Expedia Hotel Recommendation - Initial Analysis

Anandan Balaji

12 May 2016

## R Markdown

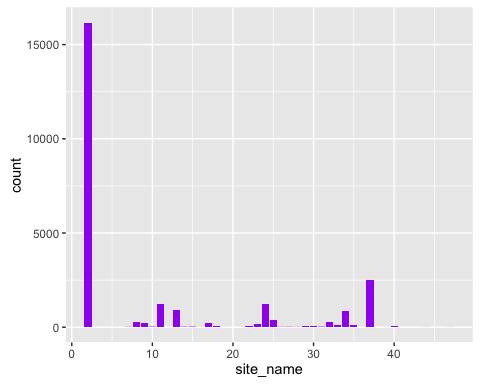
This R Markdown doucment illustrates the different plots for some of the data fields related to the Capstone Project - Expedia Hotel Recommendation.

The train data is huge - roughly 4GB. So, inorder to load and render fast, first 25K records were extracted and the following plots are generated for initial data analysis process.

# read 25 K  
train\_df <- read.csv(file = "train.25000.csv")  
  
library(ggplot2)

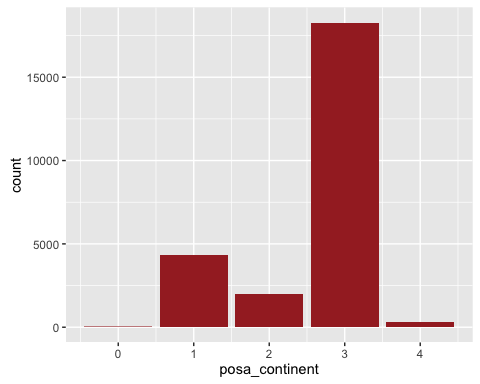
## site\_name

ggplot(train\_df, aes(site\_name)) + geom\_bar(fill = "purple")



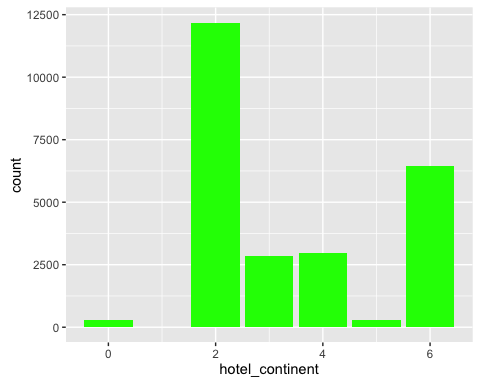
## posa\_continent - current user location

ggplot(train\_df, aes(posa\_continent)) + geom\_bar(fill = "brown")



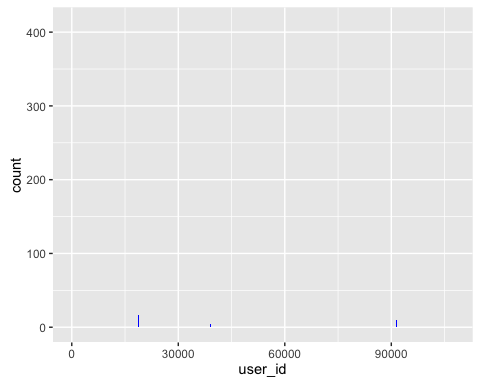
## hotel\_continent - destination location

ggplot(train\_df, aes(hotel\_continent)) + geom\_bar(fill = "green")



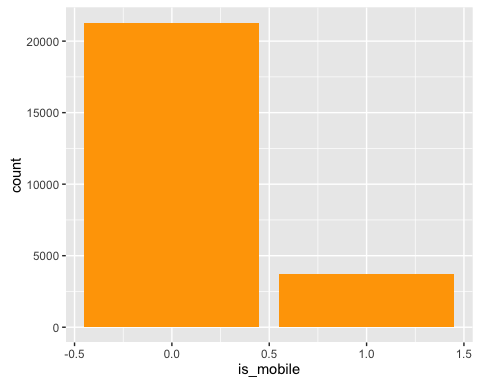
## user\_id

ggplot(train\_df, aes(user\_id)) + geom\_bar(fill = "blue")



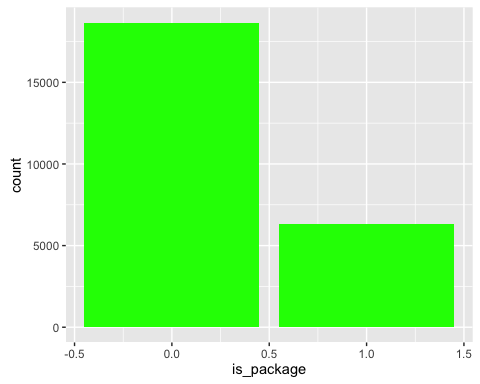
## is\_mobile

ggplot(train\_df, aes(is\_mobile)) + geom\_bar(fill = "orange")



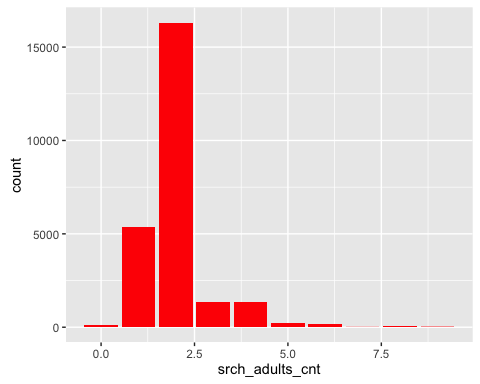
## is\_package - hotel / hotel + air

ggplot(train\_df, aes(is\_package)) + geom\_bar(fill = "green")



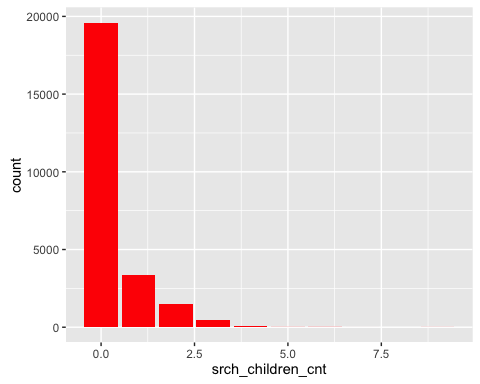
## srch\_adults\_cnt

ggplot(train\_df, aes(srch\_adults\_cnt)) + geom\_bar(fill = "red")



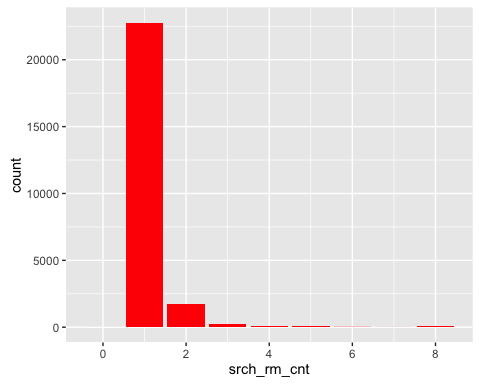
## srch\_children\_cnt

ggplot(train\_df, aes(srch\_children\_cnt)) + geom\_bar(fill = "red")



## srch\_rm\_cnt

ggplot(train\_df, aes(srch\_rm\_cnt)) + geom\_bar(fill = "red")



## hotel\_cluster

ggplot(train\_df, aes(hotel\_cluster)) + geom\_bar(fill = "blue")

