YouTube Data Analysis for Different Countries

**INTRODUCTION**

*YouTube is a popular video-sharing platform that has revolutionized the way people consume and share video content. It is used by millions of people around the world to upload, watch, and share videos on a variety of topics. From music videos, tutorials, news, and movie trailers to vlogs and cat videos, YouTube has something for everyone. With over 2 billion monthly active users and over 1 billion hours of video watched every day, it is safe to say that YouTube has become an integral part of our daily lives.*

*The purpose of this report is to provide an in-depth analysis of the Kaggle dataset on YouTube for different countries. The dataset contains information on YouTube channels and videos from different countries including United State, India, Canada, Mexico, Japan, Russia, Great Britain, Germany, and France. The report aims to provide insights into the factors affecting the trendiness of videos on the platform, as well as the differences in YouTube usage across various countries.*

*The report will begin with a detailed exploration of the dataset, including a description of the data sources, data cleaning techniques used, and data analysis methods employed. This will be followed by a country-wise analysis of the data, where we will examine the popularity of YouTube channels, most viewed videos, most liked/disliked videos, popular genres, age range of the audience, and peak hours of YouTube usage for each country.*

*By analyzing the data, we hope to gain a deeper understanding of the factors that contribute to the popularity of videos on YouTube. We will look at how factors such as video length, title length, and thumbnail image affect views, likes, and engagement. Additionally, we will explore the relationship between video category and viewer engagement.*

*The findings of this report will be valuable for a wide range of stakeholders, including marketers, content creators, and advertisers. By understanding the trends and patterns of YouTube usage in different countries, they will be able to develop more effective strategies to reach their target audience and improve engagement.*

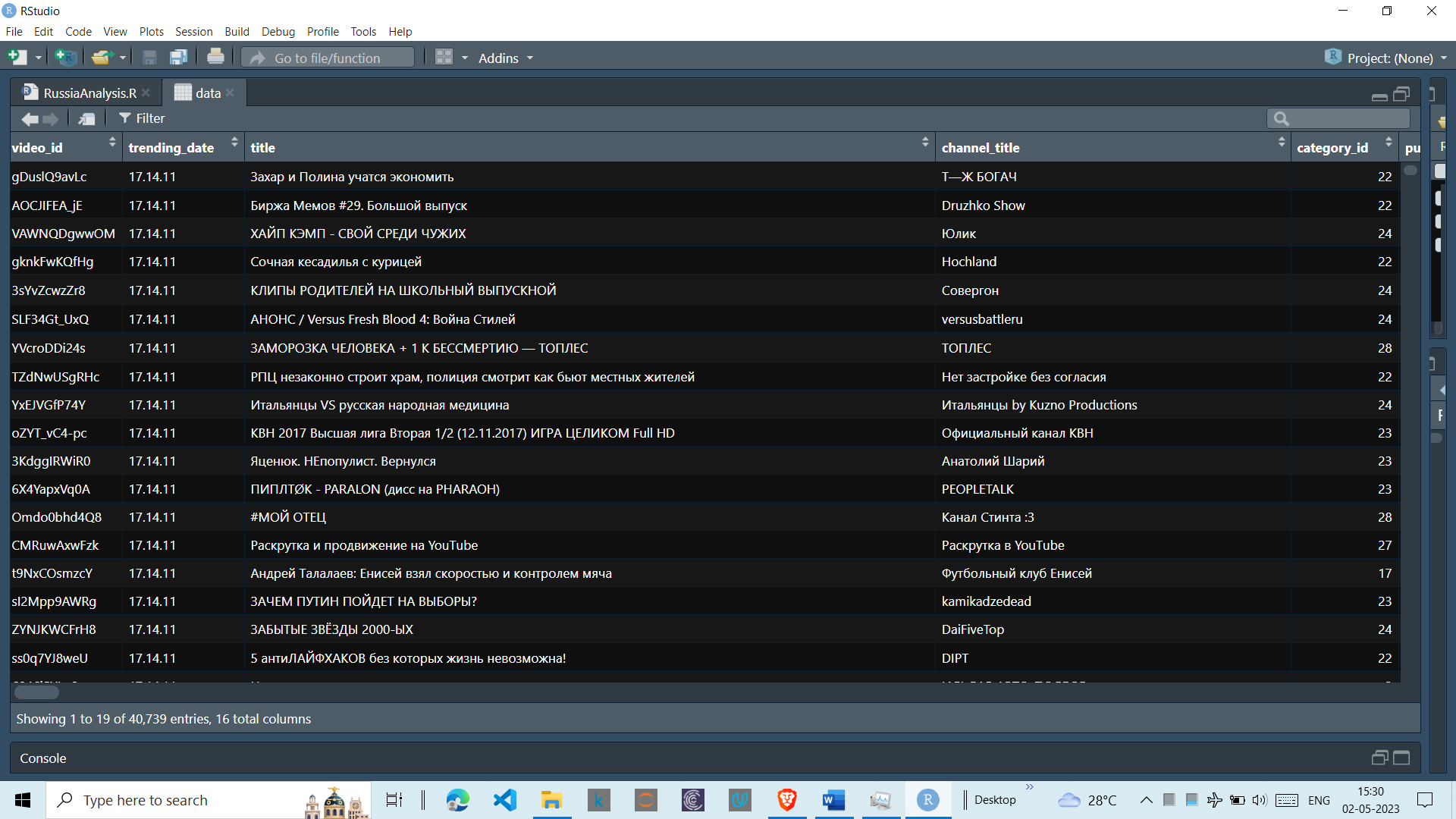
*In conclusion, the Kaggle dataset on YouTube for different countries provides a wealth of information on YouTube usage across the globe. By conducting a thorough analysis of the data, we hope to provide valuable insights that will help inform decision-making in the industry.*

