Analytics using Clickstream Data

Visualizations and Insights using R

By:

Anish Anand

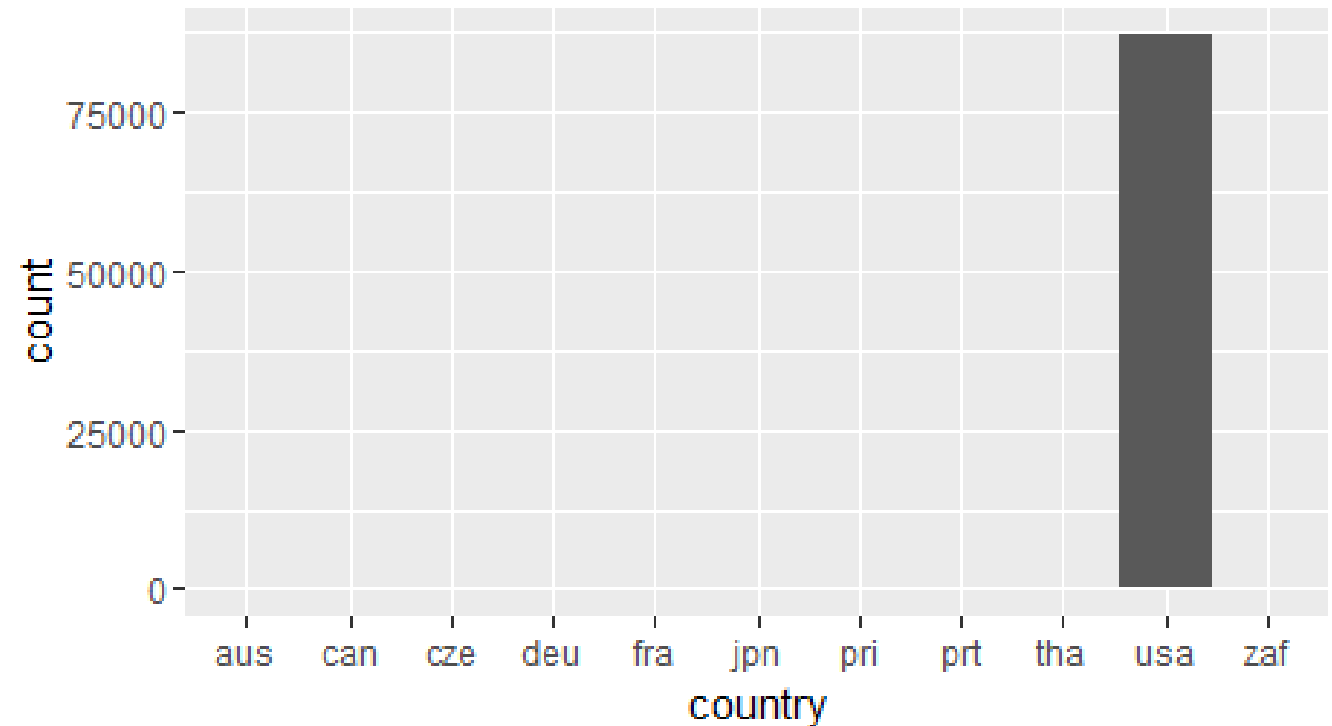# Analytics using Clickstream Data

Visualizations

Hereby, I have included visualizations done on both R and Excel 2013(powerview). The code of visualizations in R are however shared in the code at the end.

Expected Outcomes

❑ **Analyze the clickstream data by location**
The adjacent visualization shows that most of the clicks have been done from the **United States of America(usa) and then Puerto Rico(pri).**
Hence, this shall form the basis of our analysis and all the further visualizations has been done for the people and clicks obtained from usa.

