**CHAPTER FOUR**

**DATA REQUIREMENT AND ANALYSIS**

1. **Data Collection**

To achieve the study's objectives, two distinct forms of data were employed. The World Health Organization's (WHO) Global Health Observatory (GHO) data collection was used for the secondary dataset, which was drawn from the source dataset that was gathered through a survey. The main dataset examined consumer preferences and purchasing behavior with regard to alcohol use in Nigeria and the UK. A few variables that stated whether alcohol consumption is impacted by age and mood for the people in Nigeria and the United Kingdom were included in the dataset that focused on characteristics that were associated to alcohol and their buying behavior.

Given that alcohol use has been linked to a variety of health problems, including cancer, heart disease, liver disease, and mental health disorders, it is critical to analyze consumer preferences and purchasing behavior when evaluating alcohol-related concerns. By analyzing this information and spotting trends and patterns in alcohol use in Nigeria and the UK, it is possible to gauge the potential impacts of consumption and purchasing habits on health outcomes.

The World Health Organization provided the second dataset for this investigation. This dataset contains detailed information on countries, kinds of alcohol, and alcohol consumption by capital city. Alcohol use and purchasing habits are one of the key factors influencing healthy life expectancy.

1. **Datasets Description and evaluation**

The principal dataset under consideration provides information on consumer consumption and buying habits. Both Nigeria and the United Kingdom's annual income levels are covered in the data. The collection also includes data on favorite drinking establishments and retail outlets for buying alcohol. 32 columns and 215 data points make up the dataset, although not all of them are pertinent to the investigation.

The secondary dataset under review offers details on the consumption of alcohol in various nations, regions, and eras. It includes information on the four primary types of alcoholic beverages—beer, wine, spirits, and other—as well as the overall annual per-capita alcohol consumption in a particular nation. The dataset contains information about the year, the nation, and the region, all of which are coded representations.

There are 47,794 data points in the dataset and they are spread out over 23 columns. However, it is significant to highlight that several of these columns lack data and are unrelated to the study. Data from 189 distinct nations are included in the dataset, which also spans a large time span, from 1960 to 2019, and is divided into six separate areas, including the Eastern Mediterranean, Europe, Western Pacific, Americas, Africa, and South-East Asia.

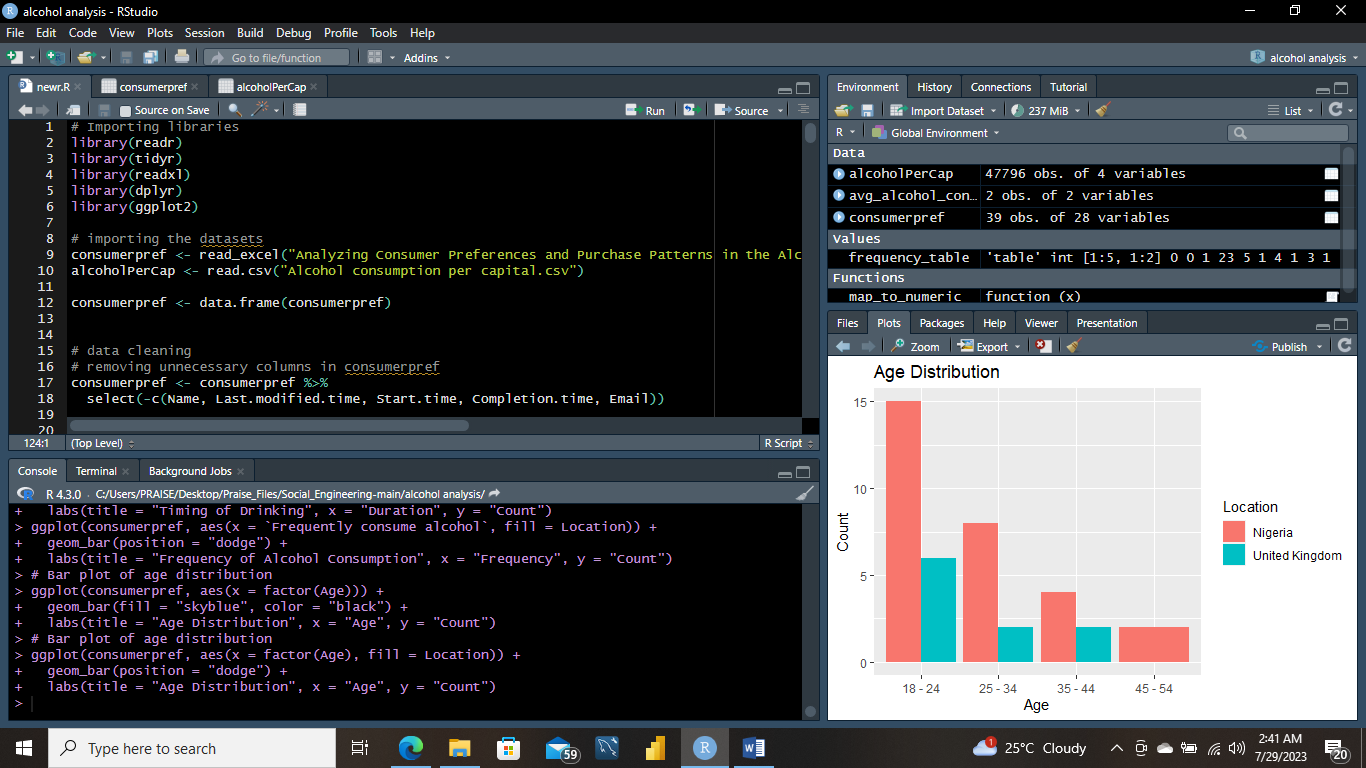
Some of features in the dataset are very important for the study. The "SpatialDimensionValueCode" denotes the name of the nation, the "DisaggregatingDimension1ValueCode" the type of alcoholic beverage, the "Value" the per-capita alcohol consumption, and the "TimeDim" the year are a few of these.

**Table 4.1:** Showing columns (attributes) present in the raw downloaded datasets

1. **DATA PREPARATION**

Data Preparation is an essential aspect in data analysis. it is the translating and arranging of raw data into a format that can be easily evaluated. It is a crucial step in the data analysis process. This procedure is essential to ensuring that the data is reliable, accurate, and consistent so that it may be used to provide insights, analyze trends and guide decisions.

