COHORT ANALYSIS

BY
KHADIJA SULTANA

INTRODUCTION

• Cohort Analysis is a method used in analytics and business intelligence to group customers or users into cohorts based on shared characteristics or experiences within a defined time-span. These cohorts are then tracked over time to observe changes in behavior, usage, or other key metrics.

• Cohort analysis is valuable for businesses as it allows them to understand user behaviour in a more granular and actionable way.

STEPS IN COHORT ANALYSIS

1. The first step is to define the cohorts based on a specific characteristic or event. For example, in an e-commerce platform, cohorts could be defined based on the month of a user's first purchase.

2. Gather relevant data for analysis.

3. Determine the time intervals you want to analyze.

4. Group users into cohorts based on the defined characteristic or event.

5. Choose the key performance metrics you want to analyze.

6. Calculate the chosen metrics for each cohort over the specified time periods.

7. Create visualizations to present your findings effectively.

ABOUT THE DATASET

The provided dataset contains user interaction data, including metrics such as the number of new and returning users, and their engagement durations on Day 1 and Day 7. The data is structured with dates, allowing for time-series analysis. Key columns in the dataset are:

1. Date: The specific dates of user interactions.
2. New Users: The count of new users for each date.
3. Returning Users: The count of users returning on each date.
4. Duration Day 1: The average duration (possibly in minutes or seconds) of user interaction on their first day.
5. Duration Day 7: The average duration of user interaction on their seventh day.

Data Preparation

1. Check for datatypes of all the columns in the data

2. The Date column is in object (string) format. For effective analysis, especially in cohort analysis, we should convert this to a datetime format.

3. Check whether the dataset has any null or duplicate values or not.

4. To perform cohort analysis, we need to create a cohort week in this case for further analysis.

Descriptive Statistics Insights

1. New Users: The average number of new users is around 3,418 with a standard deviation of approximately 677. The minimum and maximum new users recorded are 1,929 and 4,790, respectively.

2. Returning Users: On average, there are about 1,353 returning users, with a standard deviation of around 247. The minimum and maximum are 784 and 1,766, respectively.

3. Duration Day 1: The average duration on the first day is about 208 seconds with a considerable spread (standard deviation is around 65).

