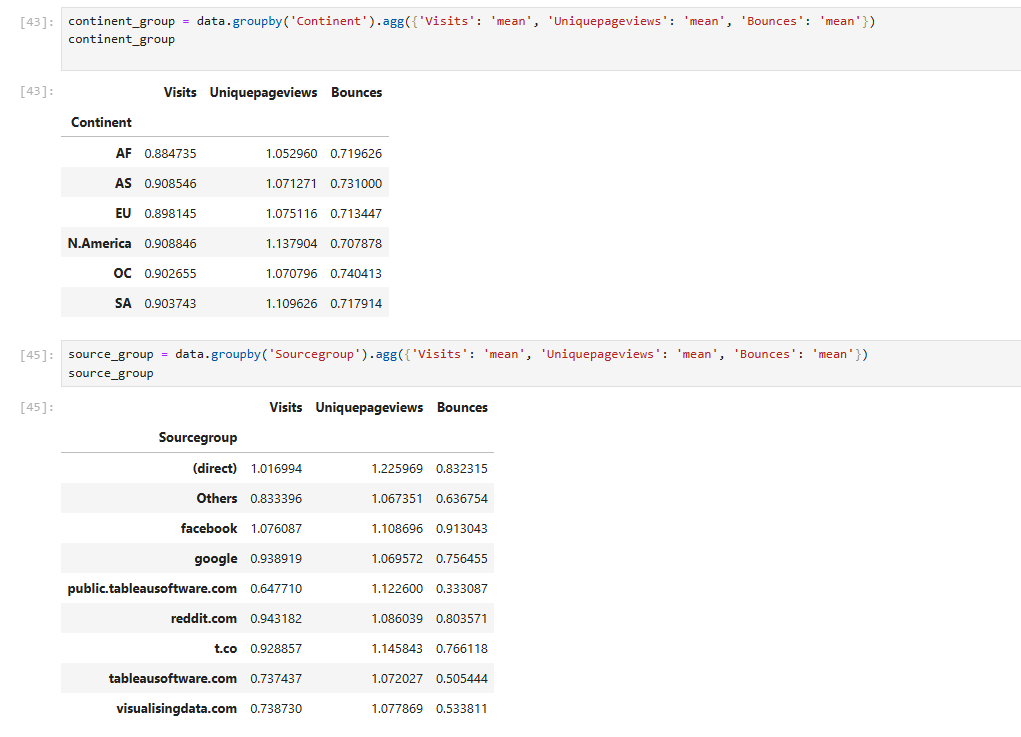
**Web Data Analysis Documentation – Pradeksha R K**

1. 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.

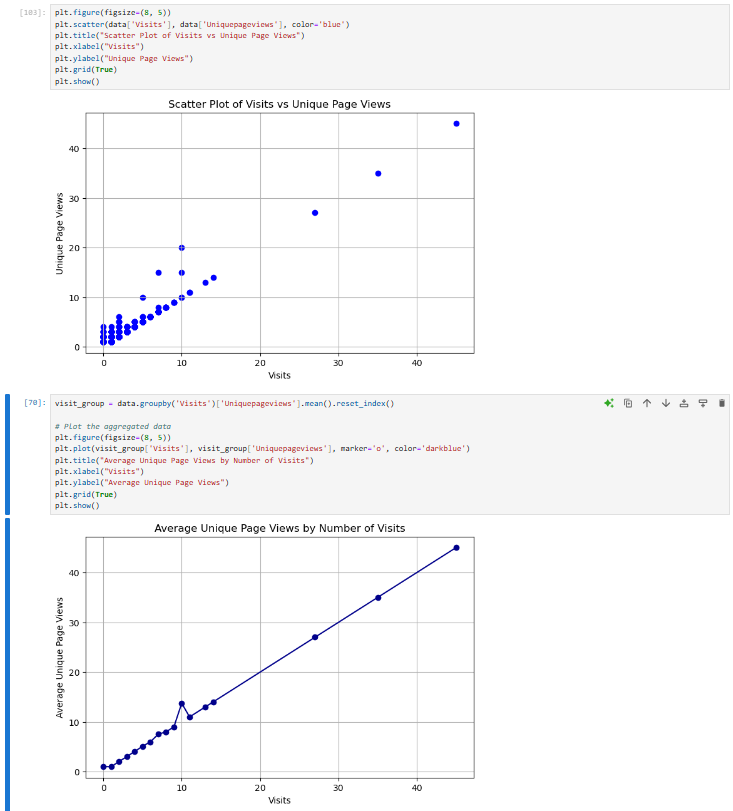
* Basic Summary of the Dataset
* Summary statistics for numeric columns
* Check for missing values
* Frequency Analysis and Plot bar plot for Continent
* Frequency Analysis and Plot bar plot for Source Group
* Group-wise Analysis for Continent and Source Group