Data preparation involves locating and fixing mistakes and inconsistencies in raw data, which is known as data cleaning. In order for data analysis to be effective, it is essential to make sure the data is accurate, dependable, and consistent.



Data preparation, also known as data preprocessing, is a crucial step in the data analysis process. It involves transforming raw data into a format that is suitable for analysis, modeling, and extracting meaningful insights. Proper data preparation helps in improving the accuracy and effectiveness of data analysis techniques. For this study here are the main steps involved in data preparation:

1. **DATA CLEANING**

A survey was created and data was gathered for the data preparation. This was used as the main information source. Additionally, we acquired information on alcohol from the World Health Organization that is pertinent to the investigation and data analysis. In order to clean up the data, mistakes, missing numbers, and inconsistencies have to be found and handled. For this data analysis, duplicates were removed, missing values were handled, data was transformed, and feature engineering was done.

In removing duplicates, any duplicate records in the dataset were eliminated. Data integration, which includes combining or connecting datasets, and feature extraction, which requires removing columns from the dataset that are not required for the research, all included checking for missing values and removing them.

The four primary stages of data preparation are frequently data cleaning, data integration, data transformation, and data reduction. Data cleaning is the process of identifying and correcting errors and inconsistencies in the data, such as missing values, duplicate entries, and outliers. Data integration involves combining data from several sources into a single dataset, as opposed to data transformation, which involves changing data into a format appropriate for analysis, such as normalizing or scaling the data.

Data reduction, for instance, is selecting a portion of the data that is relevant to the study by filtering or sampling the data.

Finding and repairing missing or erroneous data, getting rid of duplicate information, and dealing with outliers or abnormalities in the data are just a few of the crucial activities that are frequently included in data cleaning. This process could also involve fixing any inconsistencies or inaccuracies in the data as well as standardizing data formats, including converting dates or numbers into a standardized format.

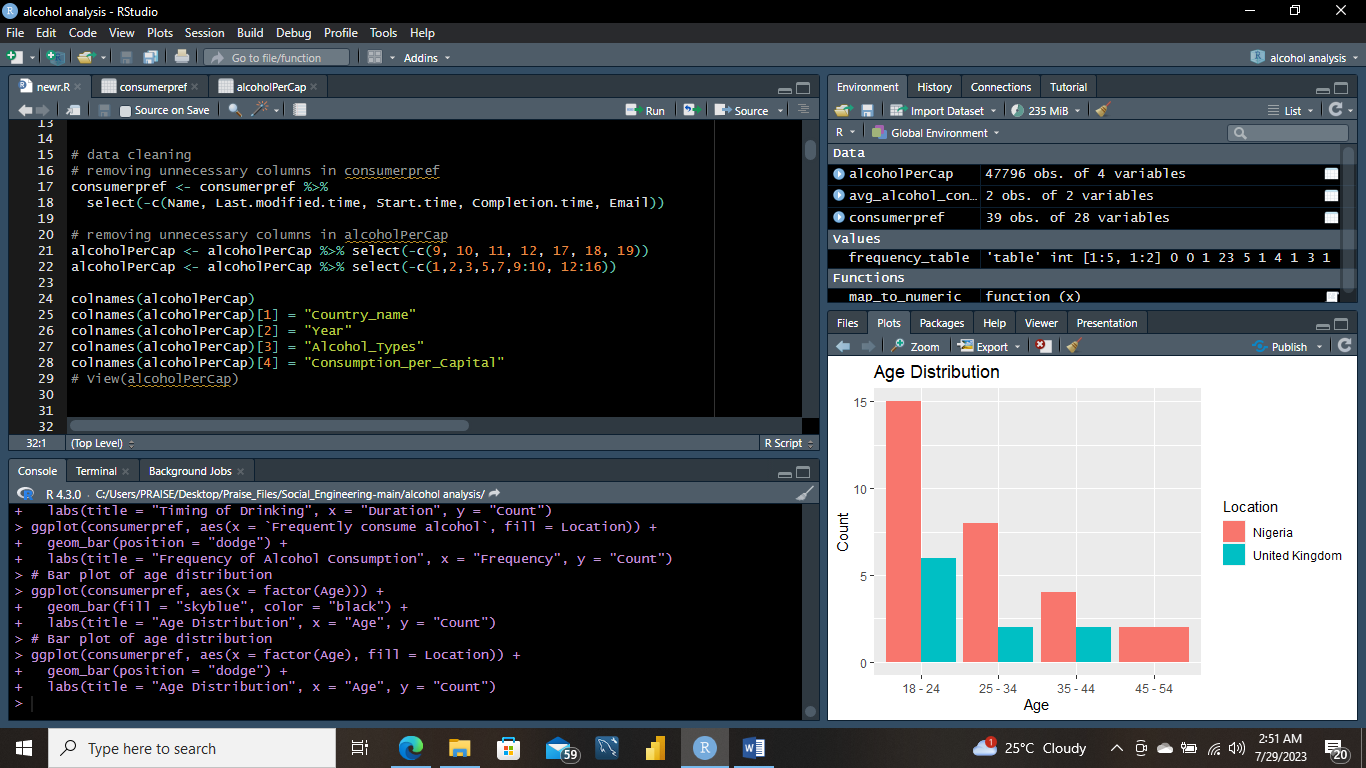
Effective data cleaning requires a methodical approach and attention to detail since mistakes and inconsistencies in the data can significantly affect the results of data analysis. By meticulously cleaning and preparing the data before analysis, analysts may make sure that their findings are accurate, reliable, and pertinent. Additionally, they can prevent any biases or mistakes that can develop as a result of using stale or contradictory data.