4. Duration Day 7: The average 7-day duration is lower, around 136 seconds, with a larger standard deviation of about 97. The range is from 0 to 304.
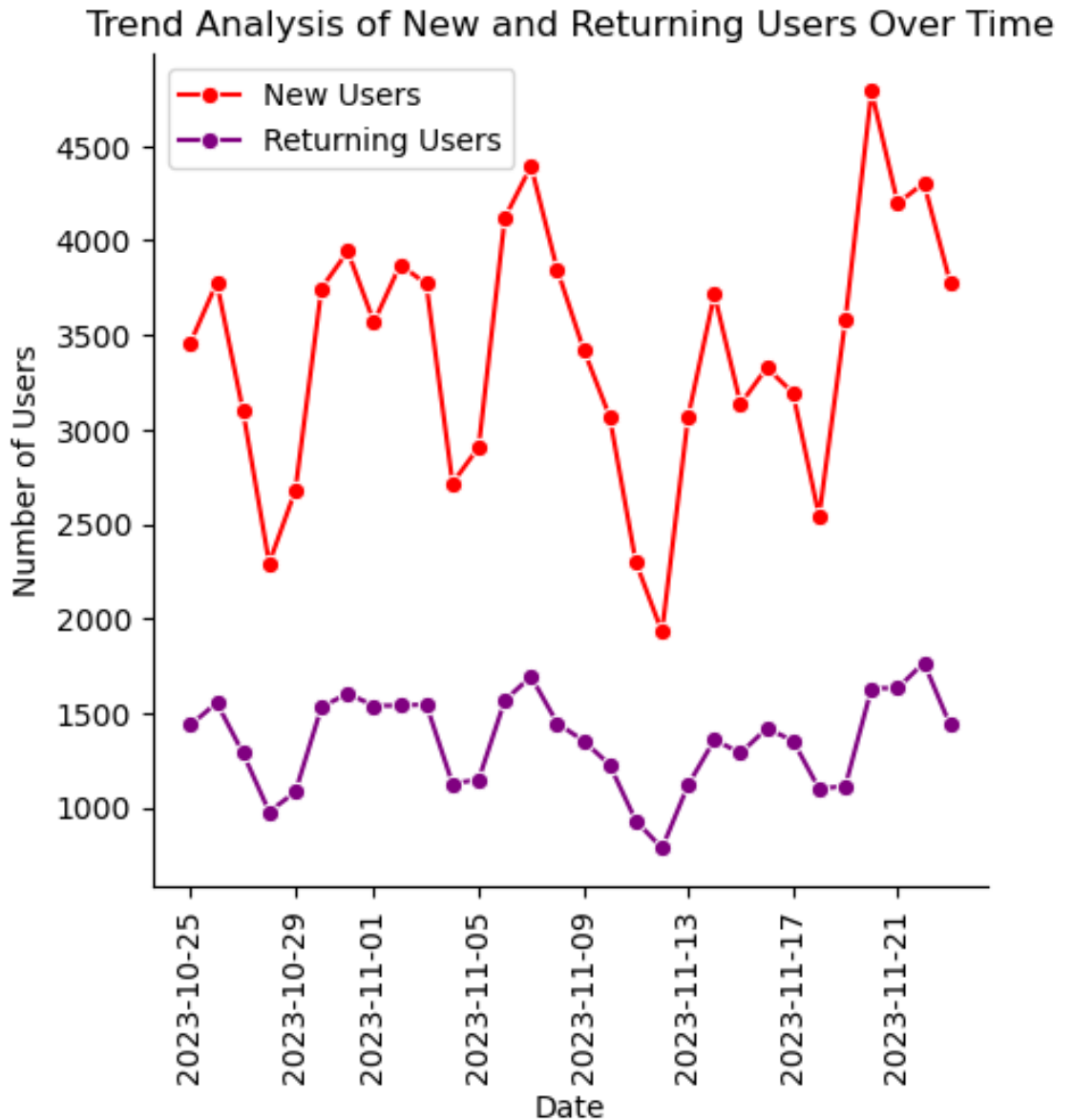
DATA DISTRIBUTIONS
AND
TREND ANALYSIS

