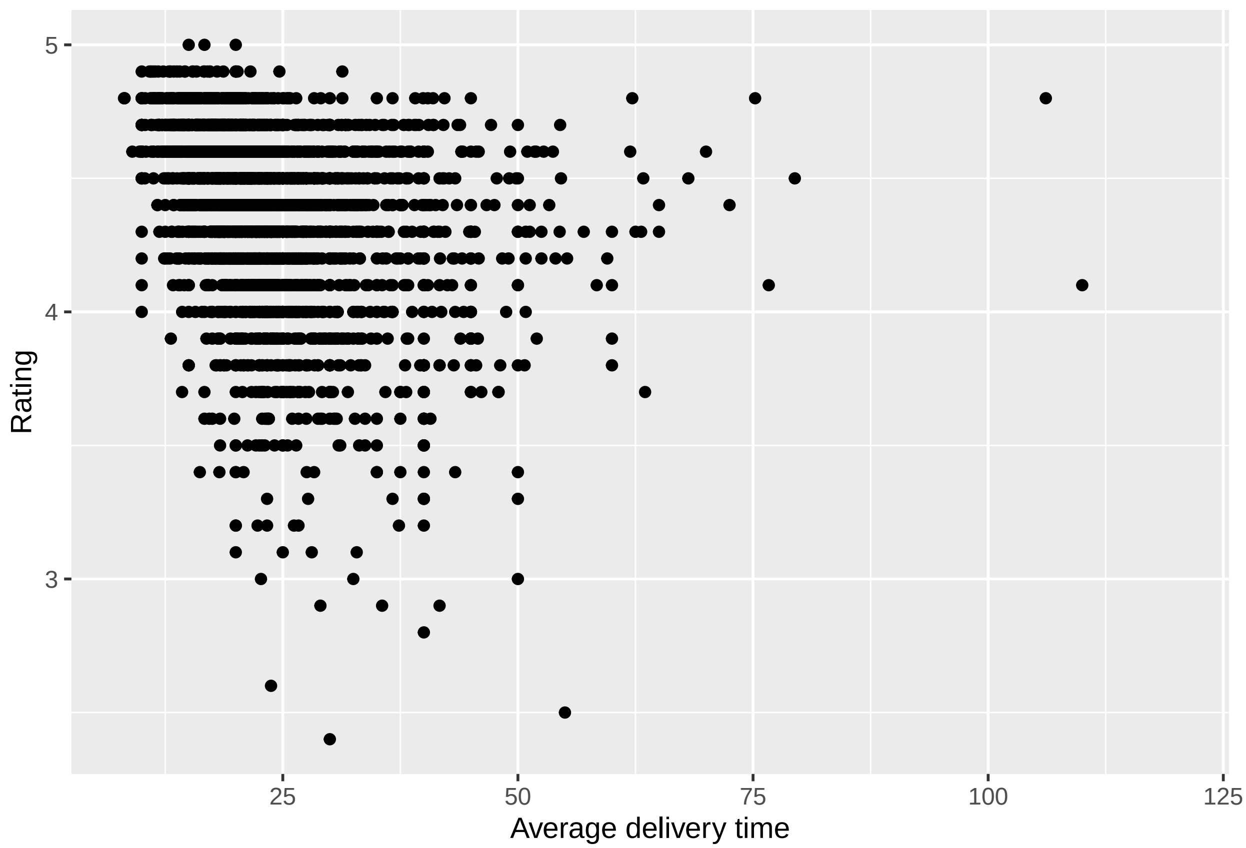
The first hypothesis is that delivery time significantly impacts ratings.

```
summary(res$rest_rating)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.800   4.200   4.400   4.355   4.600   5.000     850
```

