# EXPLORATORY DATA ANALYSIS REPORT ON ENTERPRISE GPT PRODUCT USAGE DATA

**Usman Ghani**
ughani54@gmail.com

October 2, 2024

**ABSTRACT**

This report presents an Exploratory Data Analysis (EDA) conducted on a comprehensive dataset to uncover key insights and trends. Using Power BI, interactive dashboards was created to visualize data, track performance metrics, and analyze user engagement. The analysis focuses on key metrics such as total sessions, subscription patterns, and country-wise user distribution.

The Key Insights page summarizes overall user behavior, identifying trends in product usage, frequency of sessions, and engagement levels. The Subscription Analysis page provides an in-depth look at the subscription models, highlighting patterns in user subscriptions across various plans, including the distribution of active versus inactive users. Finally, the Country-wise User Metrics page breaks down user engagement and subscriptions by geographic location, identifying regions with the highest concentration of users and engagement levels.

Through this analysis, several key findings emerged, including the identification of major user activity spikes which returns to total of 1K of session from January 2024 to September 2024, regional preferences, and subscription patterns. These insights are crucial for informing data-driven decisions and optimizing business strategies to improve user retention and growth.

## 1 Introduction

This report aims to evaluate how users interact with Enterprise GPT, the pattern of their subscription behaviour and geographical usage. The Enterprise GPT Product usage Csv file is being added to the project as an input file. This contains only a single Csv file. Our system will first process the data set employing data cleaning tool of PowerQuery of PowerBI .Then employing the Data analysis ,visualization techniques. I would try to analyze the features affecting the usage .This project's one of major aim is Analyzing subscription trends, Examine user distribution across different regions to uncover geographic trends and understand overall user activity and engagement patterns.

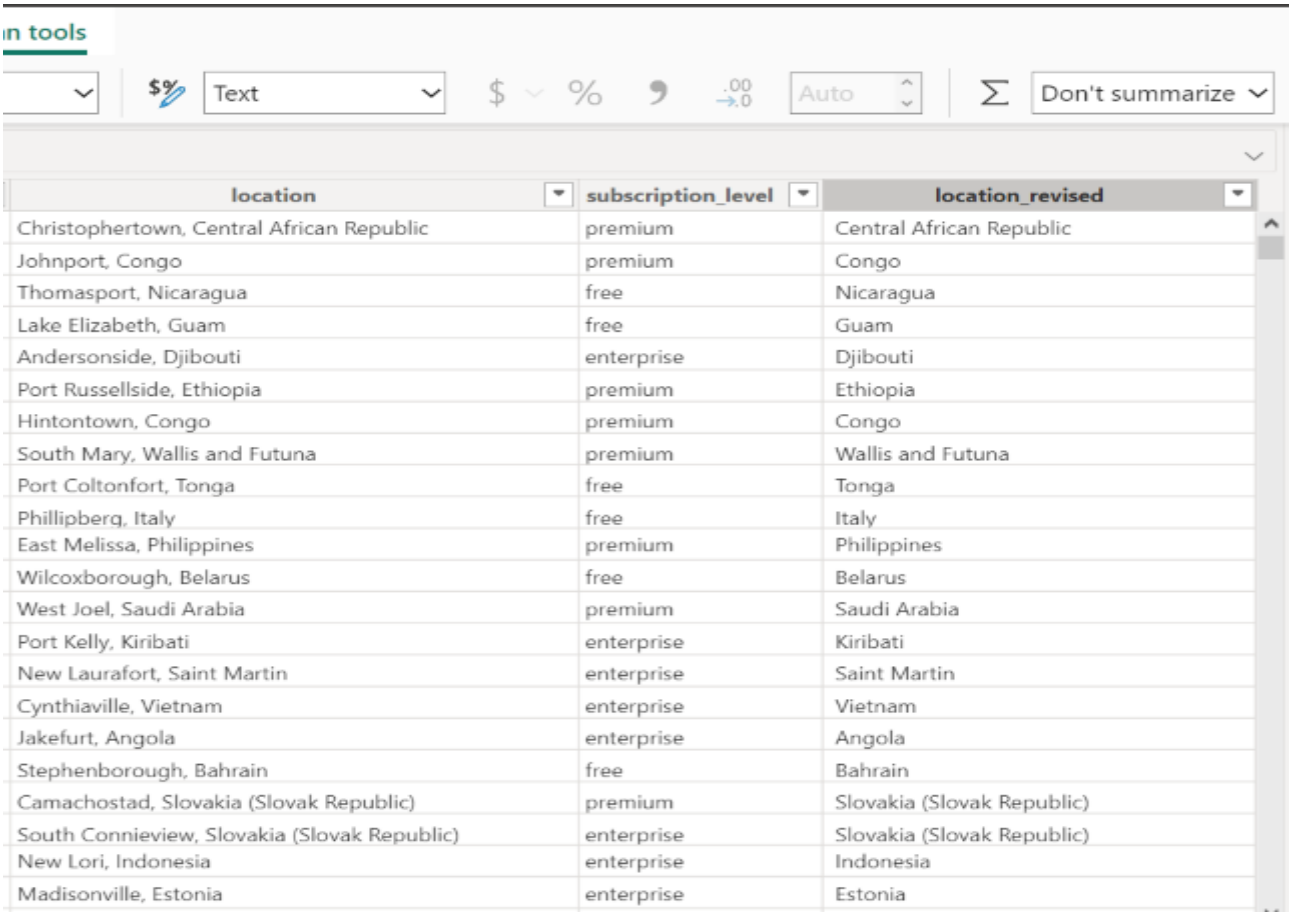
The major task can be broken down as 1) Data Cleaning 2) Data Processing 3)Analysis of features of data set 4) Answering important questions like top market where subscription are increasing and overall user activity. The following sections will provide a detailed breakdown of the data exploration process, findings from key metrics, and visual representations of the analyzed data, offering actionable insights for business decision-makers.

