

6/42 Lotto Time Series Analysis

DS150 - Time Series Analysis & Stochastic Processes - Module Exam 1

PRESENTATION BY: AGUAVIVA, CHAMORRO, VILLAPANDO

Problem Statement

This analysis utilizes time series techniques to uncover **trends and patterns** within simulated Philippine lotto data.

Examining data from all regions provides a comprehensive national perspective, allowing us to identify potential regional **variations or similarities** in winning number patterns over time.

Motivation

By analyzing simulated data for the 6/42 lotto across all Philippine regions, we hope to **gain insights into potential trends** and recurring number combinations.

Additionally, the chosen format (6/42) offers a manageable complexity for time series analysis while potentially providing a longer data record for a more robust analysis.

Data Source

Simulated PCSO Lottery Dataset Link: This link provides a downloadable .zip archive containing various simulated lottery datasets. It is important to note that this data is for educational purposes only and does not include any real information from PCSO.

lotto_draw_results.csv: This CSV file contains the simulated draw results, including winning numbers and the number of winners for each draw

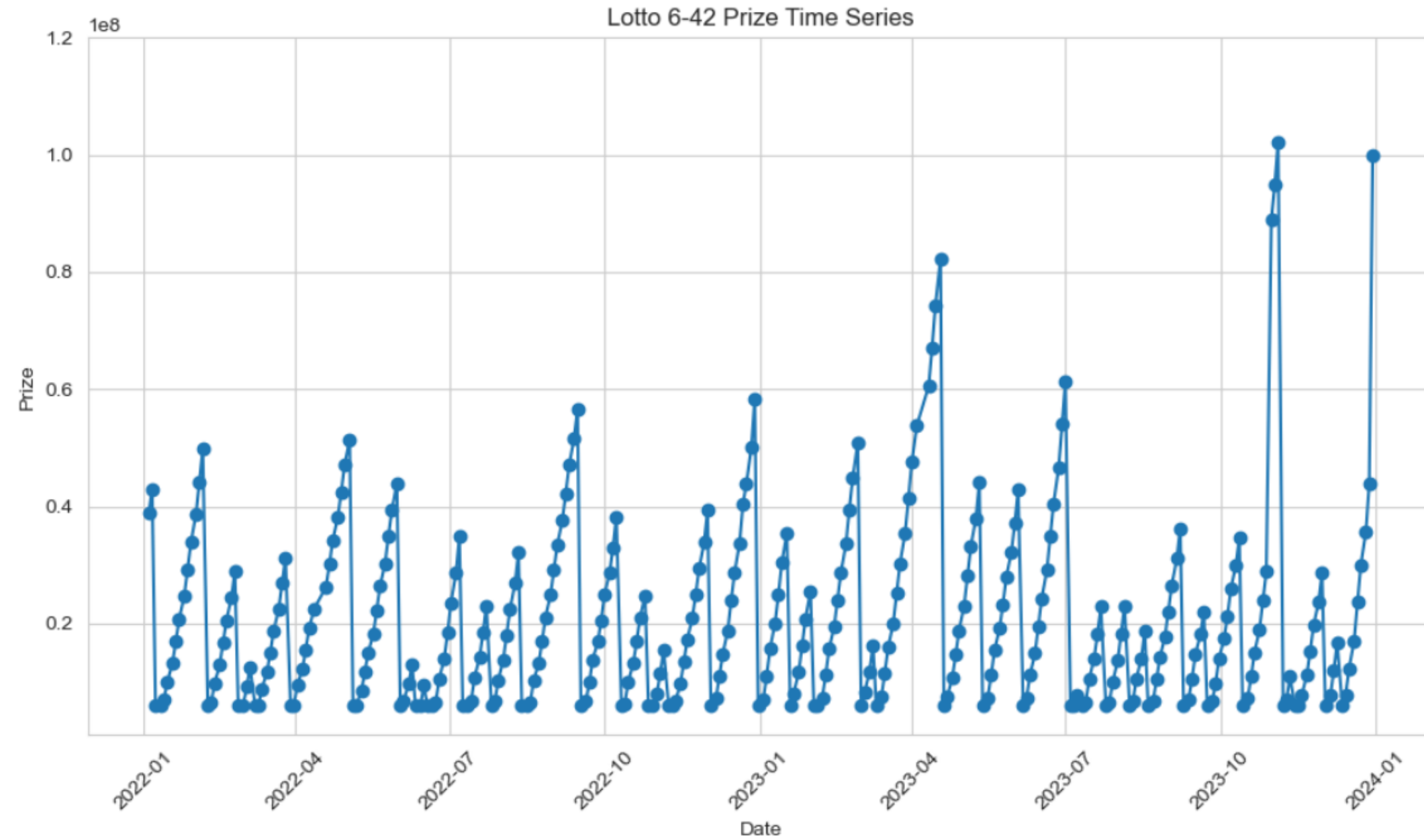
Data Dictionary

Name	Data Type	Definition
dates	datetime	Timestamp indicating when the ticket was submitted.
num1 - num6	int	lottery number selected on the ticket (between 1 and 42)
location	string	Region where the ticket was submitted ("Luzon", "Visayas", "Mindanao", or "NCR").
whole_nums	string	Formatted string representing the chosen lottery numbers (e.g., "01-02-06-16-27-31").
month	string	Month ("January", "February", etc.) extracted from the dates column.