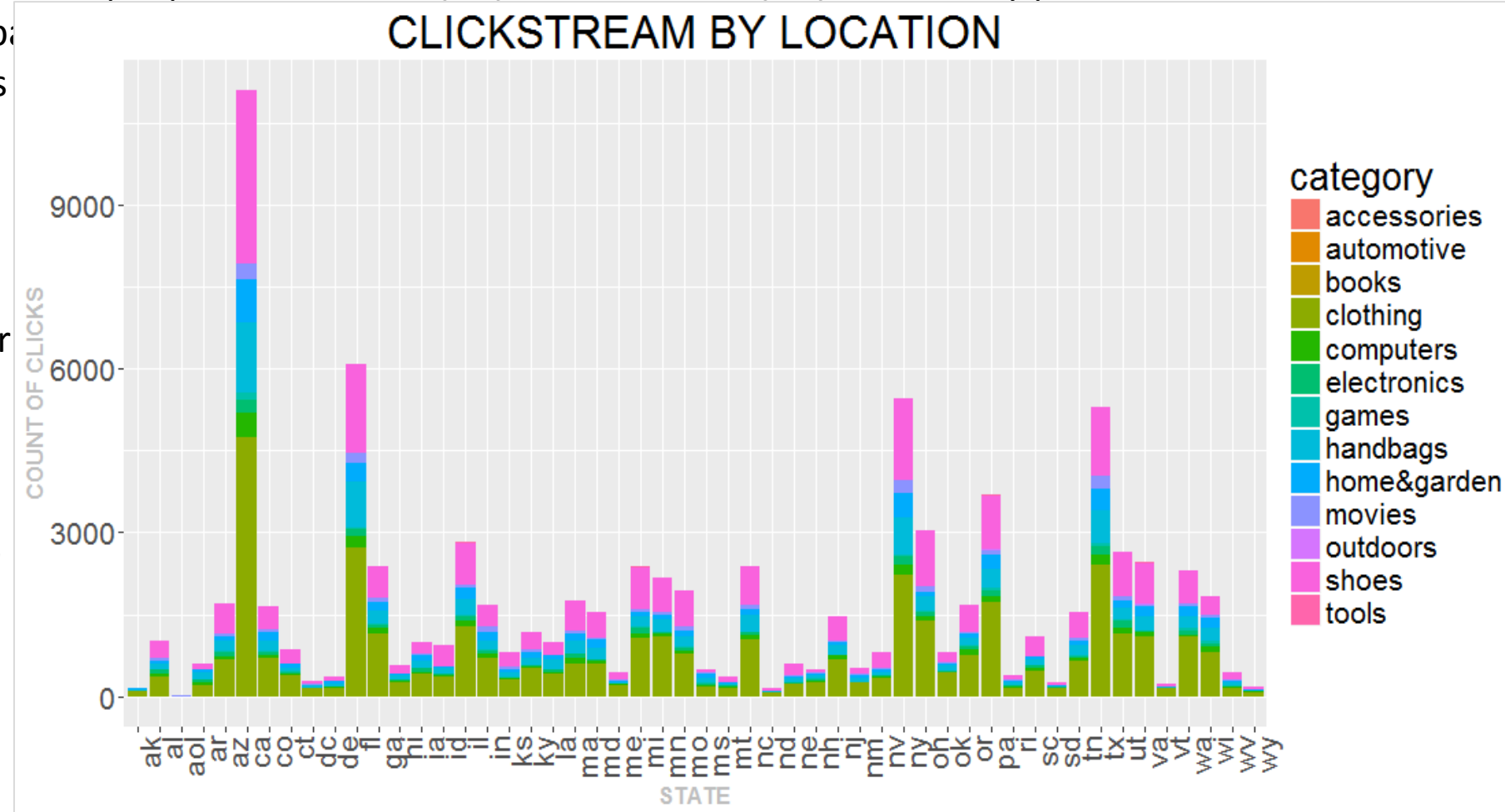# Analytics using Clickstream Data

Moreover, we can see that the product category of **clothing sites** were most visited by **people across the United States**. Moreover, one more insight is that the people of **Columbia (CA) and New York (NY)** are actively present