## 2 Data set

The Listings Data set has complete information on user id(num), session id(num), timestamp(date&time), interaction type(txt), query length(num), response time(num), user rating(num), error occurred(boolean), industry(txt), location(txt), subscription level(txt). interaction types has three categories, 1. Query 2. Error 3. feedback, user rating has been scaled from 1 to 5, Error occurred has two values True & False, Industry has three types Finance, Healthcare and education, Subscription level has three types free, Enterprise and Premium. Rows : 0 to 1000 Data columns (total 11 columns)

I then applied methods of data cleaning in PowerQuery, try to visualizing the amount of NULL VALUES but find none.

Upon reviewing the dataset, I identified discrepancies in the location column, particularly regarding incorrect city names associated with user data. To rectify this issue, I implemented a solution by creating a duplicate column, which I designated as **location_revised**. I then parsed the country data using the comma (",") delimiter, ensuring that any erroneous city entries were systematically removed from the dataset. This meticulous process left only the accurate country names, thereby enhancing the quality of the location data.



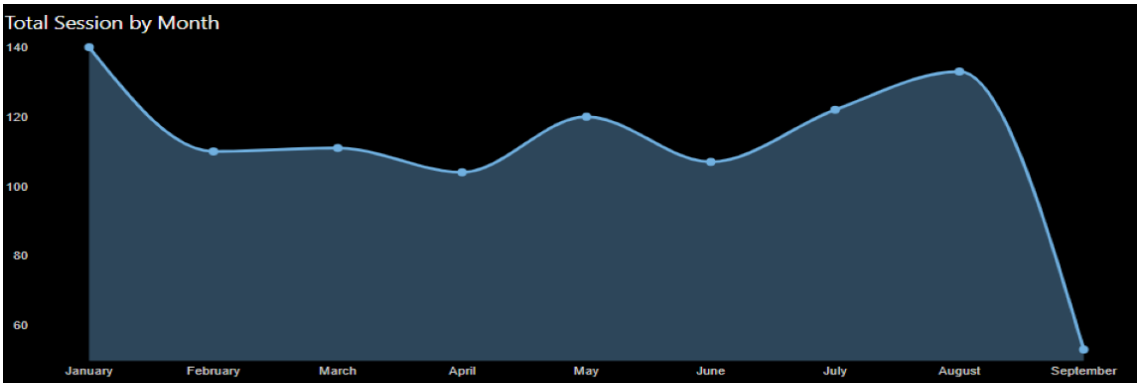

| location | subscription_level | location_revised |
|---|---|---|
| Christophertown, Central African Republic | premium | Central African Republic |
| Johnport, Congo | premium | Congo |
| Thomasport, Nicaragua | free | Nicaragua |
| Lake Elizabeth, Guam | free | Guam |
| Andersonside, Djibouti | enterprise | Djibouti |
| Port Russellside, Ethiopia | premium | Ethiopia |
| Hintontown, Congo | premium | Congo |
| South Mary, Wallis and Futuna | premium | Wallis and Futuna |
| Port Coltonfort, Tonga | free | Tonga |
| Phillipberg, Italy | free | Italy |
| East Melissa, Philippines | premium | Philippines |
| Wilcoxborough, Belarus | free | Belarus |
| West Joel, Saudi Arabia | premium | Saudi Arabia |
| Port Kelly, Kiribati | enterprise | Kiribati |
| New Laurafort, Saint Martin | enterprise | Saint Martin |
| Cynthiaville, Vietnam | enterprise | Vietnam |
| Jakefurt, Angola | enterprise | Angola |
| Stephenborough, Bahrain | free | Bahrain |
| Camachostad, Slovakia (Slovak Republic) | premium | Slovakia (Slovak Republic) |
| South Connieview, Slovakia (Slovak Republic) | enterprise | Slovakia (Slovak Republic) |
| New Lori, Indonesia | enterprise | Indonesia |
| Madisonville, Estonia | enterprise | Estonia |

## 3 Features and Analysis of Dataset drawing important inferences.

I have analyzed the listing data set and tried to answer a no of questions. Here presenting few important inferences from the dashboard.

The total number of sessions reached its highest point in August 2024, indicating a significant global upward trend in user access. This observation is particularly noteworthy given that the dataset only extends through September. The line chart illustrating total sessions by month, as shown in the first screenshot, clearly demonstrates a consistent and robust increase in user engagement over this period. This upward trajectory not only reflects growing interest in the platform but also highlights the effectiveness of the strategies implemented to enhance user experience and drive participation.

As such, it is essential to investigate the factors contributing to this surge in activity to sustain and build upon this momentum in the coming months.




