Netflix user’s analysis (Project)

Submitted BY:

Muzafar Ali

Submitted To:

Muhammad Awais Ather

Table of Contents

Dataset:

- **Netflix userbase dataset**

For conducting an analysis, it's important that we choose the right dataset. We need to explore several datasets before finding the one that suits our analysis needs. Also, we need to take the data from a certified source, which can be found through various locations including government portals, data repositories, crowdsourcing, social media Api's, and academic research papers, etc. For data analysis I choose Netflix userbase and for its dataset, I found it on Kaggle.

The dataset contains 2,500 rows and 10 columns. Here's a brief overview of each variable:

- **A user ID:** is a unique identifier assigned to an individual user within a system or platform. It helps distinguish and track users, enabling personalized interactions and data management.
- **Subscription Type:** refers to the category or level of service that a user has chosen to access within a service or platform. It typically defines the features, benefits, and limitations of the subscription plan, often varying in terms of price and available content or services.
- **Monthly Revenue:** refers to the amount of money generated monthly from a particular source, such as subscriptions, sales, or fees. It represents the income earned over a single month.
- J**oin Date:** This is the specific date on which an individual becomes a member or user of a service, platform, or community. It marks the moment when someone officially joins or starts using a particular system or group.
- **Last Payment Date:** refers to the most recent date on which a payment was made. It indicates the time when the latest financial transaction, such as a purchase, subscription renewal, or fee payment, occurred.
- **Country:** refers to a geographical region or nation where a person, entity, or location is situated. It represents a specific location on the Earth's surface and is often used to categorize and group individuals based on their geographic origin or current residence.
- **Age:** is a numerical value that represents the number of years a person has lived since their birth. It indicates the length of time that has passed since an individual was born and is often used as a demographic characteristic in various analyses and categorizations.
- **Gender:** refers to the classification of individuals based on their social, cultural, and personal identity in terms of being male, female, or another gender identity. It is a concept that goes beyond biological sex and encompasses a person's self-identification and expression.
- **Device:** refers to a physical or virtual tool, gadget, or machine used to perform specific

tasks or functions. In the context of technology, it often refers to electronic devices such as computers, smartphones, tablets, and other hardware used for communication, computation, and interaction with digital systems.

- **Plan Duration:** refers to the length of time for which a particular plan, subscription, or arrangement remains valid or active. It indicates the period during which the benefits, services, or features associated with the plan will be accessible before any renewal or expiration occurs.

	User ID	Subscription Type	Monthly Revenue	Join Date	Last Payment Date	Country	Age	Gender	Device	Plan Duration
0	1	Basic	10	15-01-22	10-06-23	United States	28	Male	Smartphone	1 Month
1	2	Premium	15	05-09-21	22-06-23	Canada	35	Female	Tablet	1 Month
2	3	Standard	12	28-02-23	27-06-23	United Kingdom	42	Male	Smart TV	1 Month
3	4	Standard	12	10-07-22	26-06-23	Australia	51	Female	Laptop	1 Month
4	5	Basic	10	01-05-23	28-06-23	Germany	33	Male	Smartphone	1 Month
...	...	...	...	...	...	...	...	...	...	...
2495	2496	Premium	14	25-07-22	12-07-23	Spain	28	Female	Smart TV	1 Month
2496	2497	Basic	15	04-08-22	14-07-23	Spain	33	Female	Smart TV	1 Month
2497	2498	Standard	12	09-08-22	15-07-23	United States	38	Male	Laptop	1 Month
2498	2499	Standard	13	12-08-22	12-07-23	Canada	48	Female	Tablet	1 Month
2499	2500	Basic	15	13-08-22	12-07-23	United States	35	Female	Smart TV	1 Month

2500 rows × 10 columns

Problem Statement:

Netflix is a popular streaming service application, which provide many of movies, TV shows and documentaries etc. There are over 200 million subscribers worldwide and is available over 190 countries. It offers different level of subscription plan which include basic, standard, and premium. Each of the plan also provide different features and pricing. Subscribers can use it through wide range of device which are TV, computers, mobile and streaming media players so on.

The objective of this research is to analyze Netflix user proportion of gender, watching device and revenue etc. Based on the data analysis, we have more insight on users' behaviors and preferences.

Questions:

1. Which Netflix subscription (Basic, Standard, or Premium) is most popular?

Analysis:

```
most_popular_subscription = users["Subscription Type"].mode()[0]
print(f"The most popular subscription is {most_popular_subscription}.")
```

```
The most popular subscription is Basic.
```

This analysis allows us to gain insights into the distribution of revenue across various subscription types. By examining this information, we can identify which subscription types contribute the most to the overall monthly revenue and which ones may have a comparatively lower impact. Such insights are valuable for businesses to adopt their strategies, marketing efforts, or product offerings based on the performance of different subscription types. From the above result we can conclude that the most popular subscription type is Basic among other subscription types.