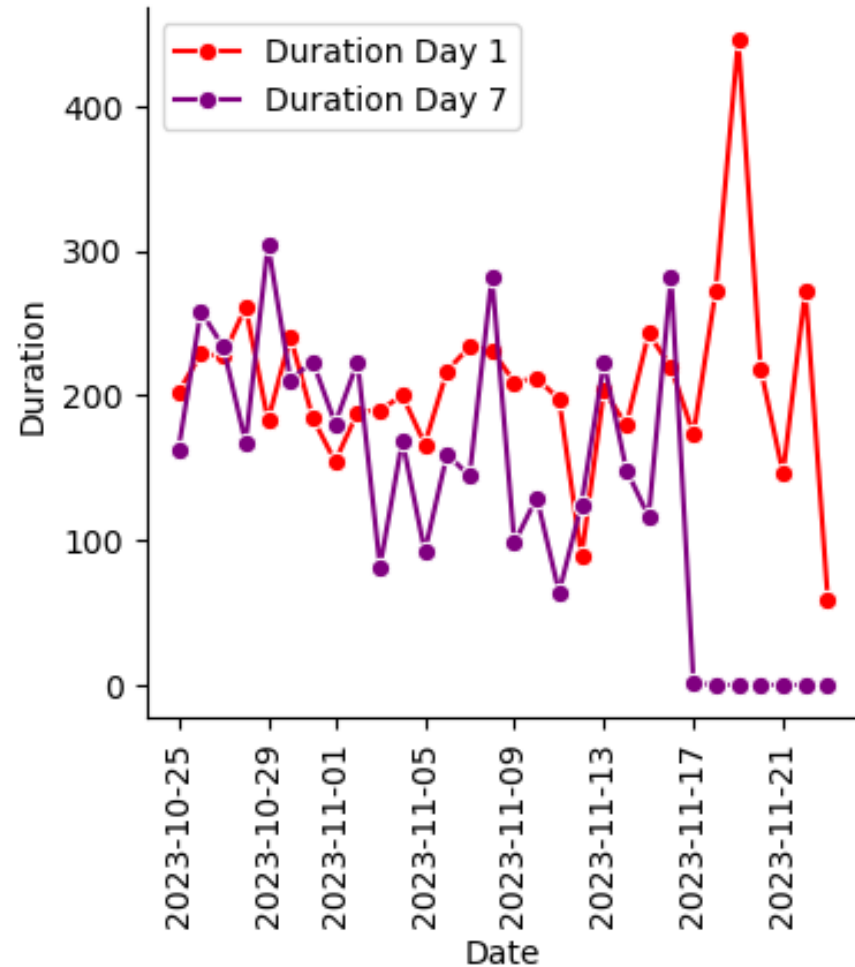# Plotting distributions of all variables
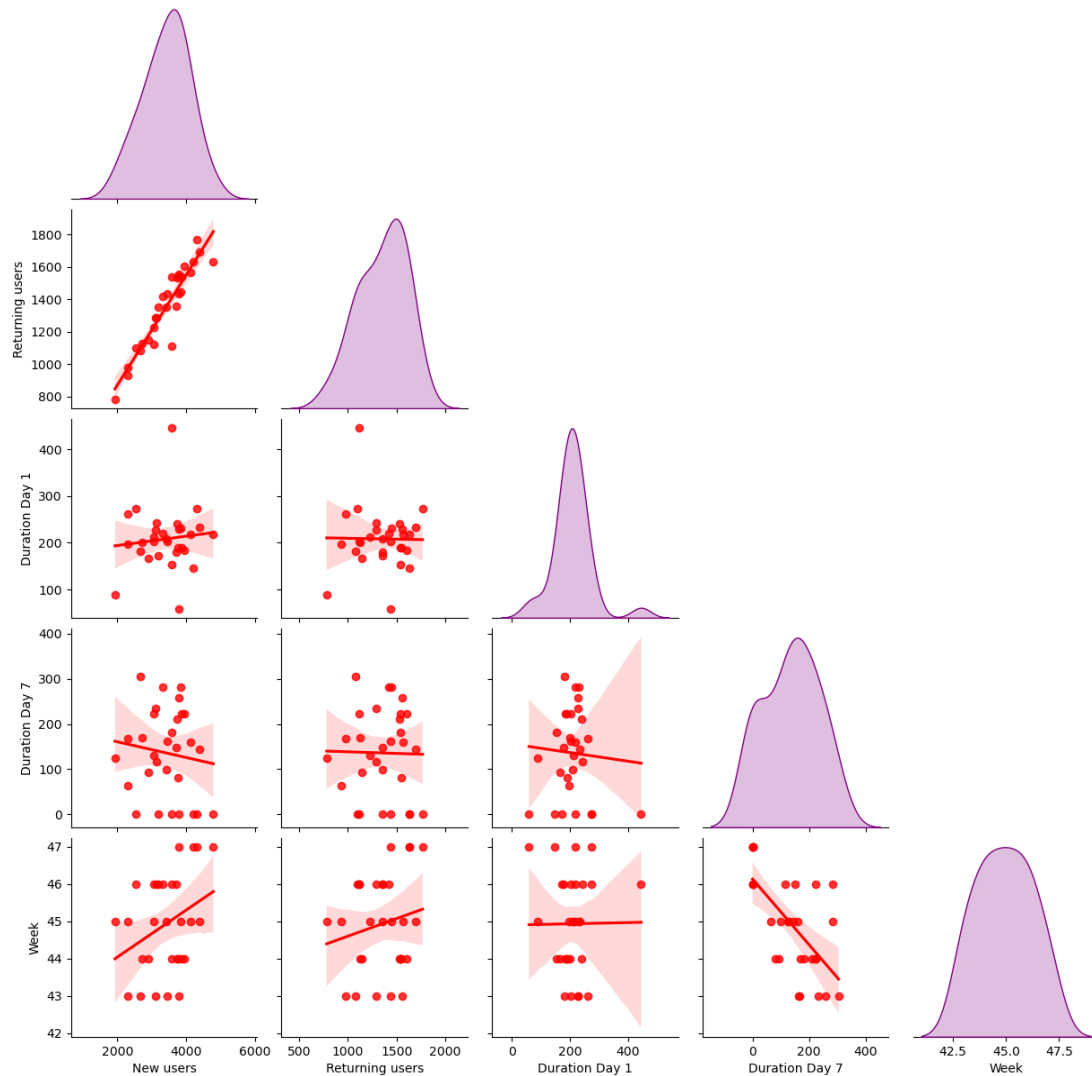