### 3.1 Relationship Between Unique Users and Average Response Time.

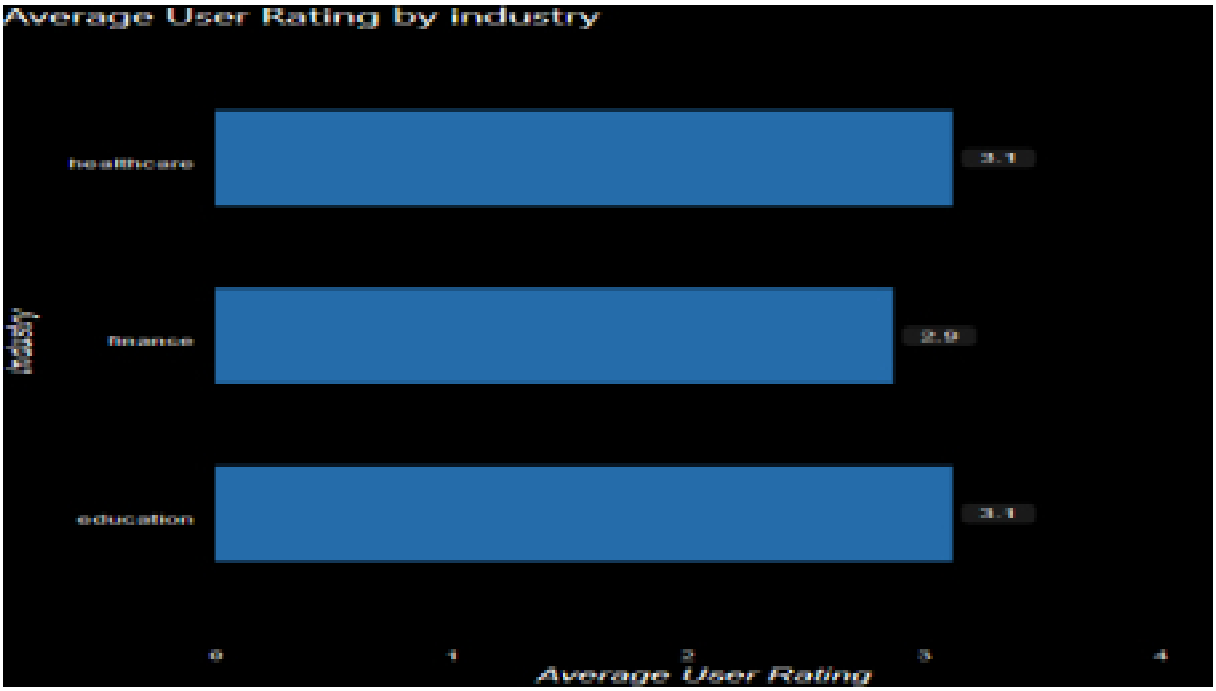
The data reveals an intriguing pattern between the number of unique users and average response time. As the number of unique users increases, the average response time tends to fluctuate. Notably, periods of significant user growth—such as in August 2024—coincide with variations in average response time, suggesting that increased demand may impact system performance.

| | | | |
|---|---|---|---|
| January | 2024 Year | 140 Unique Users | 1.92 Average Response Time |
| February | 2024 Year | 110 Unique Users | 2.01 Average Response Time |
| March | 2024 Year | 111 Unique Users | 2.07 Average Response Time |
| April | 2024 Year | 104 Unique Users | 2.05 Average Response Time |
| May | 2024 Year | 120 Unique Users | 2.00 Average Response Time |
| June | 2024 Year | 107 Unique Users | 2.04 Average Response Time |
| July | 2024 Year | 122 Unique Users | 1.99 Average Response Time |
| August | 2024 Year | 133 Unique Users | 2.03 Average Response Time |
| September | 2024 Year | 53 Unique Users | 1.91 Average Response Time |

### 3.2 Average User Rating by Industry.

This bar graph illustrates the average user rating across three key industries: Education, Healthcare, and Finance. Notably, users on both Free and Premium subscriptions provide identical average ratings across all industries, indicating a consistent level of satisfaction regardless of the subscription type. However, the Enterprise subscription reveals a distinct trend, particularly in the Finance sector, where it registers a lower average rating compared to the other industries. This discrepancy highlights a potential area for improvement within the Enterprise offering in the Finance industry, suggesting that user expectations may not be fully met. Addressing these concerns could enhance user

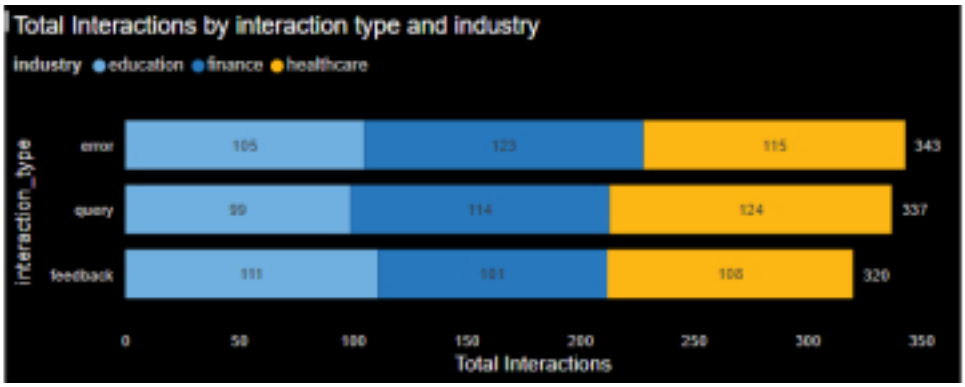
satisfaction and retention in this sector.




