

NETFLIX USERBASE DATASET ANALYSIS

INDIVIDUAL

DONE BY:

N VAMSI || AP21110010844

(SRM UNIVERSITY AP)

**Mentor’S:**

MEENAKSHI MAM

AND

NARMADA MAM

# TABLE OF CONTENTS

|  |  |
| --- | --- |
| **TOPICS** | **PAGE NO.** |
| **ABSTRACT** | **3** |
| **INTRODUCTION** | **4** |
| **SYSTEM REQUIREMENTS** | **5** |
| **USE OF DA LIBRARY**  **LOADING DATASET**  **APPLYING BASIC FUNCTIONS** | **6-10** |
| **DATA ANALYSIS AND VISUALIZATION** | **11-20** |
| **FUTURE SCOPE**  **ADVANTAGES**  **CONCLUSION**  **REFERENCE WEBSITE** | **21-22** |

ABSTRACT

This Project focuses on analyzing Netflix user data and subscription patterns to gain insights into customer behavior. The dataset includes information on user ID, subscription type, monthly revenue, join date, last payment date, country, age, gender, device, and plan duration. The main objectives of this analysis are to understand user preferences, identify trends in subscription adoption, and visualize key metrics for better decision-making.

Introduction

Netflix, a leading online streaming platform, has revolutionized the way people consume entertainment content. With millions of subscribers worldwide, it has become essential to understand user behavior and subscription patterns to optimize service offerings and enhance customer satisfaction. This study aims to analyze a comprehensive dataset containing key attributes such as user ID, subscription type, monthly revenue, join date, last payment date, country, age, gender, device, and plan duration. By conducting in-depth analysis and visualizations, we seek to uncover valuable insights that can assist Netflix in making informed decisions for its business growth.

SYSTEM REQUIREMENTS

### Operating System:

* Windows 10, macOS, or Linux (Ubuntu, CentOS, etc.)

### Software:

* Python (3.7 or higher) with necessary libraries: Pandas, NumPy, Matplotlib, Seaborn, Plotly, etc.
* Jupyter Notebook or any preferred integrated development environment (IDE) for Python.
* SQL database management system (DBMS) for data storage and retrieval.

### Hardware:

* Processor: Intel Core i5 or equivalent AMD processor (or higher).
* RAM: 8GB (or higher) for smooth data handling and analysis.
* Storage: At least 50GB of free space for storing the dataset and analysis results.
* Graphics Card: A dedicated graphics card is not strictly required for this analysis, but it can enhance the performance of data visualization, especially for large datasets.

USES OF DATA ANALYSIS LIBRARY

* The combination of NumPy, pandas, scikit-learn, Matplotlib, and Seaborn can provide a comprehensive toolkit for data analysis project. Here are some specific uses of each library within a project:

### 1.NumPy:

* NumPy provides powerful numerical operations and arrays, making it useful for mathematical computations and manipulation of large datasets.
* It offers essential functions for array manipulation, such as reshaping, slicing, indexing, and aggregations.

### 2.Pandas:

* It enables efficient handling and preprocessing of structured data, including CSV, Excel, SQL databases, and other formats.
* pandas allows data cleaning, filtering, sorting, merging, and aggregating, making it suitable for data wrangling tasks.

### 3.scikit-learn:

* scikit-learn offers various preprocessing methods, feature selection techniques, Scaling techniques.
* scikit-learn integrates seamlessly with NumPy and pandas, allowing easy data preparation and model training.

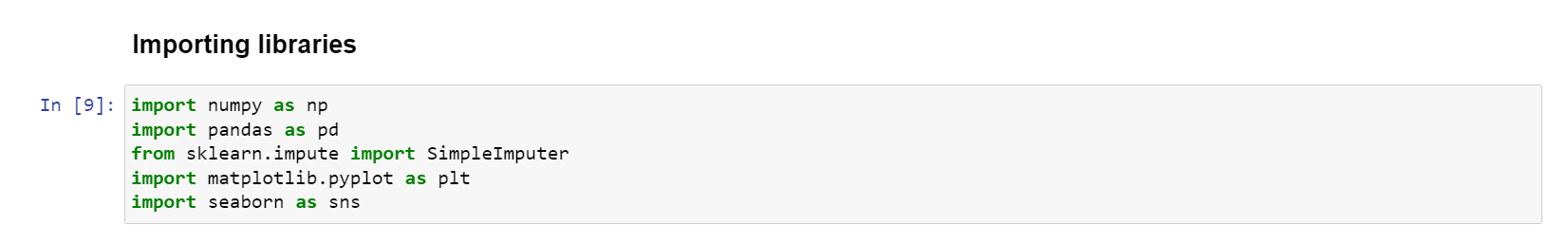
### 4. Matplotlib:

* It provides a wide range of plot types, including line plots, scatter plots, bar plots, histograms, heatmaps, and more.
* Matplotlib is highly compatible with NumPy and pandas, enabling visualization of data from these libraries.

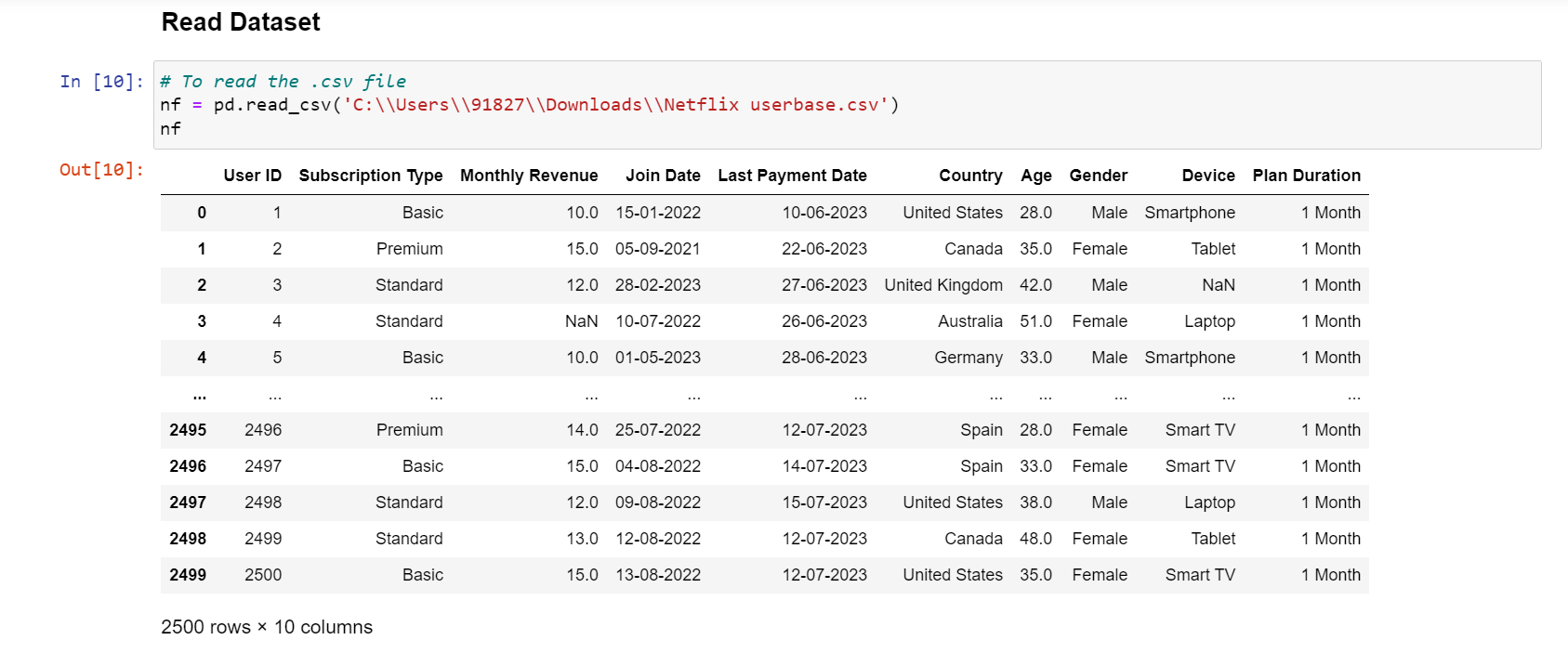
### 5.Seaborn:

* Seaborn is a statistical data visualization library that works closely with Matplotlib and enhances its capabilities.
* Seaborn includes functions for creating visually appealing plots like distribution plots, box plots, violin plots, pair plots, and correlation matrices.