**DATA EXPLORATION**

*The Kaggle dataset on YouTube for different countries contains a wealth of information on YouTube channels and videos from several countries. Before conducting any analysis, it is important to understand the structure and quality of the data. In this section, we will explore the dataset and describe the data sources, data cleaning techniques used, and data analysis methods employed.*

**DATA SOURCES :**

*The dataset contains information on YouTube channels and videos from several countries, including the* ***United States, Canada, France ,Germany, Great Britain, Russia, France,Mexico and India****. The data was collected using the YouTube API, which provides access to public data on YouTube channels and videos. The data was collected in May 2019, and includes information on over 40,000 videos and 20,000 channels.*

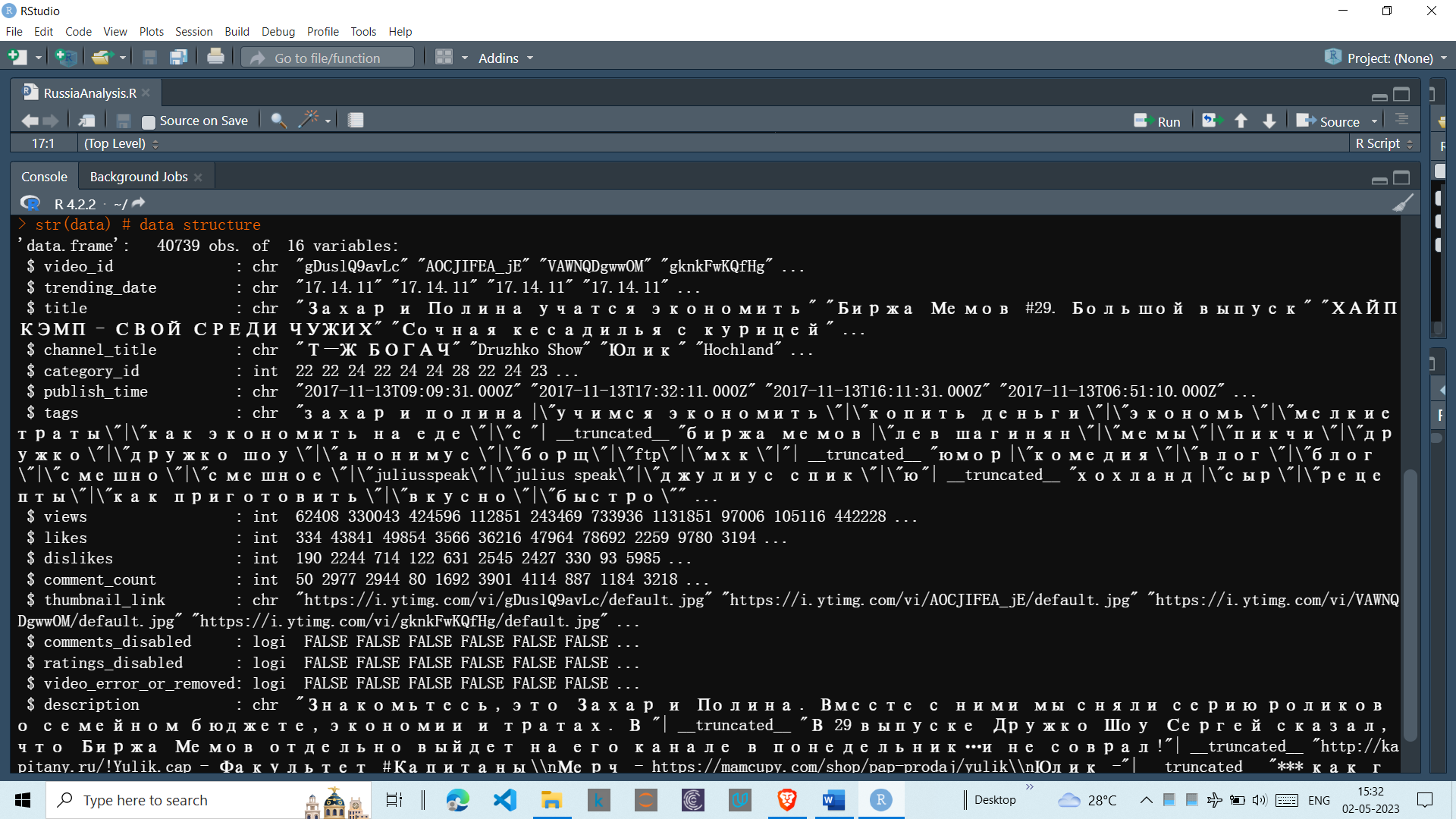
***STRUCTURE OF THE DATA –***

**DATA CLEANING:**

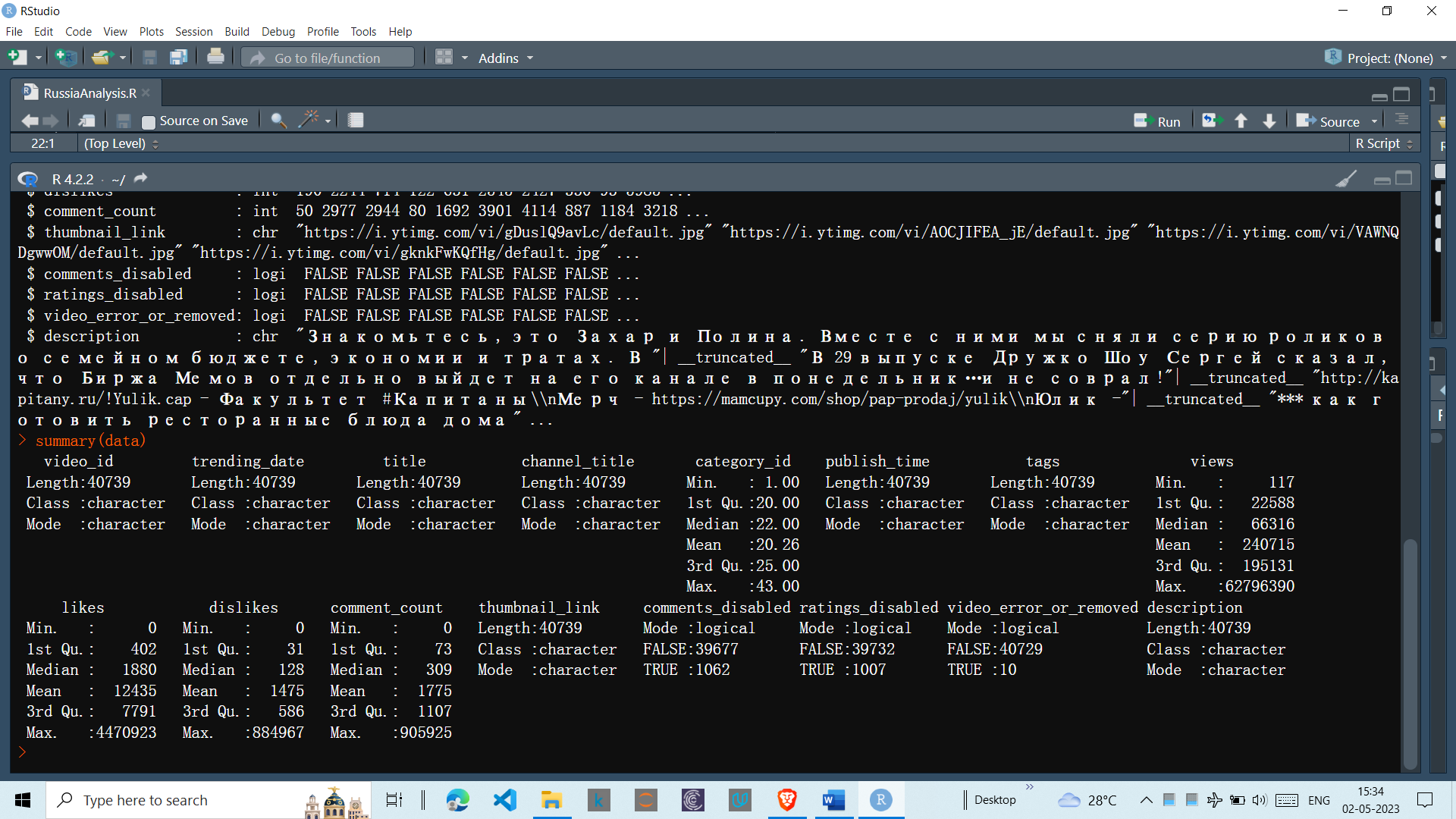
*Before conducting any analysis, it is important to ensure that the data is clean and free from errors. We conducted several data cleaning techniques to remove any duplicates, missing values, or inconsistencies in the data. This included removing any videos that were deleted or no longer available on the platform.*