2. How much monthly revenue does each type of subscription generate?

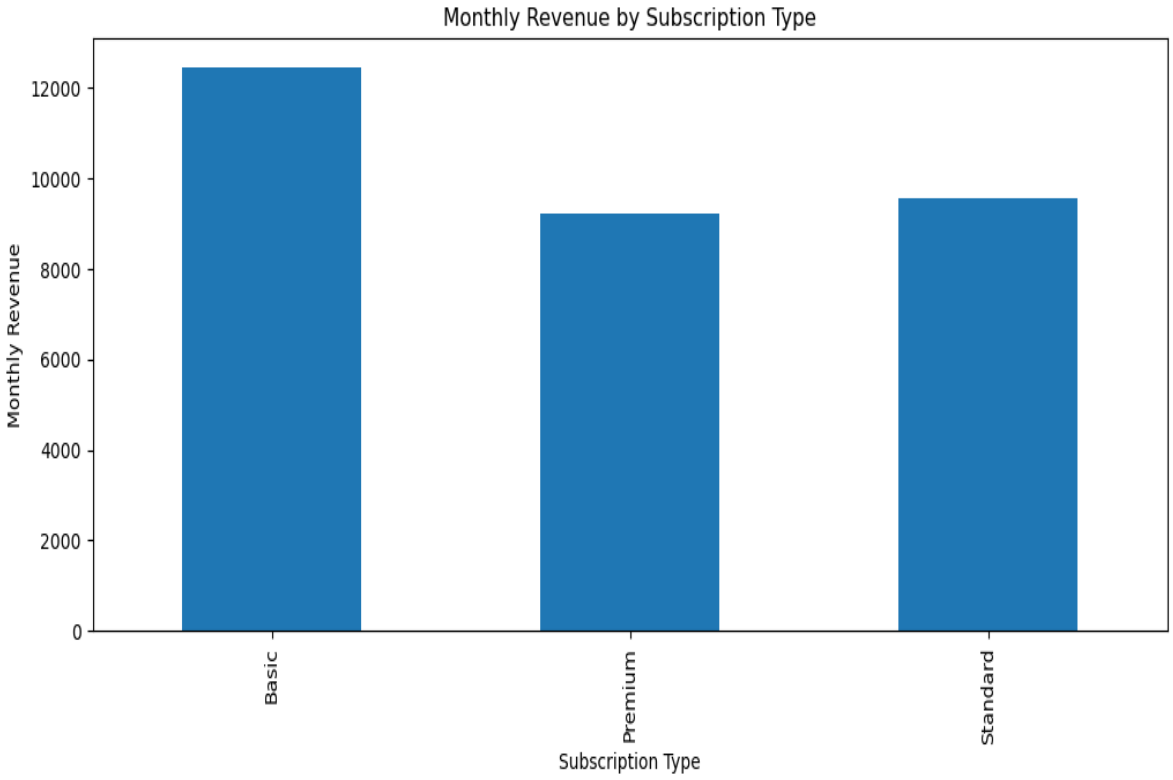
Analysis:

```
subscription_type_monthly_revenue = users.groupby("Subscription Type")["Monthly Revenue"].sum()
subscription_type_monthly_revenue
```

```
Subscription Type
Basic       12469
Premium      9229
Standard     9573
```

This analysis allows us to compute the total monthly revenue associated with different subscription types in a user dataset. The result provides valuable insights into the financial performance of each subscription category. This analysis enables businesses to discern which subscription types contribute the most to the overall monthly revenue, helping in strategic decision-making. Understanding the distribution of revenue among various

subscription models is essential for optimizing marketing strategies and resource allocation to enhance overall business success. From the above analysis we can conclude that Basic subscription type contributes the most to monthly revenue while Premium and Standard subscriptions on the other hand contributes less as compared to Basic.



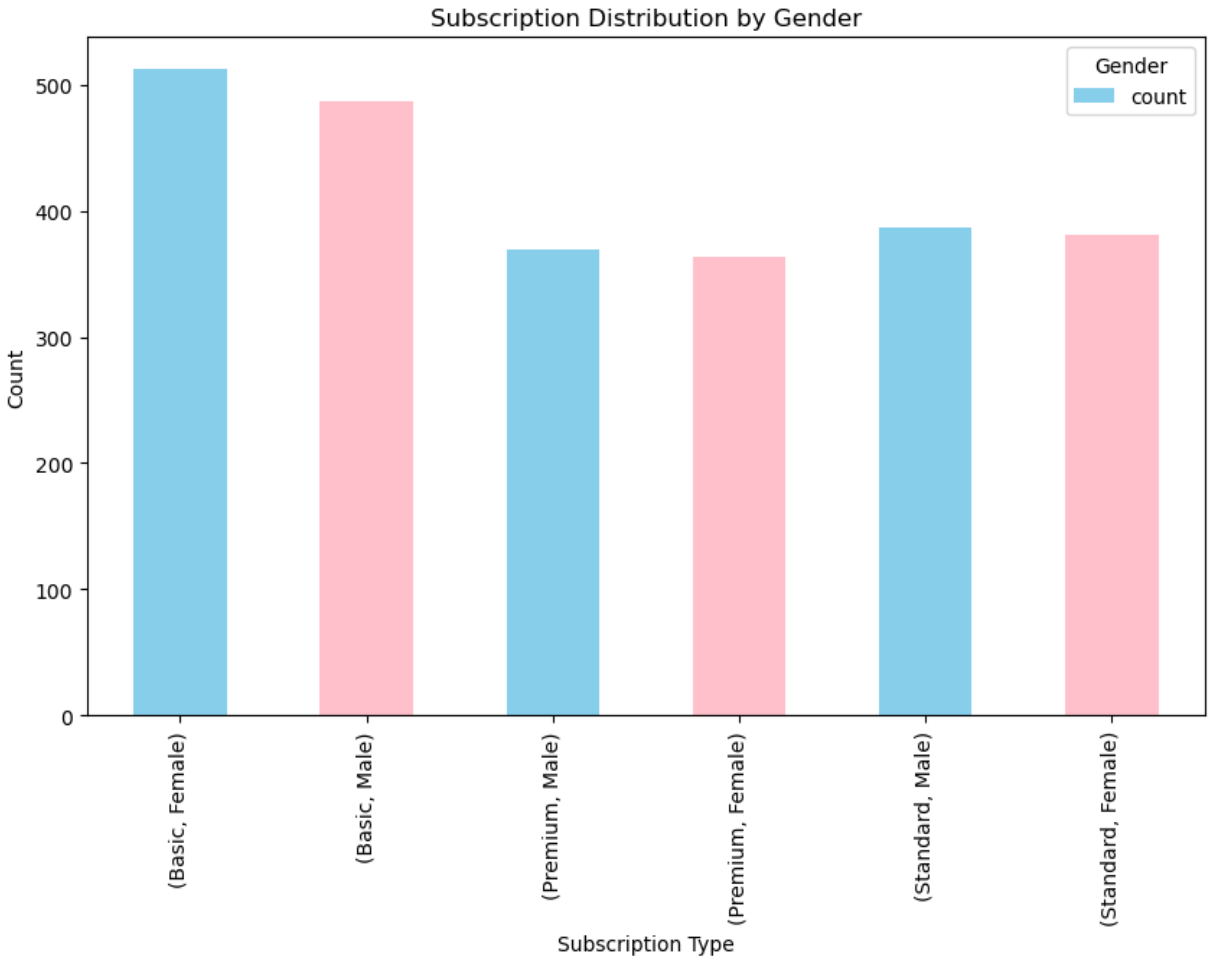
3. Are there any gender differences in preferred subscription levels?

