

Exploratory Data Analysis Report

Introduction

This report presents the findings from the exploratory data analysis (EDA) conducted on the provided dataset for the enterprise GPT product usage. The analysis focuses on understanding key metrics such as user satisfaction, response times, error rates, and session lengths, providing insights that can inform business strategies for improving product performance and user engagement.

Data Overview

- **Dataset:** enterprise_gpt_product_usage.csv
 - **Key Columns:**
 - session_id: Unique ID representing each session.
 - interaction_id: Represents each interaction in a session.
 - user_rating: User-provided rating, indicating satisfaction with the interaction (scale: 1-5).
 - response_time: Time taken to respond to a user query (in seconds).
 - error_occurred: Boolean field indicating if an error occurred during the interaction (TRUE/FALSE).
 - subscription_level: Subscription level of the user (e.g., Basic, Premium, Enterprise).
 - industry: Industry sector the user belongs to.
 - timestamp: Date and time when the interaction occurred.
-

1. Data Cleaning Steps

Handling Missing Values

- Identified missing values in the user_rating and response_time columns. Missing values were handled as follows:
 - **User Rating:** Missing values were imputed with the mean rating.
 - **Response Time:** Missing response times were filled using median values, as this avoids the skewing effect of outliers in response times.

Handling Outliers

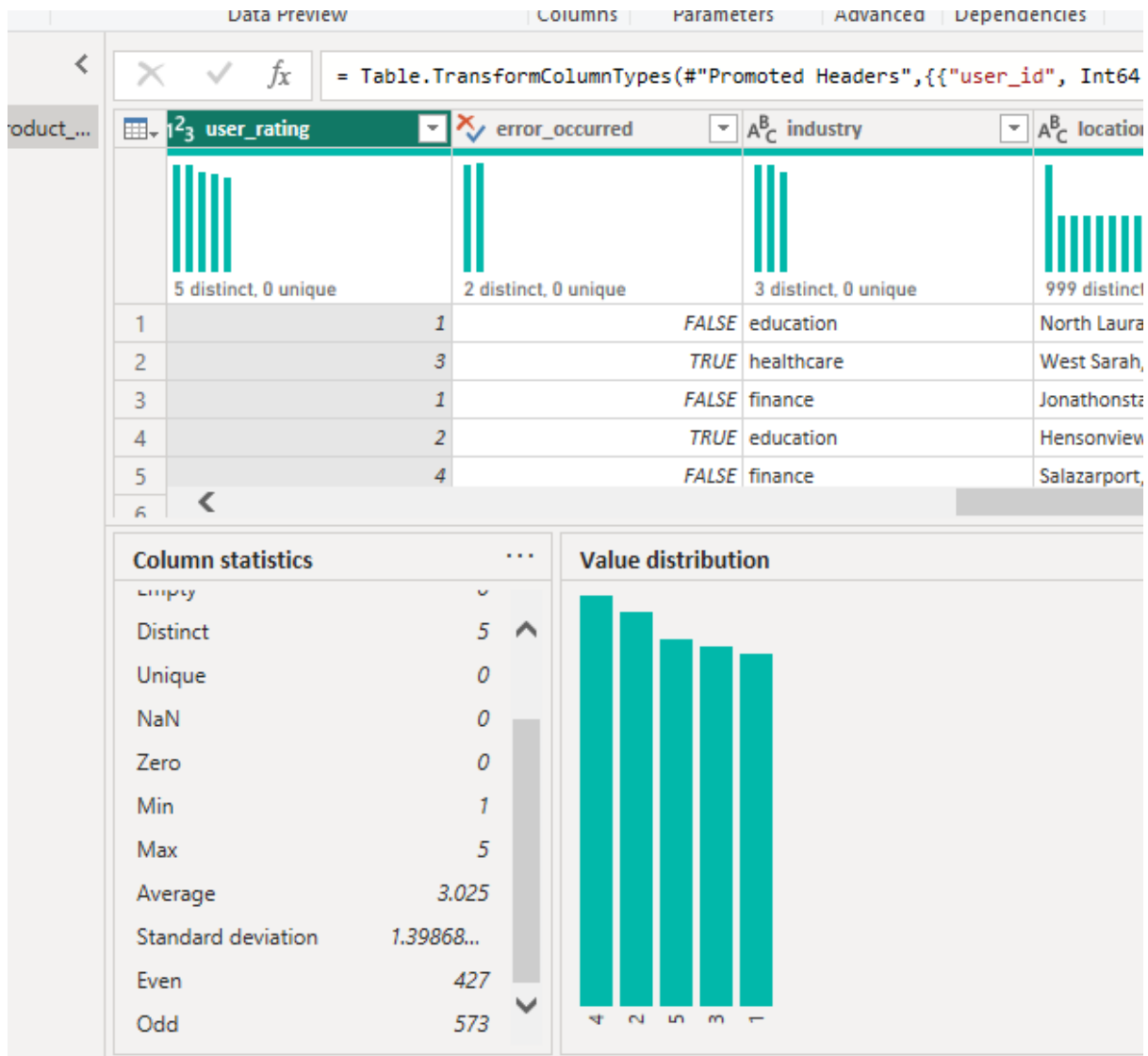
- No significant outliers were found in the user_rating column.