online, searching for products to buy. Apa[...] that, the crowd of **Arizona** shows significant interest in **electronics products** too.
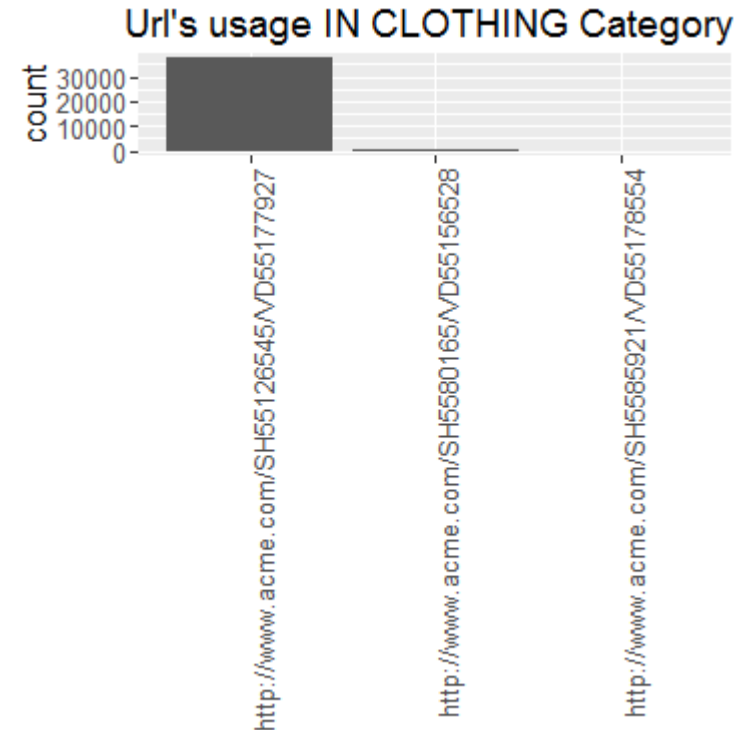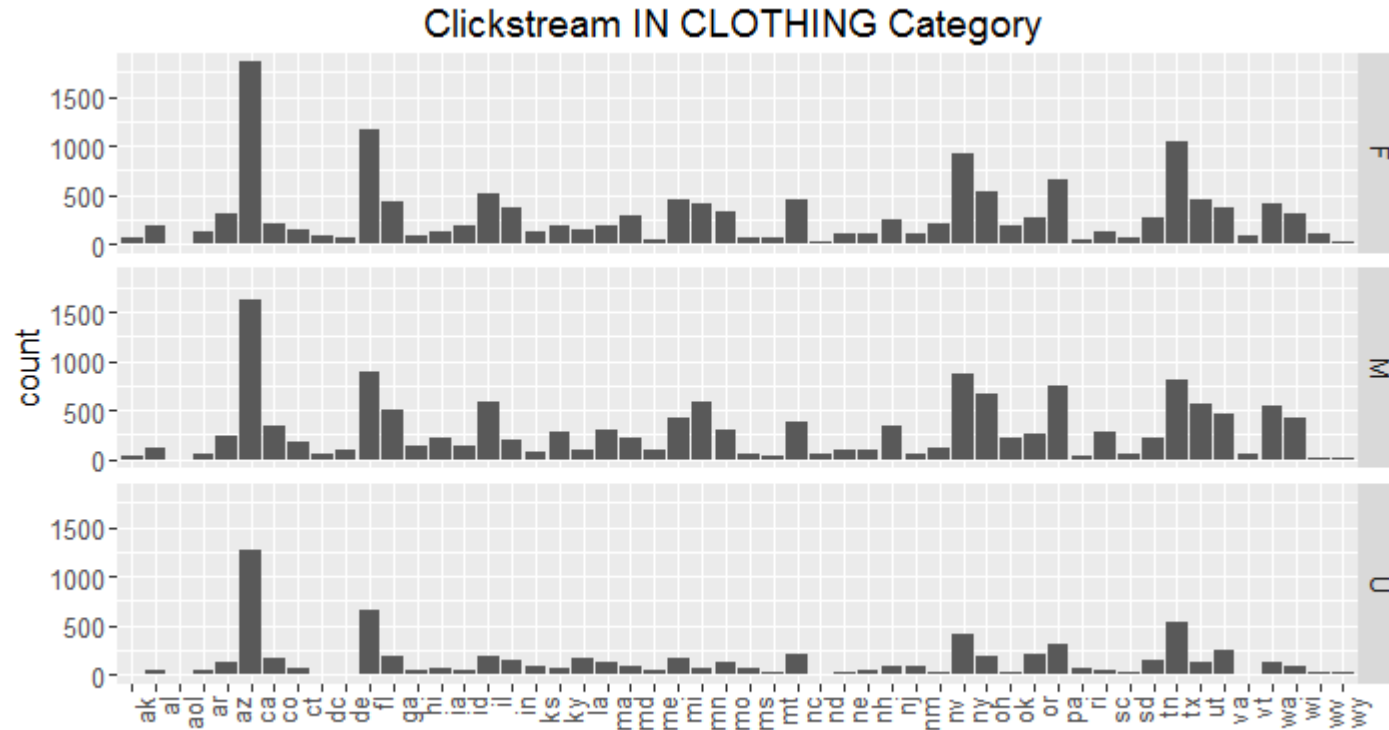
Target geographical area as per customer interest can be another insight from the visualization.