Analysis:

```
subscription_by_gender = users.groupby("Subscription Type")["Gender"].value_counts()
subscription_by_gender
```

```
Subscription Type  Gender
Basic              Female    512
                   Male      487
Premium            Male      369
                   Female    364
Standard           Male      387
                   Female    381
```

This analysis breaks down the number of users by both their subscription type and gender. It counts how many users belong to each gender within every subscription category. The result gives a clear picture of the distribution of users based on both subscription type and gender. This information is useful for businesses to understand if there are any gender-specific trends or preferences within different subscription types. It can guide marketing strategies to better cater to the needs and interests of diverse user groups.



4. How do the most popular subscriptions vary across different countries?

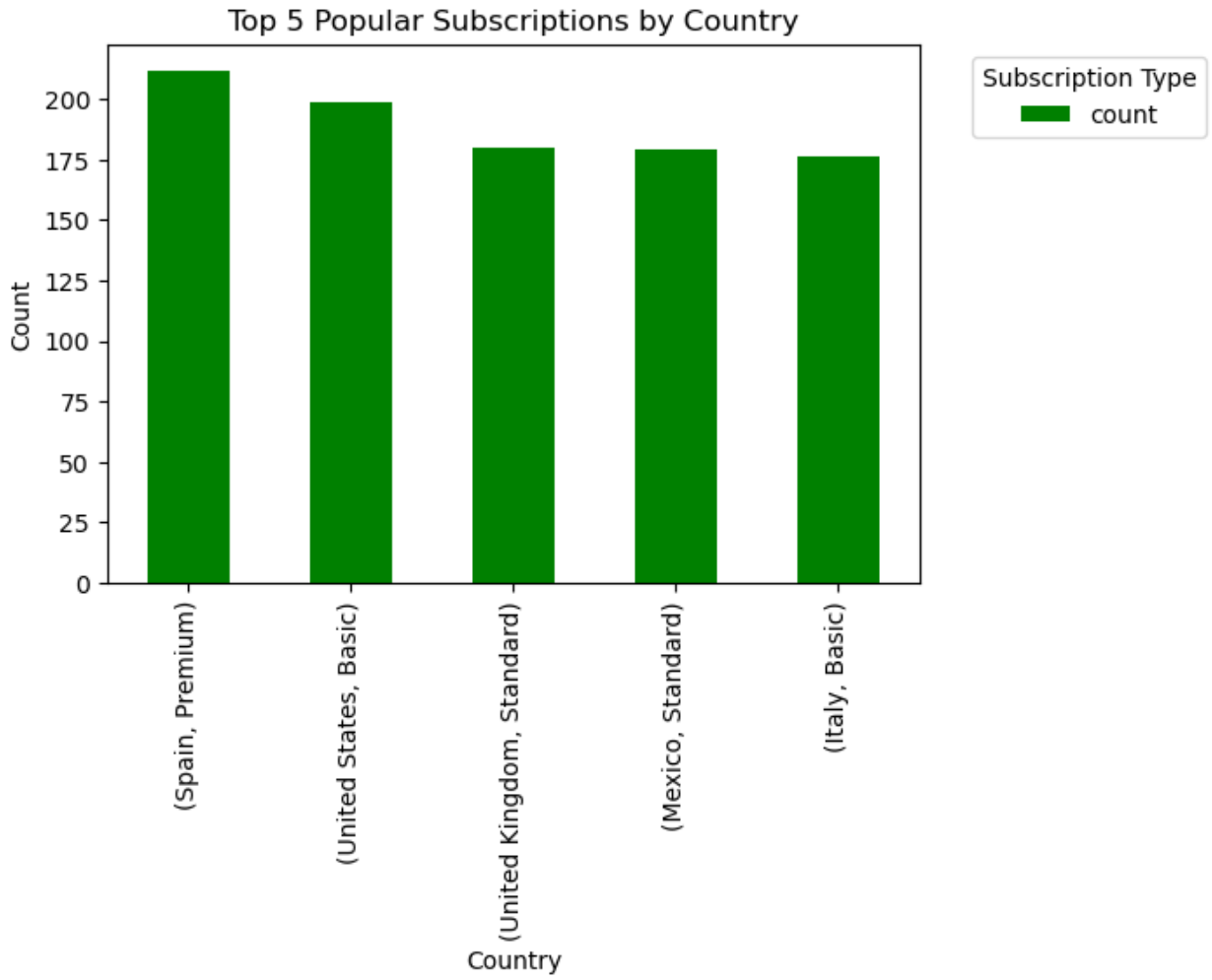
Analysis:

```
popular_subscriptions = users.groupby("Country")["Subscription Type"].value_counts().nlargest(5)
popular_subscriptions
```

```
Country         Subscription Type
Spain           Premium              212
United States   Basic                199
United Kingdom  Standard             180
Mexico          Standard             179
Italy           Basic                176
```

This analysis identifies the most popular subscription types in each country by counting the

occurrences of each subscription type within the dataset. The data is grouped by "Country," and within each country, the code tallies the occurrences of different subscription types. This outcome provides valuable insights into the subscription preferences of users across different countries. By focusing on the five most popular subscription types in each country, businesses can tailor their marketing strategies and content to cater to the prevailing trends in each region. Understanding the regional variations in subscription choices enables companies to better meet the diverse needs and preferences of their user base, contributing to more effective and targeted business decisions.



5. What is the total monthly revenue generated by Netflix?

Analysis:

```
monthly_revenue = users["Monthly Revenue"].sum()
print(f"The Total monthly revenue of Netflix is ${monthly_revenue}.")
```

```
The Total monthly revenue of Netflix is $31271.
```

The analysis calculates the total monthly revenue by summing up the "Monthly Revenue" values from the entire user dataset. It represents that the total revenue generated from all users during that period is $31271.

6. On average, how much revenue does each Netflix user generate per month?

Analysis:

```
average_monthly_revenue = users["Monthly Revenue"].mean()
print(f"The Average monthly revenue of Netflix is ${average_monthly_revenue:.2f}.")
```

```
The Average monthly revenue of Netflix is $12.51.
```

The analysis calculates the average of monthly revenue from the entire user dataset. It represents that the average monthly revenue generated from all users during that period is $12.51.