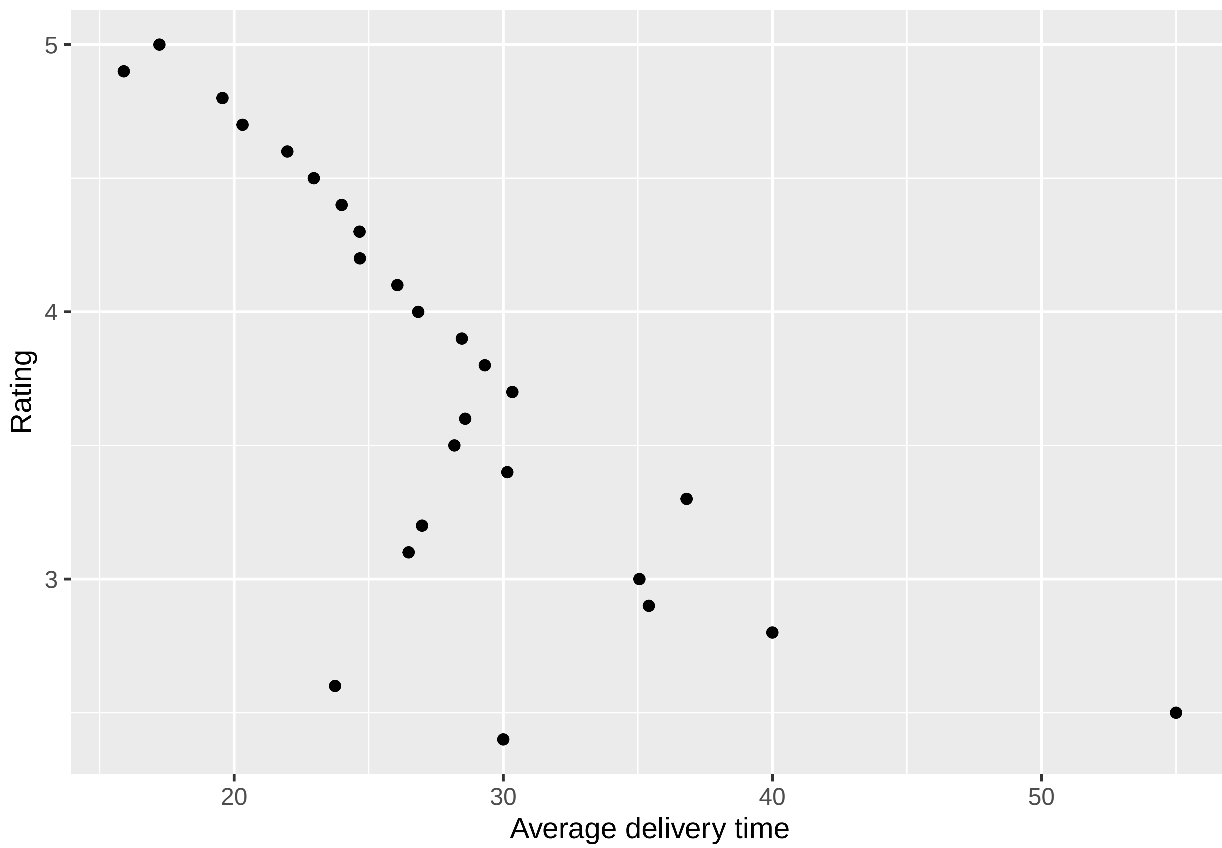
```
res2 <- res2[!is.na(res$rest_rating),]

p7 <- ggplot(res2, aes(x=avg_del_time, y = rest_rating)) +
     geom_point() +
     ylab("Rating") + xlab("Average delivery time")
p7
```

```
## Warning: Removed 1556 rows containing missing values (geom_point).
```



A very slight negative correlation might be noticed in the graph. Let's summarize data to try and have a closer look. We compute average delivery time per rating level.

```
x10 <- aggregate(avg_del_time~rest_rating, res2, mean)
p8 <- ggplot(x10, aes(x=avg_del_time, y = rest_rating)) +
      geom_point() +
      ylab("Rating") + xlab("Average delivery time")
p8
```



