Expedia Hotel Recommendation - Initial Analysis

Anandan Balaji

12 May 2016

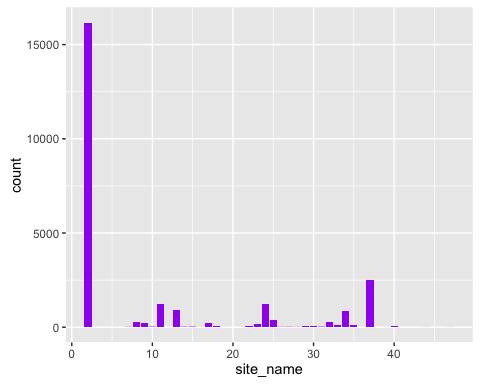
## R Markdown

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

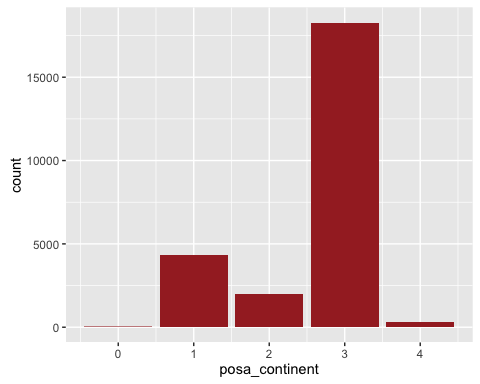
The train data is huge - roughly 4GB. So, inorder to load and render fast, first 25K records were extracted and the following plots are generated for initial data analysis process.

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

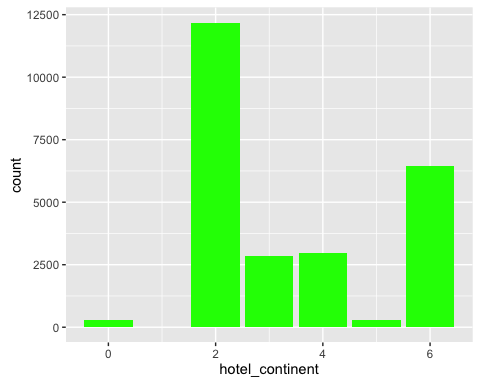
## site\_name



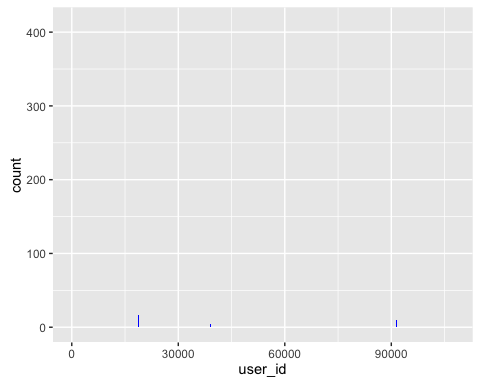
## posa\_continent - current user location



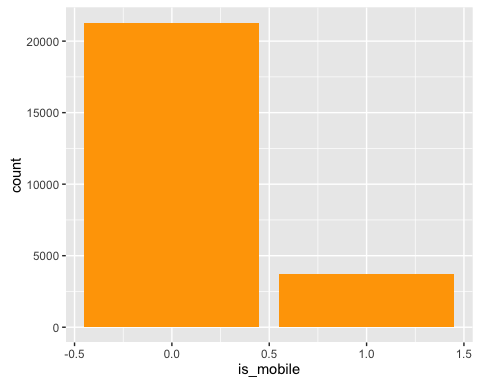
## hotel\_continent - destination location



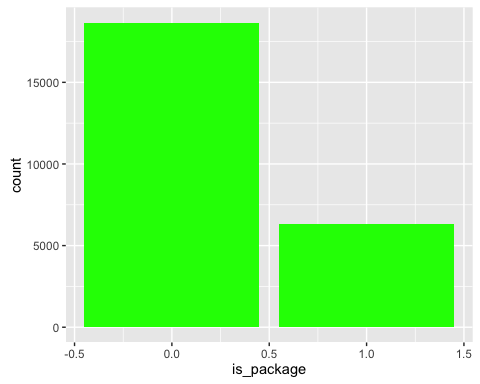
## user\_id



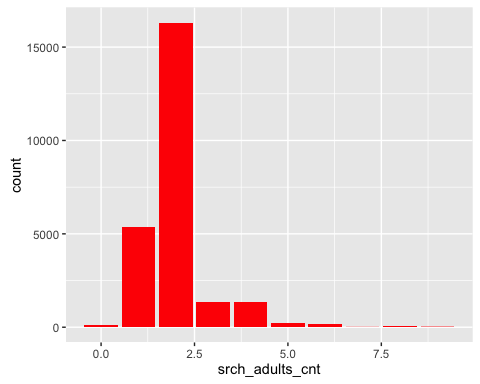
## is\_mobile



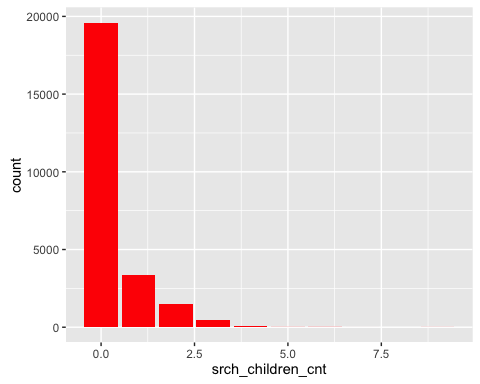
## is\_package - hotel / hotel + air



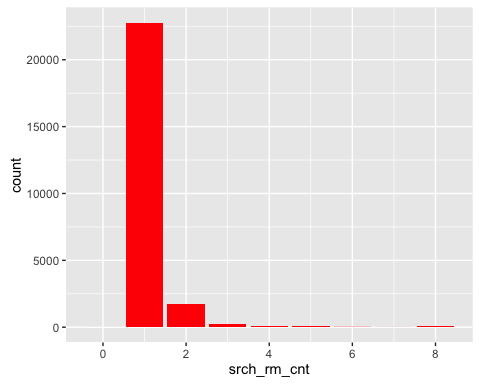
## srch\_adults\_cnt



## srch\_children\_cnt



## srch\_rm\_cnt



## hotel\_cluster