For the purposes of this investigation, the following data cleaning processes were taken into account:

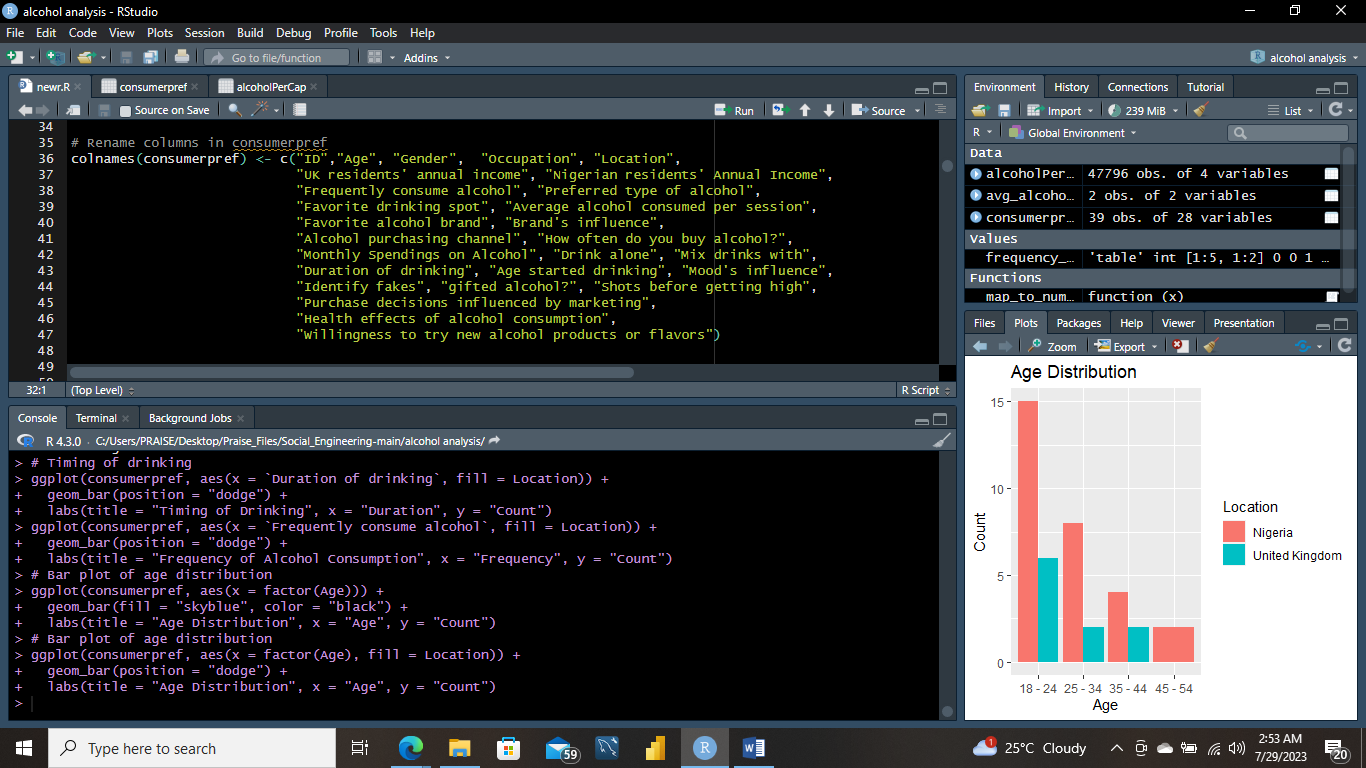
1. Removal of unneeded columns
2. Renaming columns
3. Handling inconsistences in naming convention
4. Identifying and correcting missing values
5. Converting coded names to actual names
6. Removing duplicates
7. Addressing outliers and anomaly

**Removal of unneeded columns**

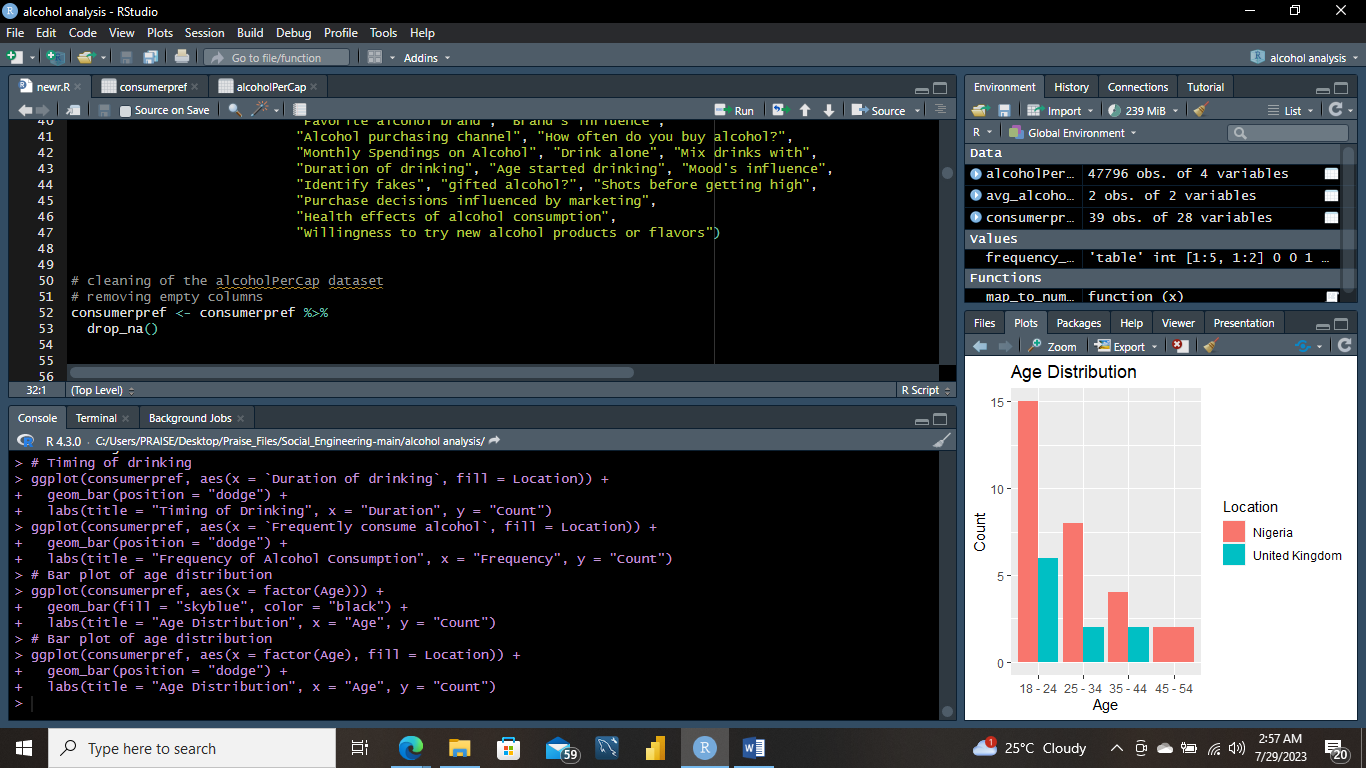
Both datasets undergo a number of data cleaning techniques to eliminate extraneous columns and rename columns to facilitate analysis.