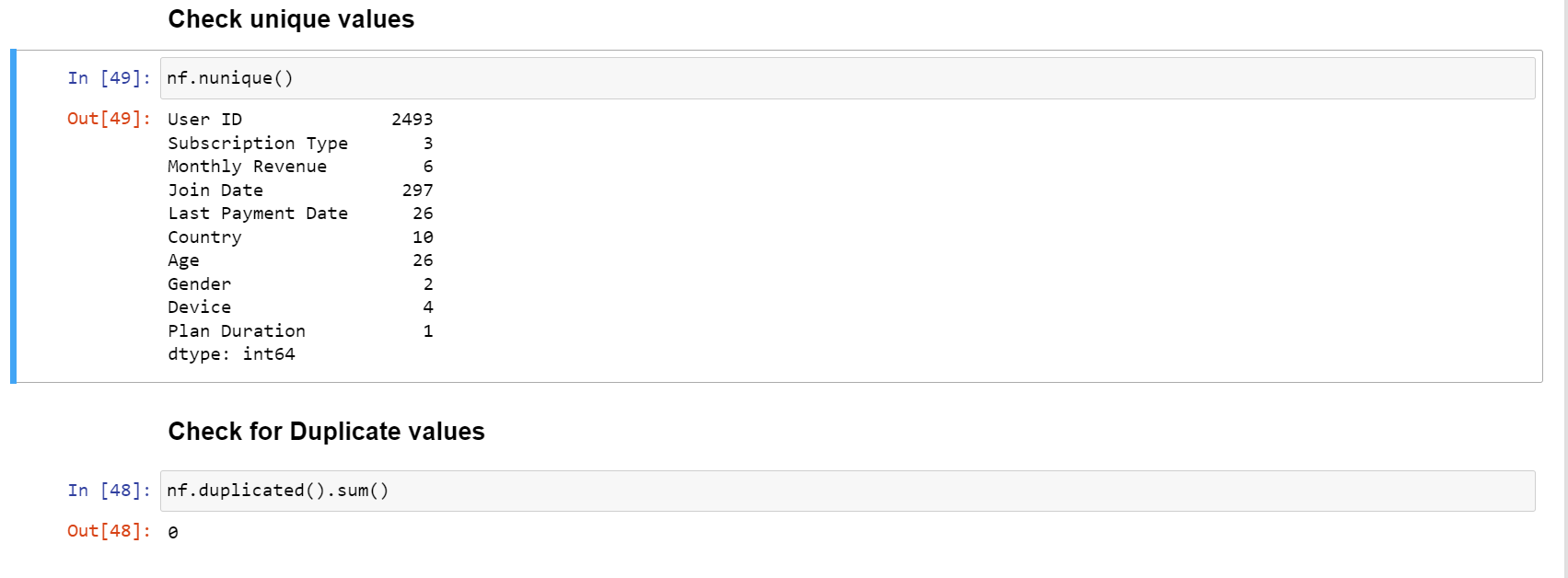
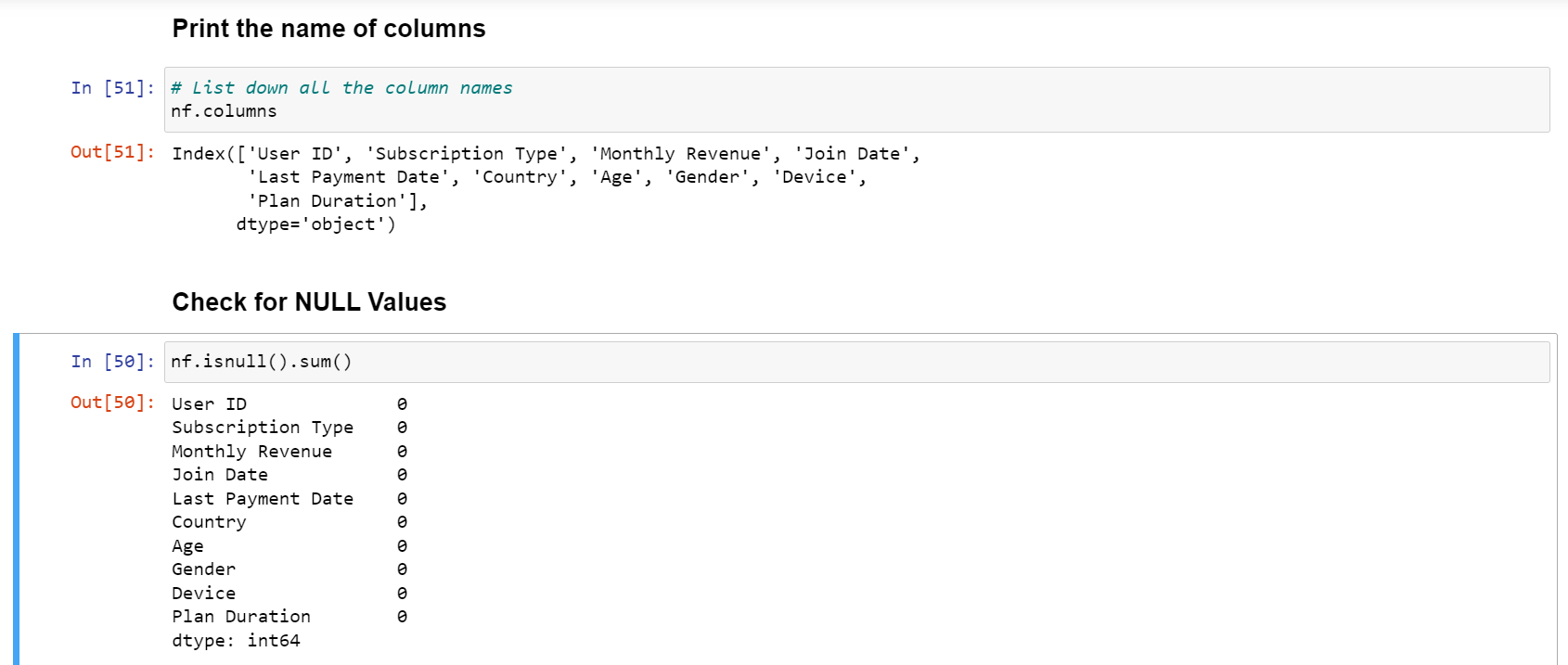
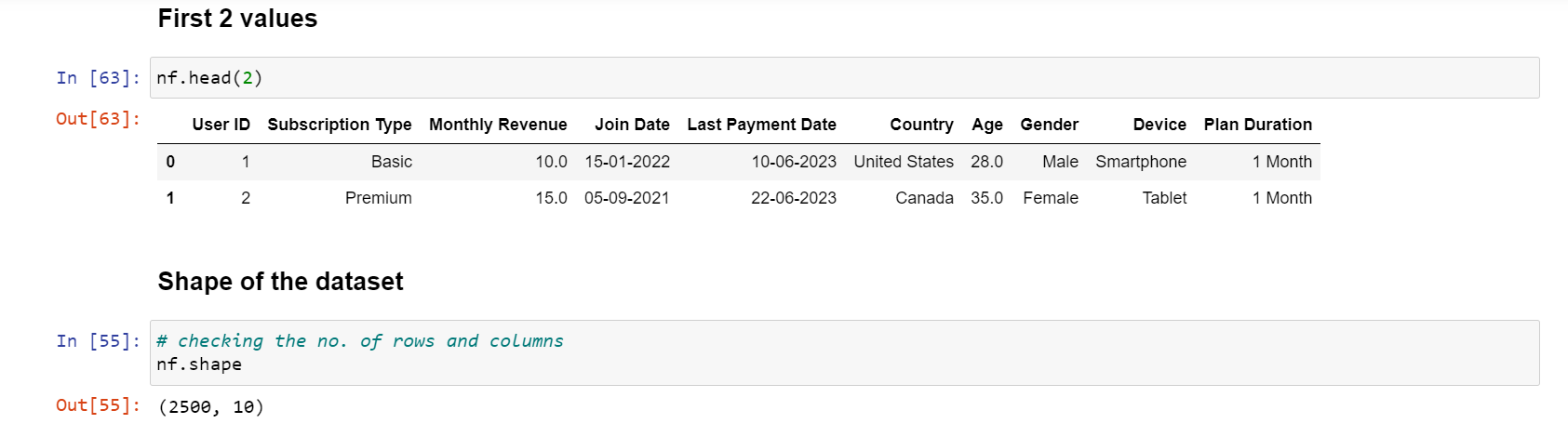
IMPORTING LIBRARY