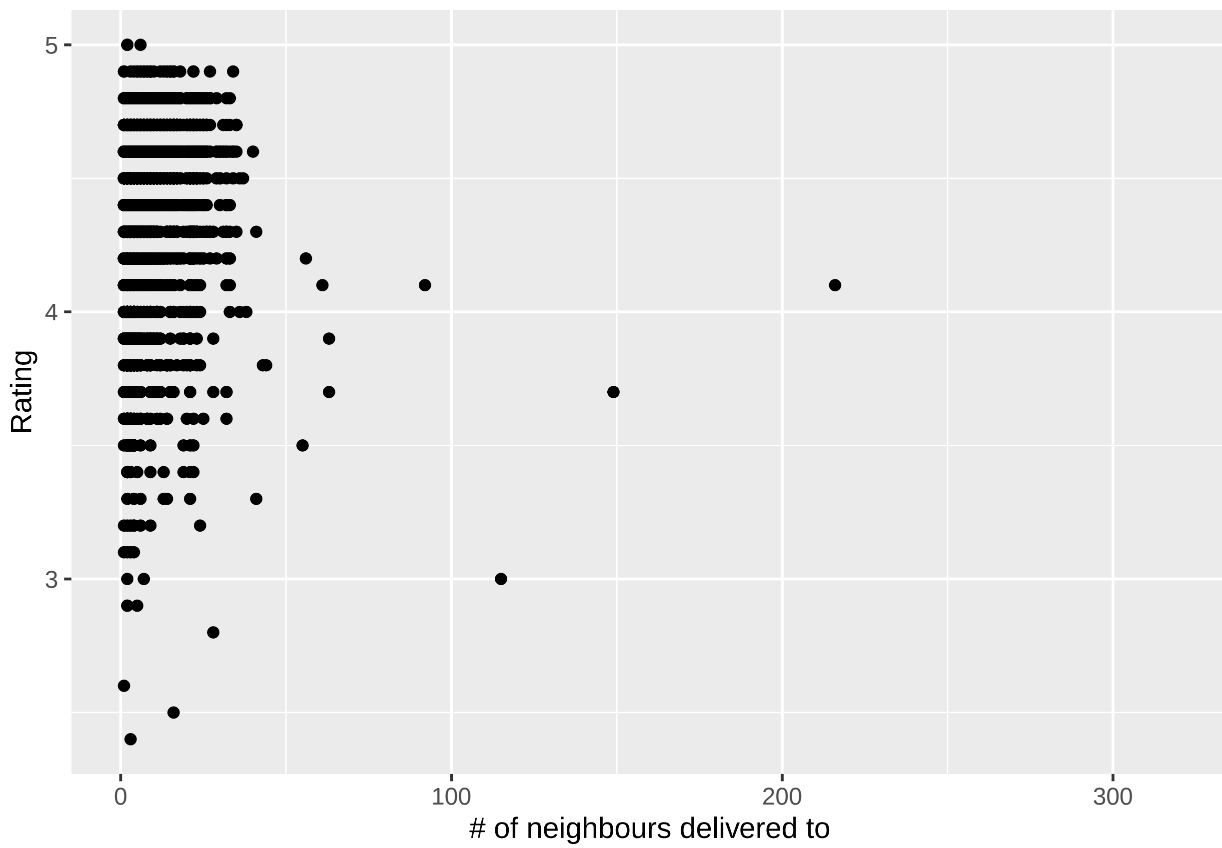
Here the negative correlation appear significantly more evident. But since size of clusters and other variables were not taken into account, we can state that our hypothesis is stronger, but it's early to reach any conclusion though. Let's continue our analysis and come back to this question later.

2) Does reach improve ratings?

