**Web Data Analysis**

Coding Challenge Python

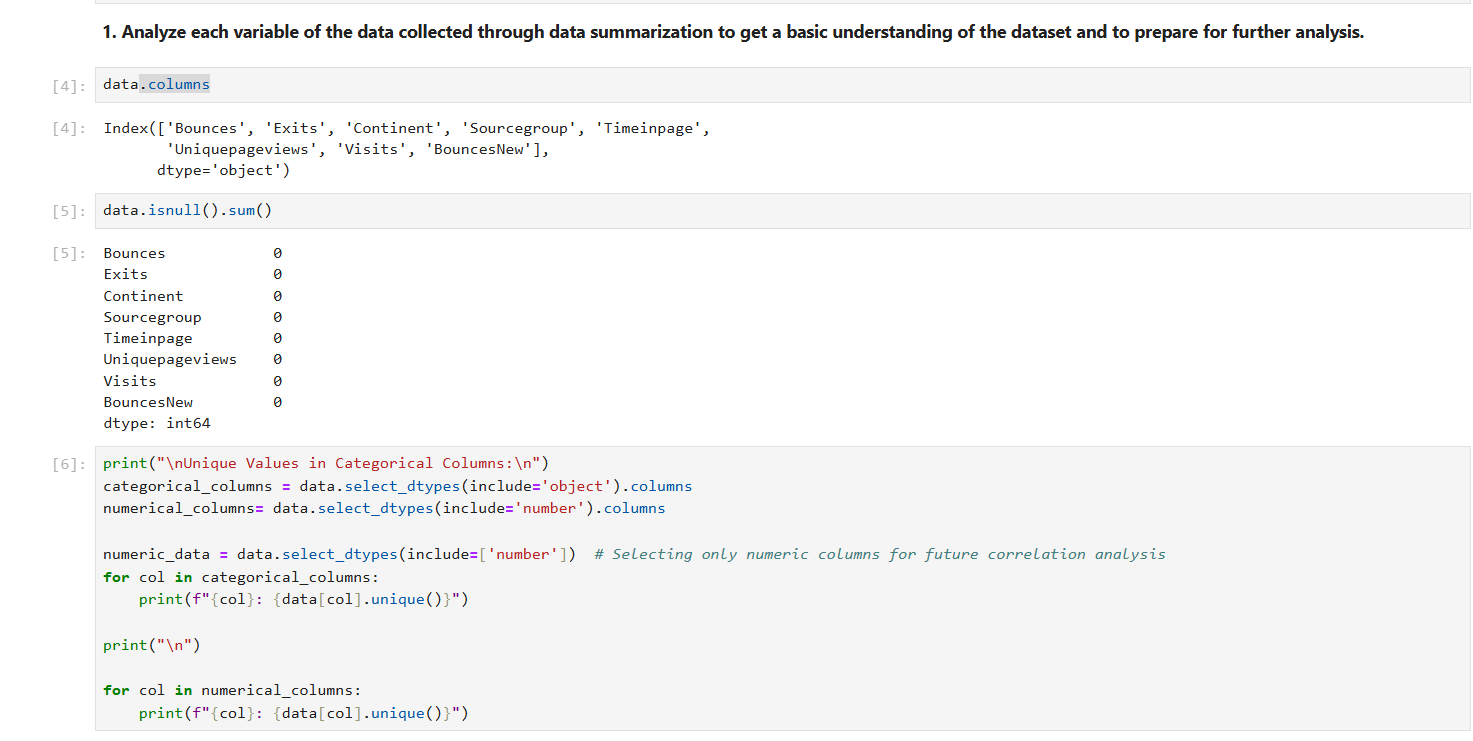
**DE128 Prit Yadav   
Batch-3**

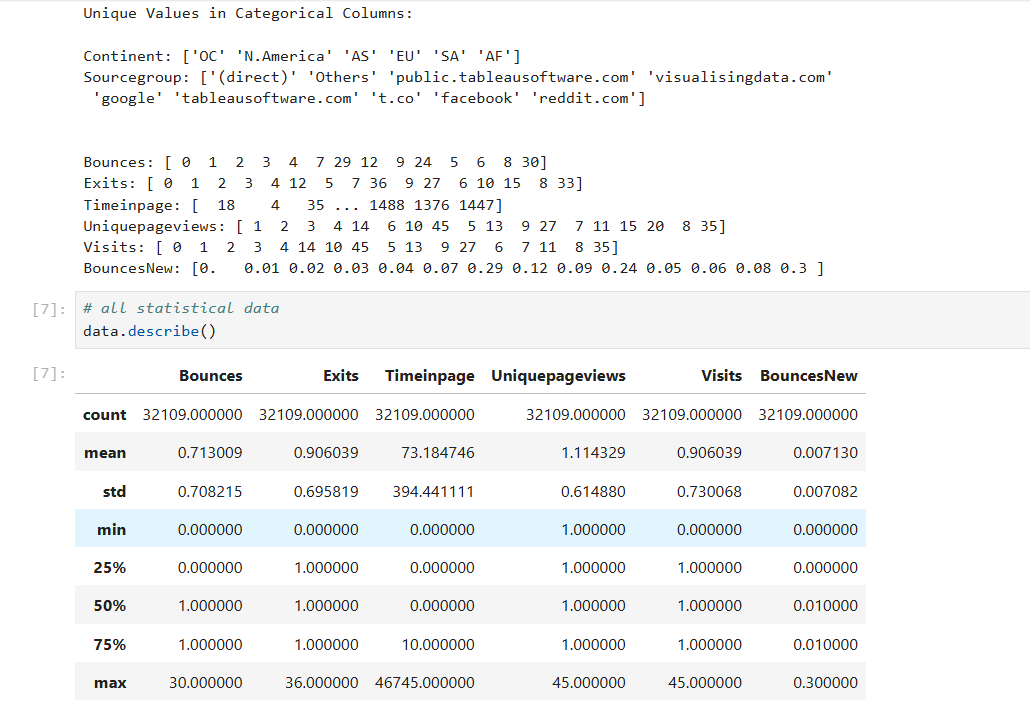
**Analysis Tasks:**

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.

1. **Analyzing the data and summarizing**



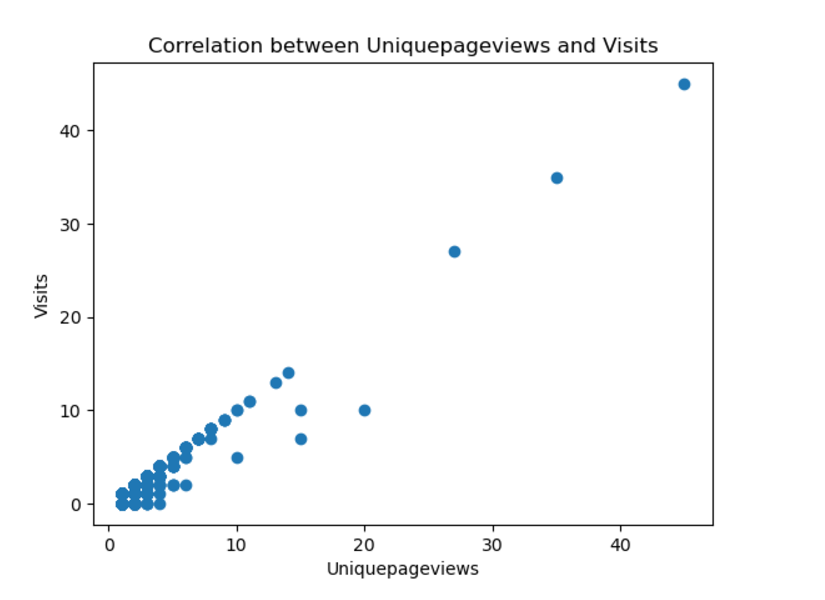


* First Import pandas library for data analysis and matplotlib for visualization