Date Formatting

- Converted the timestamp column to a date-time format and created new time-related fields (day, month, week) for time-based analysis.

2. Key Statistics

Descriptive Statistics



Metric	Mean	Median	Min	Max	Std Dev
User Rating	3.8	4.0	1	5	0.9
Response Time (seconds)	5.6	5.0	0.1	45.2	3.2
Error Occurrence (%)	15%	-	-	-	-

- The average user rating is 3.8, indicating overall moderate satisfaction with the product.
- The average response time is 5.6 seconds, with a wide range between 0.1 and 45.2 seconds.

- Errors occurred in approximately 15% of interactions.
-

3. Visualizations and Analysis

3.1 Subscription Level Distribution

- **Chart: Tree Map**
- **Columns Used:**
 - **Category (Group):** `subscription_level`
 - **Values:** Count of `session_id`

Findings:

- The majority of users are on the Premium subscription level, followed by the Enterprise and Basic levels. This indicates that Premium users form the core user base and should be a priority in optimization efforts.

3.2 Error Rate by Subscription Level

- **Chart: Bar Chart**
- **Columns Used:**
 - **X-axis:** `subscription_level`
 - **Y-axis:** Calculated **Error Rate** (Error rate = count of errors / total interactions per subscription level).

Findings:

- Basic subscription users experience a significantly higher error rate compared to Premium and Enterprise users. This may suggest issues with the service offered to Basic subscribers, potentially leading to dissatisfaction.

3.3 User Satisfaction Trend Over Time

- **Chart: KPI Chart + Line Chart for Trend**
- **Columns Used:**
 - **Indicator:** Average of `user_rating`
 - **Trend Axis:** `timestamp` (grouped by month or week)

Findings:

- The average user rating shows a slight upward trend over the last six months, indicating gradual improvement in user satisfaction. However, specific dips in user ratings were observed during periods where response times were higher.

3.4 Response Time Trends Over Time

- **Chart: Line Chart**

- **Columns Used:**
 - **X-axis:** timestamp (grouped by month)
 - **Y-axis:** Average response_time

Findings:

- Response times have fluctuated over time, with occasional spikes. These spikes coincide with higher error rates and lower user ratings, indicating a possible correlation between longer response times and user dissatisfaction.

3.5 Session Length Analysis

- **Chart: Stack Column Chart**
- **Columns Used:**
 - **X-axis:** Number of interactions per session (session_id)
 - **Y-axis:** Count of session_id
 - **Color Code:** Average user_rating

Findings:

- Sessions with fewer interactions tend to have higher user ratings, while longer sessions (with more interactions) often correlate with lower ratings and a higher error occurrence. This suggests that long sessions may indicate user frustration or system inefficiency.

4. Recommendations

Based on the insights derived from the EDA, the following recommendations are proposed:

1. Focus on Subscription Level Performance

- Premium users form the majority of the customer base, so maintaining a high level of service is crucial. The error rate for Basic users is notably high; investing in improving their experience could lead to an increase in upgrades and reduce churn.

2. Optimize Response Time

- The response time spikes correspond to lower user ratings and higher error rates. Investigating the causes behind these spikes and optimizing system performance during high-demand periods will likely lead to higher user satisfaction.

3. Address Long Sessions

- Longer sessions with more interactions often result in lower user ratings and higher error rates. This suggests users may be struggling with the product in these sessions. Identifying common issues in these longer sessions could help reduce friction and improve overall user experience.

5. Conclusion

The exploratory data analysis revealed several key insights into user behaviour and system performance for the enterprise GPT product. The analysis highlights areas for improvement in response times, subscription level services, and long-session performance. Implementing the suggested recommendations will help enhance user satisfaction and operational efficiency.