Reach of delivery (number of different locations to which each restaurant delivers) is for sure a very important factor in home deliveries. The more destinations available, the more consumers will consider the restaurant as an option, therefore the more orders opportunities.

```
x111 <- aggregate(neighborhood_name~restaurant_id, merged, uniqueN)
names(x111)[names(x111) == "neighborhood_name"] <- "neigh_count"
res2 <- merge(res2, x111, by = 'restaurant_id') #count how many neighbours to which each restaurant del

#Plot
plot_neighcount <- ggplot(res2, aes(x=neigh_count, y = rest_rating)) +
      geom_point() +
      ylab("Rating") + xlab("# of neighbours delivered to")

plot_neighcount
```

```
## Warning: Removed 103 rows containing missing values (geom_point).
```

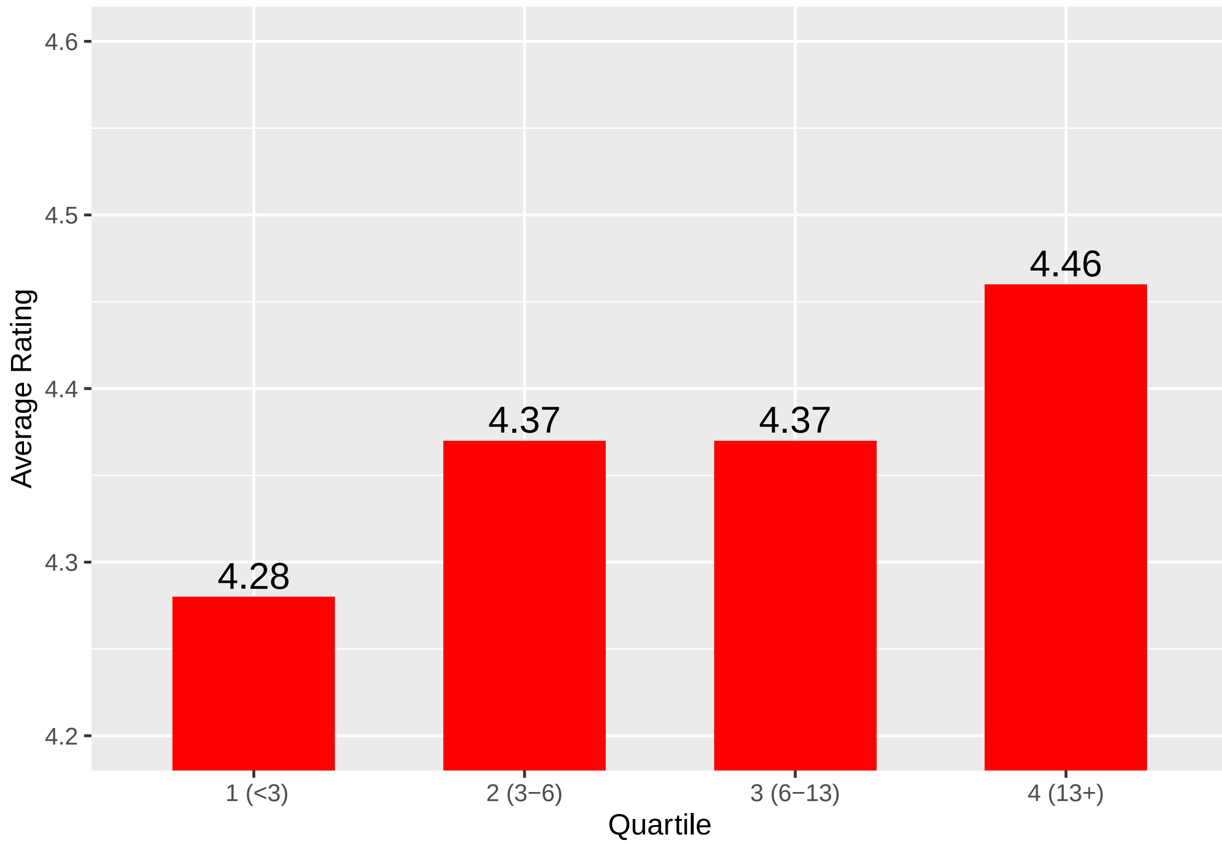


It's very hard to spot a trend in this scatter plot. Let's try to summarize data by dividing distribution of neighborhoods in quartiles and see average rating per quartile.

```
#divide quartiles

first_quart <- filter(res2, neigh_count <3)
first_quart <- first_quart[!is.na(first_quart$rest_rating),]

sec_quart <- filter(res2, neigh_count <6 | neigh_count >3)
sec_quart <- sec_quart[!is.na(sec_quart$rest_rating),]

ter_quart <- filter(res2, neigh_count >6 | neigh_count <13)
ter_quart <- ter_quart[!is.na(ter_quart$rest_rating),]

quar_quart <- filter(res2, neigh_count>13)
quar_quart <- quar_quart[!is.na(quar_quart$rest_rating),]

#Create summary table
quartile <- c(1, 2, 3, 4)
range <- c(" 1 (<3)", "2 (3-6)", "3 (6-13)", "4 (13+)" )
avg_rating <- c(mean(first_quart$rest_rating), mean(sec_quart$rest_rating), mean(ter_quart$rest_rating)
quartiles <- data.frame(quartile, range, avg_rating)
quartiles$avg_rating <- round(quartiles$avg_rating, digits = 2)
```

```
p_prova <- ggplot(quartiles, aes(x = range, avg_rating), y = avg_rating)+
            geom_bar(fill = "red", stat = "identity", width=0.6) +
```

```
        coord_cartesian(ylim=c(4.2, 4.6)) +
        geom_text(aes(label=avg_rating), vjust=-0.3, color="black", size=5)+ ylab("Average Rating")
p_prova
```



By dividing the distribution in quartiles and computing the average rating for each, we can see relevant differences. Again, we keep this result on hold for the moment. We will soon test the significance of such finding.

3) Number of deliveries