**Renaming columns and inconsistency in naming convention**



**Identifying and correcting missing values**

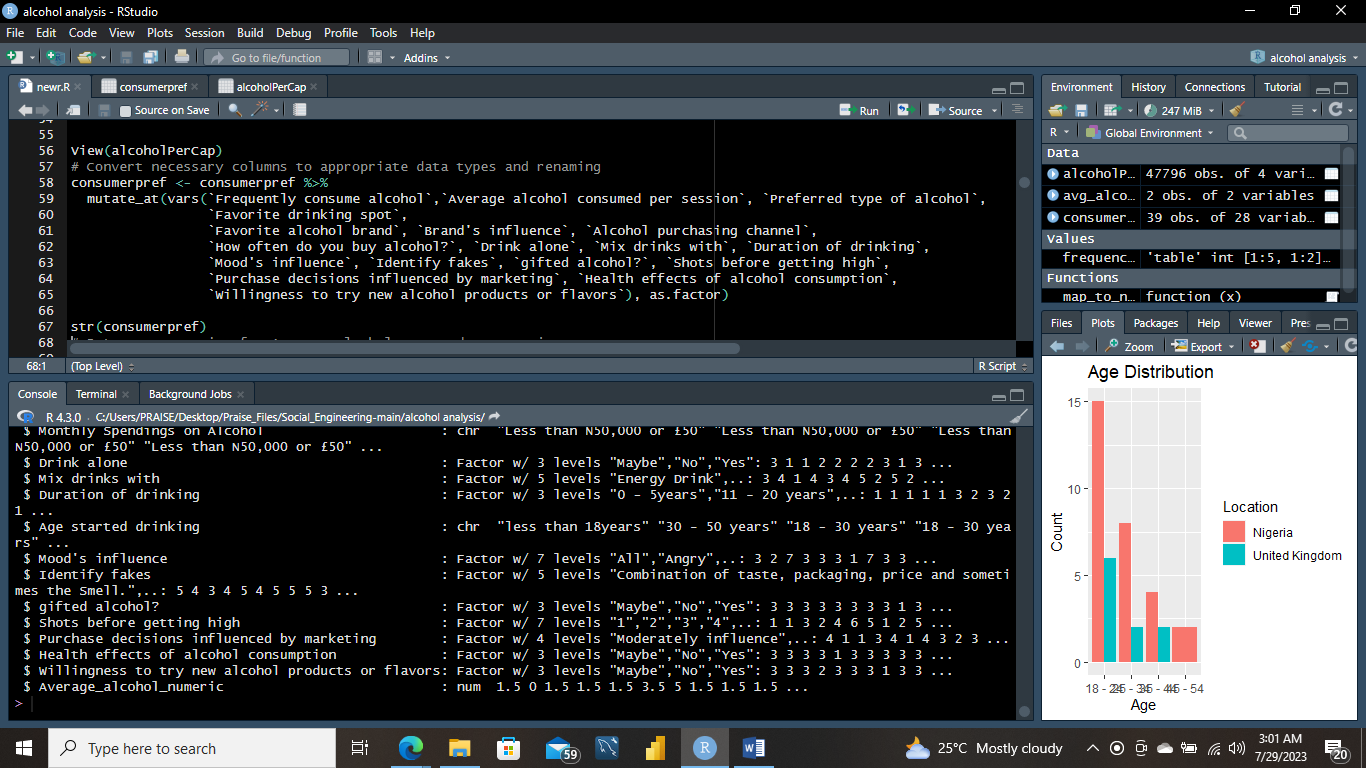


Alcohol consumption is a prevalent behavior worldwide, influenced by various factors such as culture, economics, and demographics. In this report, we analyze data from two datasets, "Analyzing Consumer Preferences and Purchase Patterns in the Alcohol Industry(1-215).xlsx" and "Alcohol consumption per capital," to explore specific patterns and preferences related to alcohol consumption in Nigeria and the United Kingdom (UK). We aim to gain insights into the influence of socio-cultural, economic, and demographic factors on alcohol consumption in these countries.

* + 1. **DATA PREPROCESSING**

After the import of the datasets, extraneous columns were deleted to concentrate on important data. To make categorical variables suitable for further analysis, we normalized column names and transformed them into the necessary data types. Following that, the datasets were cleaned and made ready for statistical analysis and visualization.

The categorical variables in the consumerpref dataset are changed to factors in some columns, according to the relevant data types.