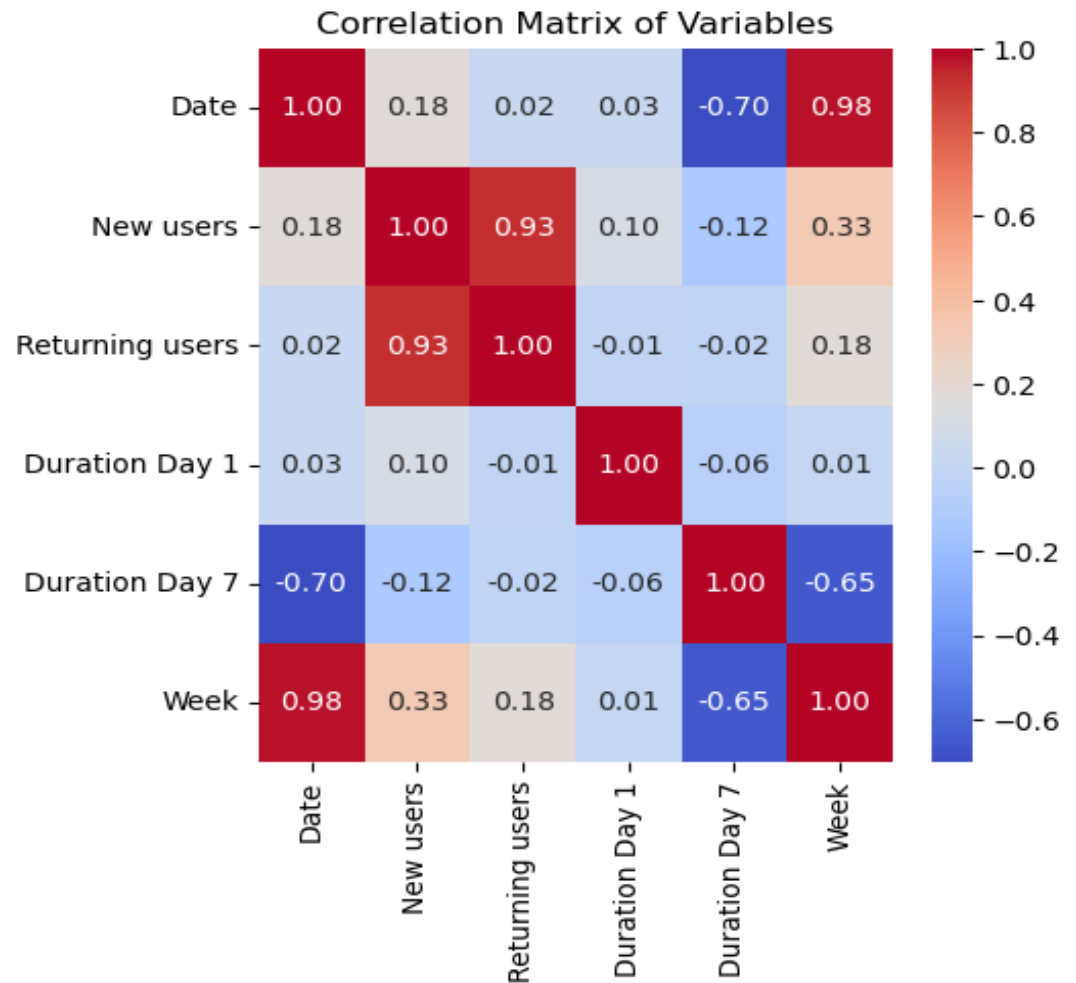The Trend Analysis for the new and returning users over time