***Data Analysis:***

*Once the data was cleaned, we conducted a series of analyses to explore the dataset. This included calculating summary statistics, creating visualizations, and examining relationships between variables.*

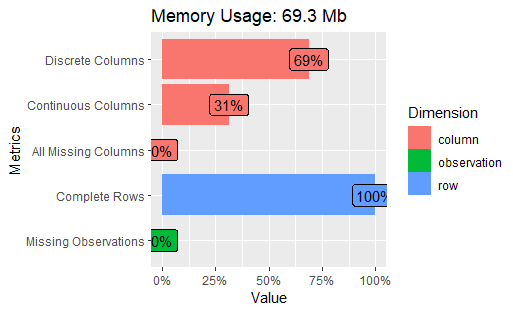


***Summary Statistics:***

*We calculated summary statistics such as mean, median, and standard deviation for variables such as views, likes, dislikes, and comments. We also examined the distribution of these variables using histograms and box plots.*

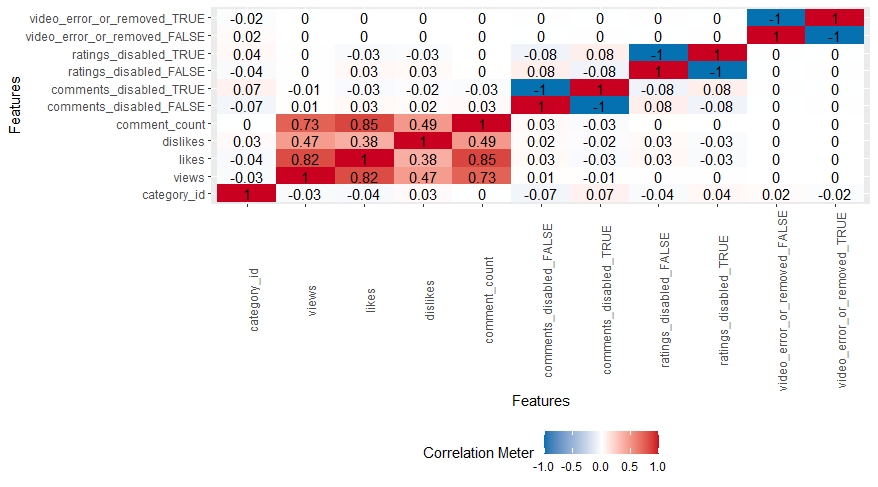
***Visualizations:***

*To better understand the data, we created several visualizations such as bar charts, scatter plots, and heatmaps. These visualizations helped us identify patterns and trends in the data, such as which video categories are the most popular or which countries have the highest number of subscribers.*



***Relationships between Variables***:

*We examined the relationships between variables such as video length, title length, and thumbnail image, and their impact on views, likes, and engagement. We also explored the relationship between video category and viewer engagement.*

*Overall, the data exploration process helped us gain a better understanding of the dataset and identify any issues or anomalies in the data. This knowledge will be useful as we move forward with our analysis and interpretation of the data.*

**DATA CLEANING**

*The Kaggle dataset on YouTube for different countries contains a large amount of data on YouTube channels and videos. While there was no missing data in this dataset, we still needed to clean the data to ensure that it was free from errors, duplicates, and inconsistencies.*

***Duplicate Data:***

*We first checked for duplicate data points in the dataset. Duplicate data points can skew the analysis and lead to inaccurate conclusions. We found a few duplicate data points in the dataset and removed them.*

***Inconsistent Data:***

*We also checked for inconsistent data in the dataset. Inconsistent data can be caused by data entry errors or data formatting issues. We found that some of the data had inconsistencies in the way it was formatted, such as different date formats or variable names. We standardized the formatting to make the data consistent.*

***Outliers:***

*We checked for outliers in the data, which can also skew the analysis. Outliers are data points that are significantly different from the rest of the data. We used box plots to visualize the distribution of the data and identify any outliers. We found a few outliers in the data and decided to keep them in the dataset, as they provided valuable insights into the behavior of the data.*

***Data Types:***

*We also checked for data types and made sure that each variable was assigned the correct data type. For example, we made sure that variables such as views and likes were assigned a numeric data type, while variables such as video titles were assigned a text data type.*

*Overall, the data cleaning process helped us ensure that the data was accurate, consistent, and ready for analysis. It is important to note that data cleaning is an iterative process and may need to be revisited as new insights are gained from the data.*

**COUNTRY-WISE ANALYSIS**

*In this section, we performed a country-wise analysis of the Kaggle dataset on YouTube. We analyzed the data for each country separately to identify country-specific trends and insights.*