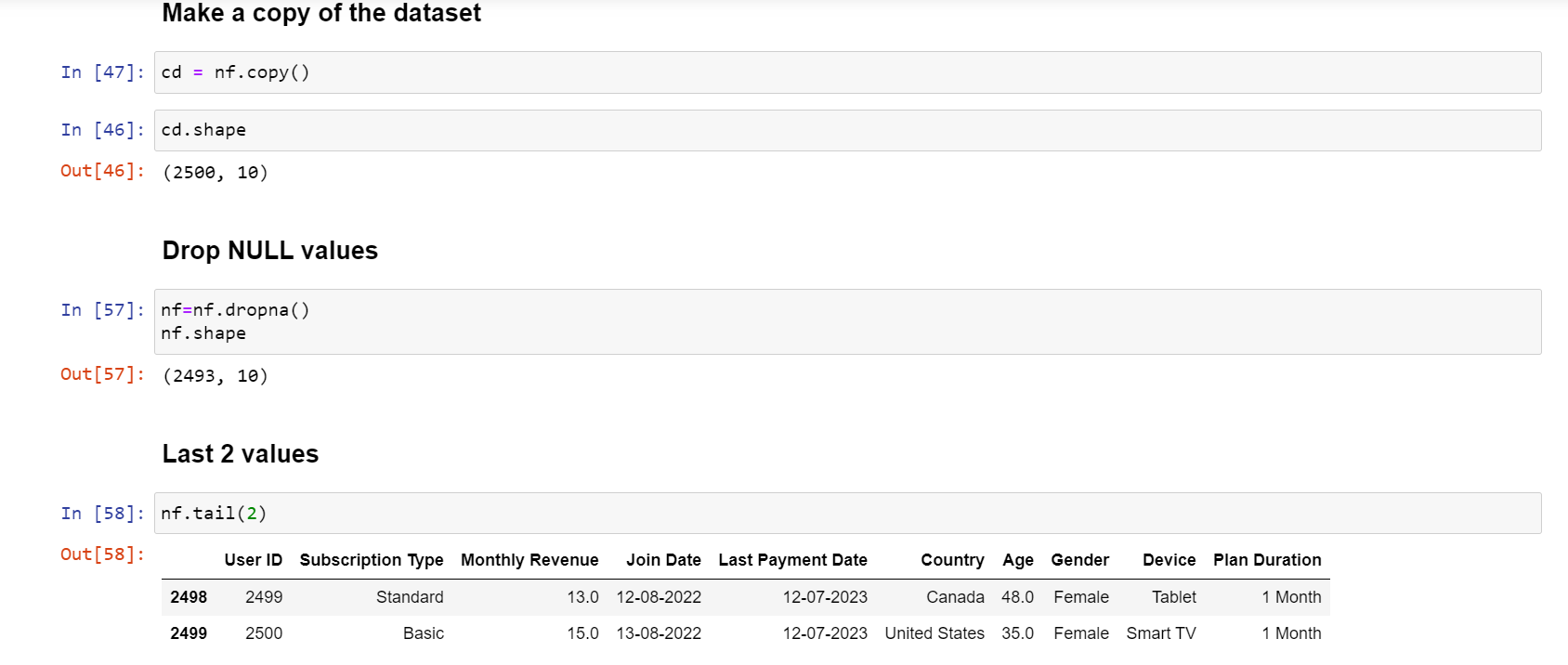


L0ADING DATASET



APPLYING BASIC FUNCTIONS

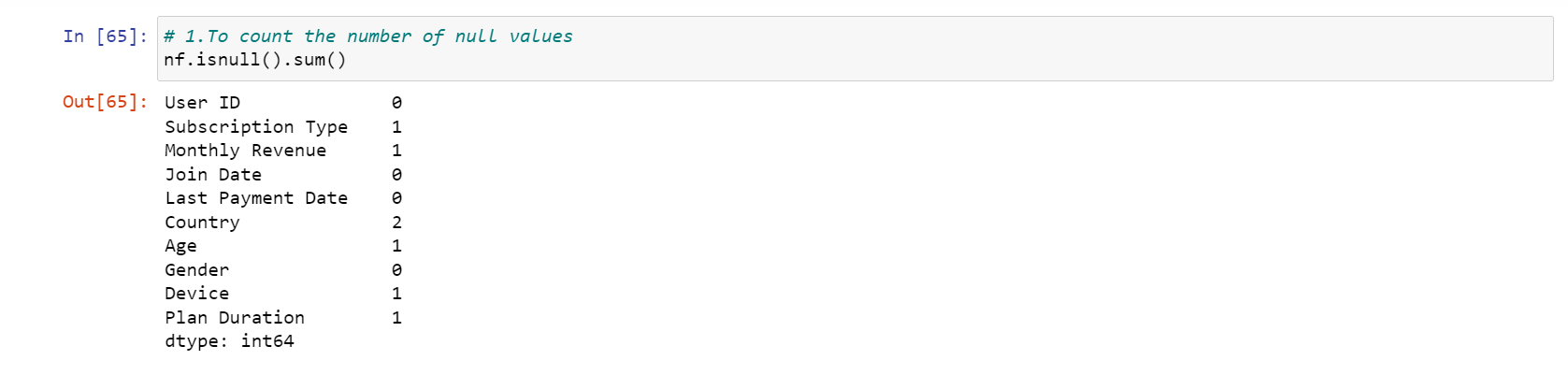


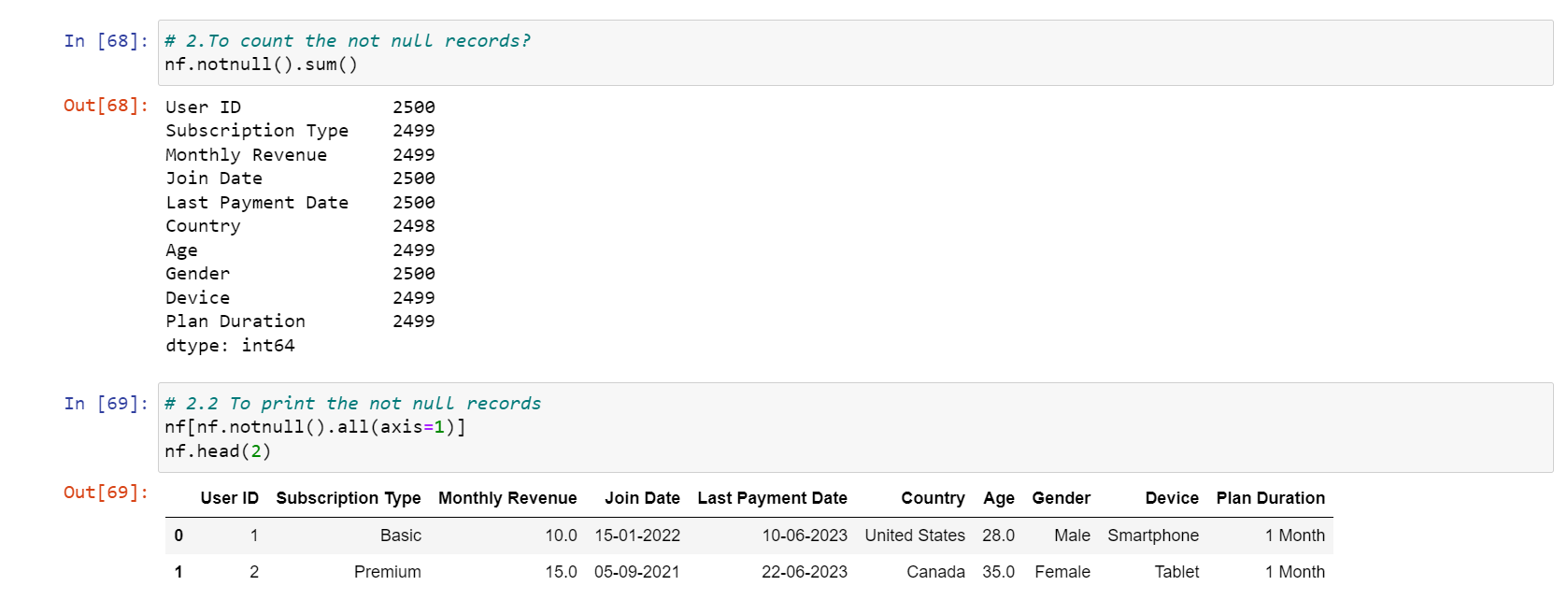


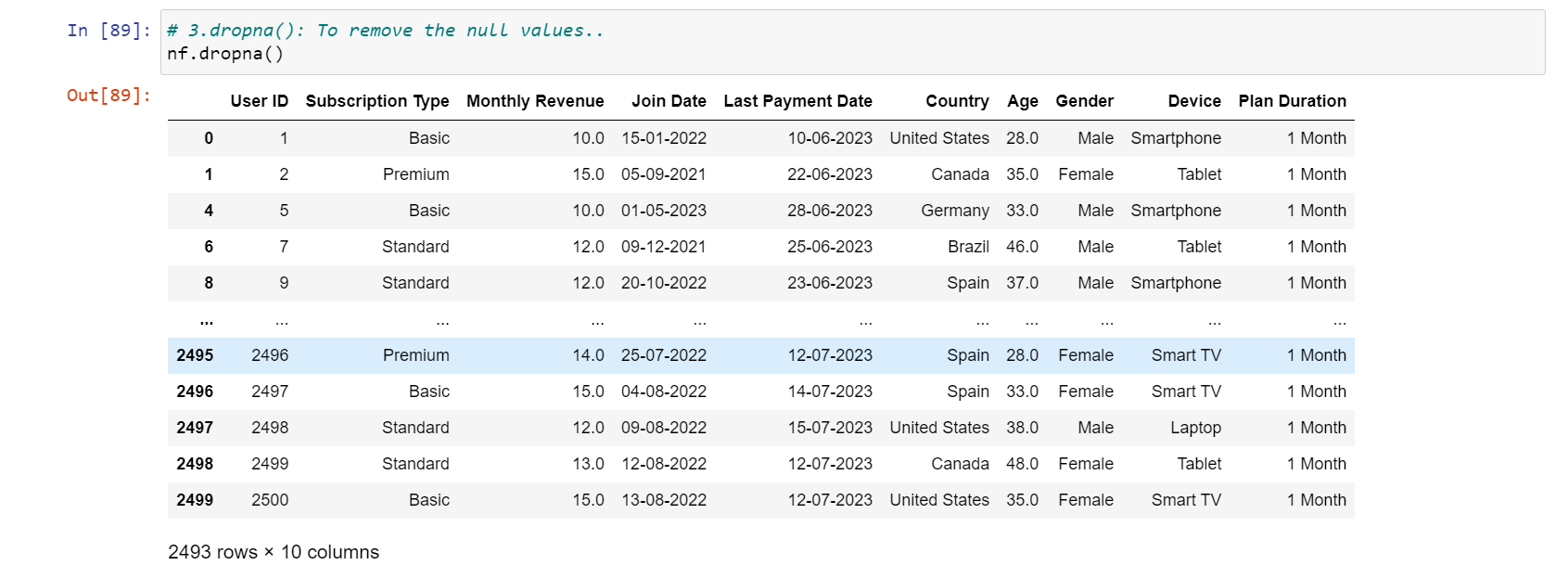
Data cleaning

Data cleaning is the process of identifying and correcting errors, inconsistencies, and inaccuracies in a dataset. It involves tasks such as removing duplicate entries, correcting typos, handling missing values, and ensuring data quality before analysis.

* 1.isnull() : To check the data having null values or not.
* 2.notnull(): To get the not null values.
* 3.dropna(): To remove the null values.
* 4.fillna(): To fill the null values with meaningfull data.
* 5.replace(): To replace the data
* 6.sort\_values(): To sort the data
* NaN : Not a number.