Moreover, one more insight is that the people of Columbia(CA), Florida (fl) and New York(NY) are actively present online searching for products to buy.

CLICKSTREAM BY LOCATION

Analytics using Clickstream Data

❑ Filter the data by product category


```
85  #filtering the data by product category. Here we are just considering categories clothing and shoes due to their
86  #high interest among buyers
87  a<-filter(list2,category=="clothing")
88  b<-filter(list2,category=="shoes")
89  hist(a$age)
90
91  #clicks by state and gender in clothing category
92  qplot(state,data=a,facets = GENDER_CD~.)+theme(text = element_text(size=12),axis.text.x =
93                          element_text(angle = 90, hjust = 1))+labs(title="Clickstream in CLOTHING Category ")
94
```
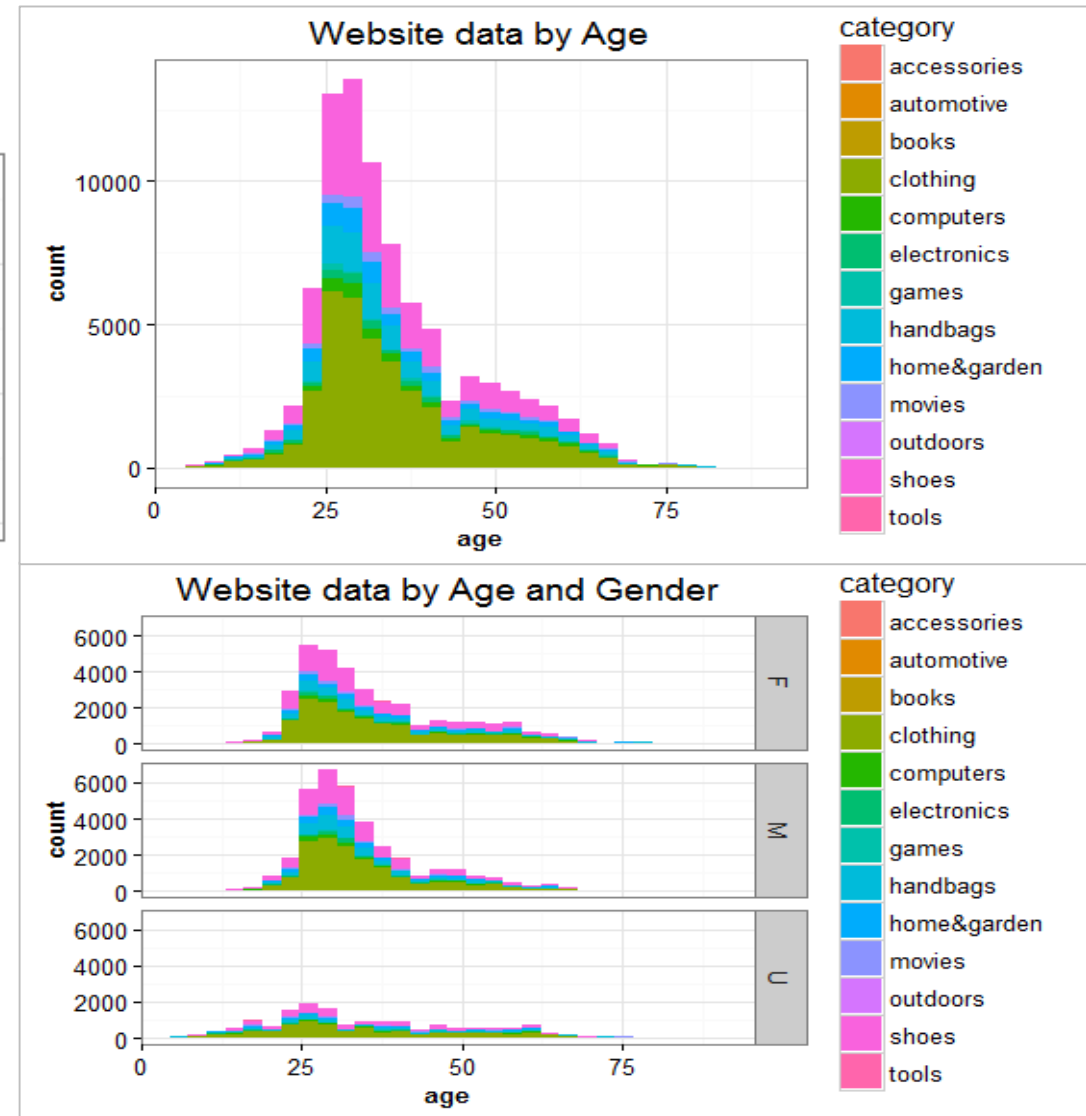
Analytics using Clickstream Data