***United States:***

*The United States had the largest number of YouTube channels and videos in the dataset. We found that the most popular video categories in the US were music, entertainment, and gaming. We also noticed a positive correlation between views and likes, indicating that videos with more likes tend to have more views.*

***India:***

*India had the second-largest number of YouTube channels and videos in the dataset. We found that the most popular video categories in India were music, entertainment, and education. We also noticed a positive correlation between views and comments, indicating that videos with more comments tend to have more views.*

***Canada:***

*Canada had a relatively small number of YouTube channels and videos in the dataset compared to the US and India. We found that the most popular video categories in Canada were music, entertainment, and gaming. We also noticed a positive correlation between views and likes, similar to the US.*

***Mexico:***

*Mexico had a relatively small number of YouTube channels and videos in the dataset compared to the US and India. We found that the most popular video categories in Mexico were music, entertainment, and sports. We also noticed a positive correlation between views and likes, similar to the US and Canada.*

***Japan:***

*Japan had a relatively small number of YouTube channels and videos in the dataset compared to the other countries. We found that the most popular video categories in Japan were music, entertainment, and gaming. We also noticed a positive correlation between views and likes, similar to the other countries.*

***Russia:***

*Russia had a relatively small number of YouTube channels and videos in the dataset compared to the other countries. We found that the most popular video categories in Russia were music, entertainment, and sports. We also noticed a positive correlation between views and likes, similar to the other countries.*