The Trend Analysis for the Duration (Day 1 and Day 7) Over Time

# Correlation between the variables

Correlation Matrix of Variables

Here, the strongest correlation is between the number of new and returning users, indicating a potential trend of new users converting to returning users.
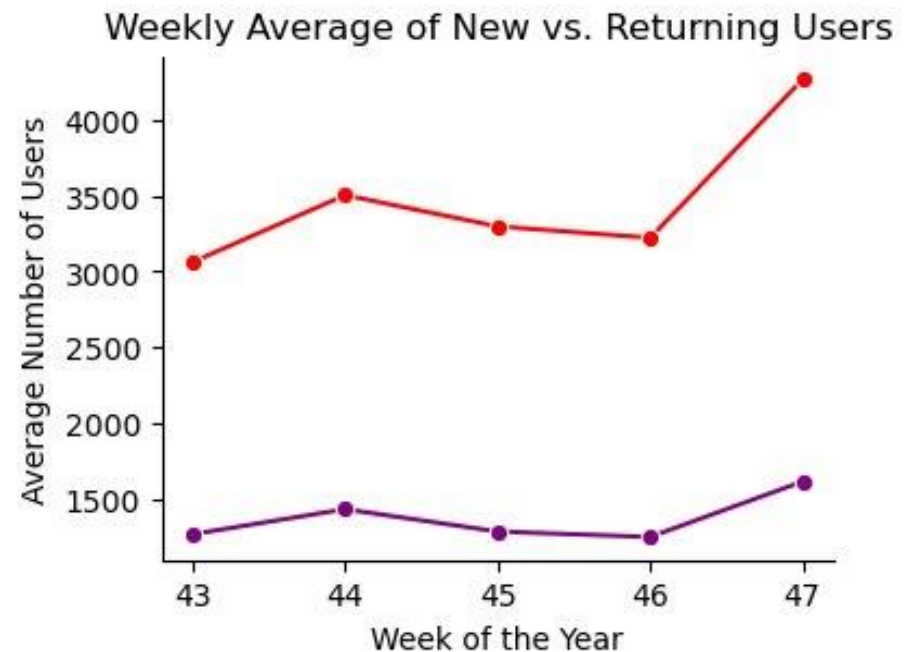
PERFORM COHORT ANALYSIS

For each cohort (week), calculating the average number of new and returning users, as well as the average of Duration Day 1 and Duration Day 7 by grouping the data by week and calculating the necessary averages.

Plotting the trend of the weekly average of the new and returning users and the duration:
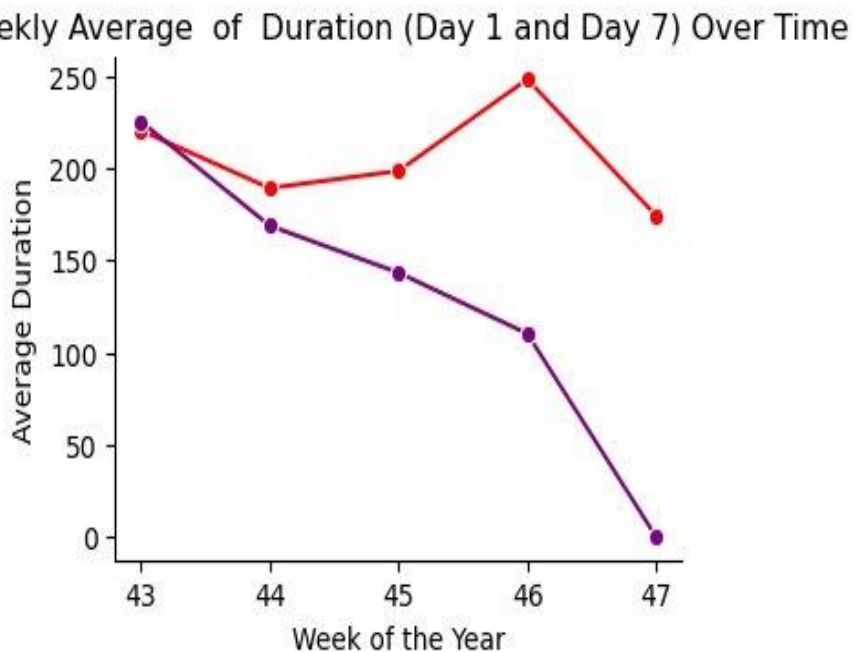
COHORT MATRIX