### 3.3 User Engagement: Total Interactions by Type and Industry

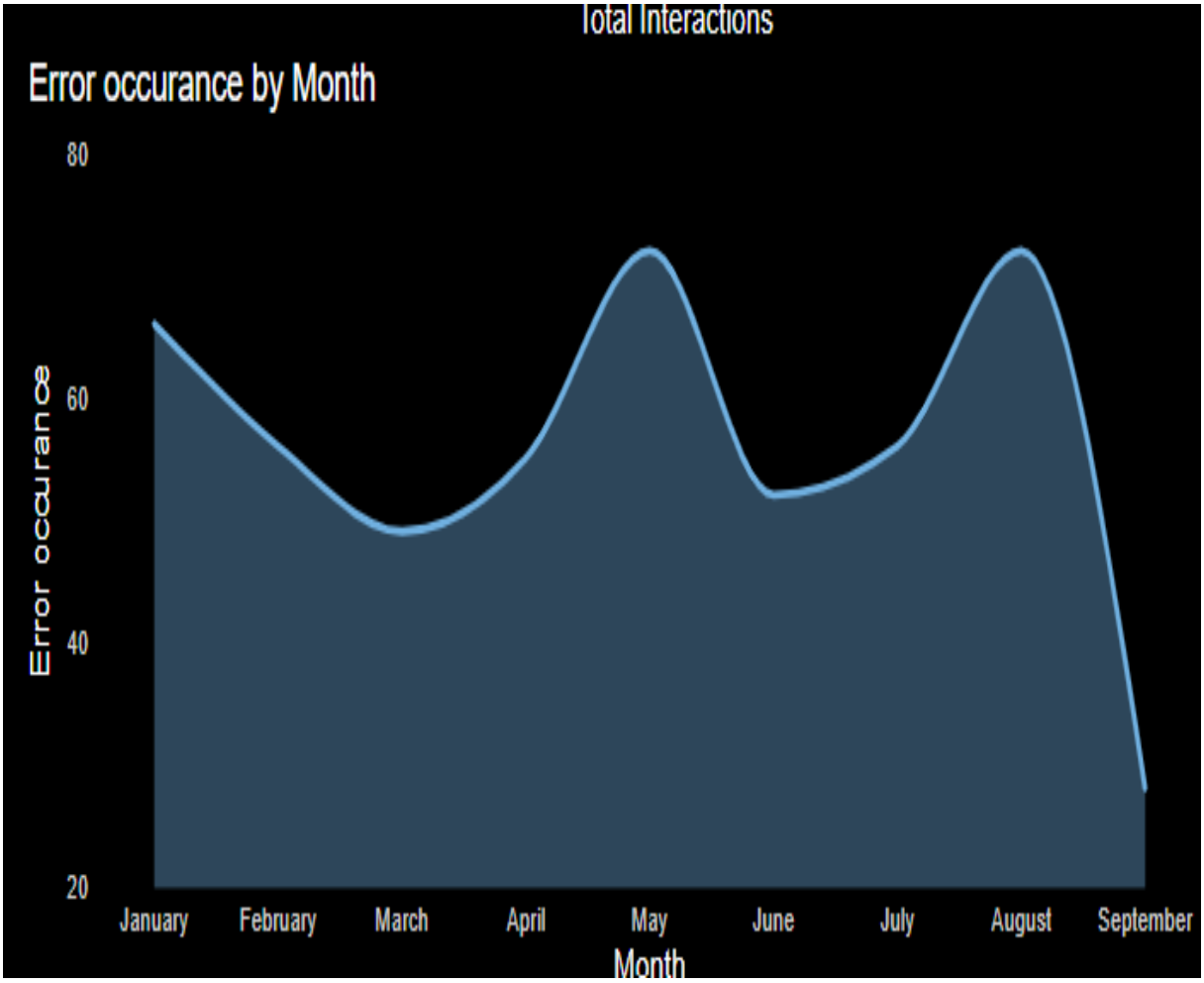
The second visual provides a detailed and comprehensive breakdown of total interactions categorized by both interaction type and industry. It delineates three distinct types of interactions: errors, queries, and feedback, each represented across the primary industries of education, healthcare, and finance. This visualization is instrumental in facilitating a nuanced understanding of user engagement patterns within these sectors, offering valuable insights into how different industries are leveraging the platform's capabilities. Upon analyzing the data, it becomes evident that both the education and healthcare sectors exhibit a relatively balanced distribution of interactions, indicating that users in these industries are actively engaging with various functionalities of the platform. In contrast, the finance sector stands out, revealing a significantly higher occurrence of error interactions. This disparity signals potential underlying issues that may be impacting user satisfaction and efficiency in the finance industry.

Furthermore, this insight is crucial for identifying specific areas within the finance sector that may require further investigation and improvement. By focusing on the trends illustrated in this visual, stakeholders can develop targeted strategies aimed at enhancing user experience across all subscription types. These strategies could include refining the interface, improving error handling processes, and increasing support resources for users in the finance sector. Ultimately, this analysis not only sheds light on current user behavior but also informs future enhancements that could lead to a more seamless and satisfying user experience.



