BKM

Atahan Şenşatar

14 11 2019

Get Data and Prepare for Analysis

We will get our data from https://bkm.com.tr/secilen-aya-ait-sektorel-gelisim/ which is given data table

First step Libraries

```
library(rvest)
```

```
## Loading required package: xml2
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------------------------
```

```
## v tibble  2.1.3     v purrr   0.3.2
## v tidyr   1.0.0     v dplyr   0.8.3
## v readr   1.3.1     v stringr 1.4.0
## v tibble  2.1.3     v forcats 0.4.0
```

```
## -- Conflicts -------------------------------------------------------------------------------
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x readr::guess_encoding()  masks rvest::guess_encoding()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x purrr::pluck()           masks rvest::pluck()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```
library(tidyverse)
```

Collecting links from website according our date request

We will change month and year sections in link with for loop and create a list.We will collect between 2018-2019(24 month)

```
x <- list()
for (i in 18:19){
  for (j in 1:12) {
    x[[paste0("20",i,".",j)]]<-paste0("https://bkm.com.tr/secilen-aya-ait-sektorel-gelisim/?filter_year=
  }
}

head(x,n=2)
```

```
## $`2018.1`
## [1] "https://bkm.com.tr/secilen-aya-ait-sektorel-gelisim/?filter_year=2018&filter_month=1&List=Listel
##
## $`2018.2`
## [1] "https://bkm.com.tr/secilen-aya-ait-sektorel-gelisim/?filter_year=2018&filter_month=2&List=Listel
```

Create blank Dataframe for our main work field

```
DF <- data.frame()
```

Creating Civilized Dataframe

In this part we will download our data link by link read it with rvest clean and bind into DF.

That chunk will give warning about LHS because our data frame contain NA's

```
for (i in seq_along(x)){
  html_monthly <- read_html(x[[i]])
  temp_df <- html_table(html_monthly,fill=T)[[4]]
  temp_df <- temp_df[-c(1,2,28,29),]
  temp_df$date <- names(x)[i]
  DF <- bind_rows(DF,temp_df)
}
```

```
## Warning in temp_df$date <- names(x)[i]: Coercing LHS to a list

## Warning in temp_df$date <- names(x)[i]: Coercing LHS to a list

## Warning in temp_df$date <- names(x)[i]: Coercing LHS to a list

## Warning in temp_df$date <- names(x)[i]: Coercing LHS to a list

## Warning in temp_df$date <- names(x)[i]: Coercing LHS to a list

## Warning in temp_df$date <- names(x)[i]: Coercing LHS to a list
```

Remove NA's from our dataframes

We don't have data after 2019.06 let's remove NA's

```
DF <- DF[complete.cases(DF),]
```

Date formatting

Our date column structure is chr we have to convert the date format for arranging

```
DF$date <- ymd(DF$date,truncated = 1)
```

Column names setup

```
colnames(DF) <- c("segment","n_cc","n_dc","sum_cc","sum_dc","date")

print(head(DF))
```

```
##                                segment       n_cc       n_dc   sum_cc
## 1                        ARABA KİRALAMA    292.290     57.920   158,36
## 2 ARAÇ KİRALAMA-SATIŞ/SERVİS/YEDEK PARÇA  2.725.497    540.511 1.920,41
## 3           BENZİN VE YAKIT İSTASYONLARI 25.260.026  8.000.259 4.641,49
## 4                     BIREYSEL EMEKLILIK  2.582.916        527   667,90
## 5                           ÇEŞİTLİ GIDA 24.041.009 12.196.630 3.483,63
## 6                     DOĞRUDAN PAZARLAMA    973.180     24.216   719,08
##   sum_dc       date
## 1  12,48 2018-01-01
## 2 100,50 2018-01-01
## 3 567,45 2018-01-01
## 4   0,19 2018-01-01
## 5 477,84 2018-01-01
## 6   5,66 2018-01-01
```

Character format and punctuation arrangement

In our dataframe current amounts format is chars and seperated with commas.We will change commas and formats for use in analysis

```
DF$sum_cc <- gsub("[,]","",DF$sum_cc)
DF$sum_dc <- gsub("[,]","",DF$sum_dc)

DF$sum_cc <- as.numeric(as.character(DF$sum_cc))
DF$sum_dc <- as.numeric(as.character(DF$sum_dc))

print(head(DF))
```

```
##                                segment       n_cc       n_dc      sum_cc
## 1                        ARABA KİRALAMA    292.290     57.920 15836.00000
## 2 ARAÇ KİRALAMA-SATIŞ/SERVİS/YEDEK PARÇA  2.725.497    540.511     1.92041
## 3           BENZİN VE YAKIT İSTASYONLARI 25.260.026  8.000.259     4.64149
## 4                     BIREYSEL EMEKLILIK  2.582.916        527 66790.00000
## 5                           ÇEŞİTLİ GIDA 24.041.009 12.196.630     3.48363
```

```
## 6                    DOĞRUDAN PAZARLAMA    973.180     24.216 71908.00000
##   sum_dc       date
## 1   1248 2018-01-01
## 2  10050 2018-01-01
## 3  56745 2018-01-01
## 4     19 2018-01-01
## 5  47784 2018-01-01
## 6    566 2018-01-01
```