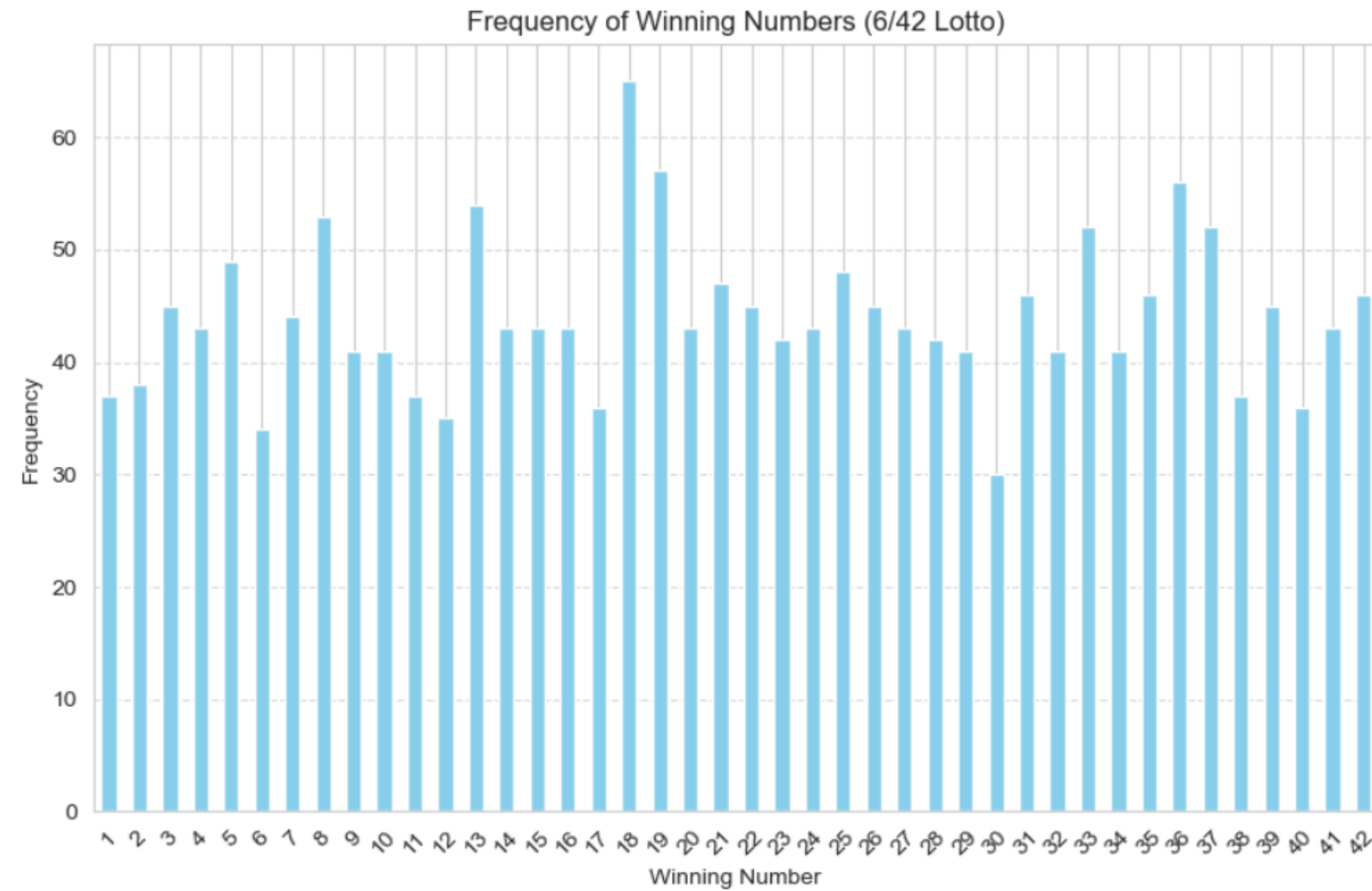
Data Dictionary

Name	Data Type	Definition
lotto_game	string	Always "6/42" since this DataFrame focuses on that specific category.
winning_nums	string	Formatted string representing the winning lottery numbers for a particular drawing (e.g., "01-02-06-16-27-31").
date	datetime	Date of the lottery drawing.
prize	float	Total prize money awarded for the winning numbers in Philippine pesos.
num_winners	int	Number of tickets that matched the winning numbers for that particular drawing.