❑ **Graph the website user data by age and gender**
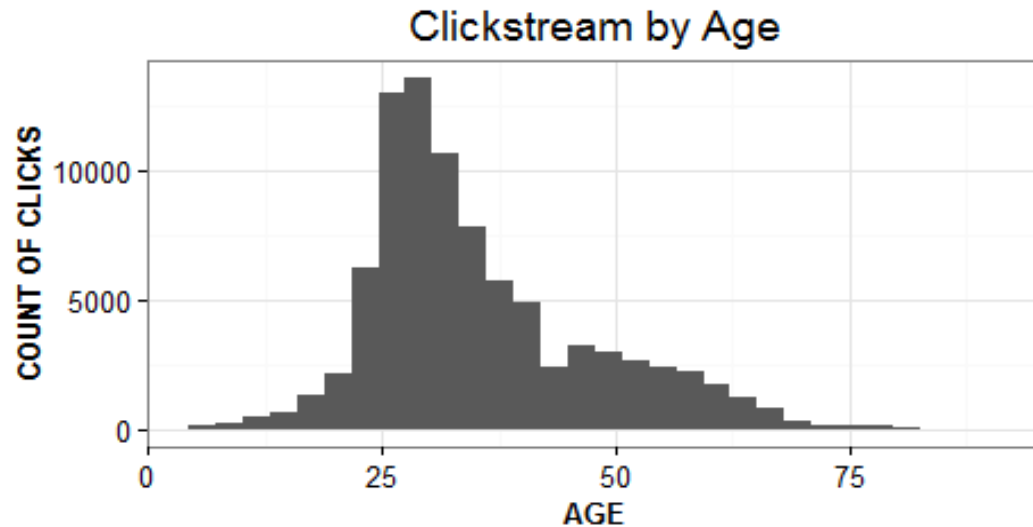
The adjacent plots are basically stacked bar plots. From the visualizations we can see that

- Clothing sites remain all time favorite destination for both males and females shoppers.
- Moreover, the handbags, home&garden and shoes websites bag the 2nd favorite place among the buyers.
- Subtle traces of interests for electronics can also be seen among the buyers
- Most shoppers are seen from 24-33 years of age(approx.)