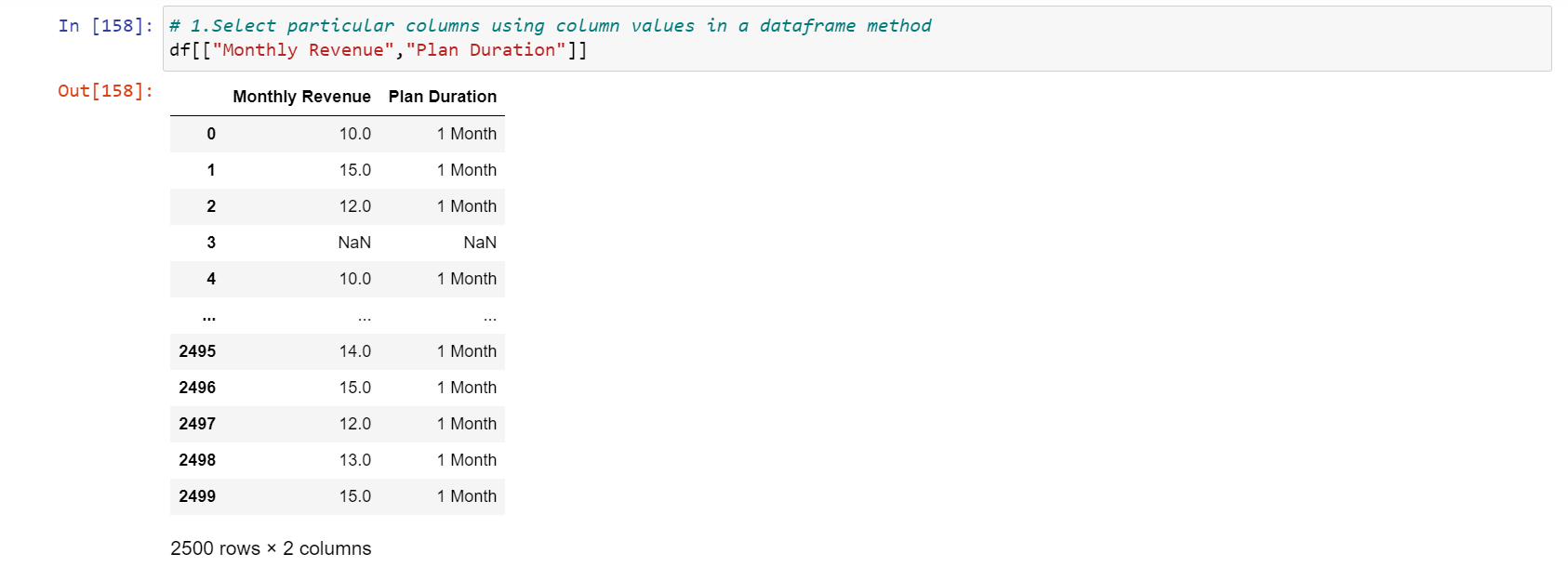


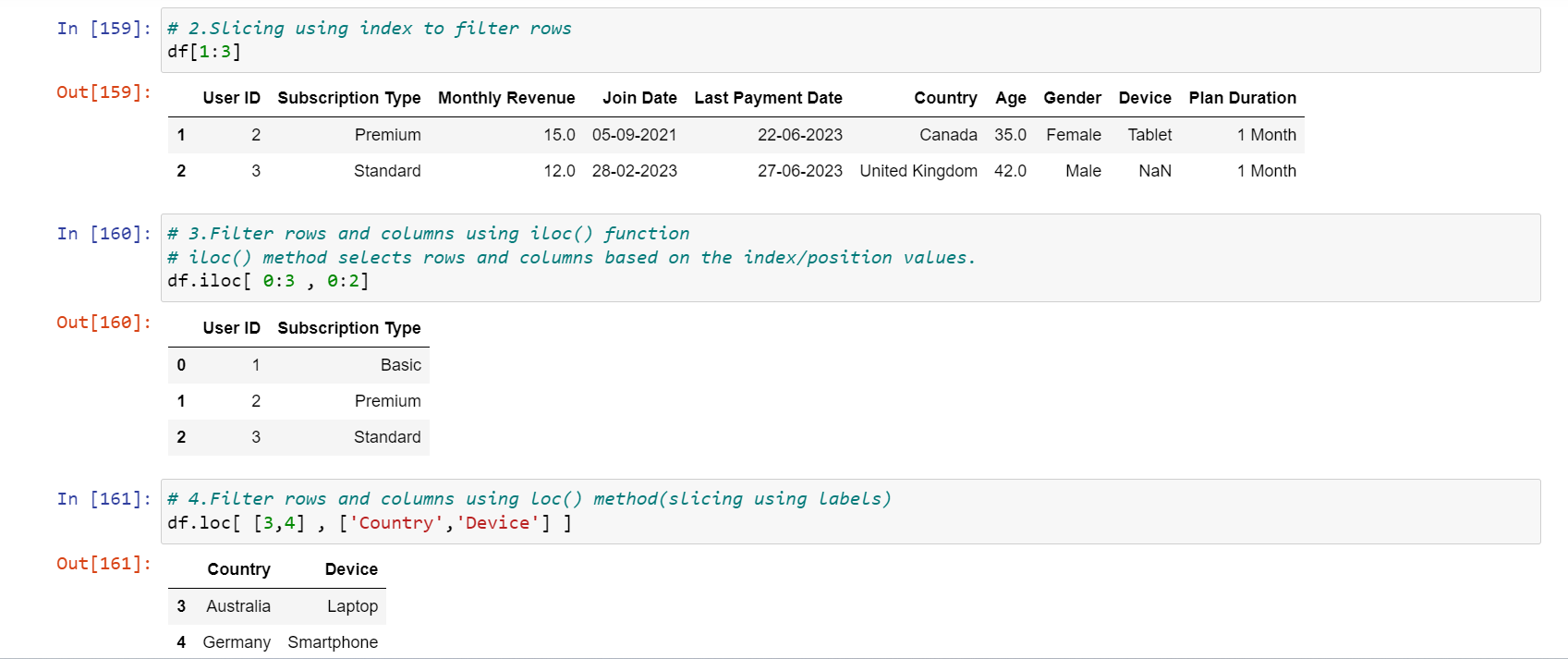




DATA FILTERING

Data filtering involves selecting a subset of data based on specific criteria. It helps in narrowing down the dataset to focus on relevant information. Filtering can be done based on conditions like date ranges, numerical values, or categorical attributes.

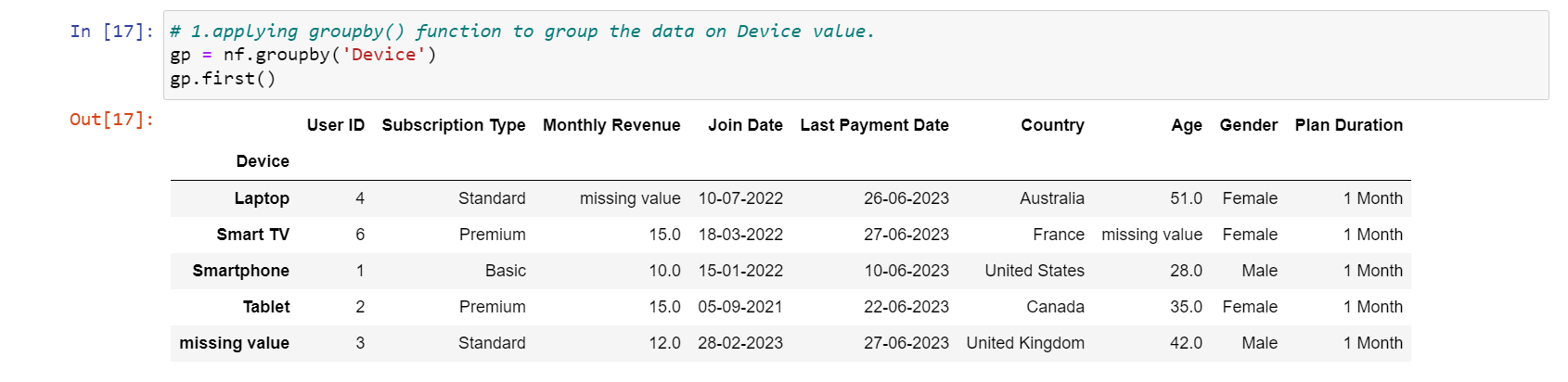






DATA GROUPING

Data grouping involves categorizing and organizing data based on common attributes. This is often done to create summaries or analyze trends within specific groups. It's a way to aggregate data together for more structured analysis.