2. 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.

* Analyze the Relationship with Correlation
* Visualize the Relationship with a Scatter Plot
* Scatter Plot with Aggregation

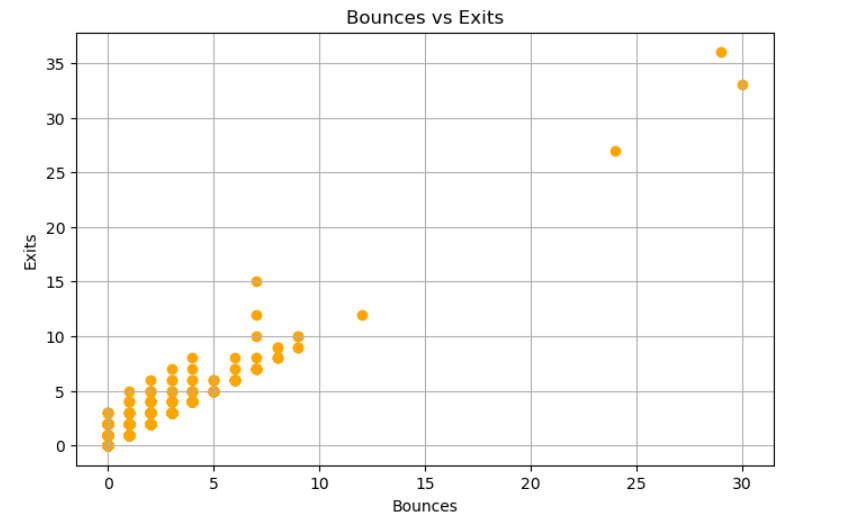
**Solution** : Yes the unique page view value depends on visits



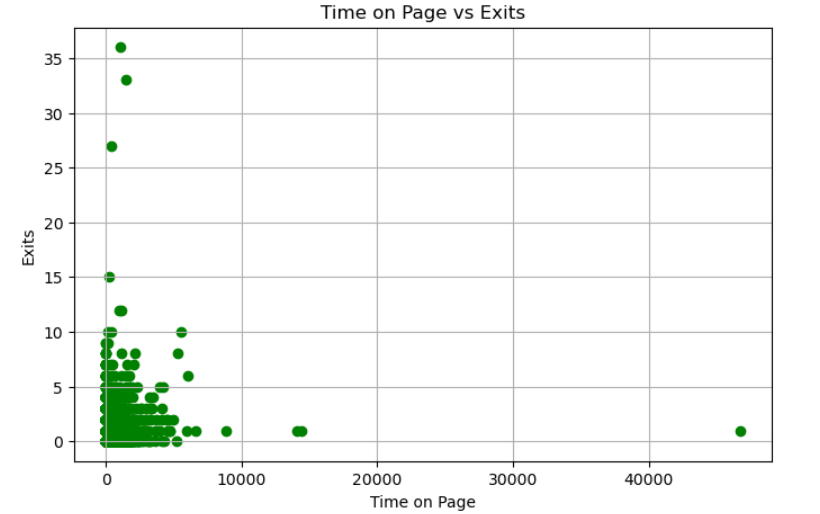
3. Find out the probable factors from the dataset, which could affect the exits. Exit Page Analysis is usually required to get an idea about why a user leaves the website for a session and moves on to another one. Please keep in mind that exits should not be confused with bounces.