7. Has the monthly revenue trend changed over time?

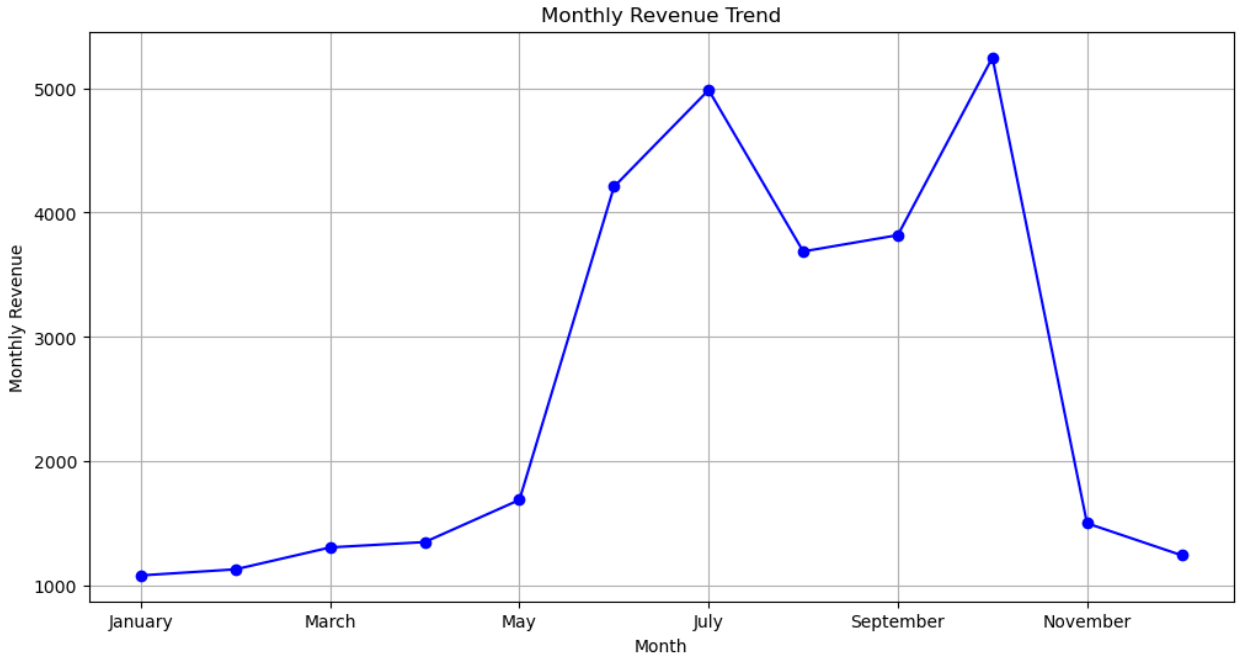
Analysis:

```
users["Month"] = users["Join Date"].dt.month_name()
custom_month_order = ["January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December"]
users["Month"] = pd.Categorical(users["Month"], categories=custom_month_order, ordered=True)
monthly_revenue_trend = users.groupby("Month")["Monthly Revenue"].sum()
monthly_revenue_trend
```

```
C:\Users\Muzafar Ali\AppData\Local\Temp\ipykernel_14116\971589936.py:4: FutureWarning: The default of observed=False is deprecated and will be changed
  monthly_revenue_trend = users.groupby("Month")["Monthly Revenue"].sum()
```

Month	
January	1085
February	1134
March	1310
April	1353
May	1692
June	4208
July	4985
August	3688
September	3819
October	5246
November	1504
December	1247

This analysis allows businesses to discern patterns and trends in revenue over the months, providing insights into the seasonality or cyclical nature of user engagement and financial performance. This information is invaluable for companies to adapt their strategies, anticipate fluctuations in user activity, and make data-driven decisions to optimize their operations throughout the year. From this result, it can see that the revenue increase sharply in May and fall sharply in December.



8. Do users on different subscription plans generate different levels of revenue?

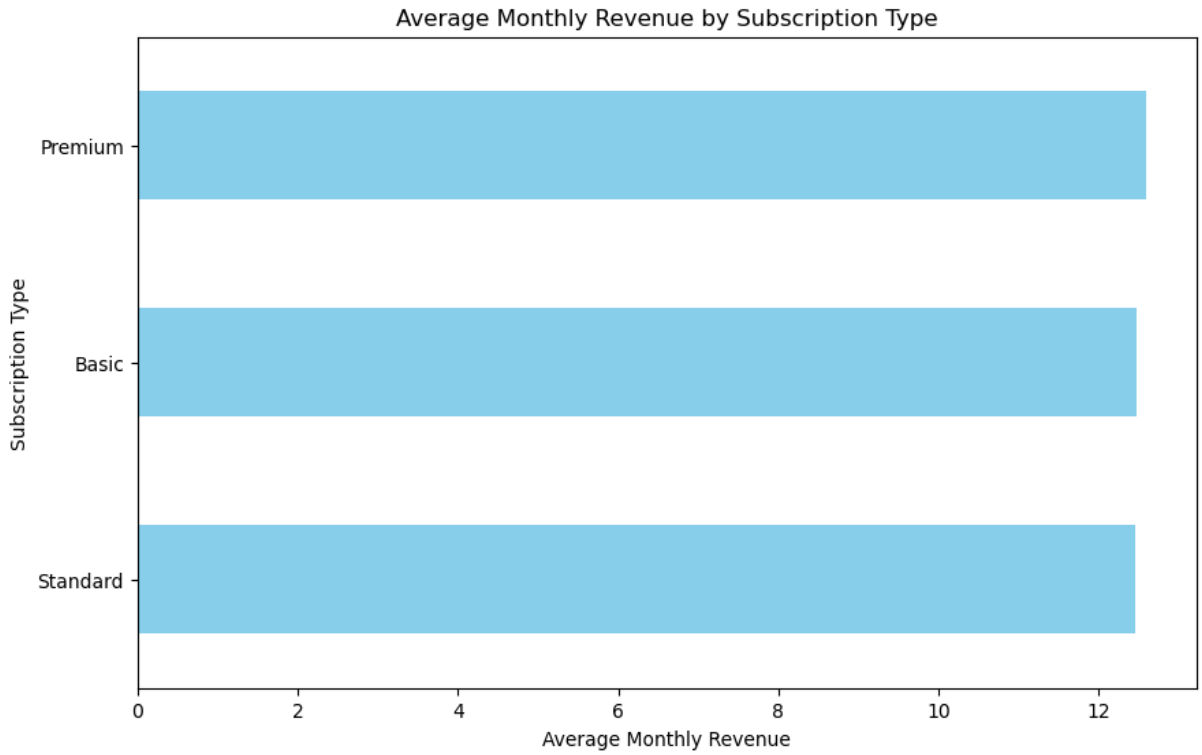
Analysis:

```
revenue_by_type = users.groupby("Subscription Type")["Monthly Revenue"].mean()
revenue_by_type
```

```
Subscription Type
Basic       12.481481
Premium     12.590723
Standard    12.464844
```