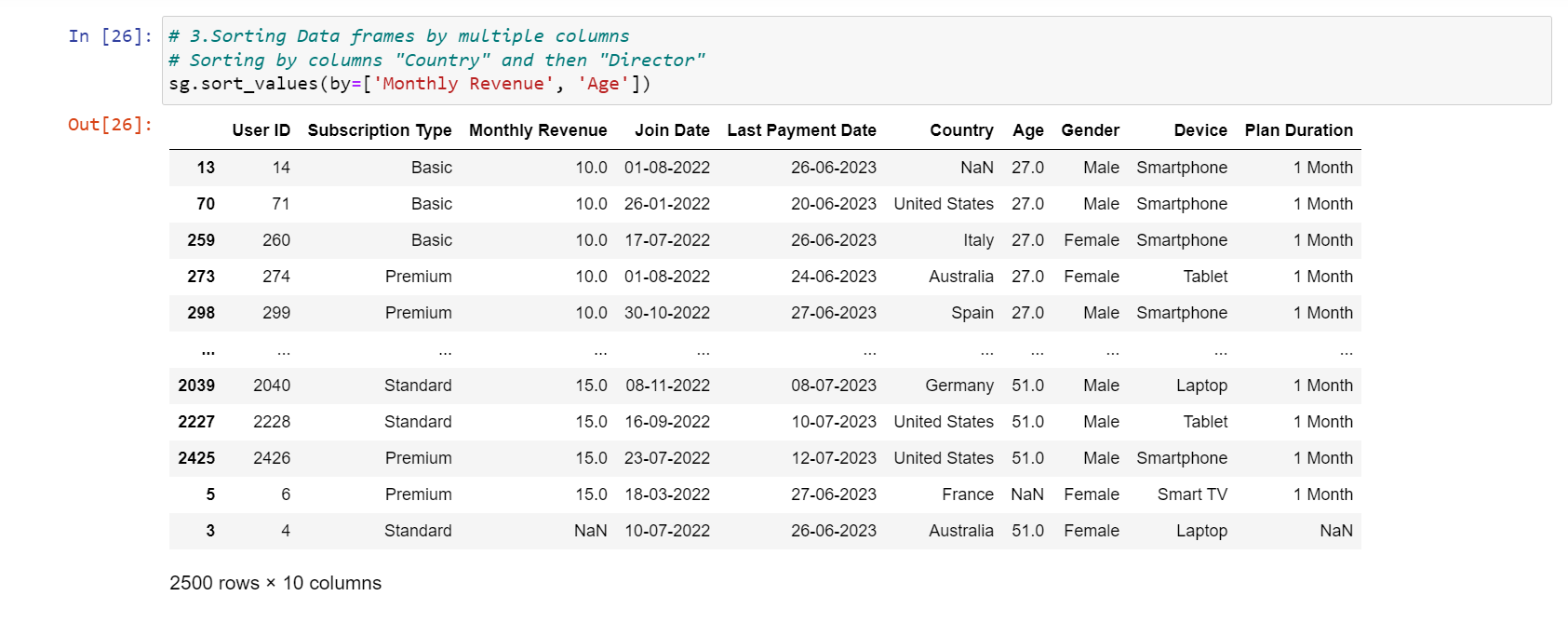


DATA SORTING

Data sorting is the process of arranging data in a particular order, typically based on the values in one or more columns. This helps in making the data more readable and organized, making it easier to identify patterns or trends.