### 3.4 Analysis of changes in occurance of Error over the month.

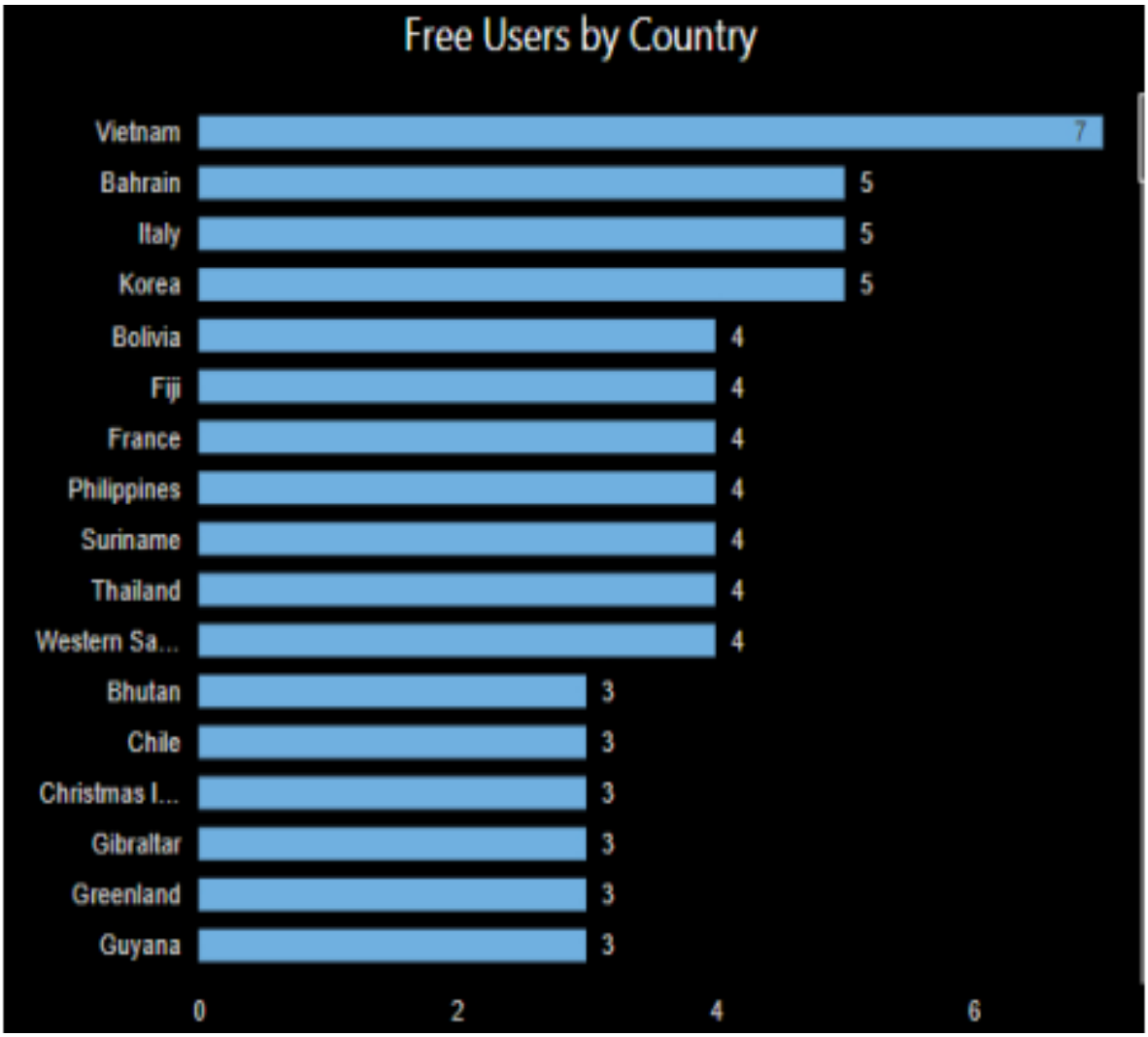
The third visual presents a month-by-month analysis of error occurrences within the platform, providing valuable insights into how errors fluctuate over time. This visualization is crucial for understanding the stability and reliability of the service, as it highlights patterns and trends related to error frequency.
Notably, the data reveals that the months of May and August experienced the most significant spikes in errors across all subscription levels. These peaks indicate potential underlying issues or changes within the platform that may have adversely affected user experiences during those periods. Identifying these months is essential for stakeholders, as it suggests the need for a thorough investigation into the causes of these errors and the implementation of corrective measures.
By focusing on the significant spikes in May and August, teams can prioritize investigations and implement corrective measures to enhance system performance, thereby improving user satisfaction across all subscription levels. Through ongoing monitoring and analysis of this data, organizations can strive to achieve a more robust and reliable user experience, ultimately fostering greater trust and engagement among their users.