Analytics using Clickstream Data
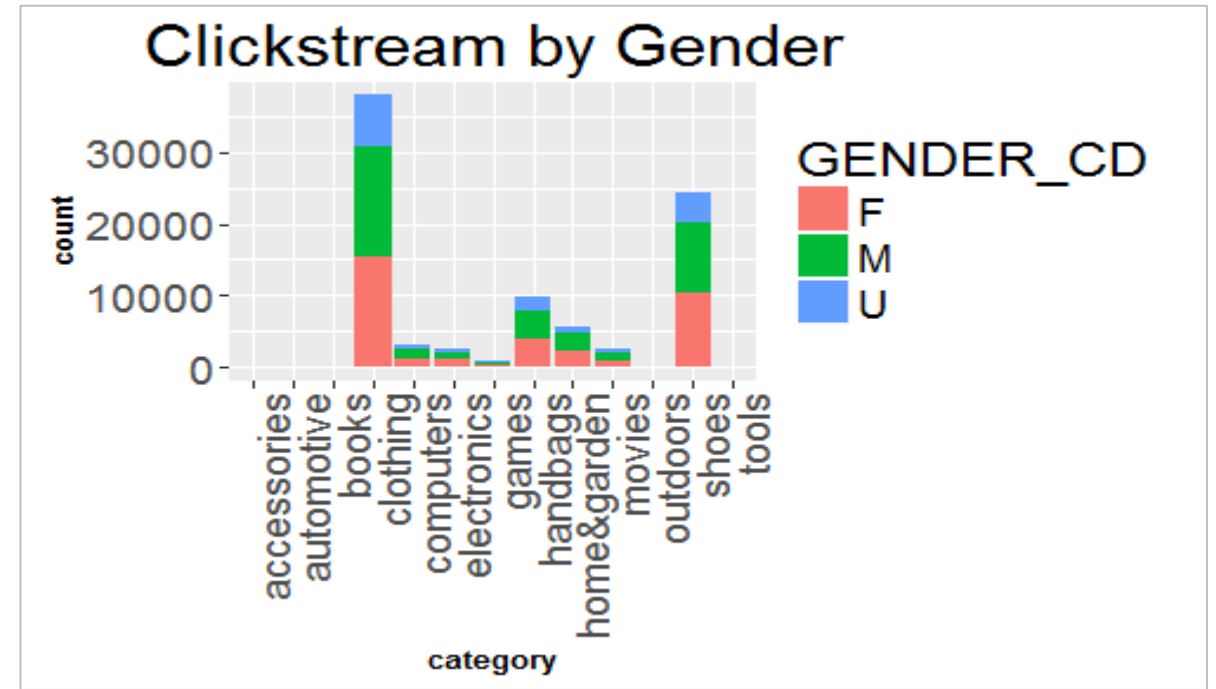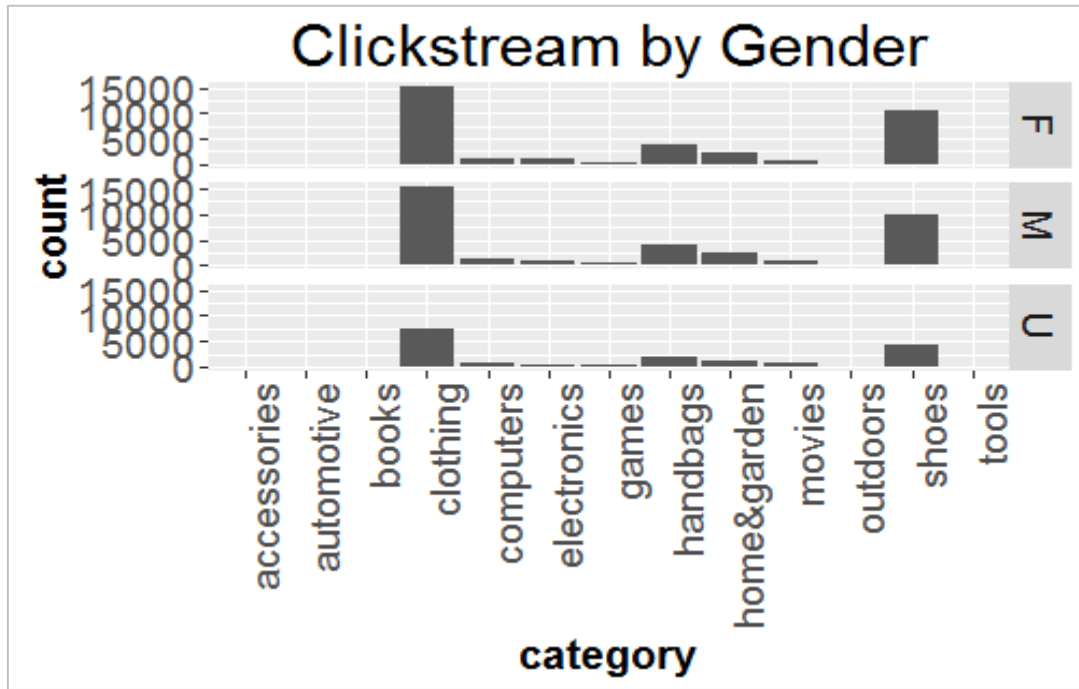
- Basically 3 gender fields are covered, namely
  - Male(M)
  - Female(F)
  - Custom gender(U).
- The males and females however continue to mark clothing sites as their favorite shopping destination.
- Moreover, most buyers of shoes are females, and the males tend to be in majority when it comes to computers and games. This perhaps is due to the psychological traits of these genders. Some websites corresponding to automotive, tools and accessories are not much in trend among the internet shoppers of usa