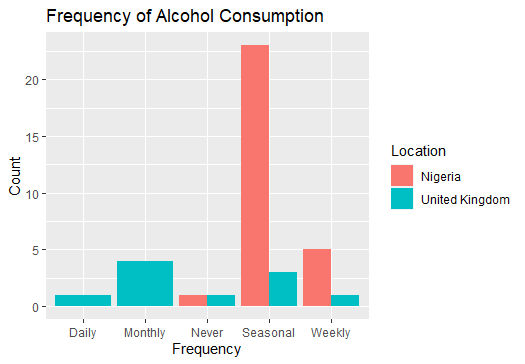


**4.3.3 DATA VISUALIZATION**

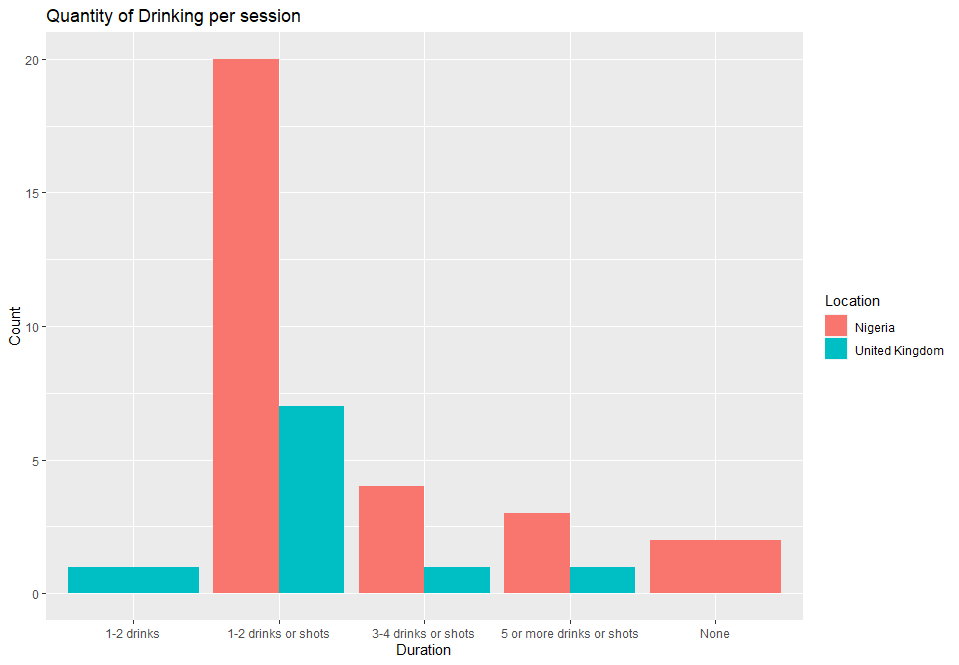
The frequency of alcohol drinking in Nigeria and the UK was our first area of investigation. The numbers of respondents from each place who regularly use alcohol were shown in the findings. We also looked at how much alcohol was typically drunk each session in each nation. To better comprehend when to drink, the length of drinking was also shown.

To study the data and obtain insights, several visualizations are made. Bar graphs and a trend plot are two examples of the visualizations. Considering the goals of this study:

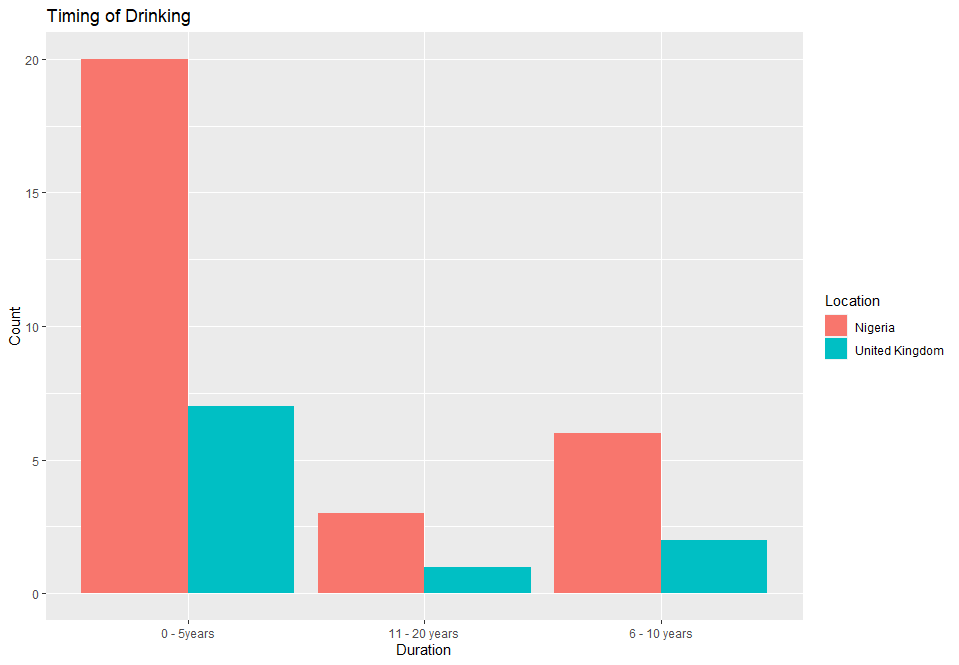
1. To identify and analyze the patterns of alcohol consumption in Nigeria and the UK, focusing on frequency, quantity, and timing of alcohol intake.



In terms of frequency, this shows that alcohol consumption in Nigeria is more seasonal than monthly as it the opposite for the United Kingdom where in the analysis showed that the residents in the United Kingdom take alcohol monthly than the rest.

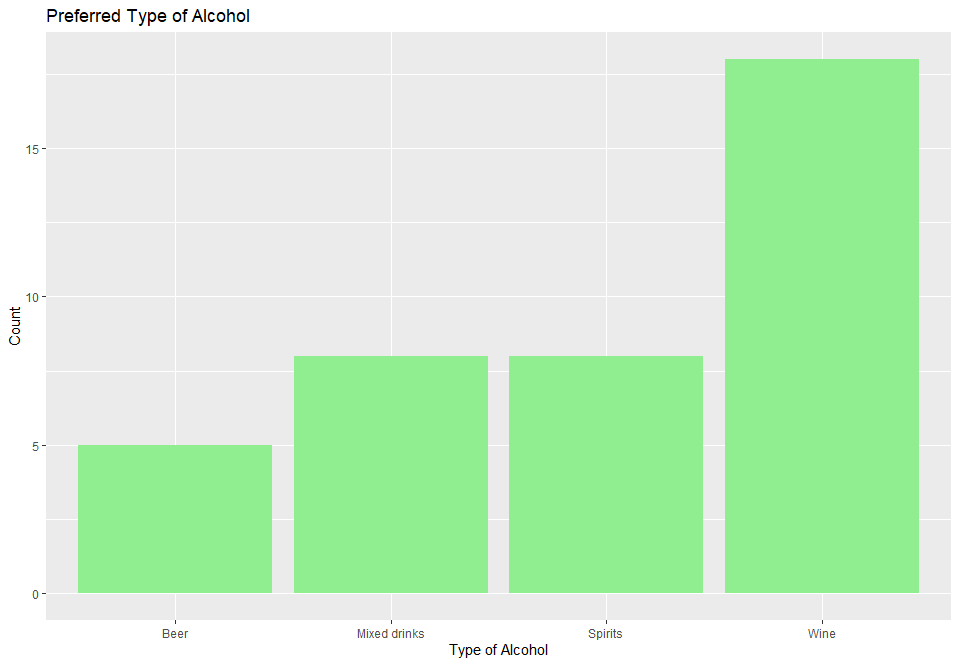


In terms of quantity of alcohol consumed, this shows that more individuals consume 1-2 drinks or shots in both the United States and Nigeria.