### 3.5 Countries With Highest Usage Of Free Subscription.

The first visual on the Country-wise User Metrics page illustrates the distribution of free users, with Vietnam, Bahrain, and Italy leading in the number of users accessing the platform at no cost. Despite these countries having relatively wealthy populations, a significant portion of users remains on the free subscription tier. This presents a unique opportunity for strategic initiatives aimed at converting free users into premium subscribers.
To effectively transition these users to a paid subscription model, several targeted strategies can be implemented. First, tailored marketing campaigns that highlight the exclusive benefits of premium subscriptions—such as enhanced features, advanced analytics, and personalized support—should be launched. By clearly communicating the added value and advantages that come with premium access, users may be more inclined to upgrade.

Additionally, offering limited-time promotions or discounts can create a sense of urgency, encouraging free users to experience the premium features at a lower initial cost. This approach can be particularly effective in regions like Vietnam, Bahrain, and Italy, where users may be more receptive to promotional incentives.
Engaging free users through webinars, tutorials, or free trials of premium features can also help demonstrate the platform's value, allowing users to experience firsthand the enhanced capabilities available to premium subscribers.
By implementing these strategies, organizations can effectively nurture their free user base in Vietnam, Bahrain, and Italy, ultimately converting them into loyal premium subscribers and enhancing overall user engagement.

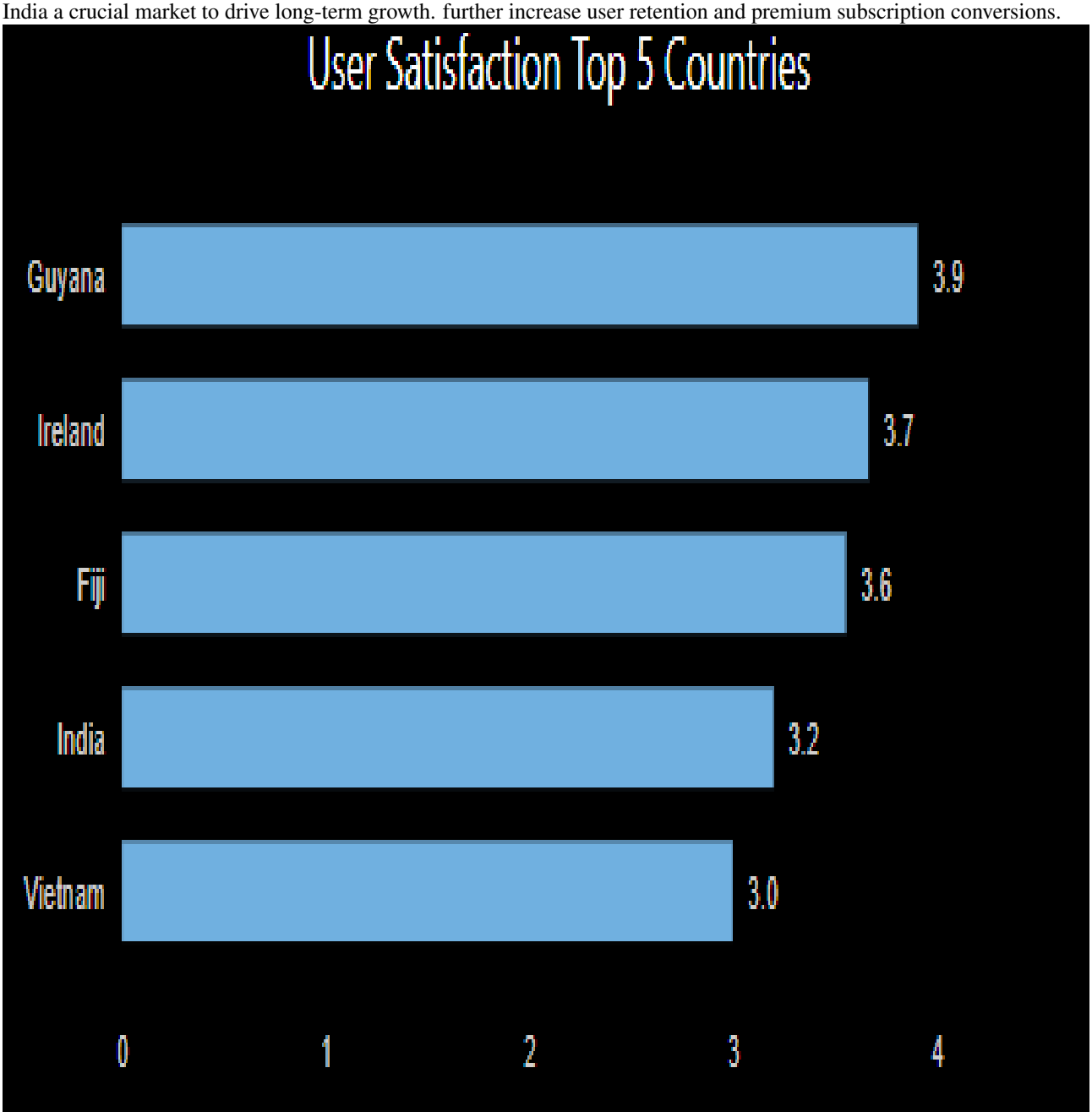



### 3.6 User Satisfaction Top 5 Markets

The second visual on the "Country-wise User Metrics" page showcases the top five countries based on average user satisfaction ratings, with Guyana, Ireland, Fiji, India, and Vietnam leading the list. Among these, India stands out as a key market with a significant user base and high satisfaction levels. The positive feedback from users in India indicates strong engagement with the product, suggesting that the features and services offered resonate well with this audience. Given India's large population and growing digital economy, this market holds considerable potential for expansion.