Analytics using Clickstream Data

❑ **Pick a target customer segment**
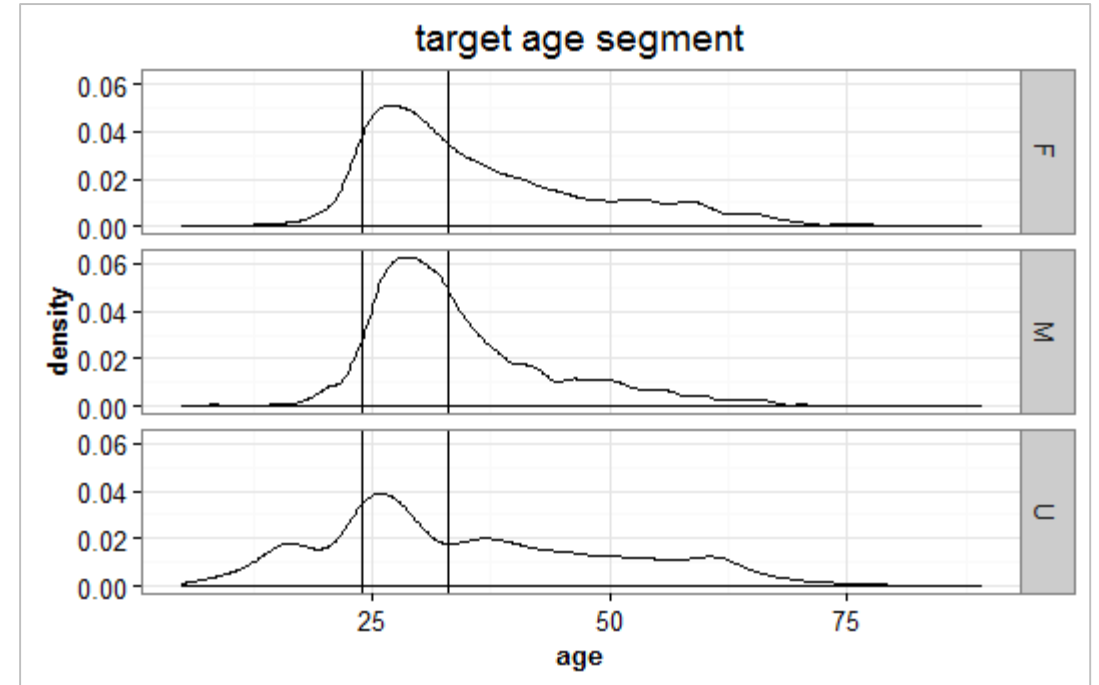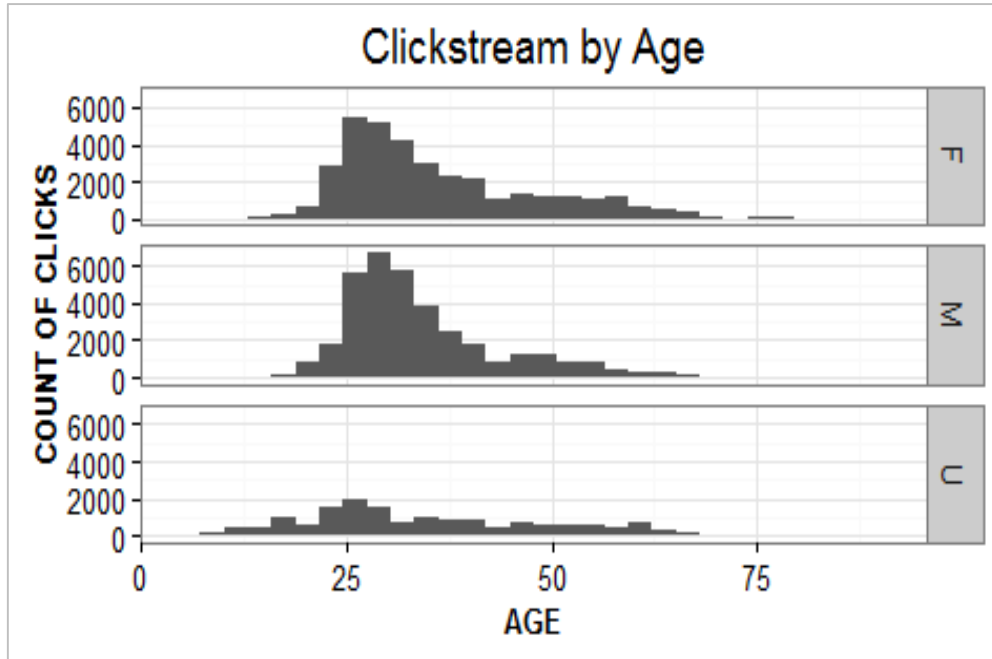
- All in all, we can say that age segment of *24-33 years* can be targeted by the sellers as these people can be actively seen surfing net for buying different product categories. The trend is almost same for males, females and custom categories (as can be seen from the density plot).
  In above plot, we can also see that most men in the targeted age segment are interested as buyers on sites.