This analysis calculates the average monthly revenue for each subscription type in the user dataset. By grouping the data based on "Subscription Type" and computing the mean of the "Monthly Revenue" for each group, the result, provides insights into the typical revenue generated by users subscribing to different plans. This information is valuable for businesses to understand the relative financial contributions of each subscription type. By knowing the average monthly revenue for each category, companies can tailor their marketing strategies, prioritize resources, and refine their services based on the preferences and spending patterns of users within each subscription type. It helps in identifying which subscription plans are more

profitable on average and can guide strategic decisions to optimize overall revenue and user satisfaction.



9. Which devices used to access Netflix generate the most revenue?

Analysis:

```
most_revenue_device = users.groupby("Device")["Monthly Revenue"].sum().idxmax()
print(f"The device which used to generate most revenue is {most_revenue_device}.")
```

```
The device which used to generate most revenue is Laptop.
```

The above analysis determines the device that has generated the highest total monthly revenue in the user dataset. This information is essential for businesses to identify the primary source of revenue among different devices. Knowing which device contributes the most financially allows companies to prioritize development, support, and marketing efforts for that specific platform. In this result, the device which used to generate most revenue is laptop.

10.How many new users join Netflix each month?

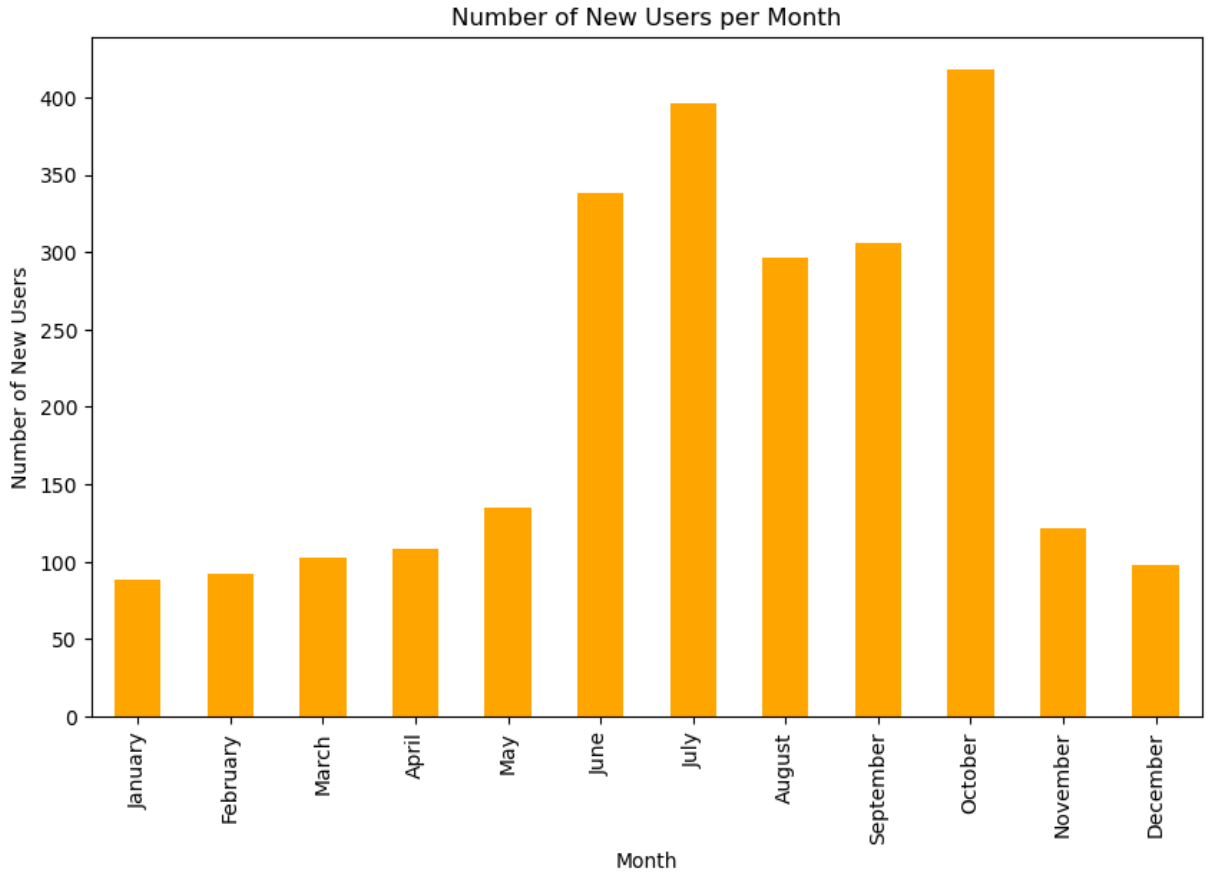
Analysis:

```
custom_month_order = ["January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December"]
users["Month"] = pd.Categorical(users["Month"], categories=custom_month_order, ordered=True)
new_users_per_month = users.groupby("Month").size()
new_users_per_month
```

```
C:\Users\Muzafar Ali\AppData\Local\Temp\ipykernel_14116\4087614707.py:3: FutureWarning: The default of observed=False is deprecated and will be chan
  new_users_per_month = users.groupby("Month").size()

Month
January      88
February     92
March       103
April       108
May         135
June        338
July        396
August      296
September   306
October     418
November    122
December     98
```

This analysis helps businesses understand when more users tend to sign up during the year. This data can be useful for adapting marketing strategies or promotions based on when user sign-ups are typically higher or lower throughout the year. It provides insights that help companies make informed decisions about when to focus efforts on attracting new users. This result describes that most of the users tend to join in second and third quarter of the year.