* Find all the columns, check if there is null value present or not in dataset
* Finding all unique values in categorical columns and get numerical values from dataset.
* data.describe(): to analyse all statistical data like mean , std, max , min etc.

1. **Check if Unique Page Views Depend on Visits**

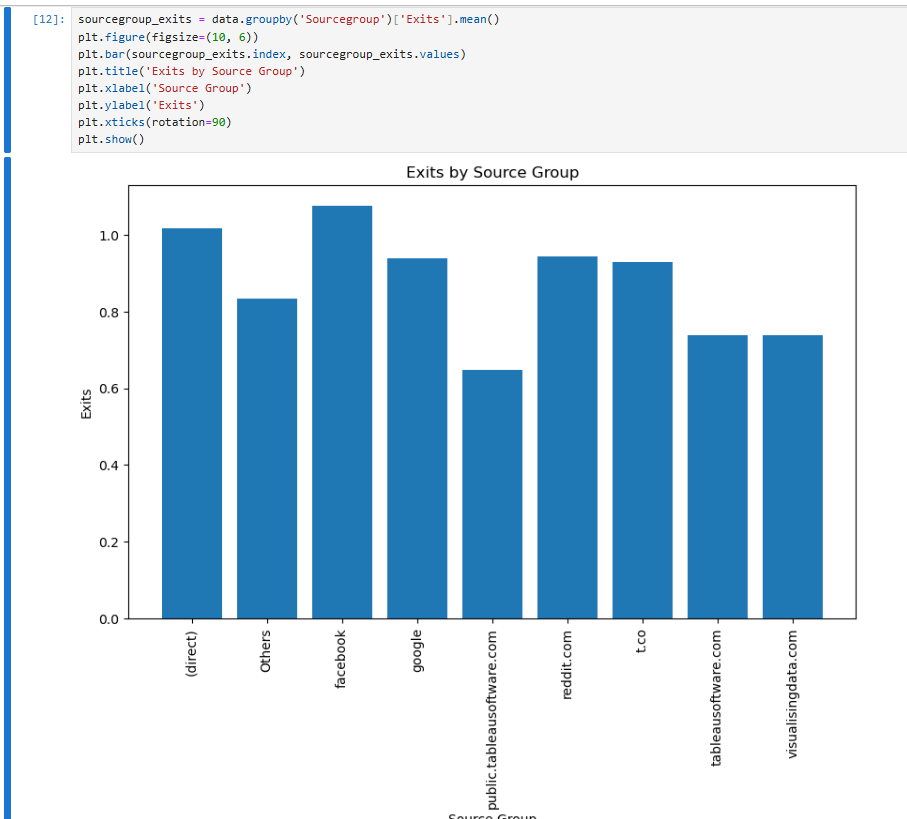
Corr function is used to find correlation matrix between variables