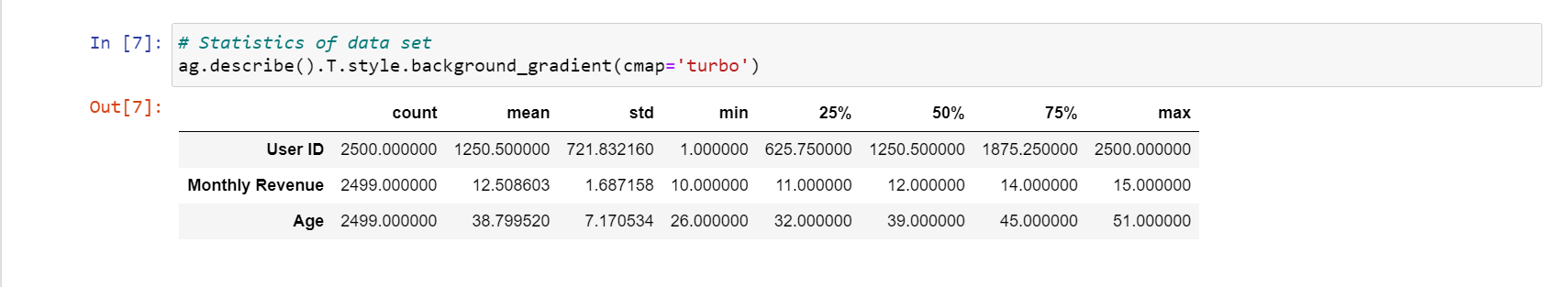






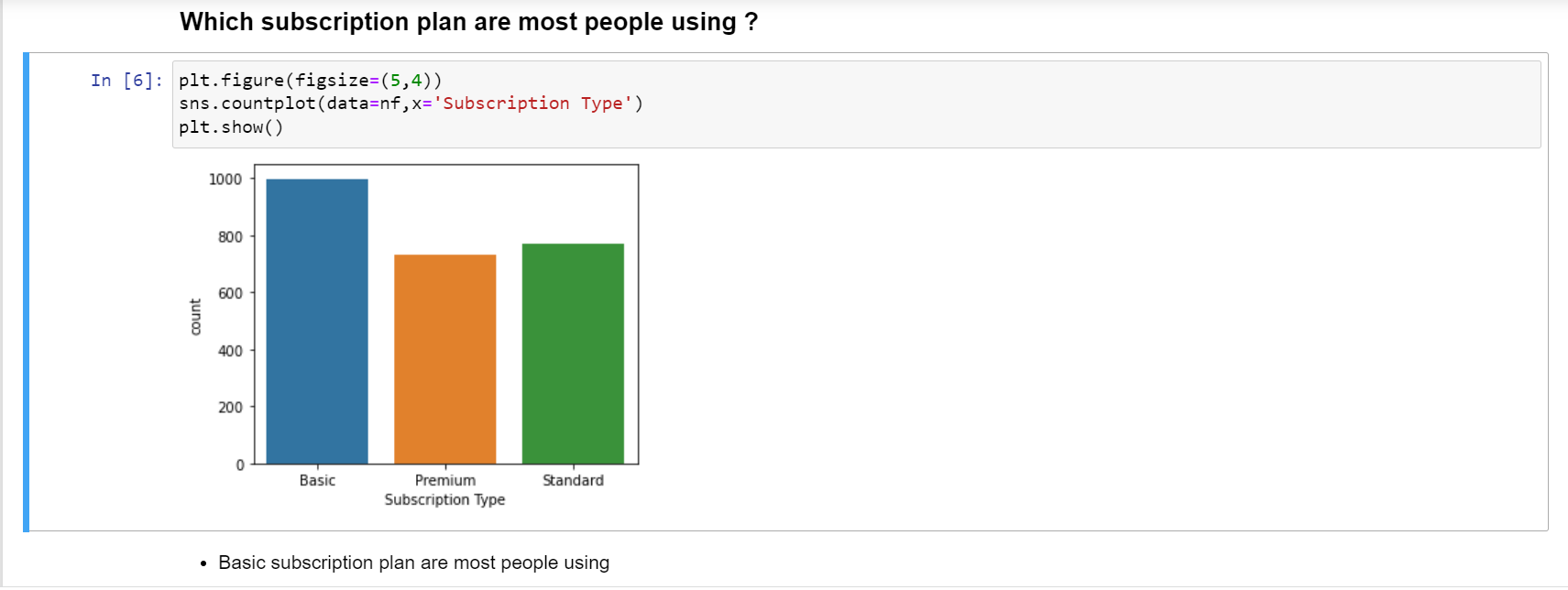
AGGREGATIONS

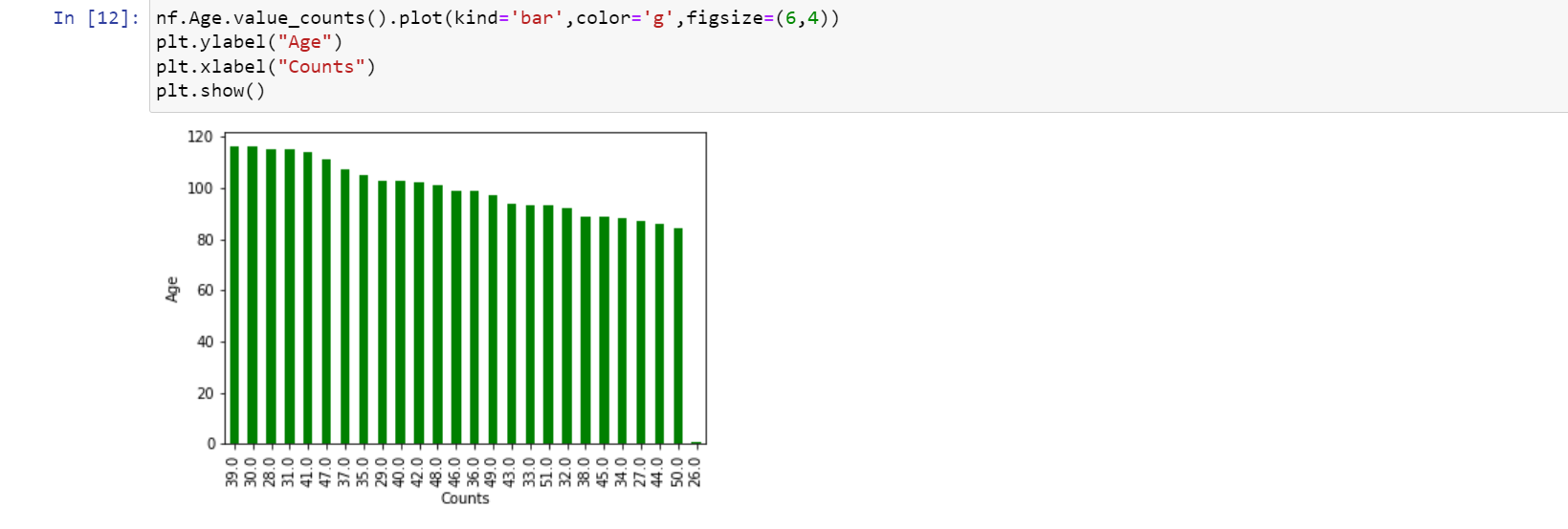
Aggregation in data analytics involves summarizing and combining data to obtain higher-level insights. It often involves functions like calculating sums, averages, counts, or other statistical measures on groups of data. Aggregation helps in simplifying complex datasets and drawing meaningful conclusions.



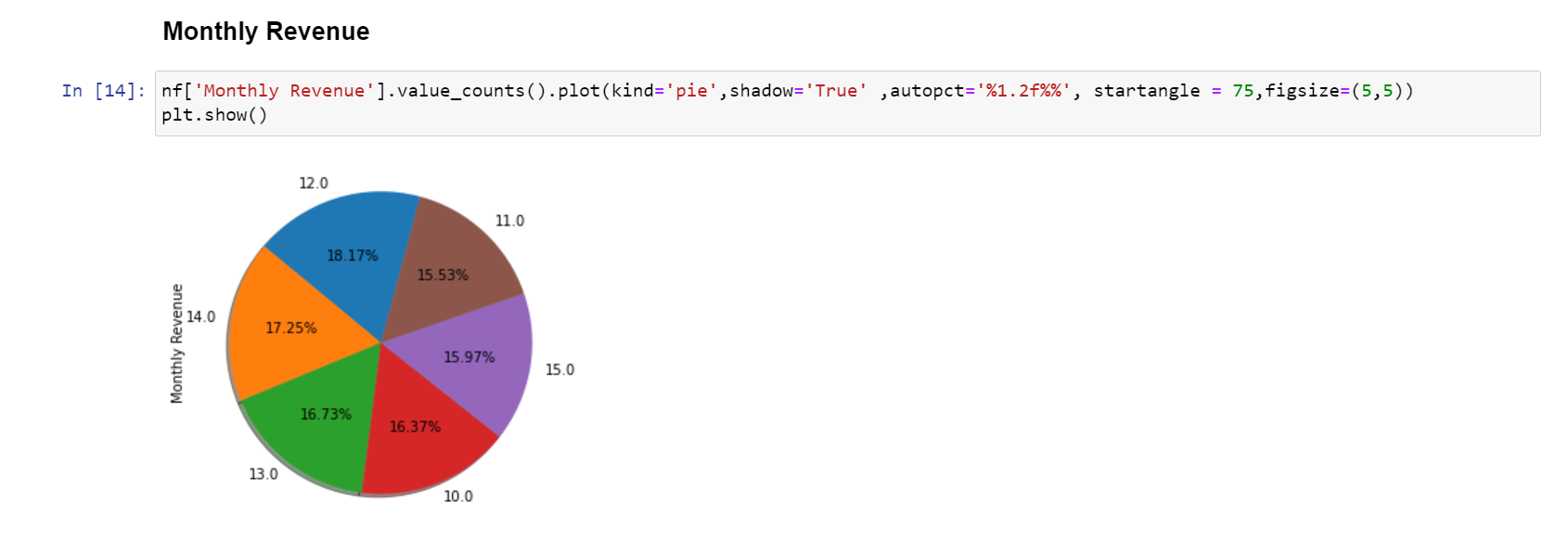
DATA VISUALIZATION

Data visualization is the graphical representation of different pieces of information or data, using visual elements such as charts, graphs, or maps. Data visualization tools provide the ability to see and understand data trends, outliers, and patterns in an easy, intuitive way

