Data Analysis Interview Challenge

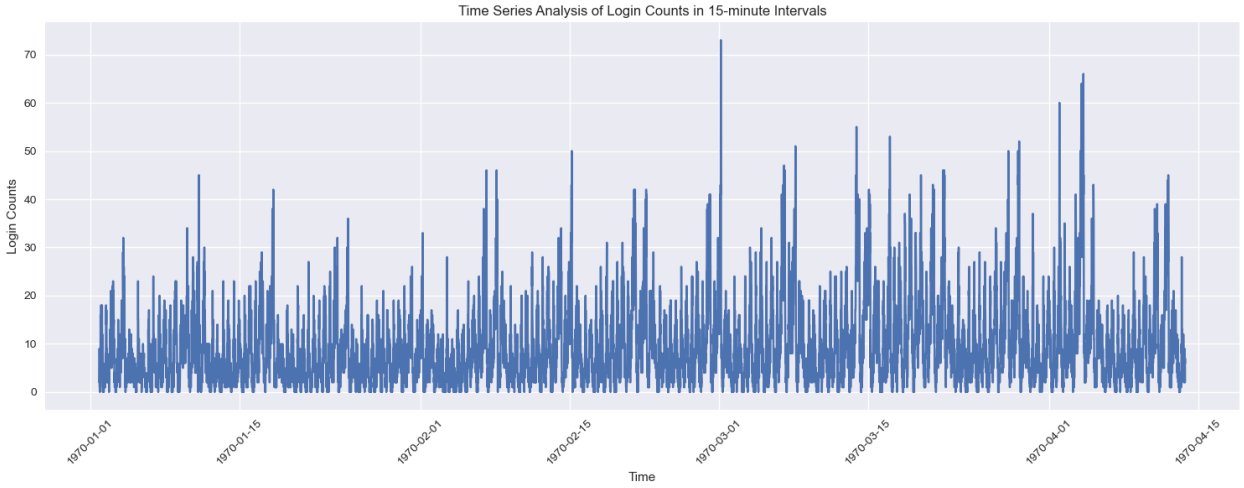
by Gulnar Armour

Part 1 - Exploratory data analysis

Findings:

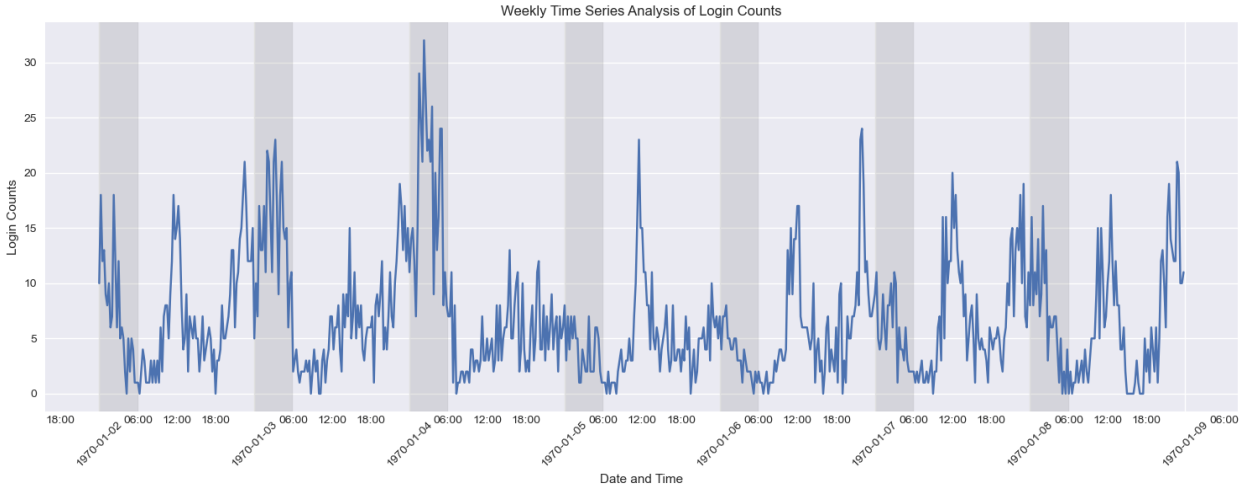
1. **15-minute Interval Analysis:**

- The plot showed fluctuations in login counts, indicating varying demand throughout the day.
- Peaks in the graph suggested specific times with higher user activity, hinting at potential peak hours.



2. **Weekly Demand Analysis:**

- In the graph below, we observed clear daily cycles.
- Nighttime hours (00:00 to 06:00) typically showed lower activity, highlighted in gray.
- Variations in daily login counts were evident, with some days showing significantly higher demand, possibly indicating a difference in weekday and weekend activity.



The analysis reveals distinct daily patterns in user login behavior, with identifiable peak and off-peak hours. The demand varies significantly throughout the day and across the week. These insights could be crucial for resource allocation, system maintenance scheduling, and understanding user engagement trends.

Part 2 - Experiment and metrics design

The Ultimate managers of city operations for Gotham and Metropolis have proposed an experiment to encourage driver partners to serve both cities by reimbursing all toll costs. This report outlines a proposed experiment to assess the effectiveness of this strategy, focusing on a key measure of success and the experimental design.

1) **Key Measure of Success:**

The primary metric for success in this experiment is the "Increase in Cross-city Trips" - the percentage increase in the number of trips where drivers cross the toll bridge to serve both cities. This metric directly reflects the objective of the experiment - to incentivize drivers to operate in both cities. An increase in cross-city trips would indicate that the toll reimbursement is effectively encouraging drivers to not limit their service to one city.

2) **Experiment Design:**

a) Implementation:

- Duration: The experiment will run for a sufficient period, suggested at 3 months, to capture variations in driver behavior.
- Groups: Drivers will be divided into two groups - the Control Group (no toll reimbursement) and the Treatment Group (with toll reimbursement).
- Data Collection: Collect data on the number of cross-city trips made by drivers in both groups, along with other relevant metrics such as total trips, hours worked, and earnings.

b) Statistical Tests:

- Primary Analysis: Use a Two-Sample T-Test to compare the mean increase in cross-city trips between the Treatment and Control Groups. This test is appropriate for comparing the means of two independent groups.
- Secondary Analysis: Conduct regression analysis to control for potential confounders such as time of day, day of the week, and seasonal effects.

c) Interpretation and Recommendations:

- Significant Results: If the T-Test shows a statistically significant increase in cross-city trips for the Treatment Group, we can conclude that toll reimbursement is effective.
- Non-significant Results: If no significant difference is found, it suggests that toll costs are not a primary barrier to cross-city trips.
- Operational Recommendations: Based on the results, provide actionable insights to the city operations team. For instance, if effective, consider a permanent toll reimbursement policy.
- Caveats: Be aware of external factors that may impact the results, such as special events, economic changes, or policy changes in the cities.

By focusing on a relevant metric and employing rigorous statistical analysis, the experiment aims to provide a data-driven basis for decision-making regarding the toll reimbursement policy.

Part 3 - Predictive modeling

Ultimate can leverage insights from the predictive model to improve long-term rider retention by identifying at-risk users and targeting them with personalized interventions, such as special promotions or enhanced service features. Insights about factors influencing retention, like user engagement in the early stages or preferences for certain services, can guide more effective marketing and operational strategies. This data-driven approach enables Ultimate to allocate resources more effectively and tailor its services to meet user needs, ultimately enhancing customer satisfaction and loyalty.