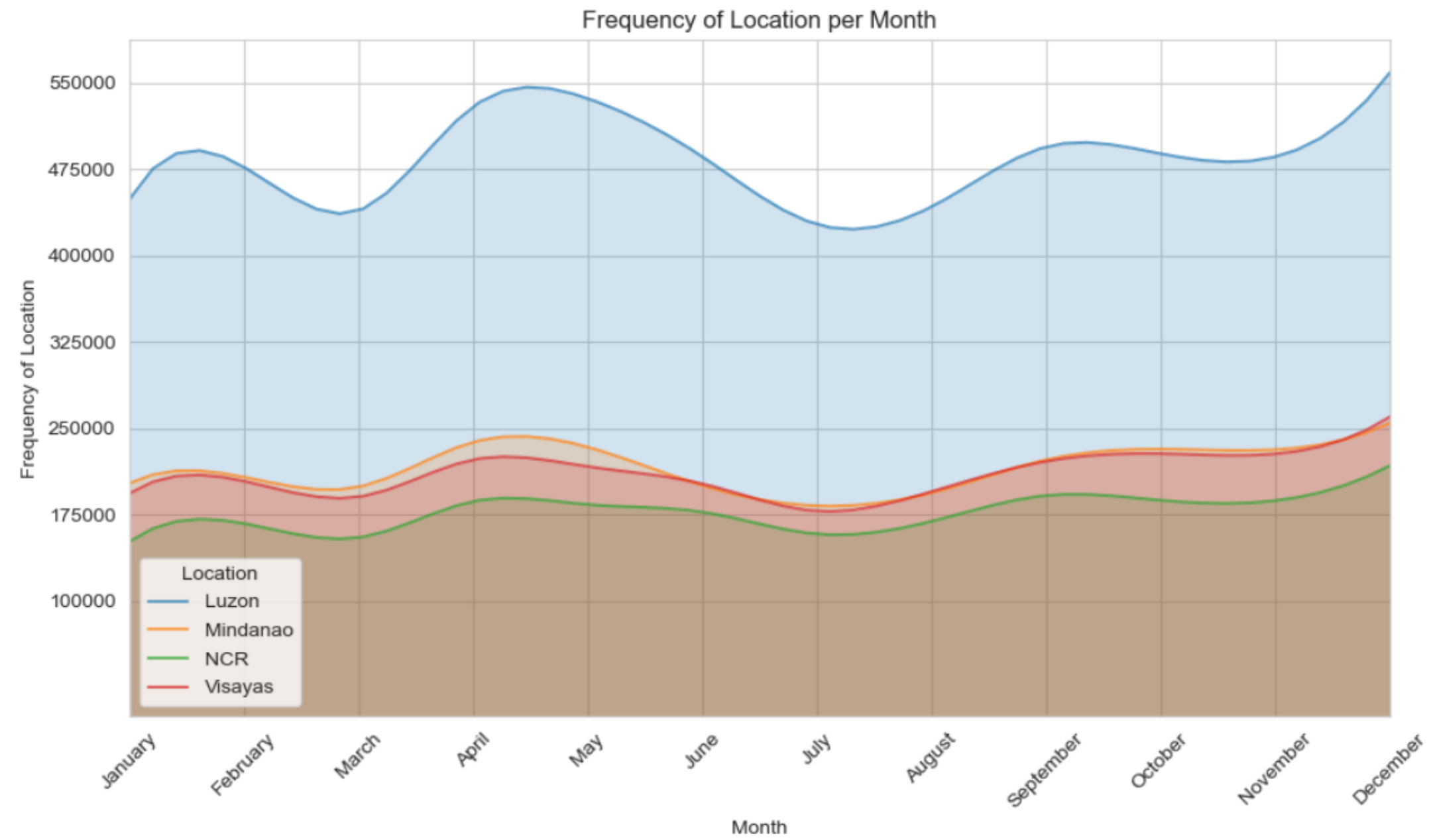
Exploratory Data Analysis



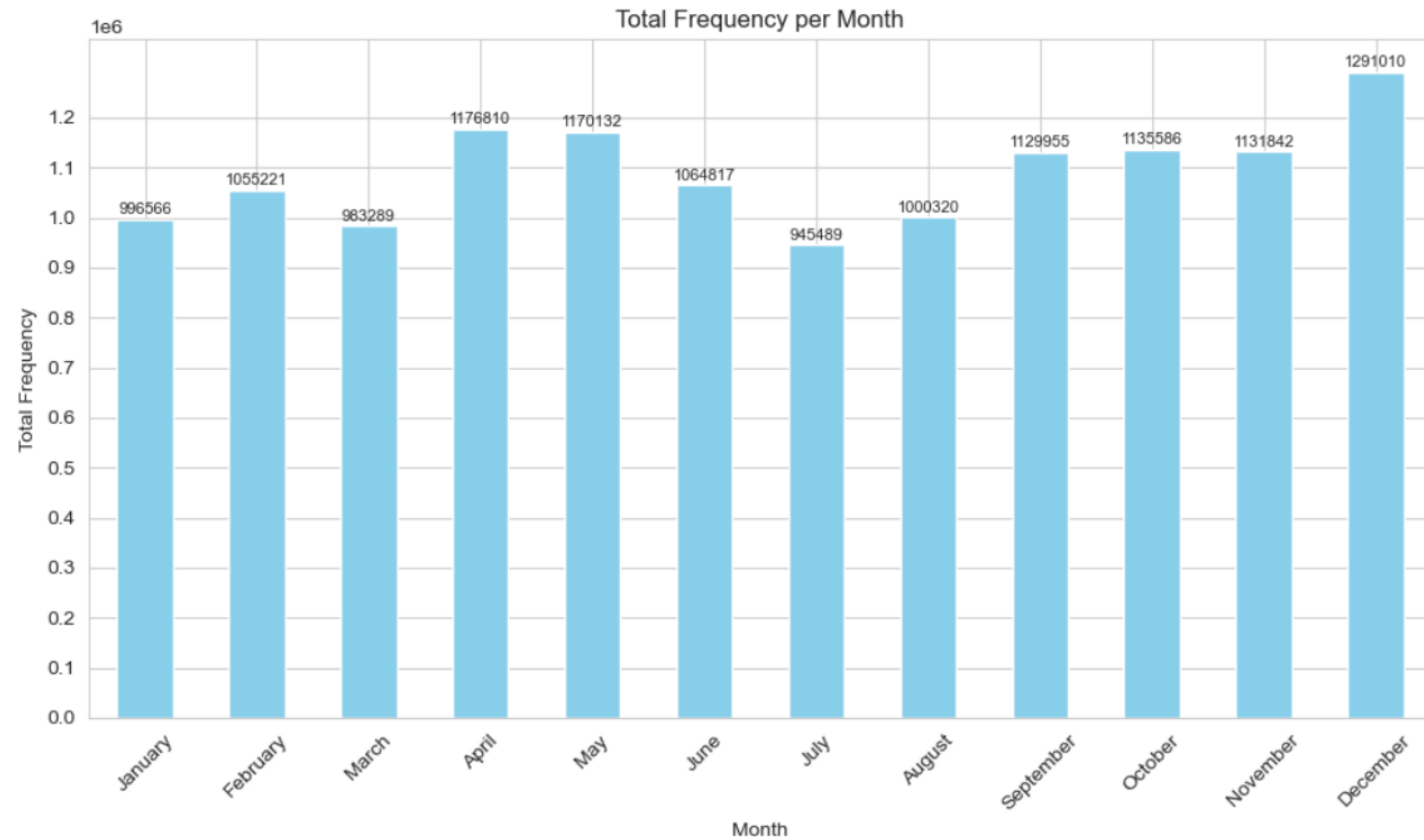
Exploratory Data Analysis



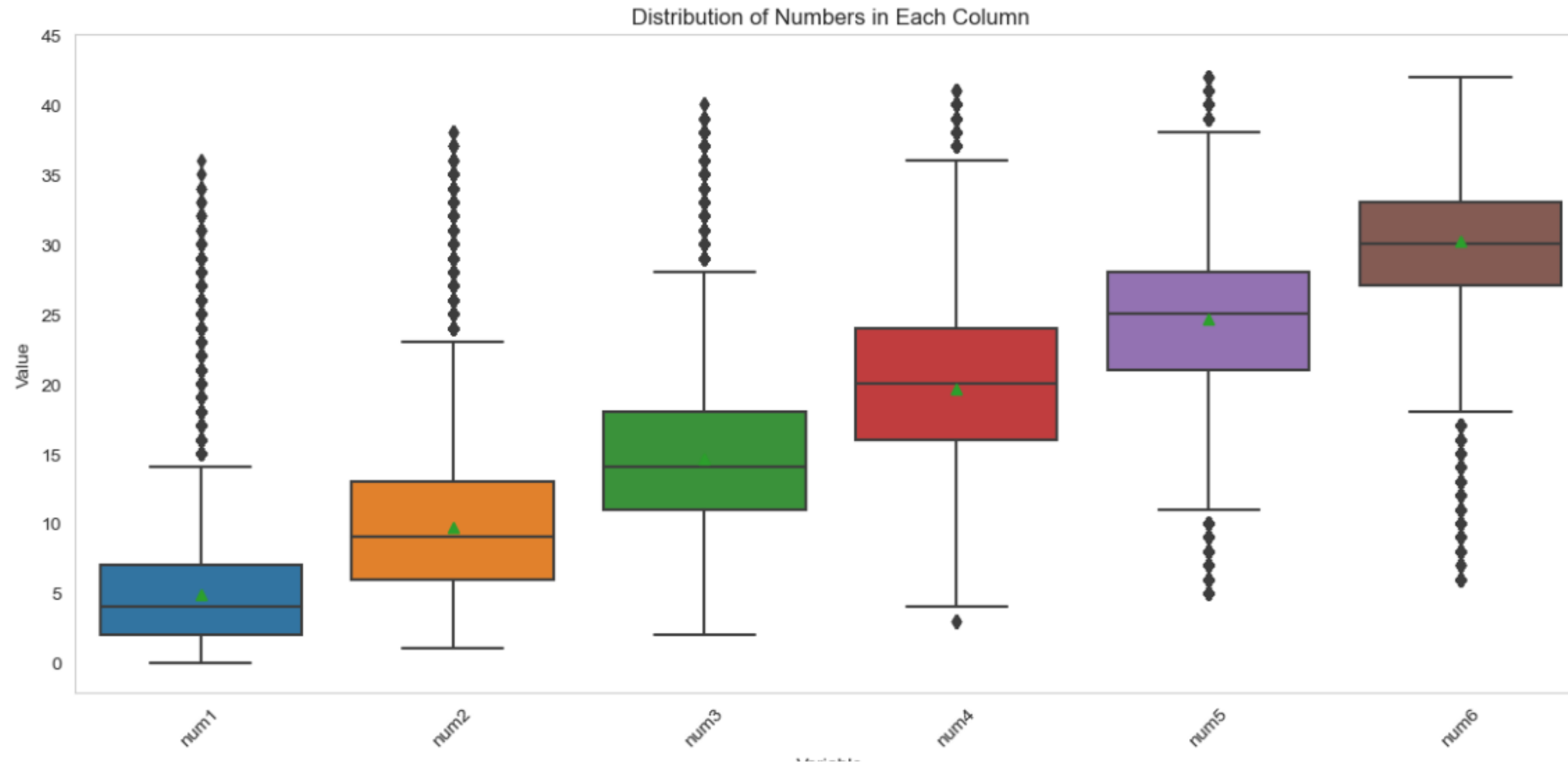
Exploratory Data Analysis



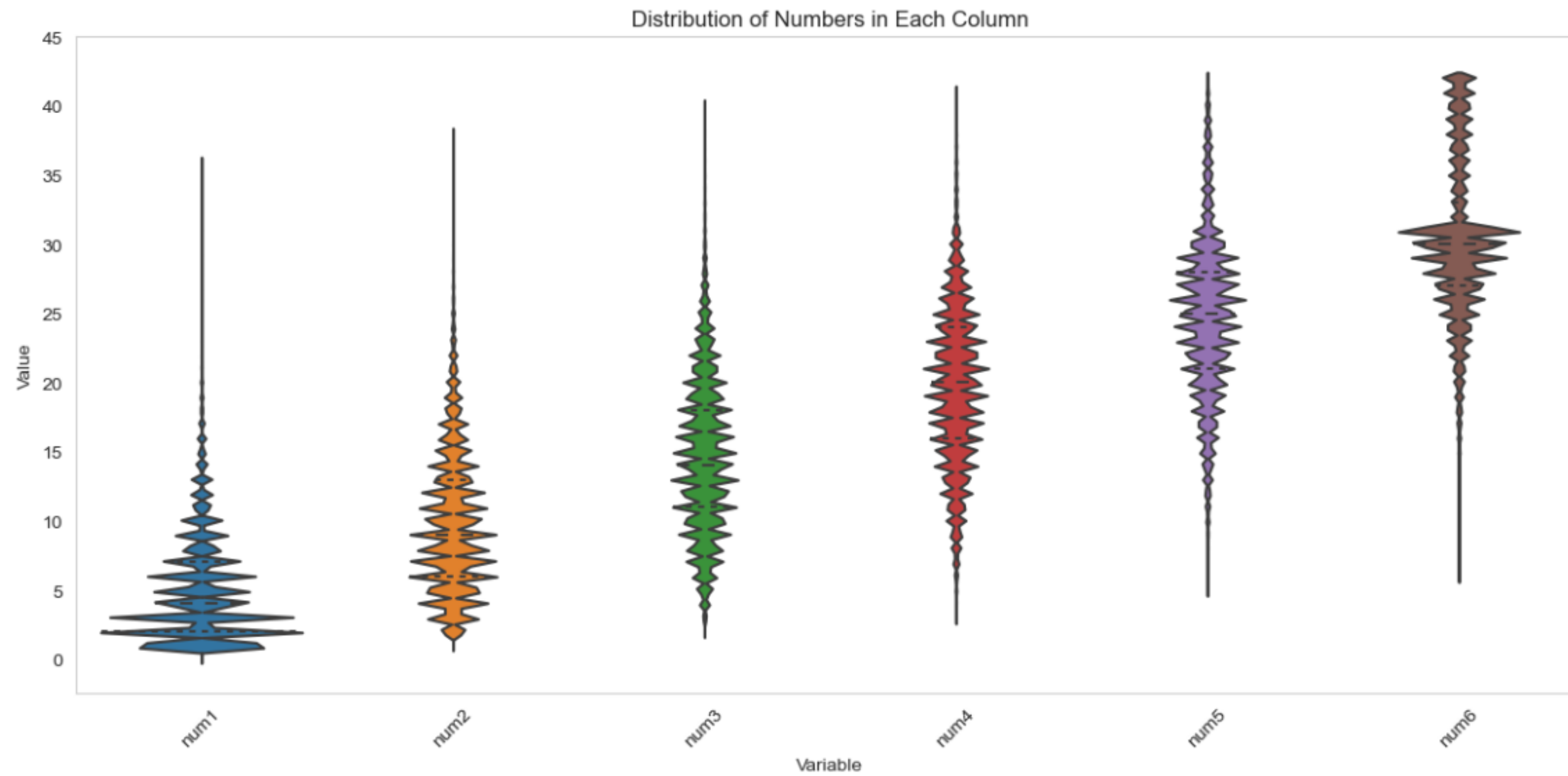
Exploratory Data Analysis



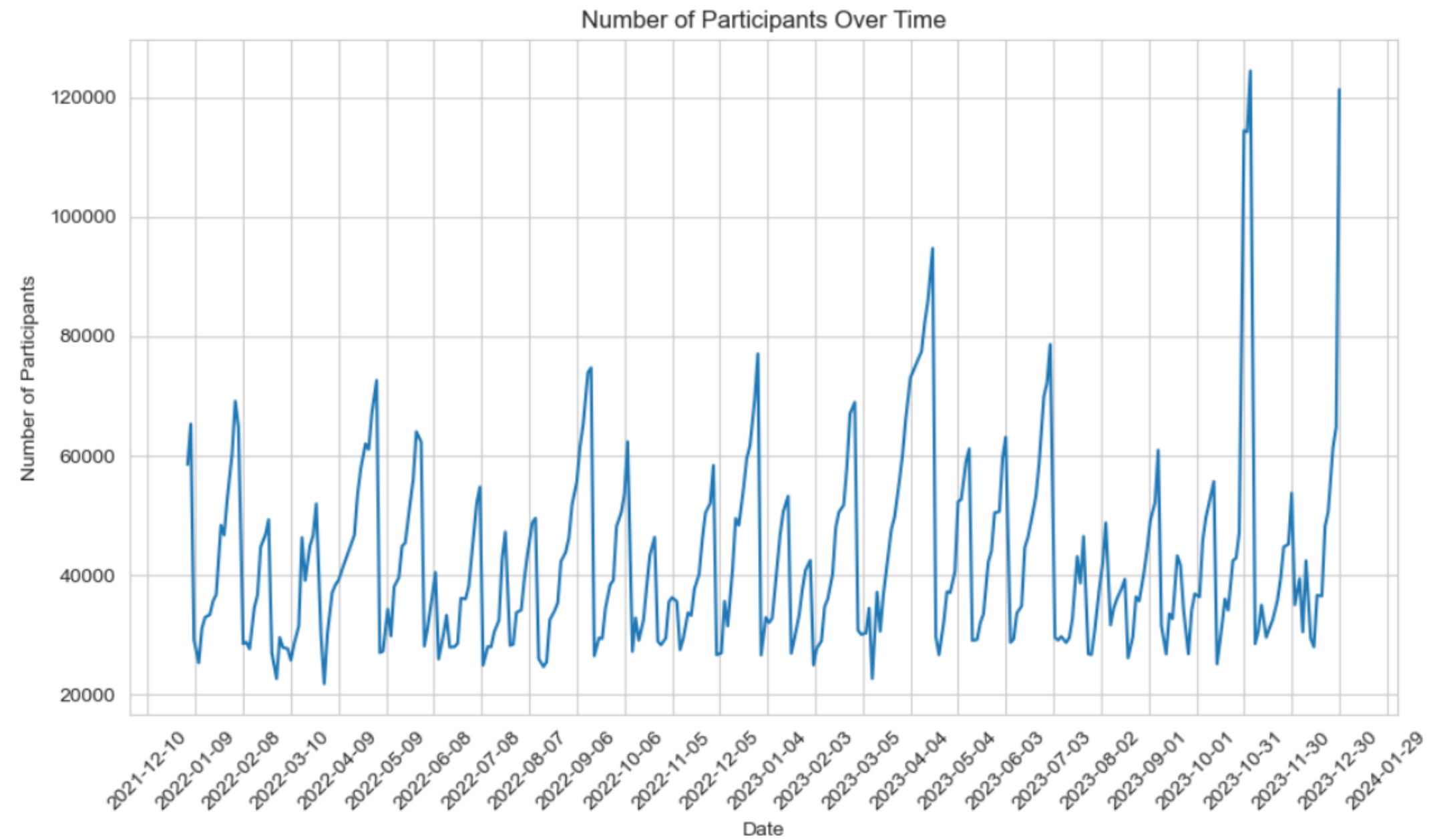
Exploratory Data Analysis



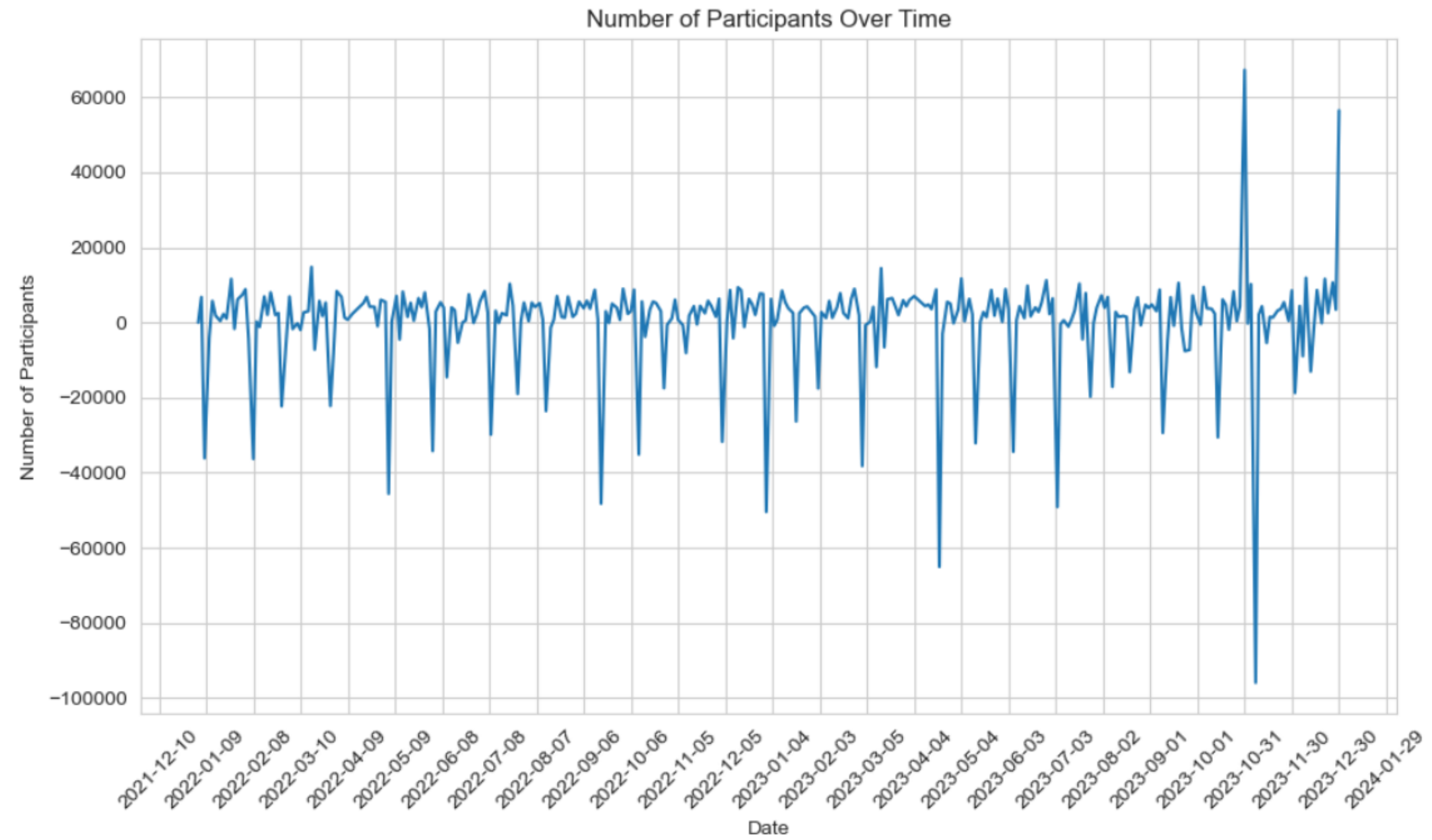