**Probable Factors Affecting Exits**

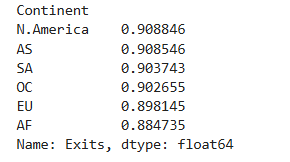
1. **Bounces**:
   * There is a clear positive correlation between Bounces and Exits, as shown in the scatter plot. The higher the Bounces, the more likely it is that users will Exit the website. This indicates that if users are not engaging with other pages on the site and instead bounce, they are more likely to leave altogether.



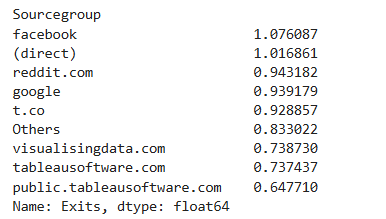
1. **Time on Page**:
   * From the scatter plot of Time on Page versus Exits, it appears that users with shorter Time on Page tend to have higher Exits. While there are a few points with long time spent on the page and high exits, the majority of high-exit data points occur when Time on Page is relatively low. This suggests that users may exit the site due to a lack of engaging or relevant content, as they are not spending much time on the pages before leaving.



1. **Continent**:
   * The data shows that certain continents, such as North America and Asia, have higher average exits, while Africa has the lowest. This suggests that users from North America and Asia may have higher exit tendencies, potentially due to differences in content relevance or language preferences.



1. **Source Group**:
   * Users accessing the site via sources like "facebook" and "direct" have higher average exit rates compared to other sources. This insight could guide marketing strategies to focus more on sources with lower exit rates or improve content targeting for high-exit sources.



4. 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.

* Correlation between Visits and Time on Page: 0.0666
* Correlation between Unique Pageviews and Time on Page: 0.1146
* Correlation between Bounces and Time on Page: -0.1091

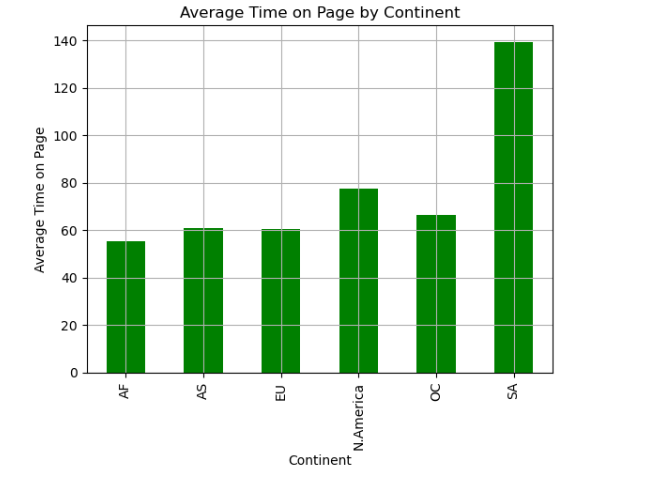
None of the variables—Visits, Unique Pageviews, or Bounces—shows a strong correlation with Time on Page, as all correlation values are close to zero. This implies that these variables do not have effect on Time on Page.

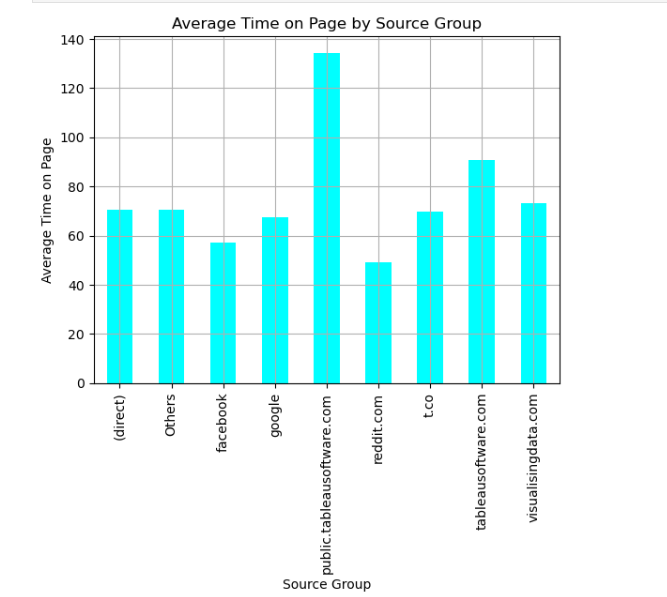
* Continent:

