Time Series Plot for Kerala Ad Data

Suhaas Ramani

9/15/2020

Upon downloading the excel file that was provided I filtered the region\_KL column and deleted the other regional areas on a new spreadsheet.

I then needed to get rid of the timestamp that was placed with each ad creation. Hence why I created two columns, creation\_date and creation\_date\_notime, respectively. For creation\_date I used the “=SUBSTITUTE” function to delete the T character and replace it with a space. For the creation\_date\_notime column I used the “=INT” function to remove the timestamp and retrieve the column in just dates.

After which I imported the excel file into R using the code below:

library(tidyverse)

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

## v ggplot2 3.3.1 v purrr 0.3.4  
## v tibble 3.0.1 v dplyr 1.0.0  
## v tidyr 1.1.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts ---------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(ggplot2)  
library(dplyr)  
library(readxl)  
Kerala\_ad <- read\_excel("C:\\Users\\Suhaas Ramani\\Desktop\\R Research Project - Facebook Ads\\Filtered\_Kerala\_Ad\_Data.xlsx")

After I imported the data, I filtered it to remove the NAs from the time variable, I then grouped the data by the time variable, and summarized the data to include a counts variable that contained the frequency for each date. This can be seen in the code below:

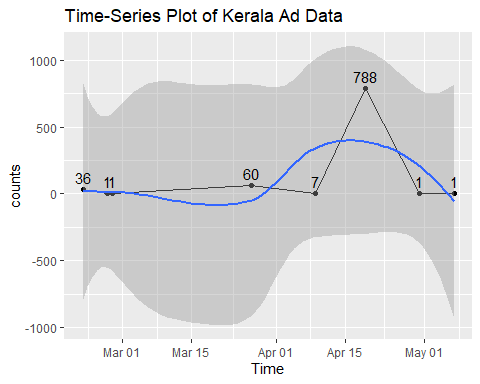
Kerala\_ad <- Kerala\_ad %>%   
 filter(!is.na(creation\_date\_notime)) %>%  
 group\_by(creation\_date\_notime) %>%   
 summarize(counts = n())

## `summarise()` ungrouping output (override with `.groups` argument)

Finally I created the plot using the ggplot2 tool in R, where I placed time on the x-axis and counts on the y-axis.

Kerala\_ad\_plot <- ggplot(Kerala\_ad, aes(x = creation\_date\_notime, y = counts)) +   
 geom\_point() +   
 geom\_line() +   
 geom\_smooth() +  
 geom\_text(aes(label = counts), vjust = -0.5) +   
 xlab("Time") +   
 ggtitle("Time-Series Plot of Kerala Ad Data")  
  
Kerala\_ad\_plot

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



As can be seen from this plot, the data shows a marked spike in ads from 7 to 788 on April 19th.