Exploratory Data Analysis



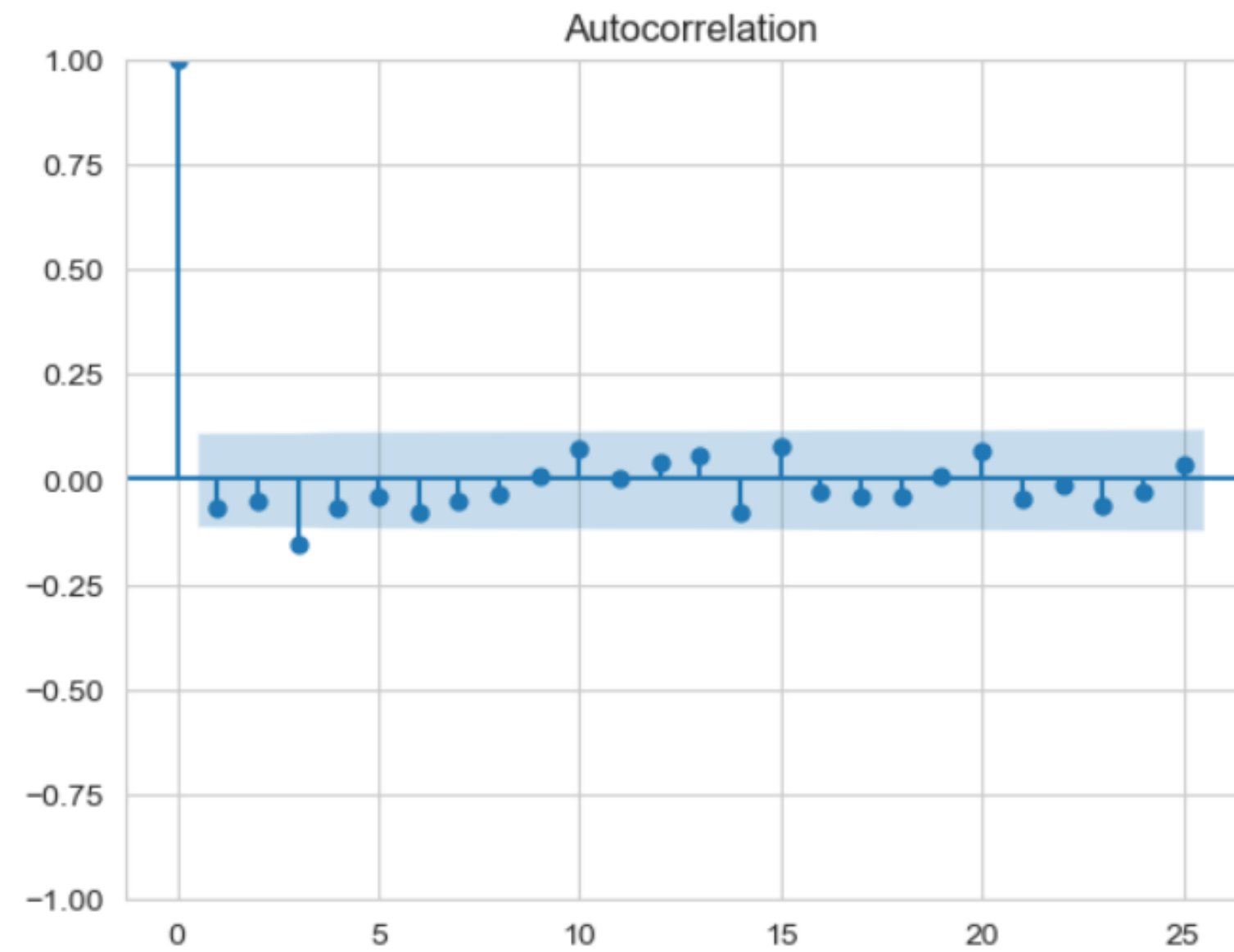
Exploratory Data Analysis



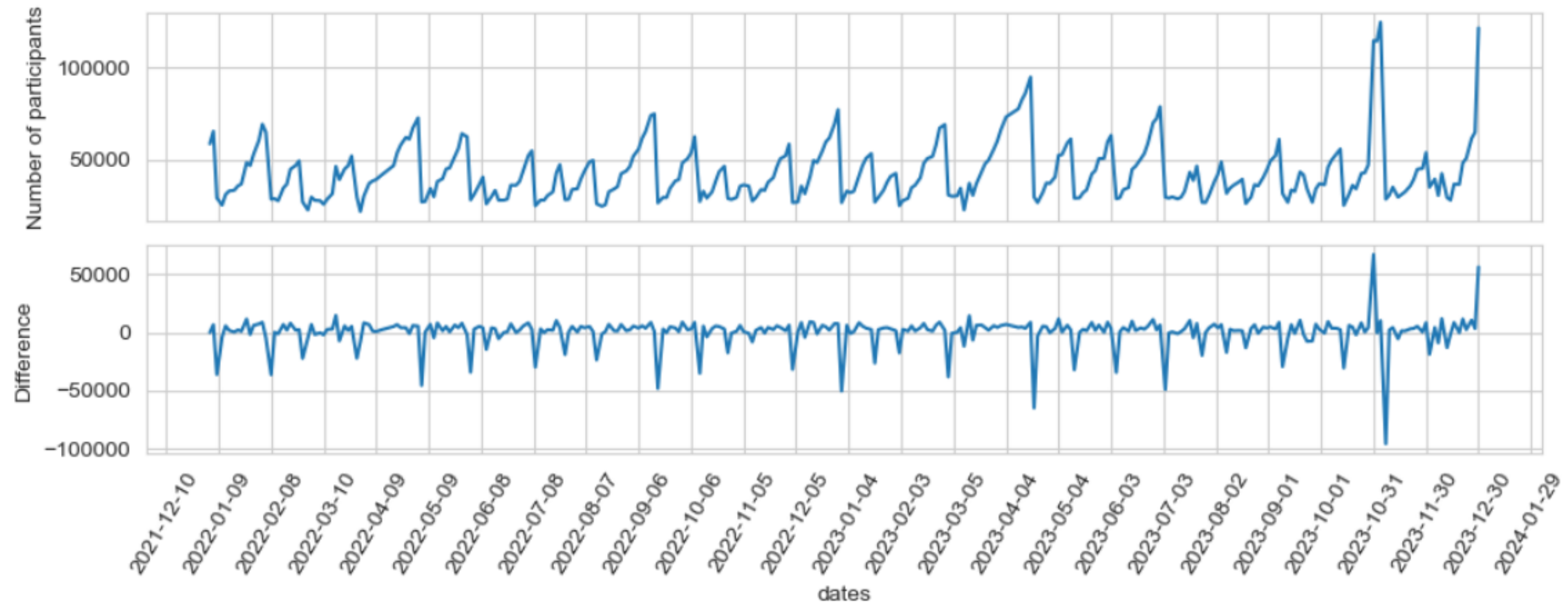
Exploratory Data Analysis



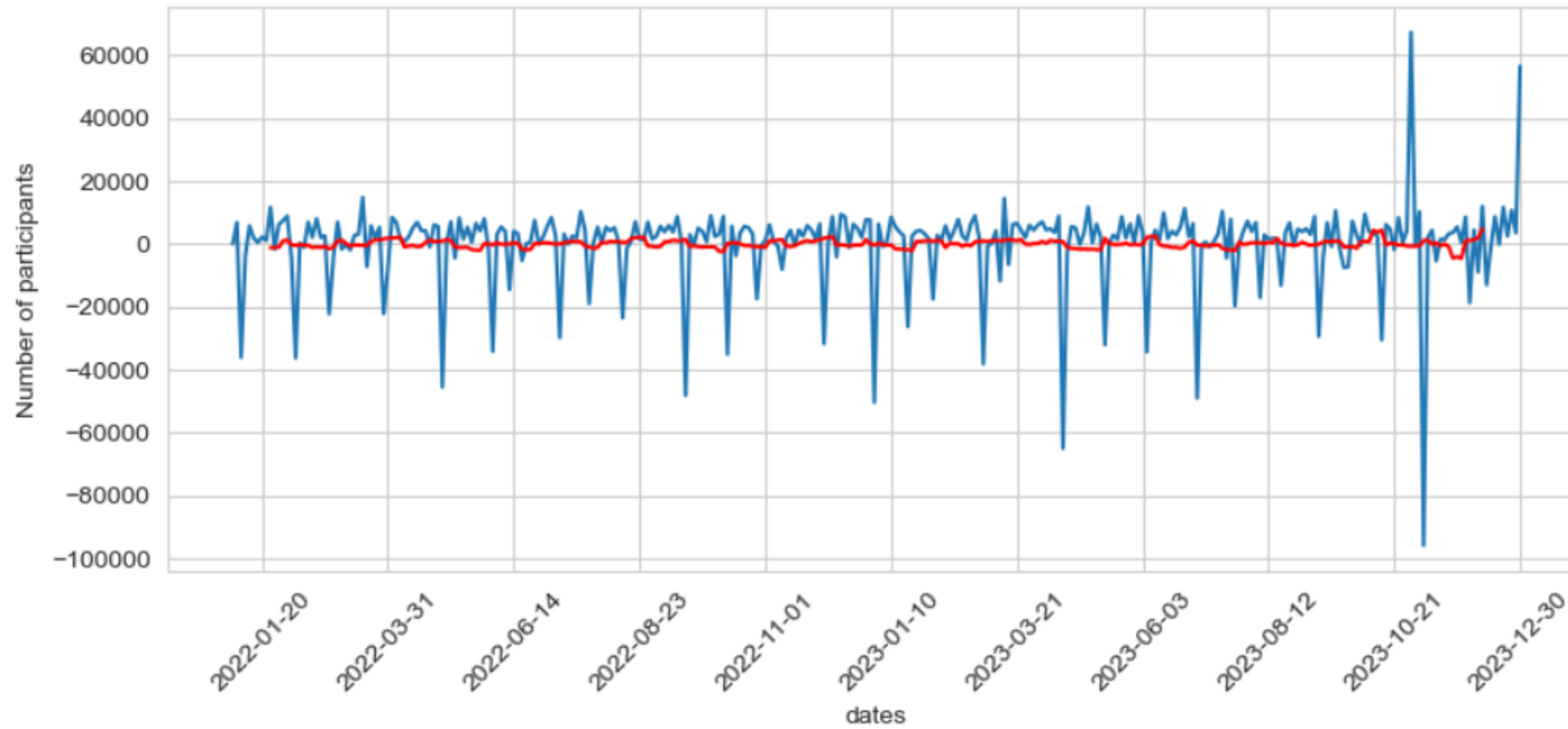
Exploratory Data Analysis



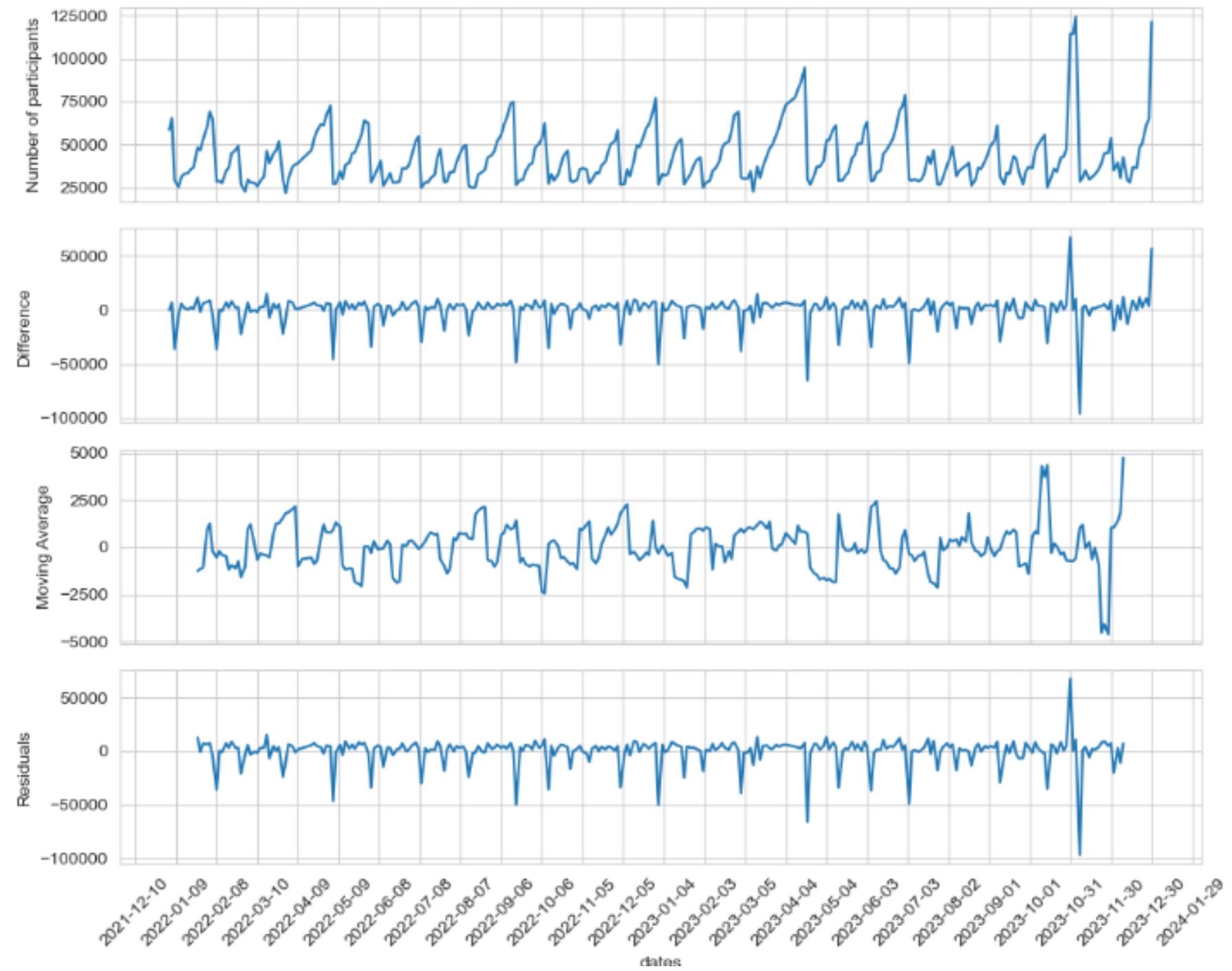
Exploratory Data Analysis



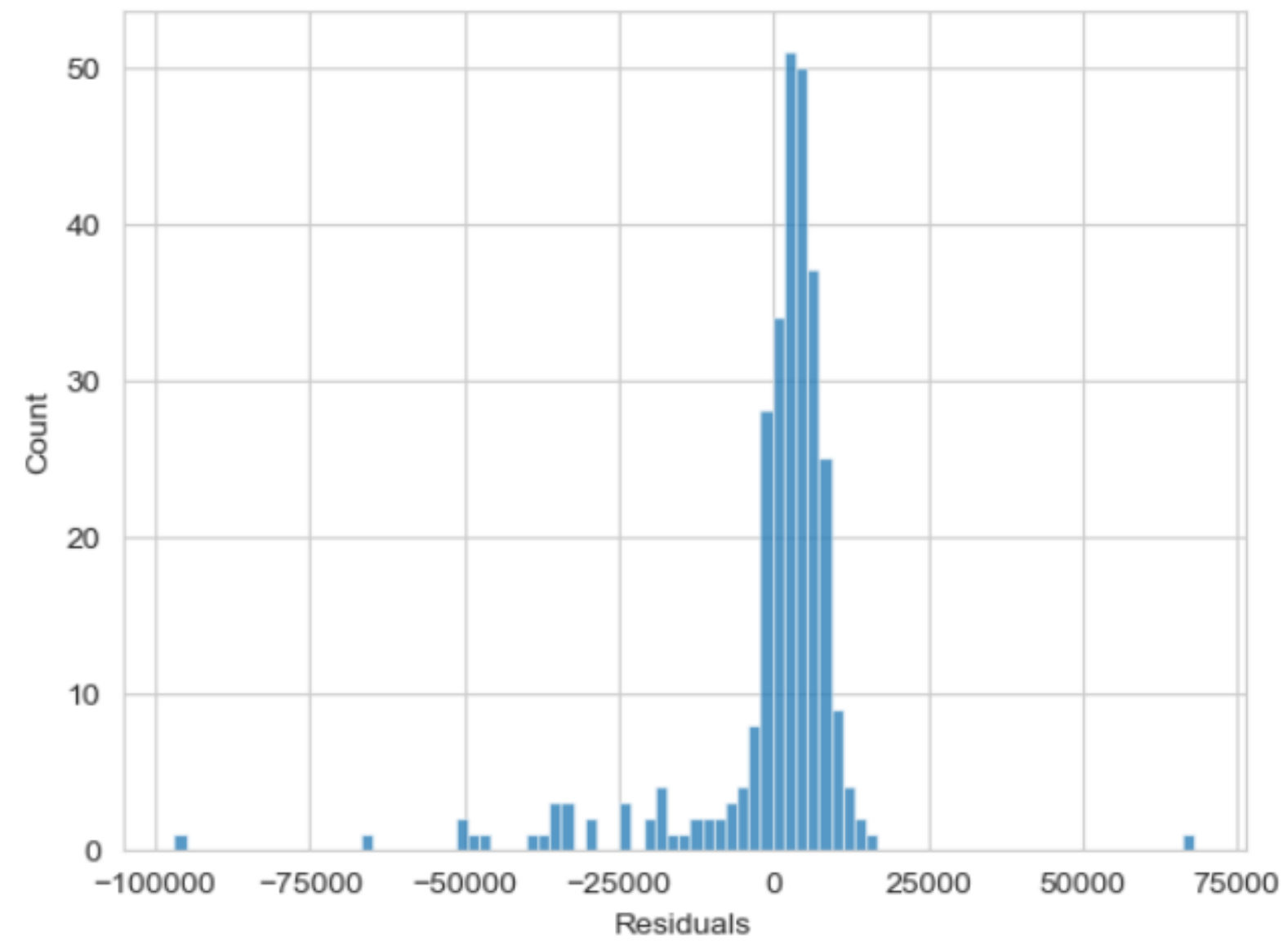
Exploratory Data Analysis



Exploratory Data Analysis



Exploratory Data Analysis



Results and Discussions



01 Overall Participation Trends: The line plot depicting the number of participants suggests a possible relationship between the prize money and number of participants

02 Daily Fluctuations: The plot of daily changes in participation highlights the same relationship towards the amount of the prize seen from the EDA.

03 Moving Average and Residuals: This figure presents a moving average of daily changes in lottery participation, aiming to smooth out short-term fluctuations and highlight underlying trends over time.

04 Distribution of Winning Numbers: The frequency distribution of winning numbers revealed that its generally random with similar distributions apart from exceptions with either greater or lower chances

05 Location-wise Participation: The visualization of ticket submissions by location suggests that the stats of NCR, Visayas and Mindanao have equal rates of submission while Luzon has greatest amount of submissions.

Recommendations

- 01 **Correlational Analysis:** Exploring potential correlations between participation levels and external factors like holidays, jackpots, or economic conditions.

- 02 **Predictive Modeling:** Developing models to predict future participation trends based on historical data and potentially identified influencing factors.

- 03 **Segmentation Analysis:** Segmenting participants based on demographics or other characteristics to understand variations in participation patterns across different groups.

THANK YOU!

PRESENTATION BY: AGUAVIVA, CHAMORRO, VILLAPANDO