By focusing on strategies to maintain and increase user satisfaction in India, the company could capitalize on this momentum, possibly converting free users to premium plans through localized offers, improved customer support, or targeted marketing campaigns. The high average user ratings reflect a strong foundation of user trust and loyalty, making

India a crucial market to drive long-term growth. further increase user retention and premium subscription conversions.

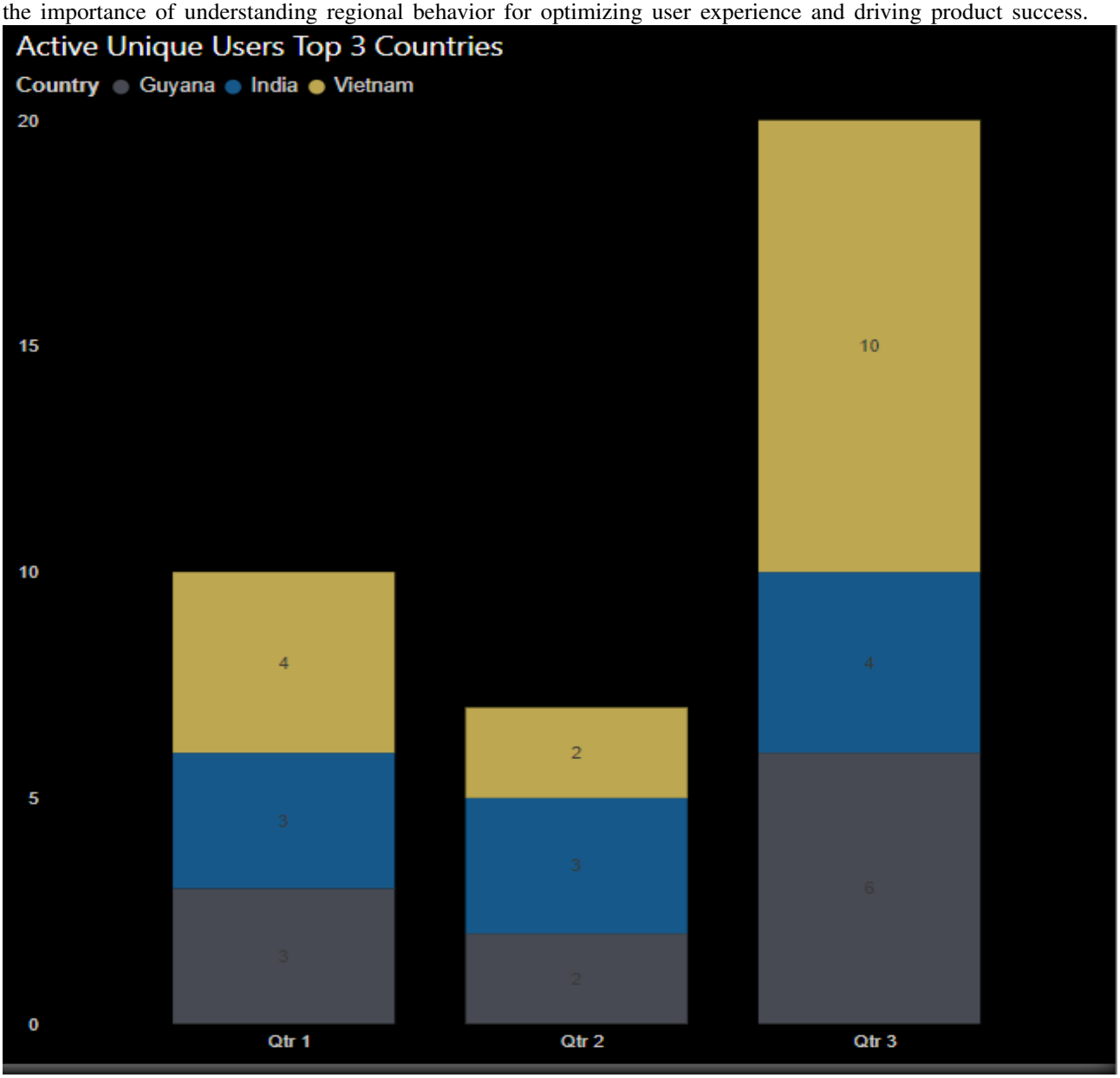



### 3.7 Top 3 Countries With Most Number Of Active Unique User

The third visual on the "Country-wise User Metrics" page highlights the top three countries with the most active users by month: Vietnam, India, and Guyana. Vietnam consistently ranks as the country with the highest number of active users, showing a strong and sustained level of engagement with the product. India and Guyana follow closely, with significant user activity observed across months.

Vietnam's dominance in user activity reflects a vibrant user base that frequently interacts with the product. This makes Vietnam a crucial market for future engagement strategies, potentially through premium offerings or tailored services. India, with its large and steadily active user base, presents an opportunity for growth, while Guyana's user engagement highlights the potential for expanding user retention strategies in smaller markets. These patterns underscore

the importance of understanding regional behavior for optimizing user experience and driving product success.