Some Analysis and Plots

First analysis i want to see the effect of dollar event in July 2018 on rent-a car services combined credit card and debit card

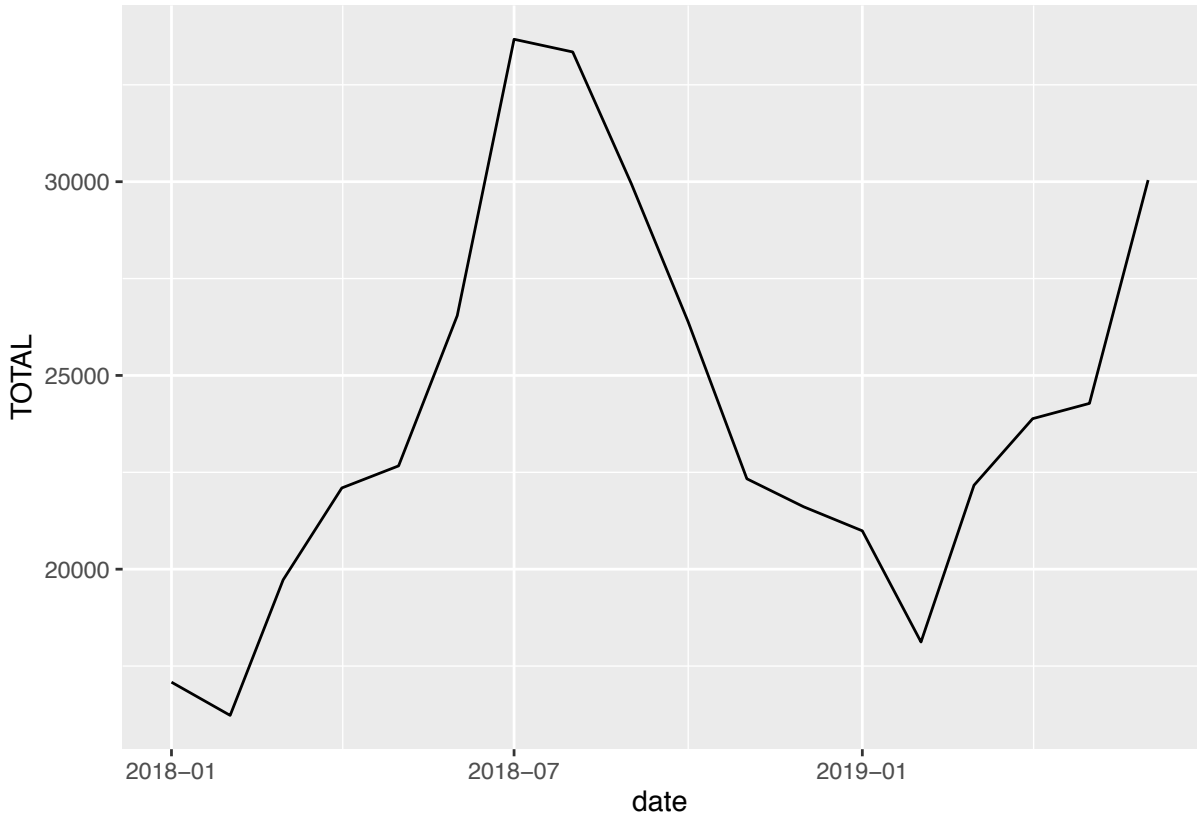
```
rent_car_analysis <- DF %>%
  arrange(date) %>%
  filter(segment %in% "ARABA KİRALAMA")%>%
  mutate(TOTAL = rowSums(.[4:5]))%>%
  select(segment,date,TOTAL)


rent_car_analysis
```

```
##           segment       date TOTAL
## 1  ARABA KİRALAMA 2018-01-01 17084
## 2  ARABA KİRALAMA 2018-02-01 16227
## 3  ARABA KİRALAMA 2018-03-01 19725
## 4  ARABA KİRALAMA 2018-04-01 22096
## 5  ARABA KİRALAMA 2018-05-01 22666
## 6  ARABA KİRALAMA 2018-06-01 26545
## 7  ARABA KİRALAMA 2018-07-01 33676
## 8  ARABA KİRALAMA 2018-08-01 33348
## 9  ARABA KİRALAMA 2018-09-01 29944
## 10 ARABA KİRALAMA 2018-10-01 26380
## 11 ARABA KİRALAMA 2018-11-01 22333
## 12 ARABA KİRALAMA 2018-12-01 21611
## 13 ARABA KİRALAMA 2019-01-01 20990
## 14 ARABA KİRALAMA 2019-02-01 18121
## 15 ARABA KİRALAMA 2019-03-01 22166
## 16 ARABA KİRALAMA 2019-04-01 23884
## 17 ARABA KİRALAMA 2019-05-01 24279
## 18 ARABA KİRALAMA 2019-06-01 30042
```

