simpli·learn
your pace, your place

Analyze the internet data of www.datadb.com

Business Analytic Foundation with R Tools- Question

Question

The web analytics team of www.datadb.com is interested to understand the web activities of the site, which are the sources used to access the website. They have a database that states the keywords of time in page, source group, bounces, exits, unique page views, and visits. The variables in the dataset are defined here for better understanding:

- Bounces: It represents the percentage of visitors who enter the site and "bounce" (leave the site) rather than continuing to view other pages within the same site.
- Exits: It represents the percentage of visitors to a site who actively click away to a different site from a specific page, after possibly having visited any other page on the site.
- Continent: It shows the continent from which the site has been accessed.
- Source group: It shows how the visitor has accessed the site.
- Time on page: It shows how long the user has spent on that particular page of the website.
- Unique page view: It represents the number of sessions during which that page was viewed one or more times.
- Visits: A visit counts all visitors, no matter how many times the same visitor may have been to your site.

To complete this project, you will require a strong understanding of the following concepts:

- Lesson 3 – Explore data using R, diagnostic analysis using R
- Lesson 4 –Logistic Regression

The team is targeting at the following issues:

- The team wants to analyze each variable of the data collected through data summarization to get a basic understanding of the dataset and to prepare for further analysis.
- As mentioned earlier, a unique page view represents the number of sessions during which that page was viewed one or more times. A visit counts all instances, no matter how many times the same visitor may have been to your site. So the team needs to know whether the unique page view value depends on visits.
- Find out the probable factors from the dataset, which could affect the exits. Exit Page Analysis is usually required to get an idea about why a user leaves the website for a session and moves on to another one. Please keep in mind that exits should not be confused with bounces.
- Every site wants to increase the time on page for a visitor. This increases the chances of the visitor understanding the site content better and hence there are more chances of a transaction taking place. Find the variables which possibly have an effect on the time on page.
- A high bounce rate is a cause of alarm for websites which depend on visitor engagement. Help the team in determining the factors that are impacting the bounce.

The data can be downloaded from the Excel file given below:

InternetCaseStudy.csv

The total time provided to complete this task is 2 hours.