11. On average, how long does it take for a user to make their second payment after joining?

Analysis:

```
users["last payment after joining"] = users["Last Payment Date"] - users["Join Date"]
average_time_last_payment = users["last payment after joining"].mean()
print(f"Average last payment days after joining: {average_time_last_payment}")
```

```
Average last payment days after joining: 326 days 02:26:52.800000
```

This analysis calculates the average time elapsed between the "Join Date" and the "Last Payment Date" for users in the dataset. This information is helpful for understanding the typical duration users remain subscribed or engaged after joining. The average time between joining and the last payment offers insights into user retention and can guide businesses in refining

their strategies to encourage sustained user engagement. A shorter average time might indicate users typically stay subscribed for a brief period, while a longer average time could suggest a higher level of user retention. In this result, the average last payment days after joining is 326 days.

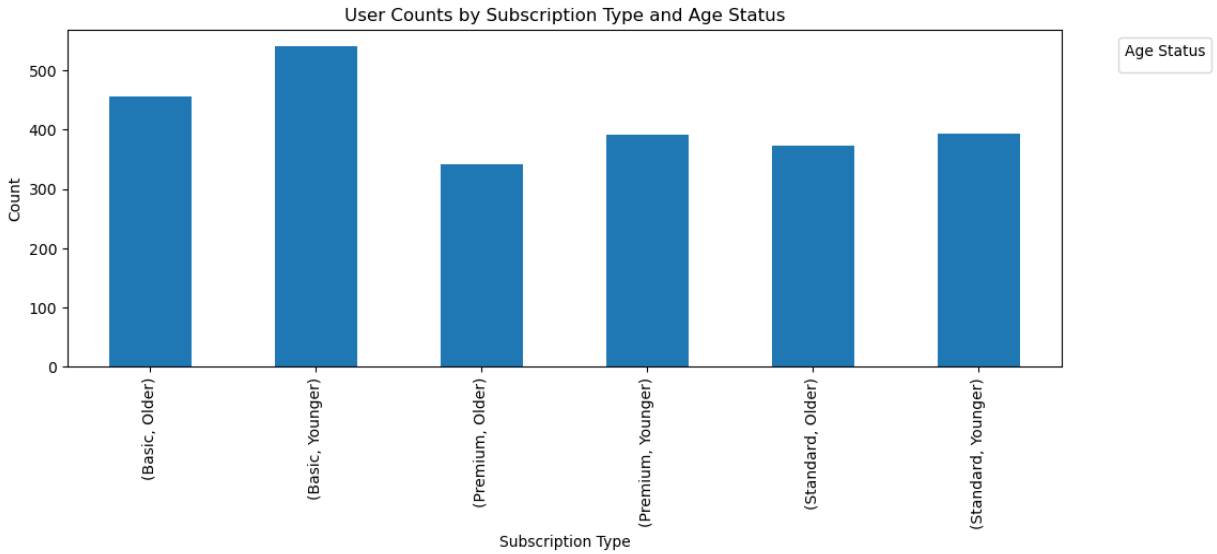
12.How is the user base distributed across different age groups and subscription types?

Analysis:

```
def categorize_age_status(Age):
    if Age < 40:
        return 'Younger'
    else:
        return 'Older'
users["Age Status"] = users["Age"].apply(categorize_age_status)
users.groupby(["Subscription Type", "Age Status"]).size()
```

```
Subscription Type  Age Status
Basic              Older         457
                   Younger       542
Premium            Older         342
                   Younger       391
Standard           Older         374
                   Younger       394
```

In this analysis, the primary objective is to understand the distribution of users across different age groups concerning their subscription types. By categorizing individuals as 'Younger' or 'Older' based on the age threshold of 40, the code aims to reveal patterns or trends in subscription preferences within distinct age brackets. The subsequent grouping and counting of users based on both subscription type and age status provide a comprehensive overview of how age may influence the selection of subscription services.



13. What is the average age of users who prefer the Premium subscription compared to other plans?

Analysis:

```
premium_users = users[users["Subscription Type"] == "Premium"]
average_age_premium_users = premium_users["Age"].mean()
print(f"The average age of premium users is: {average_age_premium_users:.0f}.")
```

```
The average age of premium users is: 39.
```

In this analysis, we are trying to determine the average age of premium users. Analysing the average age of premium users can offer valuable demographic information. It helps answer questions related to the age profile of customers who are inclined towards premium subscription services. According to this result, the average age of premium users is 39.

14.What percentage of Netflix users are under 40?

Analysis:

```
under_40_percentage = (users[users['Age'] < 40].shape[0] / users.shape[0]) * 100
print(f"The percentage of netflix users under 40 is {under_40_percentage:.2f}%.")
```

```
The percentage of netflix users under 40 is 53.08%.
```

This analysis calculates the percentage of users who are under the age of 40. This percentage can be instrumental in shaping marketing strategies, content creation, or product development, as it provides a quantitative measure of the proportion of younger users within the overall user population. In this case, the percentage of Netflix users under 40 is 53.08%.

15.Do men and women in different countries have different preferences for subscription types?

Analysis:

```
users.groupby(["Country","Subscription Type","Gender"]).size()
```

```
Country    Subscription Type  Gender
Australia  Basic              Female     16
                              Male       15
           Premium            Female     45
                              Male       56
           Standard           Female     28
                              Male       23
Brazil     Basic              Female     78
                              Male       68
           Premium            Female     16
                              Male       17
           Standard           Female      1
                              Male        3
Canada     Basic              Female     68
                              Male       77
           Premium            Female     40
                              Male       48
           Standard           Female     49
                              Male       35
France     Basic              Female     17
                              Male       19
           Premium            Female     74
                              Male       73
Germany    Basic              Female     80
                              Male       69
...
           Premium            Female     72
```

The primary objective of this analysis is to gain insights into the diversity and composition of the user base across different geographical locations, subscription types, and gender categories. Analyzing user counts within these nested groupings can reveal patterns, preferences, or trends that exist within specific segments of the user population. This information is valuable for tailoring marketing strategies, content localization, and user experience enhancements to better suit the preferences and demographics of users in various regions and demographic groups.

