Bay wheels

Melanie Guzman Lopez

Agenda

Intro to Bay Wheels

Objectives

Dataset Overview

Analysis

Conclusions

Intro to Bay Wheels

Bike share program by Lyft operating in the bay area.

Coverage: San Francisco, the East Bay, and San Jose.

Rider options:

- Casual riders: single rides available.
- Members: monthly and yearly membership plans.

Bike types:

- Electric bikes (e-bikes): can be locked at city racks for convenience.
- Classic bikes for traditional ride experiences.

Scale:

- Fleet of over 7,000 bikes.
- More than 550 docking stations across the bay area.

Objective

Evaluate the performance and usage trends of the Bay Wheels program, identify rider preferences and behaviors, and recommend strategies to enhance operations.

Guiding Questions:

Program Effectiveness:

- How is the program being utilized across different regions and stations?
- Are operational resources aligned with demand?

User Behavior and Trends:

- What are the distinct usage patterns for members vs. casual users?
- How do temporal trends (daily, weekly, hourly) impact demand?
- Which bike types (electric vs. classic) are most popular?

Revenue and Optimization Opportunities:

- What are the key revenue drivers for the program?
- How can underutilized resources and stations be optimized?
- Are there opportunities to convert casual users into members?

Dataset Overview

Source:

The dataset was obtained from the official Bay Wheels System Data page provided by Lyft.

Contents:

Includes anonymized trip-level data with details such as:

- **Trip date and time**
- **Starting station name and coordinates**

Scope:

Due to the size of the dataset, analysis was focused on four representative months:

- **January**, **April**, **July**, and **October**.
- This selection aims to capture seasonal variations and usage trends throughout the year.

Preprocessing and Feature Engineering

1. Data Integration:

- Combined multiple monthly files using SQL via psycopg2, resulting in a unified dataset:
- 1,151,729 rows
- 14 columns