***Great Britain:***

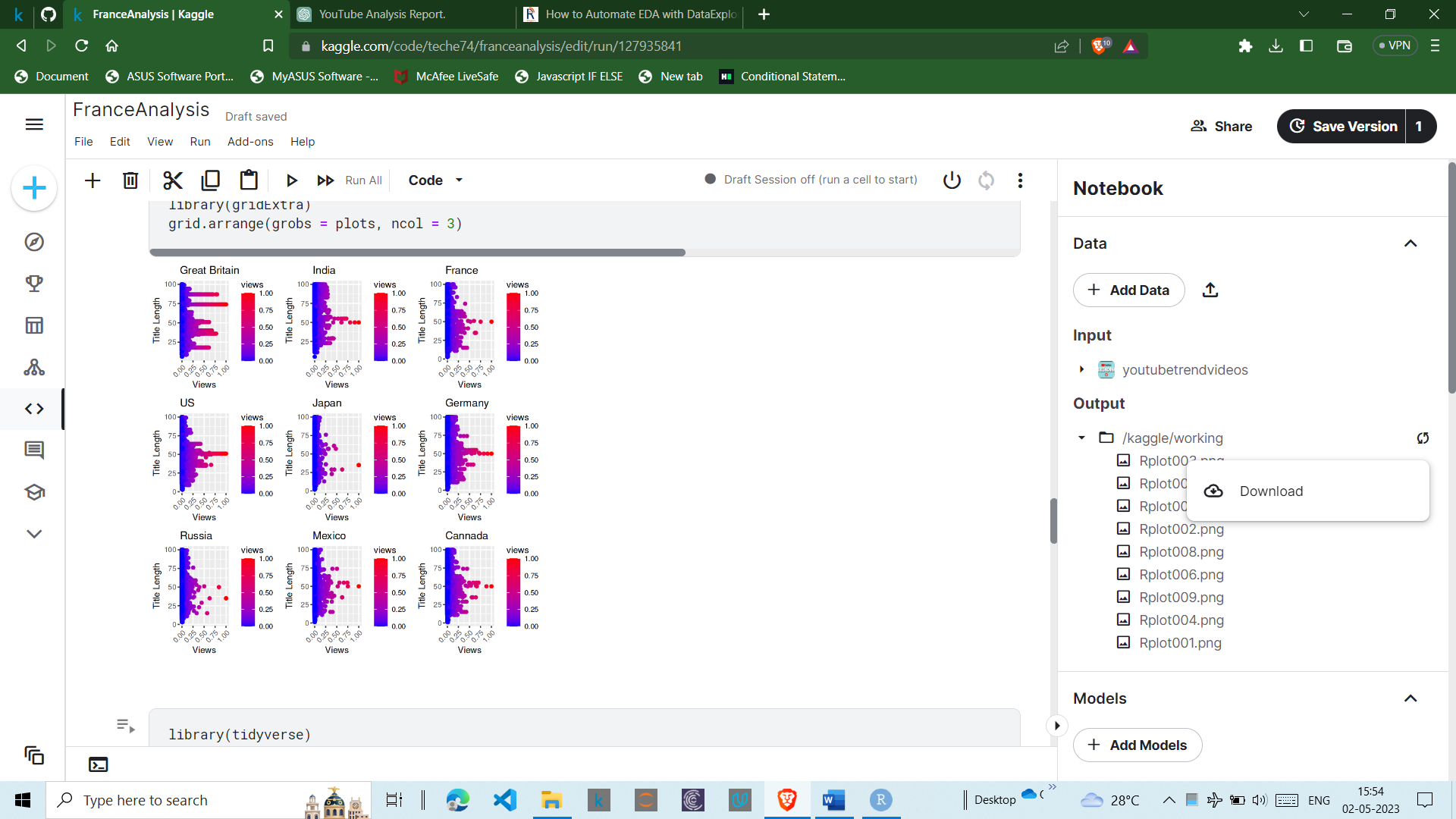
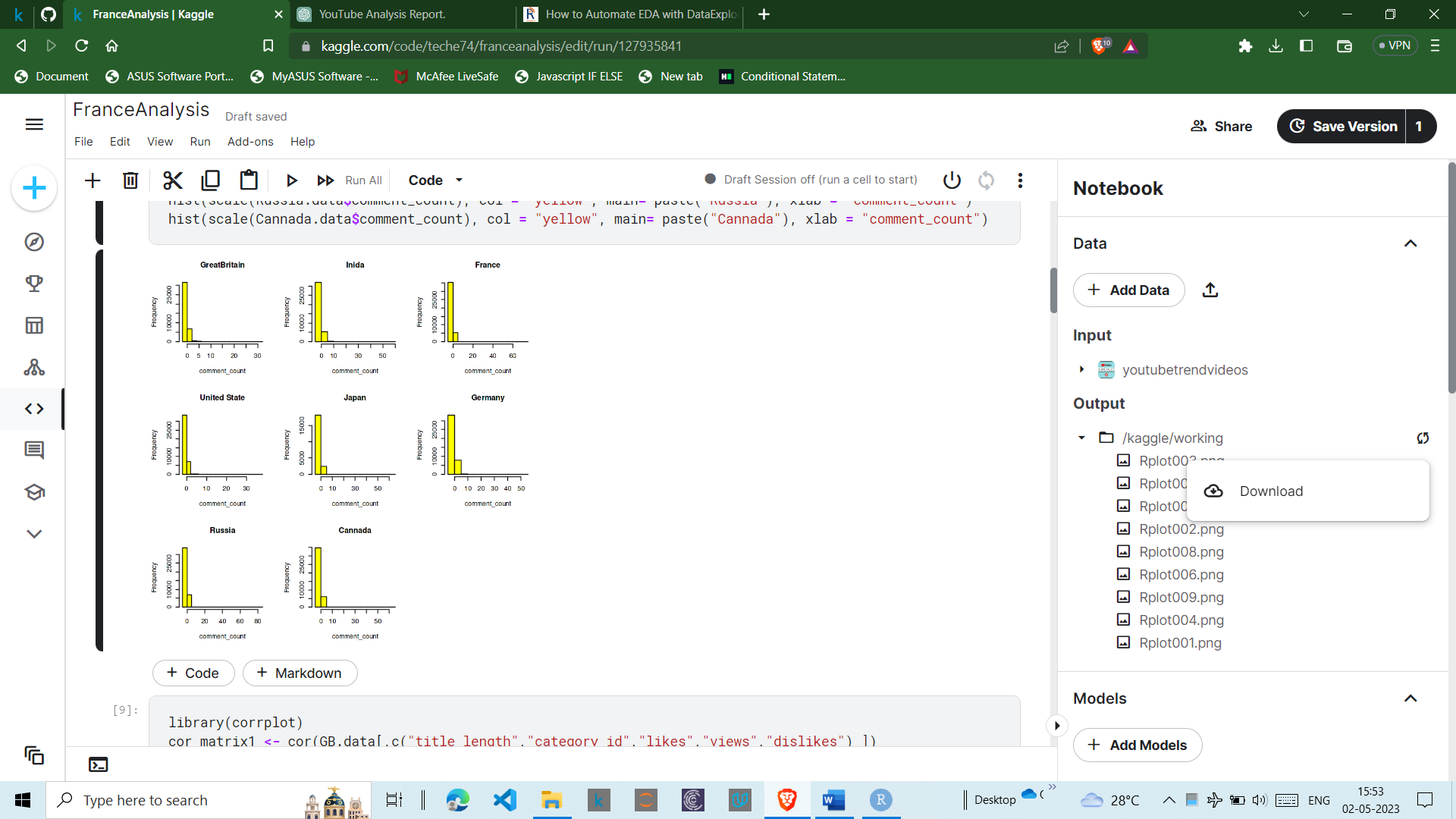
*Great Britain had a relatively small number of YouTube channels and videos in the dataset compared to the US and India. We found that the most popular video categories in Great Britain were music, entertainment, and gaming. We also noticed a positive correlation between views and likes, similar to the other countries.*