In the cohort matrix, each row will correspond to a week of the year, and each column will represent a different metric:

1. Average number of new users.
2. Average number of returning users.
3. Average duration on Day 1.
4. Average duration on Day 7.

Insights:

1. We can see that the number of new users and returning users fluctuates from week to week.

2. Notably, there was a significant increase in both new and returning users in Week 47.

3. The average duration of user engagement on Day 1 and Day 7 varies across the weeks.

4. The durations do not follow a consistent pattern about the number of new or returning users, suggesting that other factors might be influencing user engagement.

Cohort Matrix of Weekly Averages
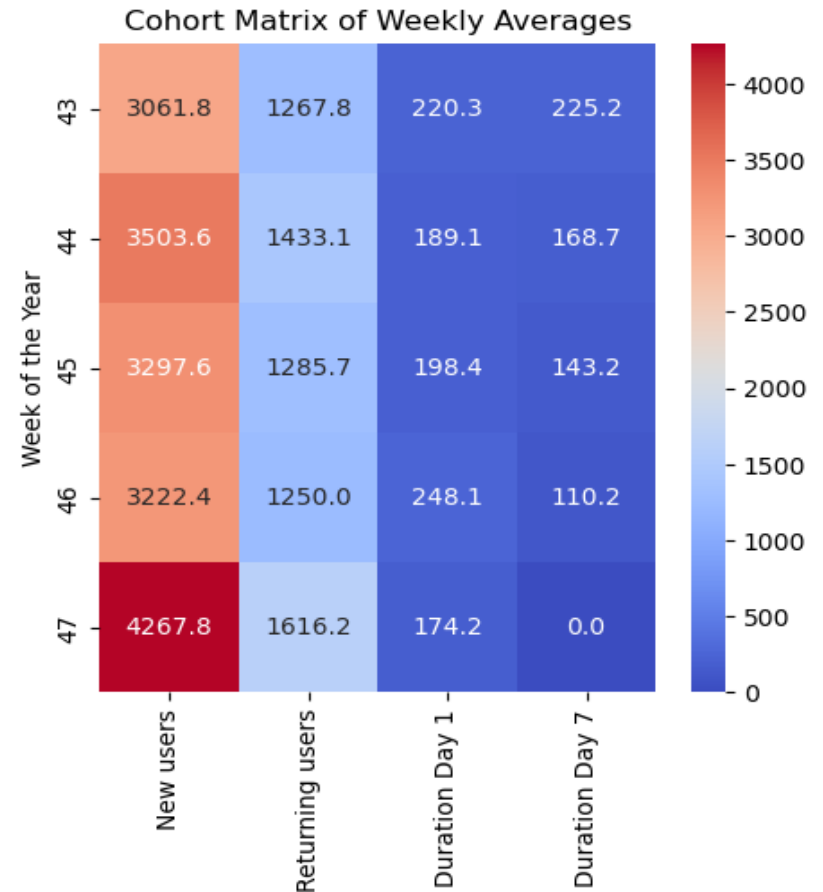
SUMMARY

Cohort Analysis is a data analysis technique used to gain insights into the behaviour and characteristics of specific groups of users or customers over time. It is valuable for businesses as it allows them to understand user behaviour in a more granular and actionable way.