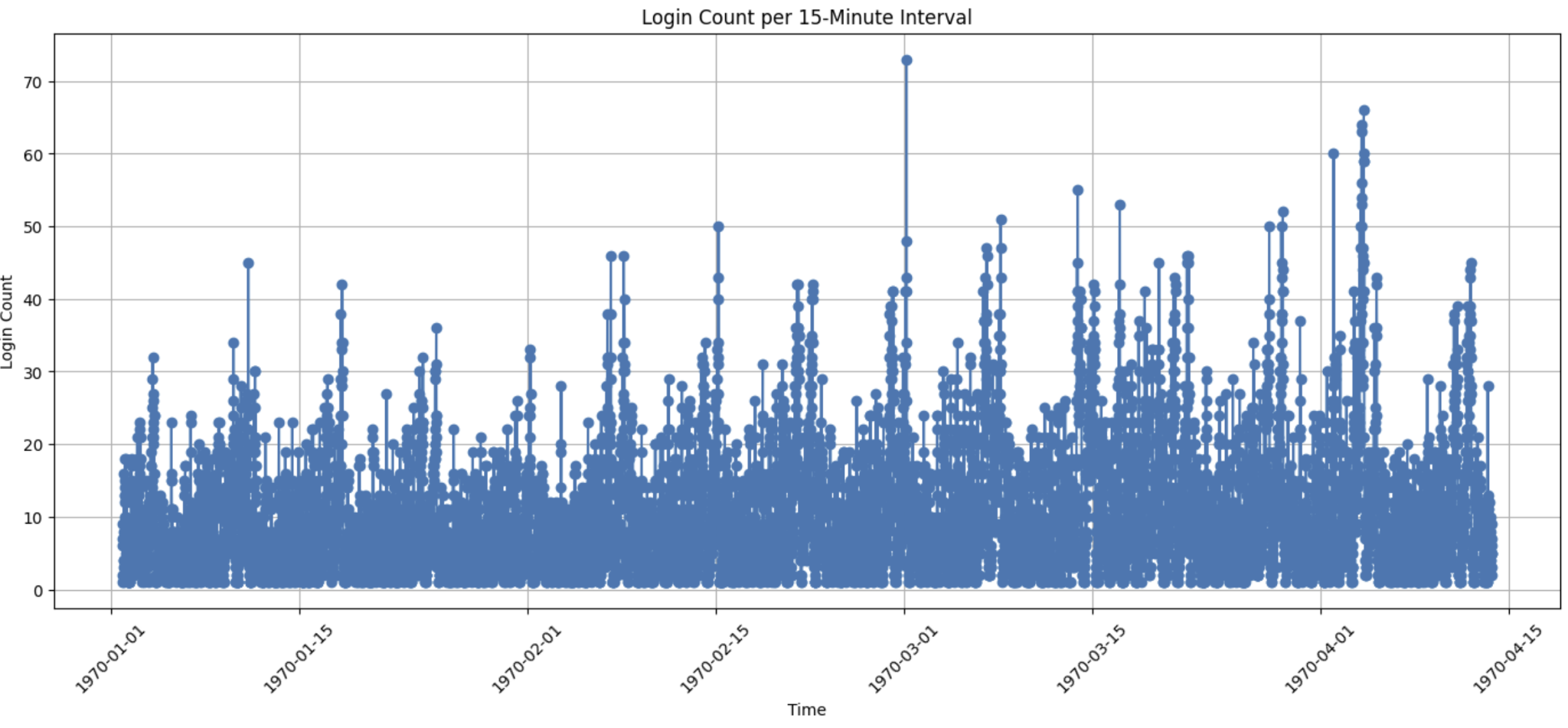
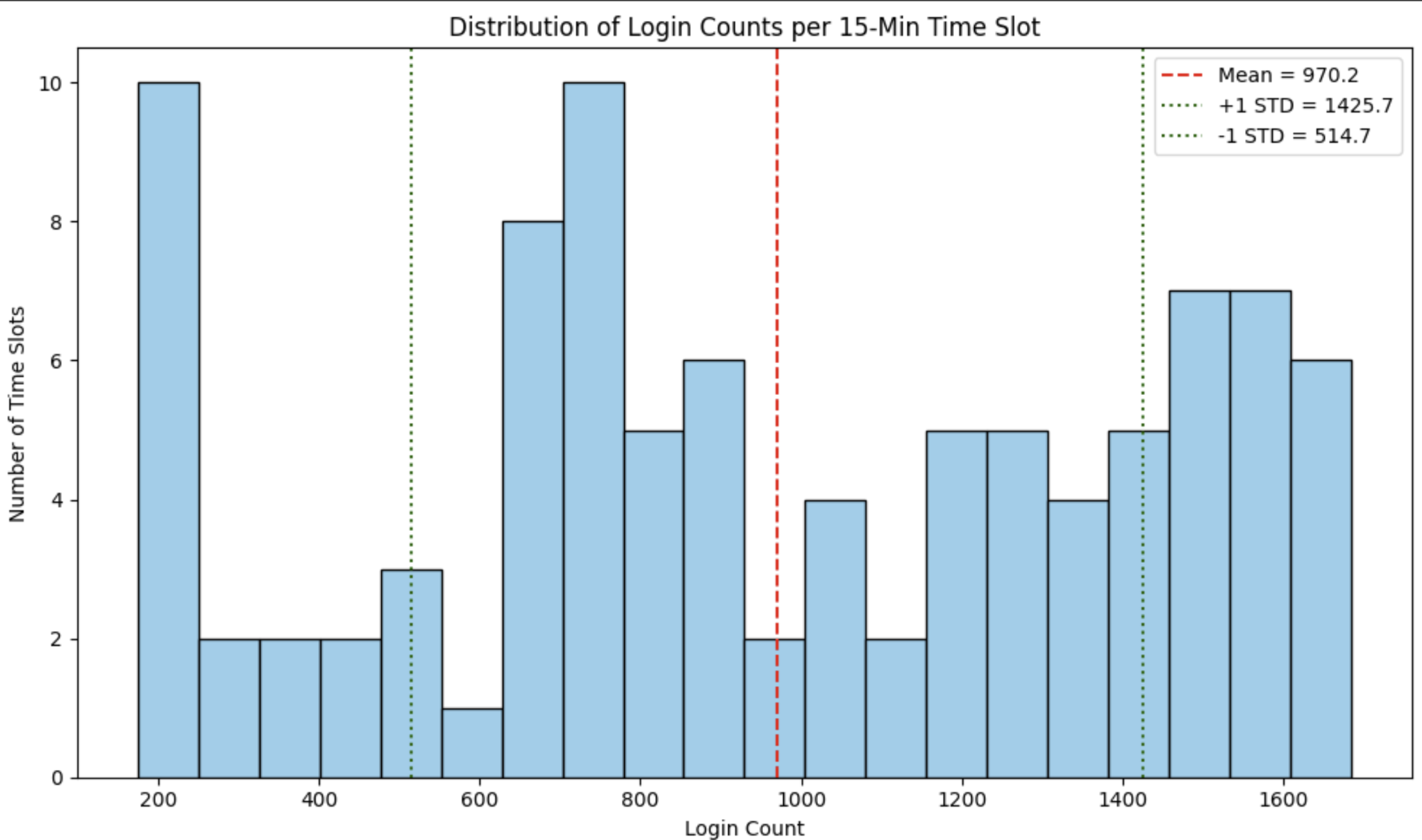
The Google Colab Notebook for Part 1 and Part 2 is [here](https://github.com/BTExpress1/ultimate-tech/blob/2a9a82bb3e32c53c85f8913bc25cd02e78aaabbe/notebooks/ultimate_tech.ipynb).