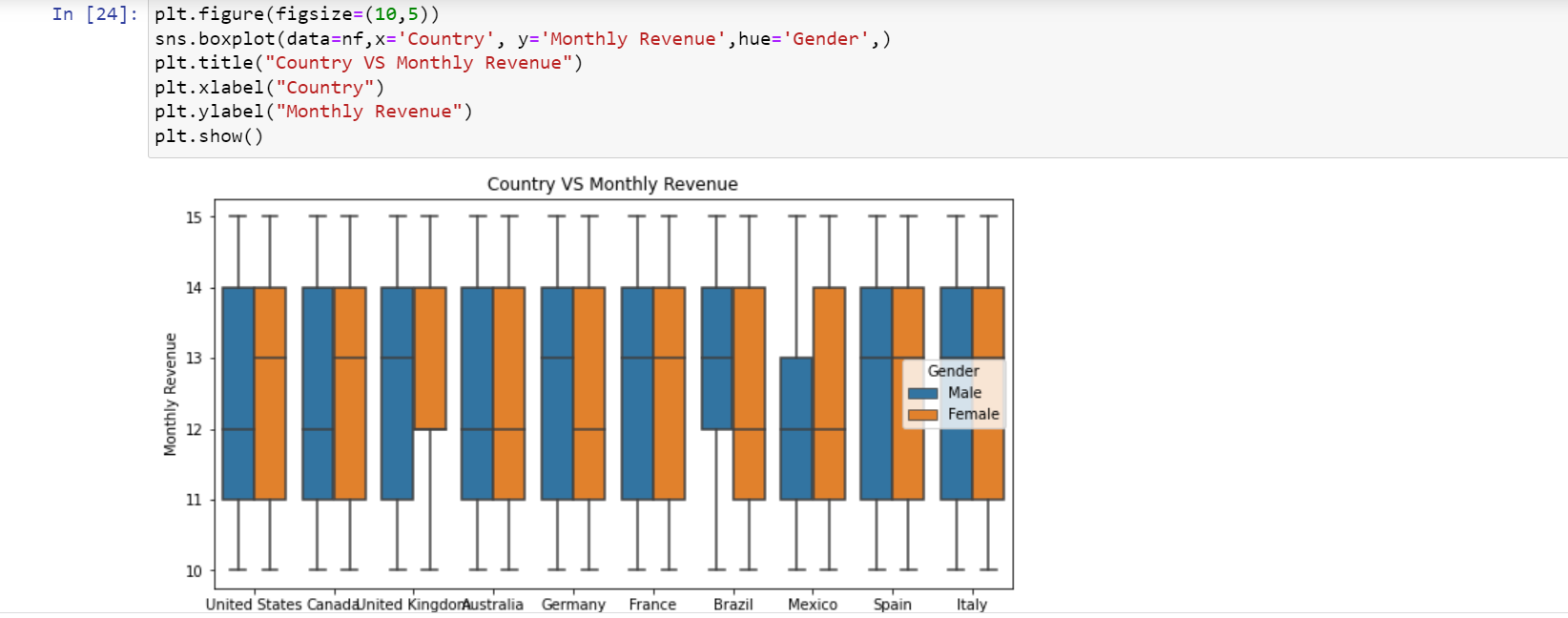


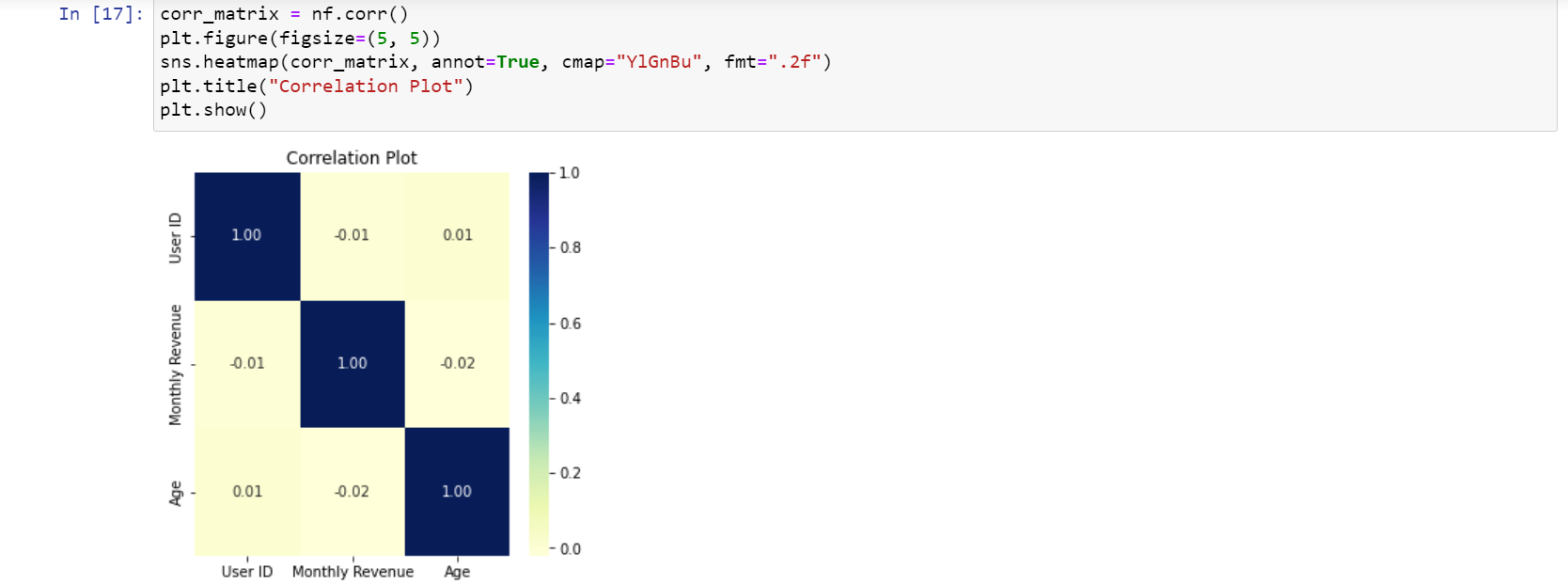












FINAL INSIGHTS

* Age group 20-29 spend more time on Netflix
* All countries have similar rate of spending time
* Age group 30-49 doing the most subscription and giving more revenue
* Basic type subscription has maximum purchasing

### Future Scope

* The future scope could involve predictive modeling for churn prevention, content recommendation enhancements, and refining subscription plans based on user segments. Additionally, using AI to personalize user experiences and refining content delivery strategies could also be part of the future scope.

### Advantages

**Understanding User Preferences**: By analyzing the subscription types chosen by users, Netflix can gain insights into which plans are most popular among their customer base. This knowledge can guide the platform in designing attractive subscription bundles and pricing models.

**Revenue Trends and Growth**: Monthly revenue analysis provides a clear view of Netflix's financial performance over time. Identifying revenue trends can help the company track its growth and assess the effectiveness of marketing campaigns or promotional strategies.

**Demographic Profiling**: Analyzing user age, gender, and country data allows Netflix to better understand the diversity of its user base. Tailoring content to different demographics can increase engagement and retention.

**Device Usage Insights**: By examining the devices through which users access the platform, Netflix can optimize its app and website experiences for various platforms, leading to improved user satisfaction.

**Plan Duration and User Retention**: Calculating the average plan duration offers valuable insights into user retention. This information can help Netflix devise strategies to enhance customer loyalty and reduce churn rates.

### Conclusion

In conclusion, this analysis and visualization of Netflix user data offer invaluable insights into user behavior, subscription trends, and revenue growth. By understanding user preferences, demographic profiles, and device usage patterns, Netflix can optimize its content offerings and user experience to better cater to its diverse customer base. The study also sheds light on user retention, helping Netflix focus on strategies to increase customer loyalty and reduce subscription cancellations. The data-driven approach presented in this analysis can serve as a foundation for Netflix's future business decisions, ensuring continuous improvement and innovation in the highly competitive streaming industry. As the digital landscape evolves, such analyses will remain crucial in maintaining Netflix's position as a market leader and providing its global audience with exceptnal streaming experiences.

### References websites

<https://www.kaggle.com/>

Greeks for Greeks