16.What are the most popular devices for accessing Netflix in each country?

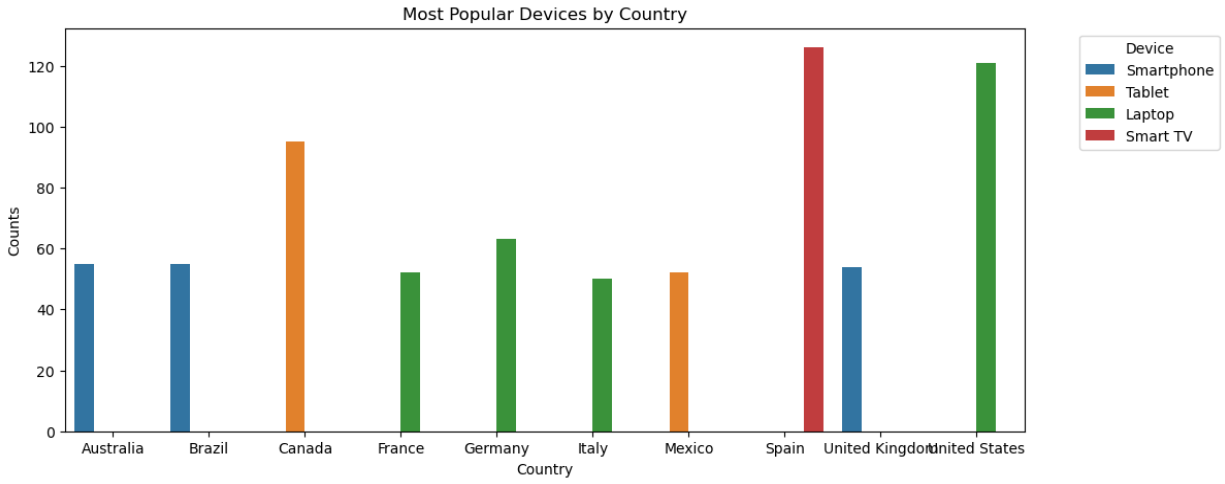
Analysis:

```
device_counts = users.groupby(["Country","Device"]).size().reset_index(name="Counts")
idx = device_counts.groupby(["Country"])["Counts"].transform(max) == device_counts["Counts"]
most_popular_devices = device_counts[idx]
most_popular_devices
```

```
C:\Users\Muzafar Ali\AppData\Local\Temp\ipykernel_14116\87718401.py:2: FutureWarning: The provided
  idx = device_counts.groupby(["Country"])["Counts"].transform(max) == device_counts["Counts"]
```

	Country	Device	Counts
2	Australia	Smartphone	55
6	Brazil	Smartphone	55
11	Canada	Tablet	95
12	France	Laptop	52
16	Germany	Laptop	63
20	Italy	Laptop	50
27	Mexico	Tablet	52
29	Spain	Smart TV	126
34	United Kingdom	Smartphone	54
36	United States	Laptop	121

This analysis seeks to uncover the prevalent devices utilized by users in different geographic regions. Understanding the popularity of specific devices in various countries is crucial for optimizing platform compatibility, tailoring marketing efforts, and potentially influencing product development decisions to cater to the preferences and device usage patterns of the diverse user base.



17.Do users who access Netflix on different devices prefer different types of subscriptions?

Analysis:

```
subscription_percentages = users.groupby("Device")["Subscription Type"].value_counts(normalize=True) * 100
print("Percentage of users with each subscription type for each device:")
print(subscription_percentages)
```

```
Percentage of users with each subscription type for each device:
Device      Subscription Type
Laptop      Basic               40.723270
            Premium             30.188679
            Standard            29.088050
Smart TV    Basic               39.016393
            Premium             30.819672
            Standard            30.163934
Smartphone  Basic               40.418680
            Standard            32.045089
            Premium             27.536232
Tablet      Basic               39.652449
            Standard            31.595577
            Premium             28.751975
```

This analysis aiming to explore the distribution of subscription types across various devices within the dataset. It does so by grouping the data based on the "Device" column and then calculating the percentage of each subscription type within each device category. Essentially, the analysis seeks to uncover patterns or trends in user preferences concerning subscription choices based on the devices they use. The percentages derived from this code offer a clear understanding of the relative popularity of subscription types within specific device categories, contributing essential insights for strategic decision-making and customer engagement efforts.

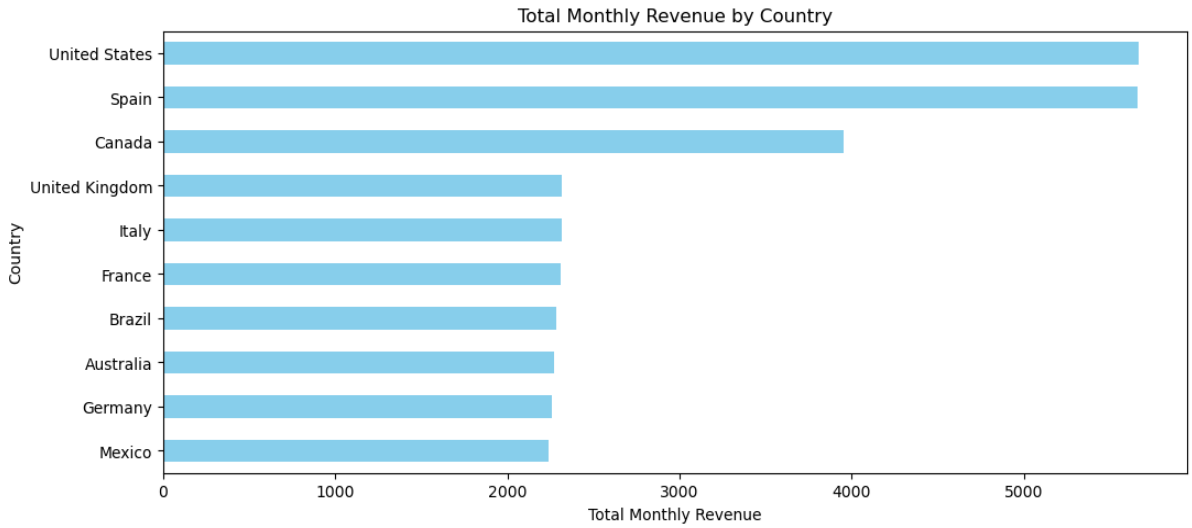
18.How much revenue does Netflix generate from users in different countries?