# Part 1

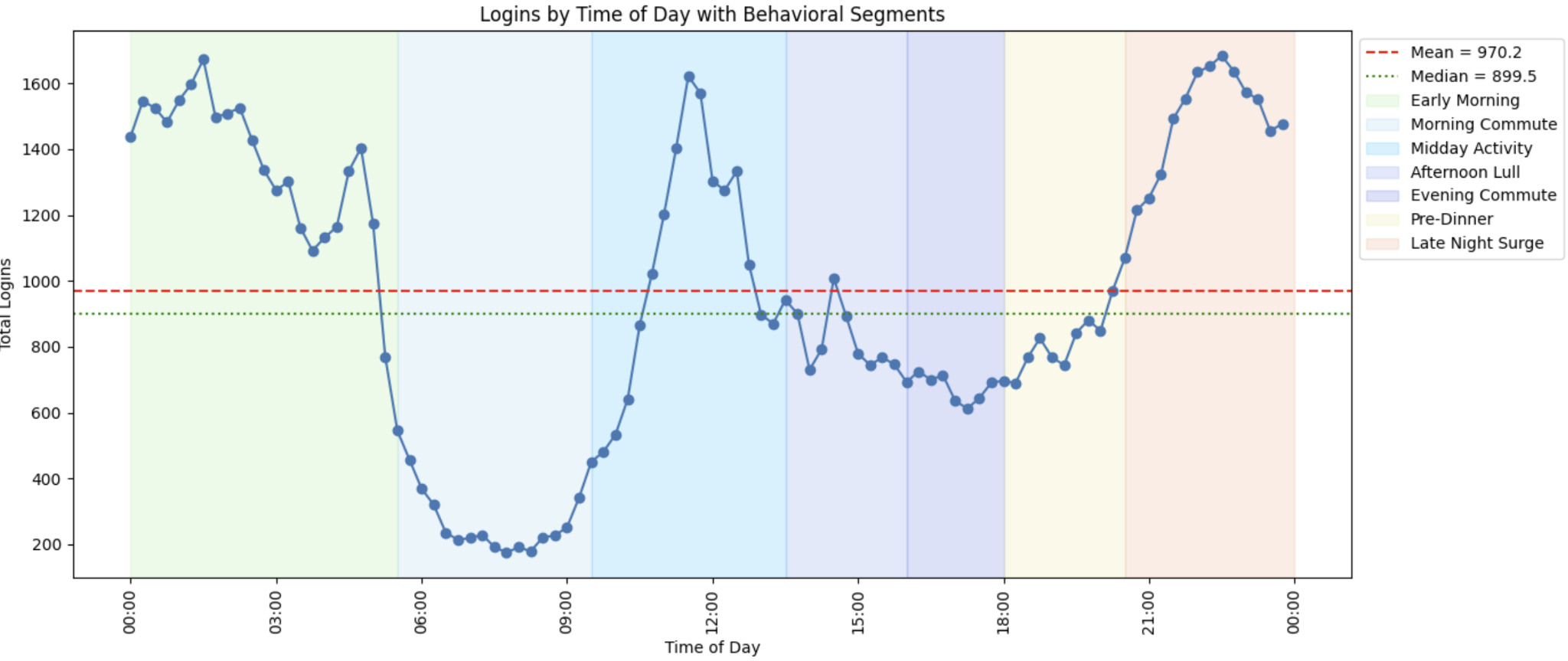
The logins.json file contains 93142 rows. There is no missing data. I made the assumption these are related to taxi service. A user login equates to a user calling for a cab. After grouping the logins by 15 min increment, the following plot shows a fairly concentrated login count day to day. There are multiple outliers and the initial assumption is that these could be related to day of week or other significant days in the month. Without additional context, it is hard to determine the cause of the spikes.