## 4 Methods used in Analysis

To conduct the Exploratory Data Analysis (EDA) and derive meaningful insights from the dataset, various methods and techniques were employed within Power BI, leveraging its data transformation, DAX, and visualization capabilities. The following key methods were implemented:

### 4.1 Data Cleaning and Transformation

#### 4.1.1 Location Parsing

I identified discrepancies in the "Location" column, where city names were incorrectly formatted. To rectify this, a new column, location_revised, was created, parsing the country data after the delimiter (","). This approach removed the erroneous city data and left only accurate country names, ensuring consistent and clean geographical information for analysis.

#### 4.1.2 Null Value Analysis

I performed an assessment of null values across the dataset. No significant null values were found, suggesting the dataset was largely complete and reliable for further analysis.

### 4.2 Creation Of Custom Mesures

Several measures were created using DAX (Data Analysis Expressions) to extract and aggregate data insights effectively:

#### 4.2.1 Total Session

Created the measure Total Sessions C̄OUNTROWS $'enterprise\_gpt\_product\_usage'$ to count the total number of user sessions. This was crucial in understanding user engagement trends and identifying monthly peaks, such as in August 2024, where the number of sessions significantly increased.

#### 4.2.2 Average User Rating

Another measure was developed to calculate the average user rating, providing a deeper understanding of user satisfaction across different industries and countries. This measure was particularly useful in identifying patterns in regions with high satisfaction levels, such as Guyana, Ireland, and India.

#### 4.2.3 Error Occurance By Month

To analyze system performance, I created a measure that tracked the occurrence of errors over time. This allowed for the visualization of spikes in errors, particularly in May and August, highlighting areas where platform stability could be improved.

### 4.3 Interactive Visualization

Various visualizations were created to interpret the data effectively:

#### 4.3.1 Slicer For Subscription Type

A slicer was added to filter data based on subscription type (Free, Premium, Enterprise), allowing for a segmented analysis of user engagement, satisfaction, and error occurrences based on the type of subscription users hold.

#### 4.3.2 Bar Chart and Line Graph

These visualizations were used to represent key metrics, such as total interactions by industry and error occurrences by month. Bar charts allowed for easy comparison across industries, while line graphs helped identify temporal trends in user activity and errors.

## 5 Insights: It's Relevance and Practicality

The insights drawn from the analysis provide actionable information that directly addresses key business objectives. The data reveals patterns in user behavior, subscription preferences, and geographic engagement that are crucial for making informed decisions. Key takeaways include:

### 5.1 Subscription Insights

A large number of users in economically strong countries like Vietnam, Bahrain, and Italy are currently on the free subscription plan. This presents an opportunity to implement targeted strategies for converting free users into paid subscribers, such as offering limited-time promotions, exclusive features for premium users, or incentivized upgrades.

### 5.2 User Satisfaction and Engagement

The top five countries by average user satisfaction (Guyana, Ireland, Fiji, India, Vietnam) offer opportunities for localized marketing campaigns to boost engagement and retention. India, in particular, shows strong user satisfaction, which can be leveraged to drive further growth. By understanding where satisfaction is high, efforts can be focused on

maintaining this momentum, while countries or industries with lower satisfaction levels—such as the enterprise finance sector—can be targeted for improvements.

### 5.3 Error Occurrence Trends

The spikes in error occurrences during May and August, across all subscription types, suggest potential issues during these periods. This is a valuable insight for the technical team, as addressing these issues can improve platform stability and enhance the overall user experience. Preventing future error spikes could lead to reduced churn rates and higher user satisfaction.

## 6 Strength of Recommendations

The recommendations are grounded in data and focus on directly addressing the trends observed:

### 6.1 Subscription Conversion

The suggestion to focus on converting free users in countries like Vietnam, Bahrain, and Italy is well-supported by the data. These countries show a high volume of free users despite their economic capacity for paid plans. This provides a clear opportunity to push for subscription upgrades through targeted campaigns.

### 6.2 Localized Strategy

The data reveals high user satisfaction in specific countries, particularly India. Stakeholders can capitalize on this by developing localized strategies to engage these users further, ensuring continued satisfaction while growing the user base in these regions.

### 6.3 Technical Improvements

The error spike insights provide actionable information for the technical teams. By investigating the root causes of the spikes in May and August, resources can be allocated to address these issues, improving the platform's reliability.

## 7 Data-Driven Support for Insights

The findings and recommendations are directly supported by the data collected and analyzed. Visualizations clearly depict the patterns, and custom measures provide precise metrics that stakeholders can rely on for decision-making.

### 7.1 Subscription Data

The breakdown of subscription types and user distribution provides clear guidance on where to focus efforts for conversions and engagement. The data shows not just the volume of users but also their behaviors and trends across different subscription models.

### 7.2 User Engagement

Insights on user satisfaction and interaction patterns are reinforced by visual representations, which make the trends easy to interpret. This enables stakeholders to quickly grasp where the business is performing well and where improvements are needed.

### 7.3 Error Data

The error trends by month provide an immediate red flag for when the platform may be under performing. This data-driven insight allows stakeholders to preemptively allocate resources to mitigate these issues and enhance overall platform performance.

**In summary, the insights and recommendations are practical, actionable, and well-supported by the data. They provide a clear road map for enhancing user satisfaction, improving platform stability, and driving subscription growth. Stakeholders can confidently use these findings to make informed, data-backed decisions.**

### 7.4 GITHUB LINK FOR REPORT :

https://github.com/aashirwad01/Project_knn.git