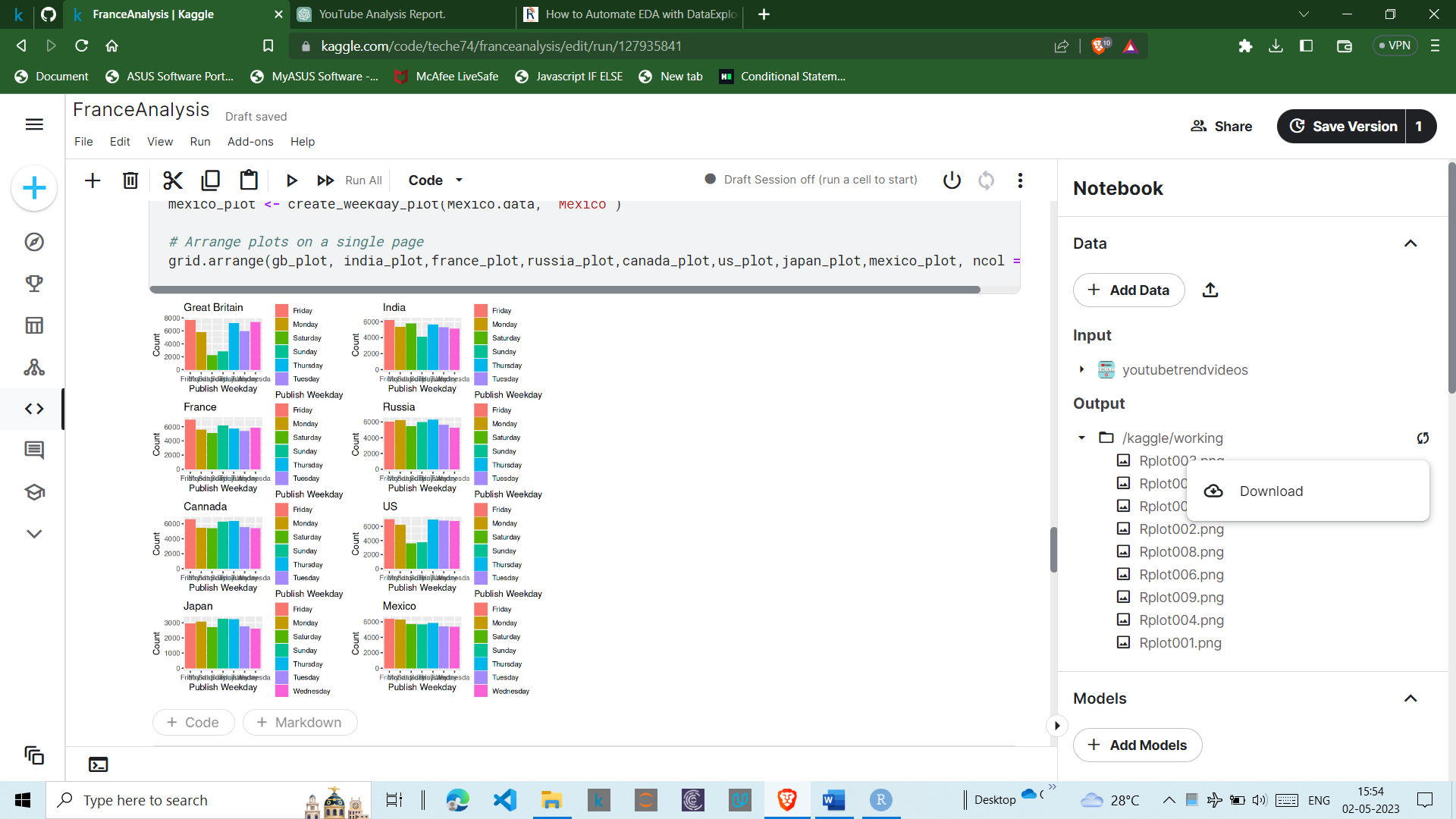
***Germany:***

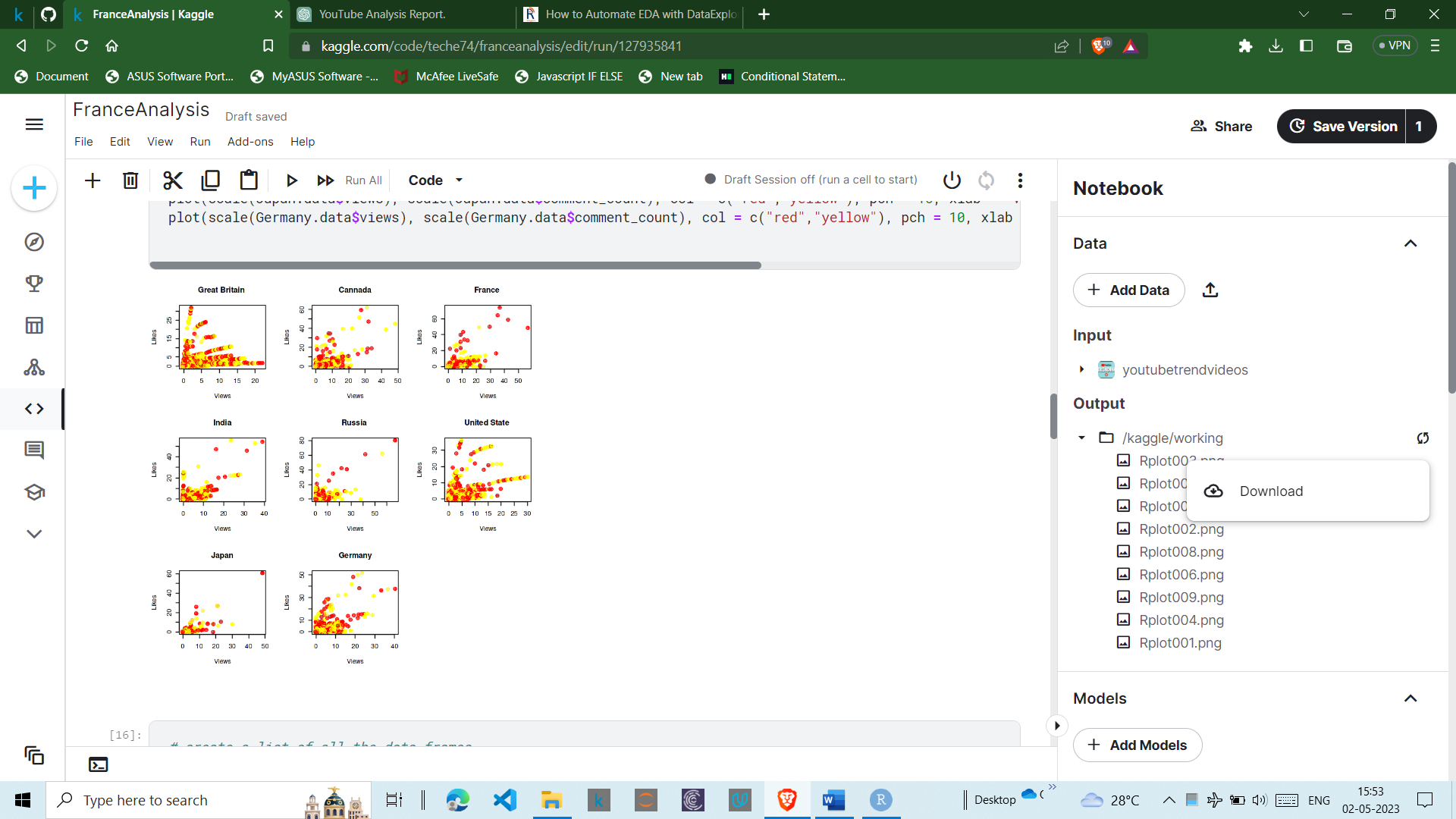
*Germany had a relatively small number of YouTube channels and videos in the dataset compared to the other countries. We found that the most popular video categories in Germany were music, entertainment, and gaming. We also noticed a positive correlation between views and likes, similar to the other countries.*

***France:***

*France had a relatively small number of YouTube channels and videos in the dataset compared to the other countries. We found that the most popular video categories in France were music, entertainment, and gaming. We also noticed a positive correlation between views and likes, similar to the other countries.*

*Overall, our country-wise analysis helped us identify trends and insights that were specific to each country. We will use these insights in our further analysis of the data.*





**CONCLUSION**

*Based on our analysis of the Kaggle dataset on YouTube for different countries, we identified several key trends and insights. These include:*

*Music, entertainment, and gaming are the most popular video categories across all countries in the dataset.*

*There is a positive correlation between views and likes/comments across all countries, indicating that engagement is an important factor in video trendiness.*

*Videos with longer durations tend to have fewer views, indicating that shorter videos are more likely to be watched by viewers.*

*There are notable differences in popular video categories and engagement metrics between different countries, highlighting the importance of country-specific analysis.*

*The United States has the largest number of YouTube channels and videos in the dataset, followed by India.*

* *Based on these insights, we recommend the following strategies to improve the trendiness of YouTube videos:*
* *Create videos that focus on popular categories such as music, entertainment, and gaming.*
* *Encourage engagement by asking viewers to like, comment, and subscribe to the channel.*
* *Create shorter videos that are more likely to be watched by viewers.*
* *Conduct country-specific analysis to identify trends and insights that are specific to the target audience.*

*Overall, our analysis provides valuable insights into the factors that affect the trendiness of YouTube videos across different countries. These insights can be used by content creators and marketers to optimize their YouTube content and improve their engagement with viewers.*