The distribution of login counts per 15-min time slot (interval) show a multi-modal distribution. Once again no specific trend to glim from it. The histogram below is the visual representation. However, the distribution suggests a high traffic period that could be tied to user behavior. The need to look at the logins by time of day is explored.



Breaking down the data into time of day, there is a trend we are able to identify. There are three times of the day that traffic is high. During early mourning hours (midnight to 5:30 am), midday (9:30 am to 1:30 pm), and late night (10:30 pm to midnight).The segments are color coded in the chart below. During the early morning hours, people are making their way home from late night outings, e.g. at bars. This segment is followed by the late-night block, during which I assume people will be heading to activities by taxi, allowing them to indulge freely in alcohol. Therefore, the return home requires a similar taxi ride. The midday block, could be tied to people running errands by taxi during their lunch or extended break at work. Conversely, during to and from office commute hours, and family dinner time, the taxi traffic is low.



I identified the data containing duplicate rows. Based on the assumption that multiple people could call for a cab during the same time slot, I decided not to remove duplicates. This data might be useful in the future when additional analysis is performed.

# Part 2

**Pilot Program:** Toll Reimbursement Initiative for Cross-City Ride-share Expansion between Ultimate Gotham and Ultimate Metropolis

**Objective**

Ultimate Gotham and Ultimate Metropolis are two cities separated by a two way toll bridge. Ultimate operations finds that drivers in each city prefer to operate on their side of the bridge . The goal is to implement a toll reimbursement program to encourage drivers to serve both Gotham and Metropolis, aiming to balance supply across cities and increase cross-city trip availability.

**Methodology**

* **Design:** Pre-post analysis using **Difference-in-Differences (DiD)** statistical modeling, OLS regression.
* **Comparison:** Treatment group (drivers eligible for toll reimbursement) vs. Control group (drivers not reimbursed).
* **Success Measures:**
  + Increase in active drivers.
  + Increase in cross-city trips.
  + Higher average distance traveled.
  + Volume of reimbursement requests.
  + Trip volume and driver retention.

**Pilot Implementation**

* Marketing campaign: Announce the program externally by focusing on higher earnings potential and toll reimbursement benefits.
* In-App notification: Push real-time reminders to drivers, especially when demand spikes in the other city.
* Signup incentives: Offer a bonus for new drivers who join during the program window and cross cities.
* Cross-city activity bonus: Offer a performance bonus: e.g., serve both cities 5+ times in the first 30 days to earn an extra cash bonus.
* Weekly highlight emails: Celebrate top earners or top cross-city drivers each week - keep momentum and social proof.
* Reimbursement process clearly communicated to drivers. In the future, this should be automated.

**Expected Key Results**

* Statistically significant increases observed in key metrics post-implementation during the pilot phase:  
  + 10% increase in active drivers.
  + 50% increase in cross-city trips.
  + 50% increase in average distance traveled.
  + 100% of reimbursement requests per eligible driver.
* Assumption is that control and treatment groups showed **parallel trends pre-implementation**, validating the causal inference.

**Post Pilot Recommendation**

Transition from **pilot** to **run-state program**, if the results support measurable improvements in driver engagement and cross-city service levels.

**Caveats**

* **Driver Incentive Fatigue:** Impact may diminish over time if driver excitement wanes.
* **Toll Profitability Risk:** Rising tolls or lower fares could erode net driver profits even with reimbursements.
* **Parallel Trends Risk:** Natural market shifts could cause post-implementation divergence between groups.
* **Gaming the System:** Drivers may attempt to maximize reimbursements through artificially short cross-city trips.
* **Administrative Complexity:** Scaling manual reimbursement processes could strain operations unless automated.
* **Market Shifts:** External changes (e.g., new bridges, traffic reroutes) could distort program impact measurements.

**Next Steps**

* Automation of reimbursement processing.
* Expand marketing and in-app notification campaigns.
* Introduce tiered bonuses for sustained cross-city engagement.
* Monitor KPIs continuously post-rollout.

# Part 3