Analysis:

```
revnue_by_country = users.groupby("Country")["Monthly Revenue"].sum()
revnue_by_country
```

```
Country
Australia        2271
Brazil           2285
Canada           3950
France           2307
Germany          2260
Italy            2317
Mexico           2237
Spain            5662
United Kingdom   2318
United States    5664
```

This analysis is helping us figure out how much money we are making from users in different countries. It groups our data by country and adds up the monthly revenue for each country. This is important because it helps us see which countries are contributing a lot to our overall income. Knowing this helps us plan better, like deciding where to focus our marketing efforts or allocate resources. In simple terms, this analysis helps us understand where our money is coming from and where we might want to pay more attention to grow our business.



Analysis:

The following criteria were used for analysis:

- The most popular subscription type.
- Revenue from different subscription type.
- Gender differences in preferred subscription levels.
- monthly revenue generated by Netflix.
- devices used to access Netflix.
- different plan durations (monthly vs. Annual).
- age distribution of Netflix users.
- Gender preferences for subscription types.

Issues:

Common issues such as identifying missing data (empty cells), misspellings, duplicate data were sorted to ensure the integrity of the data. Columns were formatted with prices set to the currency, join date, and last payment date columns set to the short date format and others converted to correlate with their corresponding data types. New fields were created to get the important numbers as the numbers given could not generate the insight needed alone.

Methodology:

For this project, I used Python pandas for exploratory data analysis and matplotlib for visualization and I also used MS Power BI for visualization dashboard of analysis.

Data acquisition:

• Source identification: Identify and gather data from the Netflix userbase dataset on Kaggle.

• Data collection: Download the dataset and extract relevant information into a central data repository.

• Data cleaning: Clean and pre-process the data to address missing values, outliers, inconsistencies, and data quality issues.

Data exploration:

• Data overview: Gain an overall understanding of the data by summarizing key variables, identifying data types, and checking for missing values and outliers.

• Data visualization: Create required descriptive visualizations, such as histograms, scatter plots, and bar charts, to make visual representation the data.

Data Analysis:

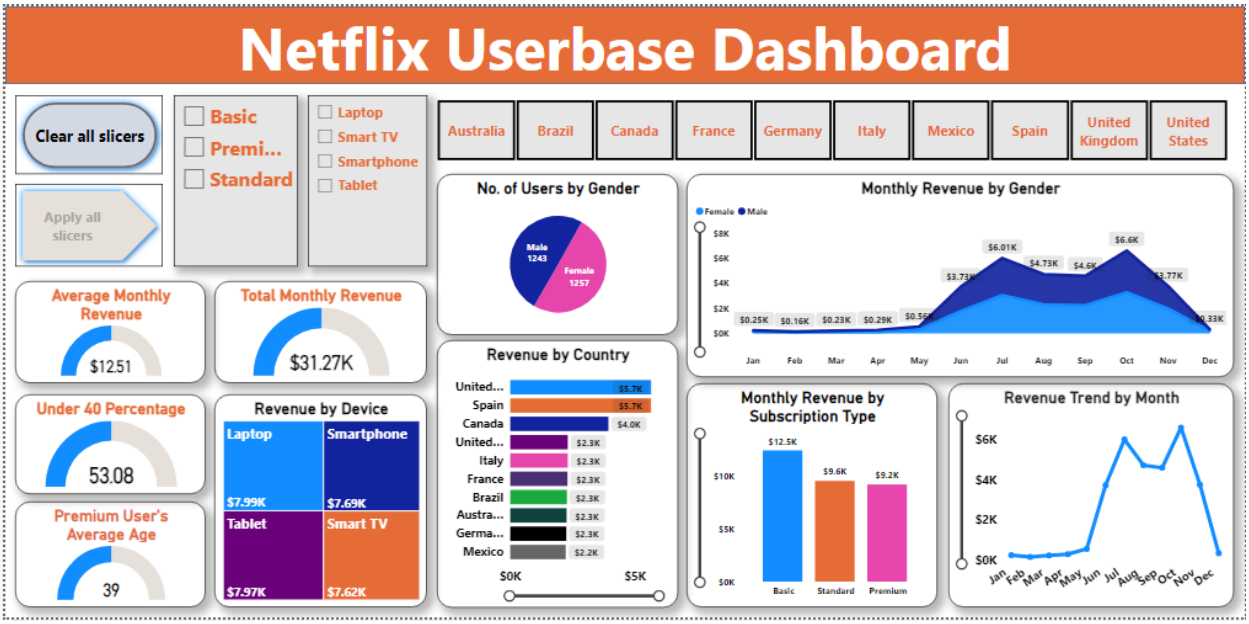
• Pivot tables: Utilize pivot tables to summarize data across different dimensions, such as product sections, regions, and time periods.

• Formulas: Create calculated fields and formulas to derive new insights from existing data, such as profit, sales trends across different variables, conditions, functions, etc. using Python Libraries.

Reporting and visualization:

• Dashboards: Develop interactive dashboards on MS Power BI that combine data visualizations, tables, and charts to provide a comprehensive overview of sales performance and trends.

• Conditional formatting: Employ conditional formatting techniques to highlight key data points, identify outliers, and visualize trends effectively.



Results:

- Basic subscription type is the most popular, contributing significantly to the overall monthly revenue.
- Both genders prefer the basic subscription type mostly.
- Spain prefers the premium subscription type.
- The total monthly revenue of Netflix is **$31271**.
- The average monthly revenue of Netflix is **$12.51**.
- The revenue increase sharply in May and fall sharply in December.
- The device which used to generate most revenue is laptop.
- Most of the users joins Netflix in second and third quarter of the year.
- The average last payment days after joining is **326** days.
- The average age of premium users is **39**.
- The percentage of Netflix users under 40 is **53.08%.**

References:

- www.kaggle.com
- www.chat.openai.com