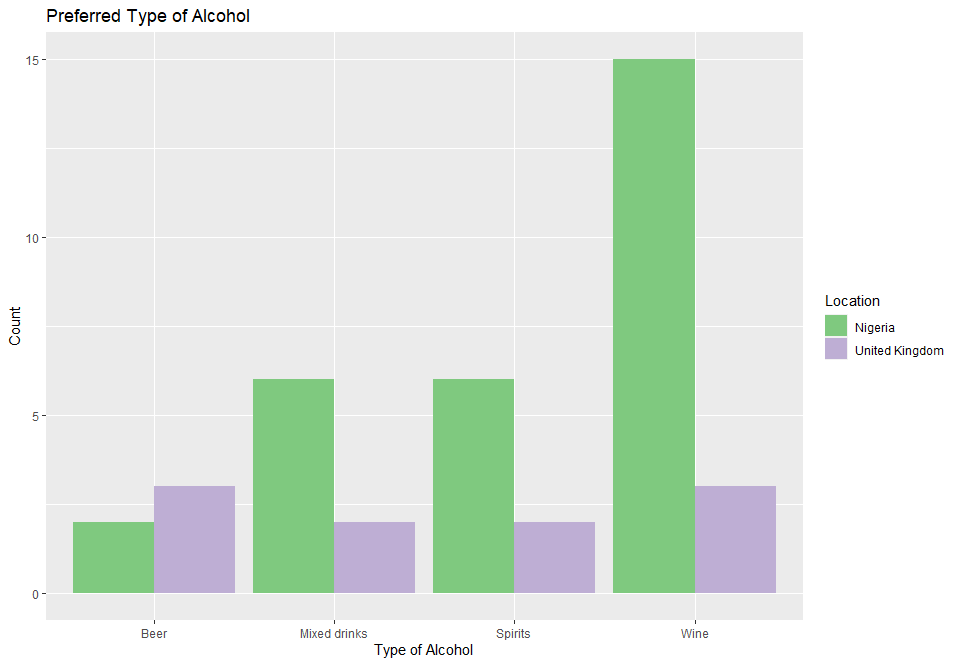


This shows that more individuals have been drinking for 0-5 years than 6-20 years.

1. To investigate the types of alcoholic beverages preferred in both countries and the reasons behind these preferences.



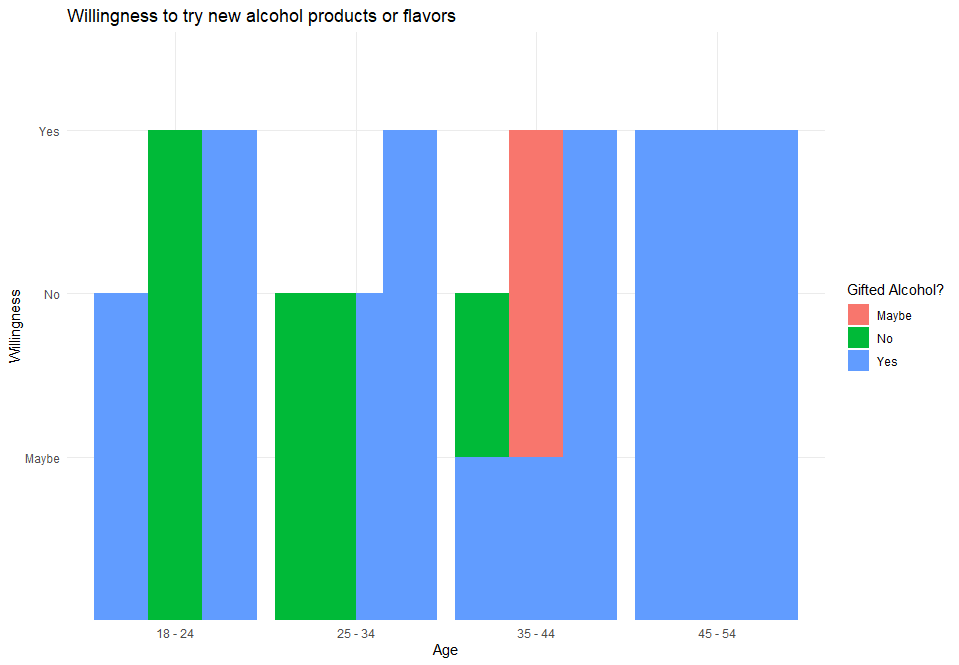
This shows that residents in both countries – Nigeria and the United States prefer wine than Beer, Mixed drinks and Spirits.



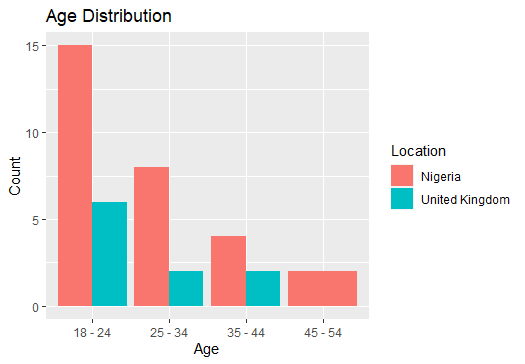
This shows that residents in Nigeria, more individuals prefer wines to beers, mixed drinks and spirits while in the United Kingdom, the residents relatively prefer wine and beer to mixed drinks and spirits.

1. To examine the socio-cultural factors, such as traditions, customs, and societal norms, that influence the choice of alcoholic beverages and consumption patterns in Nigeria and the UK.

Social and cultural norms, traditions, and conventions affected people's alcohol use choices and habits were examined. The favored types of alcohol and location were compared using a cross-tabulation technique. The results were subjected to a chi-square test for association.