Restaurant's results can be measured by counting the total number of deliveries made. We do not have details about the time frame in which delieveries occurred, but looking at descriptive statistics we can resonably assume a time frame of a single day. Even if time frame is just as assumption, it will be the same for all records.

To answer this question, we first create a definitive dataframe with all necessary information, and there directly apply multiple linear regression to study coeffients significancy

```
del_numb <- sort(table(merged$restaurant_id), decreasing=TRUE)
del_numb <- as.data.frame(del_numb)
del_numb_table <- del_numb[1:10,]

names(del_numb)[names(del_numb) == "Freq"] <- "deliveries"
names(del_numb)[names(del_numb) == "Var1"] <- "restaurant_id"
```

```r
complete <- merge(res, del_numb, key='restaurant_id') #add deliveries column
complete <- merge(complete, x8, by='restaurant_id') #add avg_del_time column

#change rest_brand to dummy
complete$rest_brand[!is.na(complete$rest_brand)] <- "1"
complete$rest_brand[is.na(complete$rest_brand)] <- "0"
complete$rest_brand[complete$rest_brand == "Get drinks delivered"] <- "0"
complete$rest_brand <- as.numeric(complete$rest_brand)

#remove unuseful variables, rename existing ones
complete <- merge(complete, x111, by='restaurant_id') #add neighboord count column
rownames(complete) <- complete$restaurant_id
names(complete)[names(complete) == "avg"] <- "average_item_cost"
names(complete)[names(complete) == "count"] <- "items_on_menu"
complete$average_item_cost <- as.numeric(complete$average_item_cost)
complete$items_on_menu <- as.numeric(complete$items_on_menu)
complete <- select(complete, -c(restaurant_id, rest_name, rest_postcode, rest_neighborhood, rest_menu_i
complete <- complete[!is.na(complete$rest_rating),]
complete[1:5,]
```

```
##    rest_brand rest_rating average_item_cost items_on_menu deliveries
## 3           1         4.6          7.265741           108         38
## 5           1         4.5         10.756209           153         40
## 8           1         4.7          8.360360           111         32
## 10          0         4.8          8.862637            91         29
## 15          1         4.7          6.253691           149         14
##    avg_del_time neigh_count
## 3         19.61           9
## 5         17.12          10
## 8         18.28          10
## 10        18.62           9
## 15        21.79           4
```

To follow the order of business questions, we first compute linear regression using ratatings as dependent variable. We go back to number of deliveries in a moment.