2. Missing Values:

- 59,650 rows with missing start_station_name and start_station_id.
- 73,196 rows with missing end_station_name and end_station_id.
- 290 rows with missing end_lat and end_lng.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1151729 entries, 0 to 1151728
Data columns (total 14 columns):
 #   Column              Non-Null Count    Dtype
---  ------              --------------    -----
 0   ride_id             1151729 non-null  object
 1   rideable_type       1151729 non-null  object
 2   started_at          1151729 non-null  datetime64[n
 3   ended_at            1151729 non-null  datetime64[n
 4   start_station_name  1092079 non-null  object
 5   start_station_id    1092079 non-null  object
 6   end_station_name    1151439 non-null  object
 7   end_station_id      1078161 non-null  object
 8   start_lat           1151729 non-null  float64
 9   start_lng           1151729 non-null  float64
 10  end_lat             1151439 non-null  float64
 11  end_lng             1151439 non-null  float64
 12  member_casual       1151729 non-null  object
 13  city                1151729 non-null  object
dtypes: datetime64[ns](2), float64(4), object(8)
memory usage: 123.0+ MB
```

Preprocessing and Feature Engineering

Imputation:	Matched rows using nearest coordinates (end_lat, end_lng) to impute missing end_station_name values.
Deletion:	Removed **290 rows** where end_station_name, end_lat, and end_lng remained null (~0.025% of the dataset).
Result:	Final dataset contains **1,151,439 rows** with no null values.

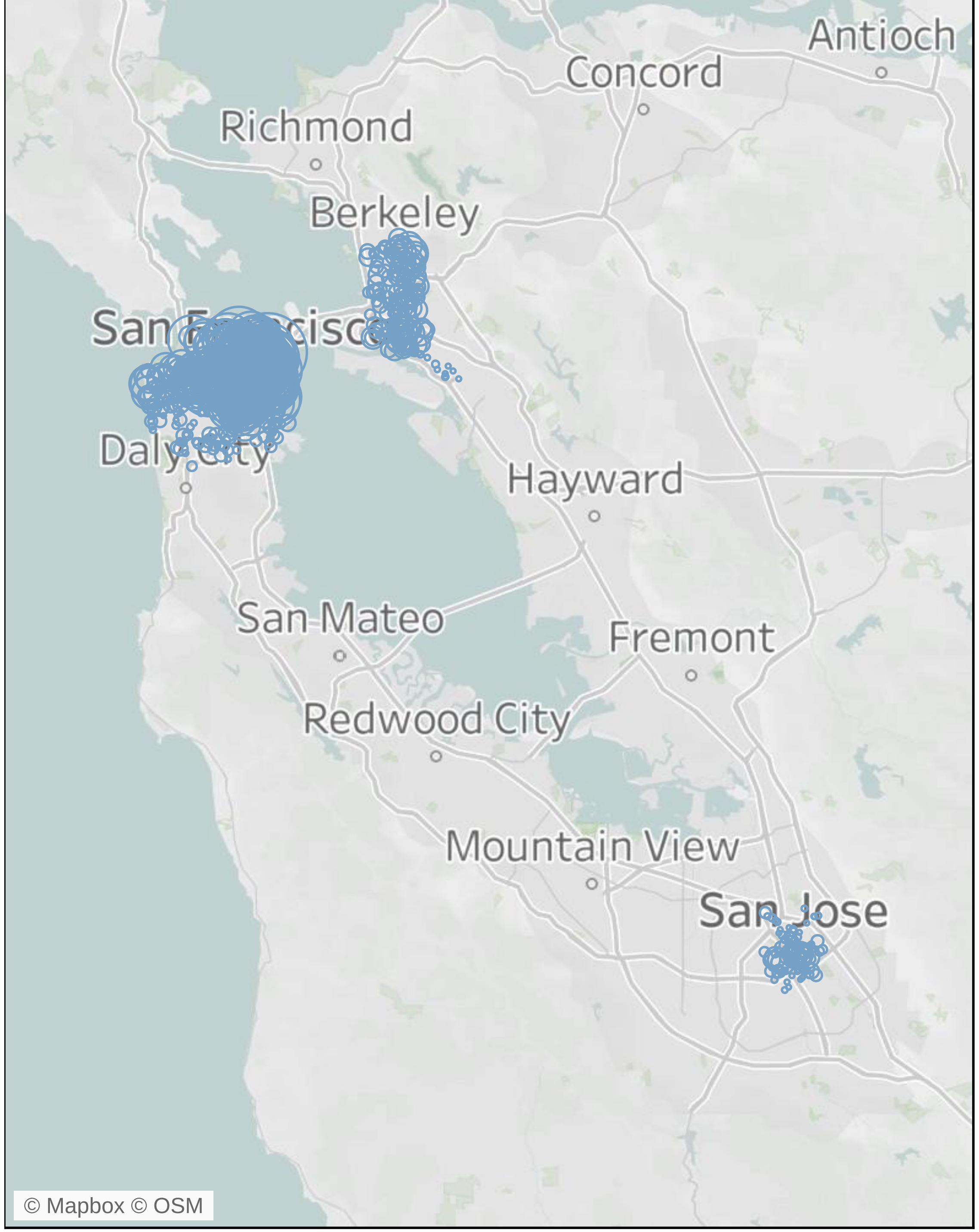
Preprocessing and Feature Engineering

Dropped rows where ride duration was:	**Under 1 minute**: Likely errors. **Over 24 hours**: Unrealistic durations.
Retained rides up to 2 hours	(~5% of data) could represent actual trips.
Dropped rides exceeding **3 hours:**	Considered extreme outliers based on IQR thresholds. Represented a negligible portion of the dataset.

Preprocessing and Feature Engineering

- Created derived columns:
  - **Ride Duration** (minutes), **Day Type** (Weekday/Weekend), **Time of Day**(Morning/Afternoon/Evening/Night).
- **New Features:**
  - **Ride Cost**: Based on trip duration, membership, and bike type.
- **Column Cleanup:**
  - Removed redundant or irrelevant columns (e.g., duplicate ride duration, unused IDs).

Analysis



Station Distribution

Key Observations

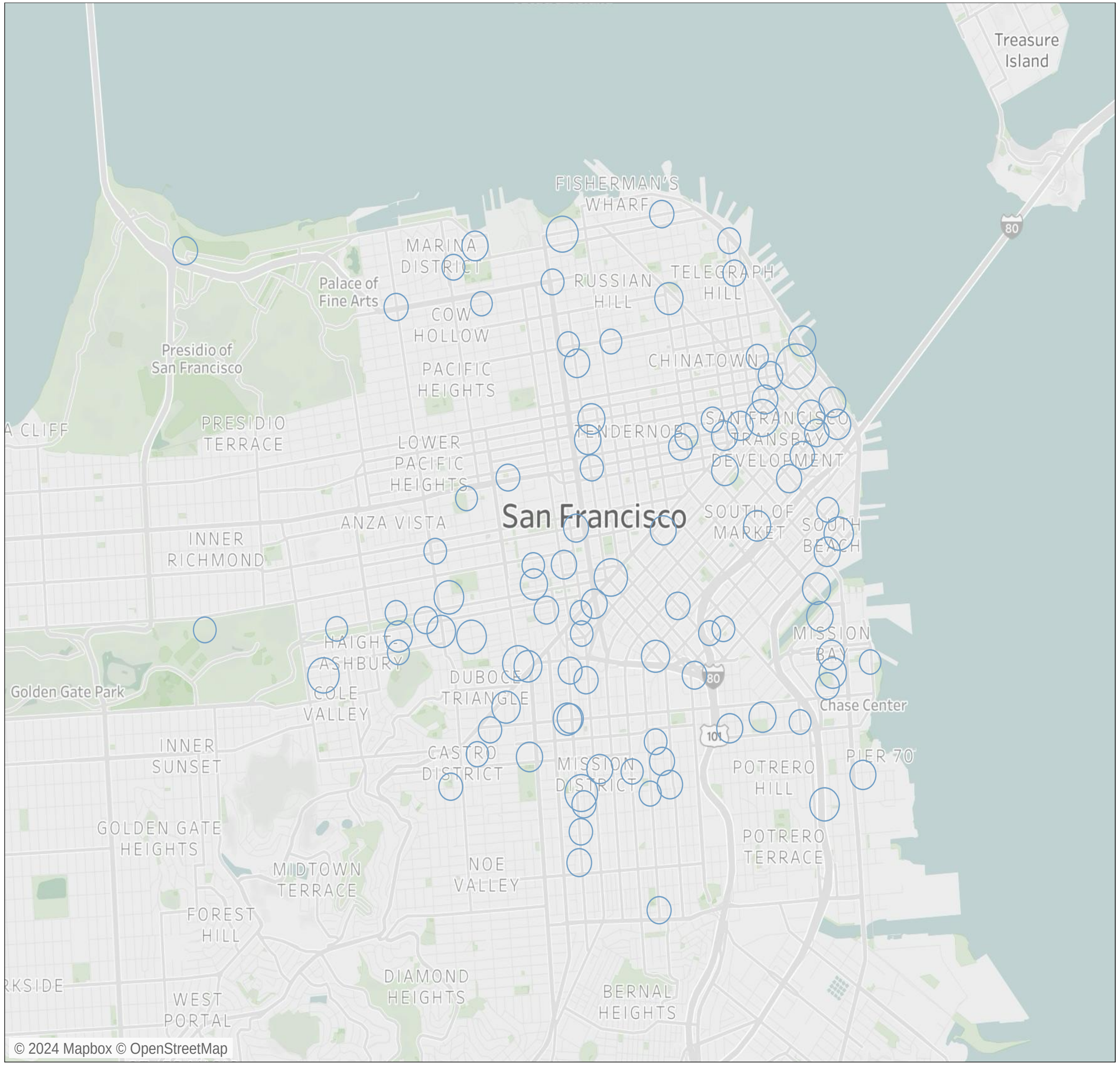
