

Coding Challenge

DESCRIPTION

Background and Objective:

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 the page, source group, bounces, exits, unique page views, and visits.

Domain: Web

Dataset Description:

The variables in the dataset are defined here for better understanding:

Attribute	Description
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.
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 pageview	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.

Analysis Tasks:

The team is targeting 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.
-
- To summarize the data we use the method `describe()` so that it shows the basic information like mean, standard deviation, maximum ,minimum values of the each column.
 - To know the probable factors that depend on one column we need to find the correlation matrix If the correlation co-efficient value is nearer to 1 it represents that that particular column is highly dependent.
 - To simply understand the correlation between the columns we use heat map. Here with the colour variations we can find the correlation. The colour of each cell represent the correlation coefficient between the corresponding columns.
 - The colour scheme is 'coolwarm' with the blue representing negative correlation and red representing positive correalation.