What impacts ratings?

```r
lmRatings = lm(rest_rating~rest_brand + neigh_count + avg_del_time + deliveries + items_on_menu + averag
summary(lmRatings) #Review the results
```

```
##
## Call:
## lm(formula = rest_rating ~ rest_brand + neigh_count + avg_del_time +
##     deliveries + items_on_menu + average_item_cost, data = complete)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.90061 -0.15046  0.04487  0.19579  1.20223
##
```

```
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.537e+00  1.733e-02 261.766  < 2e-16 ***
## rest_brand         8.768e-02  9.197e-03   9.534  < 2e-16 ***
## neigh_count        5.026e-03  1.347e-03   3.731 0.000194 ***
## avg_del_time      -1.108e-02  5.372e-04 -20.617  < 2e-16 ***
## deliveries        -7.941e-04  3.343e-04  -2.375 0.017581 *
## items_on_menu     -3.389e-04  4.543e-05  -7.459 1.07e-13 ***
## average_item_cost  1.143e-02  9.278e-04  12.324  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2875 on 3958 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1784, Adjusted R-squared:  0.1772
## F-statistic: 143.2 on 6 and 3958 DF,  p-value: < 2.2e-16
```

COMMENT: all coefficients are significant (pvalue < 0.001), meaning that there is statistical evidence to affirm correlation among dependent and independent variables.

rest_brand, neigh_count and avg_del_time have high B values. A one unit negative change in delivery time results in an average increase of 0.14 rating points. A one unit change in average item cost results in an average increase of 0.15 rating points.

What impacts number of orders?

```r
lmOrders = lm(deliveries~rest_brand + rest_rating + avg_del_time + neigh_count + items_on_menu + averag
summary(lmOrders)
```

```
##
## Call:
## lm(formula = deliveries ~ rest_brand + rest_rating + avg_del_time +
##     neigh_count + items_on_menu + average_item_cost, data = complete)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -149.927   -4.260    1.211    4.901  285.944
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.386146   3.523530   0.110  0.91274
## rest_brand         1.255931   0.441465   2.845  0.00447 **
## rest_rating       -1.792453   0.754617  -2.375  0.01758 *
## avg_del_time      -0.080350   0.026827  -2.995  0.00276 **
## neigh_count        3.719844   0.024789 150.063  < 2e-16 ***
## items_on_menu      0.022791   0.002143  10.635  < 2e-16 ***
## average_item_cost  0.268360   0.044712   6.002 2.12e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.66 on 3958 degrees of freedom
```

```
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.8568, Adjusted R-squared:  0.8566
## F-statistic:  3948 on 6 and 3958 DF,  p-value: < 2.2e-16
```

All coefficients are significant (pvalues < 0.05). This means that all variables are correlated with the dependent variable. Ratings appear to be negatively correlated to the number of deliveries. This is unexpected, since one would assume that higher ratings could drive more orders. This can be explained by the growing use of delivery apps, where cousine type and delivery time are the biggest drivers of choice. Indeed, delivery time appears to be negatively correlated to number of deliveries. Being part of a brand significantly impacts the number of deliveries: about +1.25 on average. Ultimately, delivery reach (number of neighbours in which the restaurant delivers) appears to have the biggest impact on number of orders: +3.72 for a unit increase in average number of neighbors.

#Takeaways

Initial hypothesis are confirmed:

RATINGS:

- Delivery times have a significant negative impact on ratings
- Reach has a significant (but small) impact on ratings and…
- Being part of a chain significantly impacts ratings

NUMBER OF ORDERS:

- Delivery times have a slight negative impact on number of orders
- Reach has a significant positive impact on number of orders and…
- Being part of a chain significantly impacts number of orders.