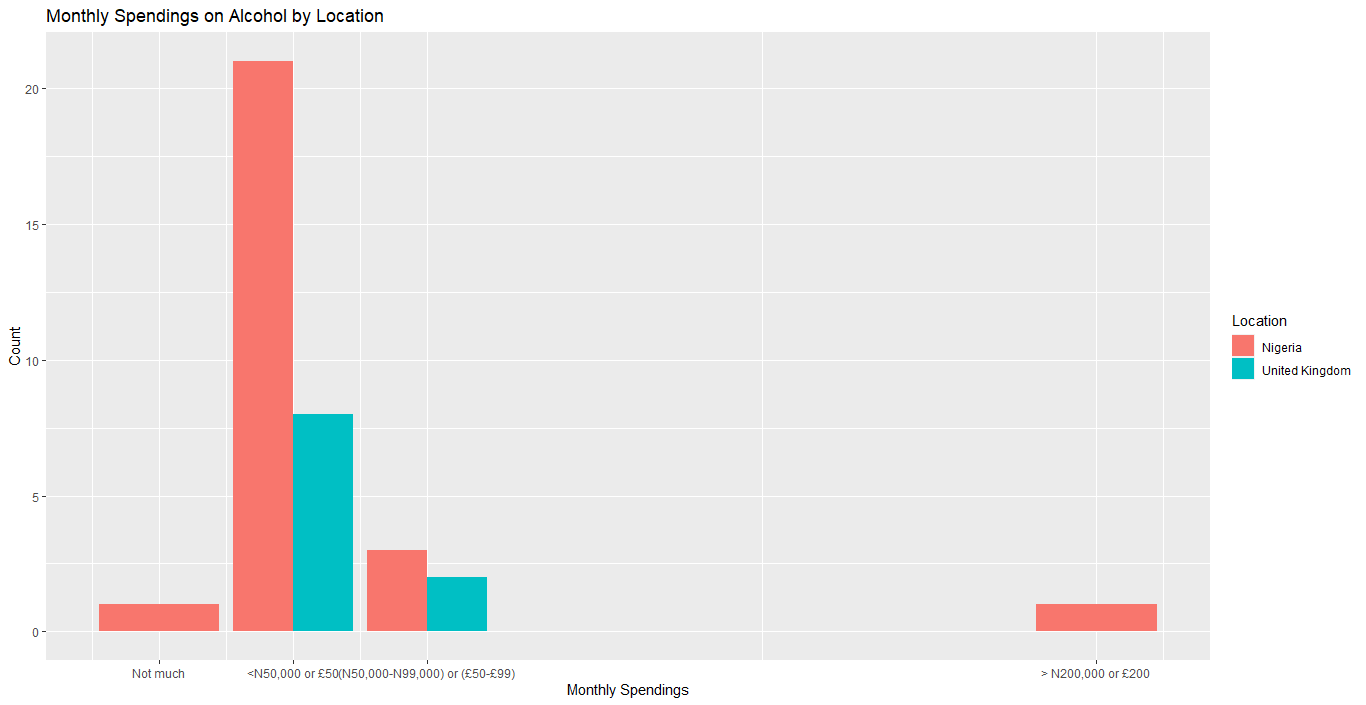
This shows that individuals gifted alcohol has a higher probability to taking more alcohol than individuals that haven’t been gifted alcohol. The age range of 45-54 and 18-24 has more willingness to try new flavors of alcohol.



This shows that the Nigeria have more alcohol consumption for age distribution ranging from 18 to 24 years old than the United Kingdom.

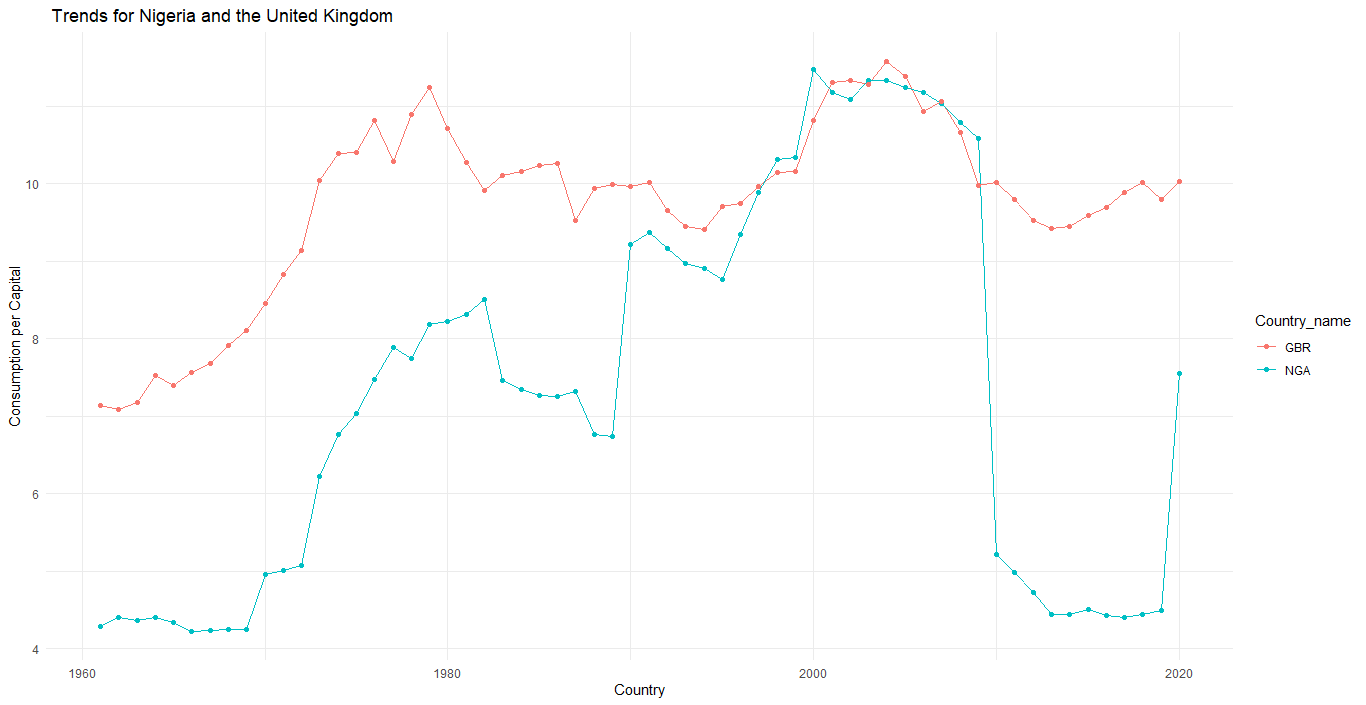
1. To explore the role of economic factors, such as income levels, pricing, and availability, in shaping the choice of alcoholic beverages and consumption patterns in both countries.