Analytics using Clickstream Data

❑ Identify a few web pages with the highest bounce rates
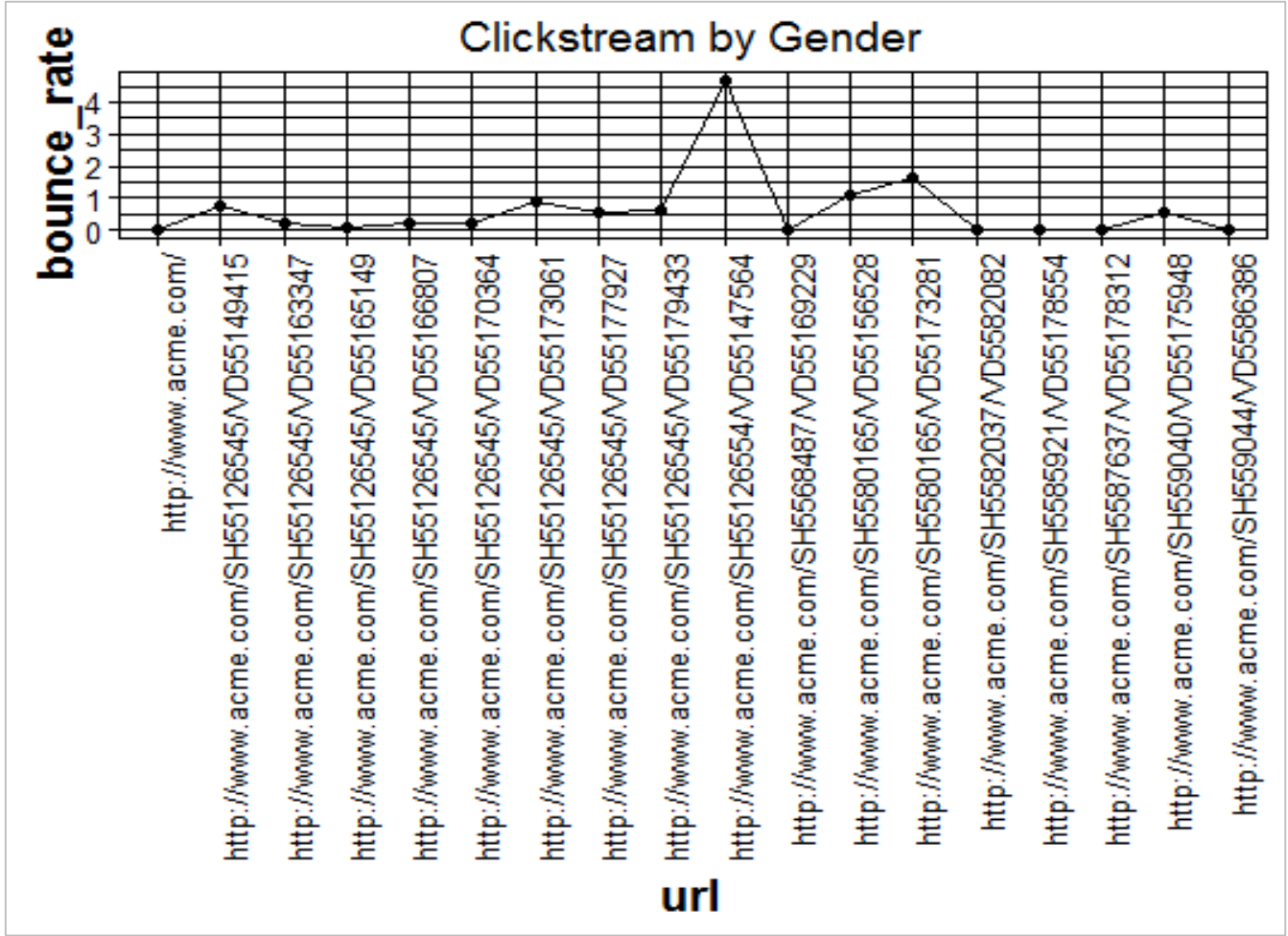
Bounce rate (sometimes confused with exit **rate**) is an Internet marketing term used in web traffic analysis. It represents the percentage of visitors who enter the **site** and then leave ("**bounce**") rather than continuing on to view other pages within the same**site**.The url having maximum bounce rates corresponds to the site which least interests the customers and proper changes are required on the site so as to lure in customers or buyers.

In our scenario the url http://www.acme.com/SH55126554/VD55147564 has the highest bounce rate which corresponds to the category of outdoor related sites.