we can see that visitors from South America spend the longest time on the page (139.42 seconds), followed by North America (77.56 seconds) and Oceania (66.24 seconds). This suggests that the Continent of the visitor could be a factor impacting the time spent on the page. Visitors from certain regions may find the content more relevant or engaging.

* Source Group:

visitors coming from public.tableausoftware.com have the highest average time on page (134.32 seconds), followed by those from tableausoftware.com and visualisingdata.com. In contrast, visitors from social media platforms like facebook and reddit.com tend to spend less time on the page, possibly due to lower engagement or a quicker browsing style.





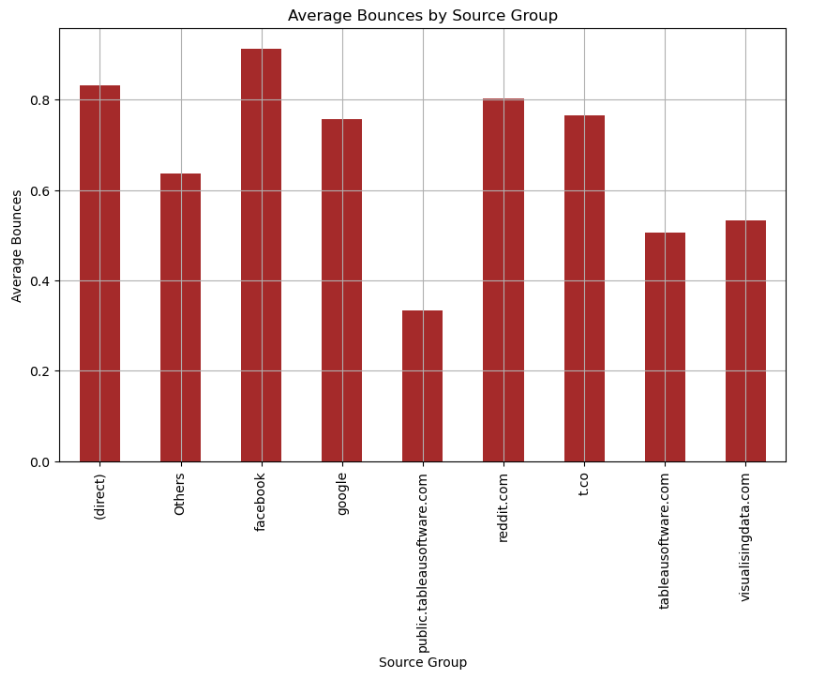
**Solution** : So Continent and source group are the potential variables affecting the time in page. While bounce rate is not a strong indicator ,a lower bounce rate may suggest higher engagement, possibly leading to a longer time on page.

5. 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.

**Key Factors Impacting Bounce Rate**

* **Source Group**:

Social media platforms (like Facebook and Reddit) have the highest bounce rates, while specialized sources (like public.tableausoftware.com) have lower bounce rates. This suggests that visitors from sources closely aligned with the website’s content are less likely to bounce.



* **Visits and Unique Pageviews**:

This high positive correlation indicates that as the number of visits increases, the bounce rate also tends to increase. here is also a moderate positive correlation between unique pageviews and bounces. The weak negative correlation here suggests that pages with longer time on page experience slightly lower bounce rates. However, the effect is minimal and likely not a primary factor affecting bounces.