* **1**: Strong positive relationship (increasing together).
* **0 to 1**: Positive correlation (but weaker as it gets closer to 0).
* **0**: No correlation.
* **-1 to 0**: Negative correlation (the larger the negative value, the stronger the inverse relationship).
* **-1**: Strong negative relationship (increasing one leads to a perfect decrease in the other)

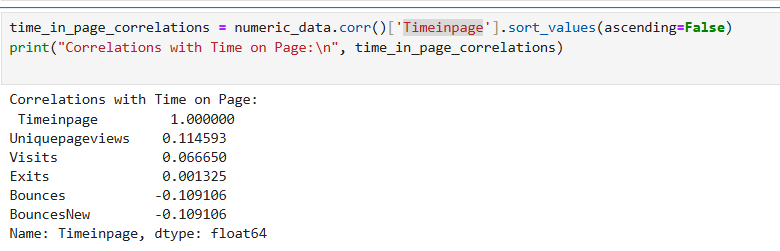
To visualize the data, scatter plot is used show the relationship  


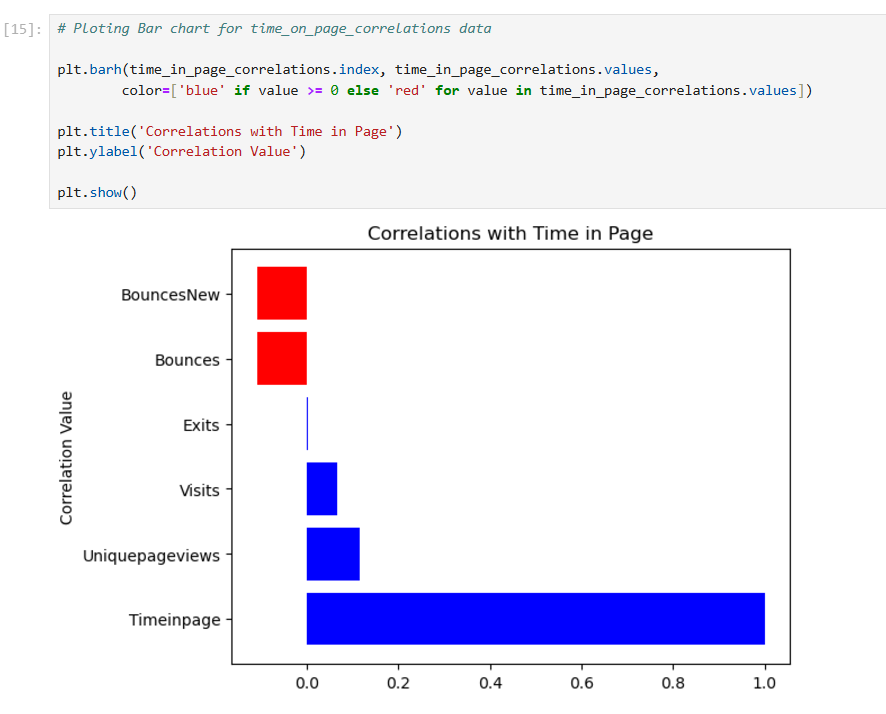
1. **Factors which could affect the Exits**

To find the factors which can affect the Exits, we first find the correlation value between Exits and other columns.  
  


Thus, we can see Bounces is strongly co-related with Exits.   
- To visualize, we can plot bar graph for source group to understand more. Group by is used to calculate total exits for every source groups  


1. **Analyzing Factors Affecting Time on Page.**

To find the factors affecting time on page, we can find all correlation value between time\_on\_page and other columns.  


From the output, we can see Uniquepageviews is strongly correlated.  
To understand more, we can plot horizontal bar chart.  


1. **Determining Factors impacting Bounce Rate**

To determine the factors impacting bounce rate, first find the correlation value between Bounce rate and other columns.  
Then plot the horizontal bar graph for visualization.  
  
