UrbanRoll Bike-Sharing Data Analysis Report

Using Google BigQuery & SQL

---

1. Introduction

UrbanRoll operates a city-wide bike-sharing service designed to support urban mobility. As competition and operational complexity increase, data-driven decision-making becomes essential to ensure efficiency, customer satisfaction, and sustainable growth.

This report presents findings from an **end-to-end SQL analysis conducted in Google BigQuery**, covering over **15,000 ride records**. The objective is to uncover actionable insights related to **user behavior, operational performance, station logistics, and growth dynamics**, and to translate those insights into practical business recommendations.

The analysis relies exclusively on **SQL**, demonstrating how structured querying alone can deliver insights typically associated with more complex analytics stacks.

---

2. Data Overview and Preparation

2.1 Dataset Scope

The dataset captures detailed ride-level activity, including:

- Ride start and end timestamps
- Start and end station identifiers
- User classification (Subscriber vs Casual)
- Membership information
- Ride duration metrics
- Station metadata

This breadth of information enables both **operational analysis** (stations, flows, timing) and **behavioral analysis** (user personas, retention, usage intensity).

2.2 Data Quality Assessment

Before analysis, extensive validation checks were performed to ensure reliability:

Null Values

Key analytical fields—such as station names, timestamps, and user types—were inspected for missing values. Null entries were either excluded or handled appropriately to prevent skewed aggregations.

Result:
The majority of nulls were concentrated in non-critical descriptive fields, allowing the core dataset to remain robust.

False Starts

False starts were defined as rides with:

- Zero or negative duration
- Missing start or end points

These records were filtered out to preserve analytical accuracy.

Impact:
Removing false starts ensured that ride duration averages, peak-hour analysis, and net-flow calculations reflected real user behavior.

---

3. Platform Scale and Baseline Metrics

A high-level aggregation was performed to establish UrbanRoll's operational footprint.

Key Metrics Identified:

- Total number of rides
- Total unique users
- Total active stations

These baseline metrics provide context for deeper analysis and serve as reference points when evaluating growth, demand concentration, and station performance.

---

4. Operational Demand Patterns: The “Double Hump” Effect

4.1 Temporal Ride Distribution

Hourly aggregation of rides revealed a **distinct bi-modal demand pattern**, often referred to as the *double hump*:

- **Morning peak:** ~7:00 AM
- **Afternoon peak:** ~3:00 PM

This pattern strongly indicates that UrbanRoll functions primarily as a **commuter mobility solution**, rather than a leisure-first service during weekdays.

4.2 Operational Implications

The existence of predictable peak periods creates operational pressure on:

- Bike availability
- Dock capacity
- Station balance

Recommendation

To minimize peak-hour shortages:

- **Rebalancing crews should be scheduled between 10:00 AM and 2:00 PM**, a period of reduced demand.
- This timing allows the system to reset before the afternoon commute surge.

---

5. User Segmentation: Subscriber vs Casual Riders

5.1 Identification of User Personas

The analysis reveals two clearly differentiated user segments:

Subscriber (Commuter Persona)

- Short, consistent trips
- Average ride duration: ~15 minutes
- High frequency and predictability
- Responsible for the majority of weekday volume

Subscribers represent the **operational backbone** of UrbanRoll.

Casual User (Explorer Persona)

- Longer, less predictable trips
- Average ride duration: ~35 minutes
- Lower frequency, higher time-on-bike per ride
- Often associated with leisure or weekend usage

Casual users consume more bike time per ride but contribute less to daily ride counts.

5.2 Strategic Insight

These personas have **different value propositions**:

- Subscribers optimize volume and predictability
- Casual users maximize utilization per ride

Recommendation

Introduce a **"Weekend Explorer Pass"** targeted at casual users:

- Encourages higher off-peak and weekend frequency
- Increases revenue without straining weekday commuter capacity

---

6. Station Popularity and Network Usage

Station-level aggregations identified:

- High-traffic stations
- Consistently underutilized locations

This information is critical for:

- Infrastructure investment decisions
- Station expansion or consolidation
- Targeted maintenance and marketing efforts

Popular stations often align with:

- Business districts
- Transit hubs

- High-density residential areas

---

7. Rebalancing Crisis: Net Flow Analysis

7.1 Net Flow Concept

Using a **Net Flow Common Table Expression (CTE)**, each station's balance was calculated as:

Net Flow = Arrivals − Departures

This allowed stations to be classified as:

- **Sources:** Stations that frequently run out of bikes
- **Sinks:** Stations that accumulate excess bikes

7.2 Key Findings

The analysis showed persistent imbalances:

- Certain stations consistently act as sinks
- Others repeatedly face bike shortages

These imbalances:

- Degrade user experience
- Increase operational costs
- Exacerbate peak-hour congestion

Recommendation

Adopt **dynamic rebalancing incentives**, such as:

- Small ride discounts
- Loyalty points for ending rides at source stations

This leverages user behavior to assist operational logistics at a lower cost than manual redistribution alone.

---

8. Growth, Retention, and Power Users

8.1 Growth Trends

Month-over-Month (MoM) analysis demonstrates:

- Consistent growth in ride volume
- No evidence of stagnation or decline

A **7-day moving average** was applied to smooth daily fluctuations, confirming that growth is structural rather than noise-driven.

8.2 Retention and Power Users

Retention analysis uncovered a critical insight:

- The **top 1% of users ("power users") account for a disproportionate share of rides and revenue**

These users are:

- Highly engaged
- High lifetime value
- Essential to revenue stability

Strategic Implication

Losing even a small portion of power users could have an outsized negative impact.

Recommendation

- Introduce loyalty rewards
- Offer premium membership tiers
- Provide early access or usage perks

---

9. Summary Statistics and Overall Performance

Descriptive statistics reinforced earlier findings:

- Clear differences in ride duration by user type
- High variance in station usage

- Predictable temporal demand cycles

These metrics validate both the segmentation and operational analyses.

---

10. Executive Takeaways

1. UrbanRoll is fundamentally a **commuter-driven platform**
2. Subscribers drive **volume**, while casual users drive **ride duration**
3. Station imbalance is a **material operational risk**
4. Growth is steady and sustainable, powered by a **small but critical power-user segment**

---

11. Conclusion

This project demonstrates that **SQL alone—when applied rigorously—can deliver deep, business-ready insights**. By combining time-series analysis, segmentation, net-flow modeling, and retention metrics, UrbanRoll gains a clear understanding of:

- How users behave
- Where operational bottlenecks exist
- Which growth levers matter most

The findings provide a strong analytical foundation for **operational optimization, targeted marketing, and long-term strategic planning**.