Economic considerations, such as income levels, cost, and accessibility, can have a big influence on how much alcohol people drink. We used a side-by-side bar plot to compare the distribution of the "Monthly Spendings on Alcohol" variable in the two nations in order to comprehend this.



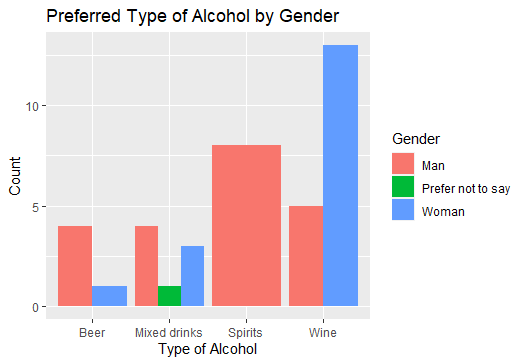
This shows that the more Nigeria residents and United Kingdom residents spend N50,000 or £50 on alcohol monthly, with Nigeria residents topping the chart.

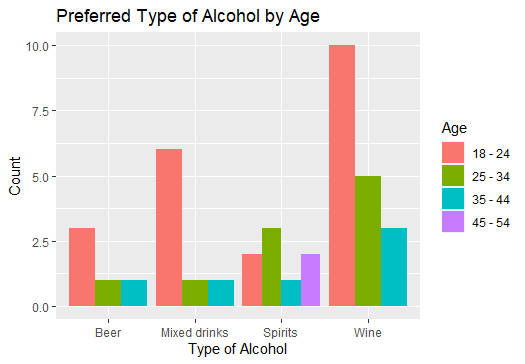
Data on annual per capita alcohol consumption in Nigeria and the UK were extracted and combined. To see how the patterns of alcohol use have changed over time in both nations, a trend map was created.

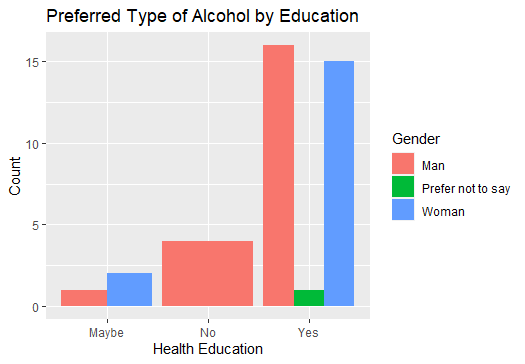


1. To understand the impact of demographic factors, such as age, gender, and education, on alcohol preference and consumption patterns in Nigeria and the UK.

Alcohol preferences and consumption habits can also be influenced by demographic characteristics including age and gender. To depict the preferred alcoholic beverages depending on gender and age categories in Nigeria and the UK, bar plots were developed.





We may utilize the consumerpref dataset to assess the preferred kind of alcohol by type. Each respondent's favourite kind of alcohol is listed in the column favourite type of alcohol. To see the distribution, we may count the different types of alcohol and make a bar chart.

We can apply the consumerpref Variable to assess the preferred kind of alcohol based on education level. Each respondent's chosen alcohol type is listed in the chosen Type of Alcohol column, while their educational background is listed in the Education column. To see the distribution, we may count the number of people who prefer each sort of alcohol for each educational level and make a bar chart to represent the results.

**CONCLUSION**

Important information about the variables that affect people's decisions regarding alcohol intake after performing an extensive data study of consumer preferences and consumption trends in Nigeria and in the UK. The consumer preference' dataset, which includes crucial data on respondents' age, gender, employment, geography, income, alcohol preferences, and other pertinent characteristics, was used to conduct the analysis.

The study yields numerous important conclusions. First off, we noticed some intriguing patterns in the quantity and timing of alcohol use. The results showed that a sizable proportion of respondents in both Nigeria and the UK claimed to routinely drink alcohol, with different preferences for the length of drinking sessions. For stakeholders in the alcohol business and politicians, knowing consumer behavior is essential to developing specialized marketing plans.

The most popular alcoholic beverage categories in Nigeria and the UK were also investigated. The investigation revealed differences in the favored types of alcohol between the two nations. This study has significance for beverage producers and marketers that want to appeal to certain geographical preferences and tastes.

The influence of sociocultural influences on alcohol preferences has also been studied. The data analysis produced intriguing findings on the impact of society norms, traditions, and customs on the selection of alcoholic drinks. Having an understanding of these cultural quirks might make it easier for businesses to operate in other markets.

Additionally, it was shown that economic factors including income levels, pricing, and availability had a significant influence on alcohol use habits. The research of monthly expenditures on alcohol provided insight into the disparities between consumers' spending patterns in Nigeria and the UK. Businesses looking to adjust their pricing and advertising methods to the economic climate of each nation may find this information to be useful.

Age, gender, and education were revealed to be major demographic variables that influence alcohol preferences and consumption patterns. Different age groups and educational levels' alcohol preferences were shown to follow various patterns, according to the research. This information might be useful in creating innovative products and targeted marketing strategies.

As a result, this data analysis offers insightful information on the many factors that affect consumer preferences and alcohol use in Nigeria and the UK. The results emphasize how crucial it is to comprehend cultural, economic, and demographic factors in order to properly customize products and marketing methods. These insights may be used as a roadmap for decision-makers as the alcohol business develops to help them make wise and responsible decisions and make sure their goods are in line with customer tastes and demands. However, it is crucial to note the data and analysis's shortcomings, such as possible biases and missing factors, which should be taken into account in next studies and decision-making procedures.