Let's look in line graph

```
ggplot(rent_car_analysis,aes(x=date,y=TOTAL))+
  geom_line()
```

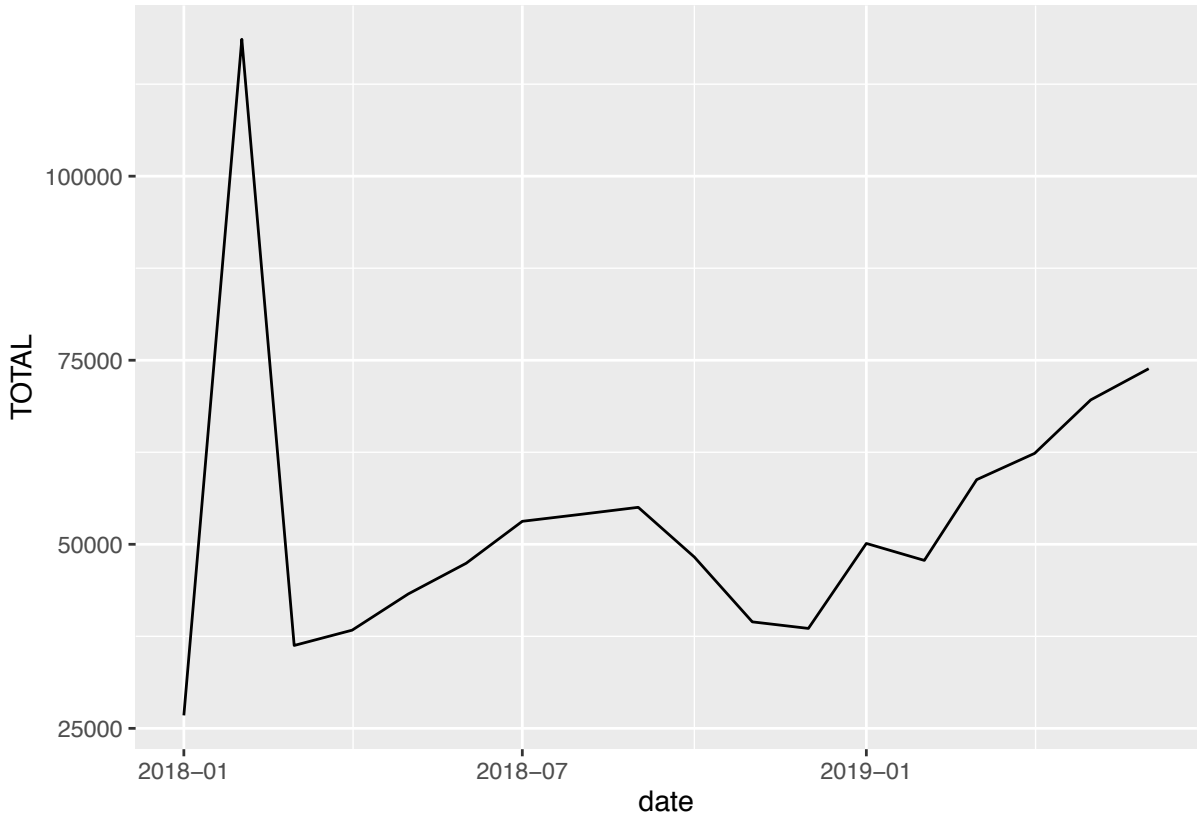


It seems a sharp fall after July 2018

Let's look air ticket spending

```
air_ticket_analysis <- DF %>%
  arrange(date) %>%
  filter(segment %in% "HAVAYOLLARI")%>%
  mutate(TOTAL = rowSums(.[4:5]))%>%
  select(segment,date,TOTAL)

ggplot(air_ticket_analysis,aes(x=date,y=TOTAL))+
  geom_line()
```



It doesn't look like there is a significant change on air tickets