- San Francisco: Highest density of stations.
- East Bay (Oakland, Berkeley, Alameda): A moderate density of stations.
- San Jose: Fewer stations compared to San Francisco and the East Bay.

High-Performing Stations:

Key Observations

The most active stations are concentrated in downtown San Francisco, particularly around:

- Financial District: Likely driven by commuters.
- Embarcadero and Market Street Area: A hub for both tourists and professionals.
- These stations have the highest trip counts, indicated by larger bubbles with darker colors (e.g., 13,284 trips).



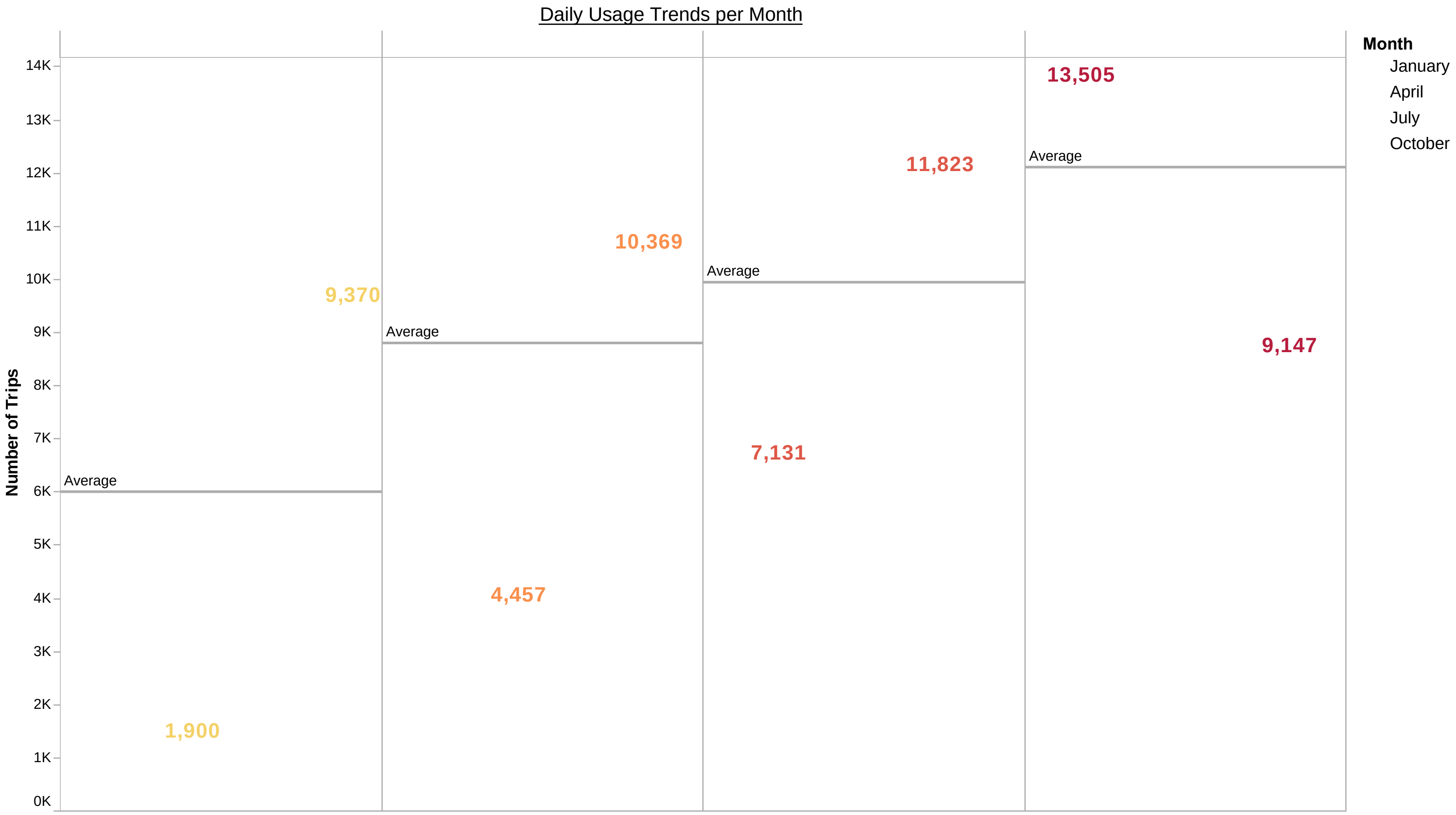
Key Observations

Winter (January): Low ridership.

Spring (April): Moderate increase in trips.

Summer (July): High usage, with a steady trend of high daily trips.

Fall (October): The highest daily usage, possibly due to pleasant weather.



Week Trends

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Casual User	34.92% 45,326	21.75% 35,300	20.78% 41,290	20.47% 37,964	22.77% 39,739	26.35% 39,550	35.78% 47,740
Member	65.08% 84,472	78.25% 127,023	79.22% 157,385	79.53% 147,461	77.23% 134,800	73.65% 110,557	64.22% 85,679

Key Observations

Weekdays Drive Demand:

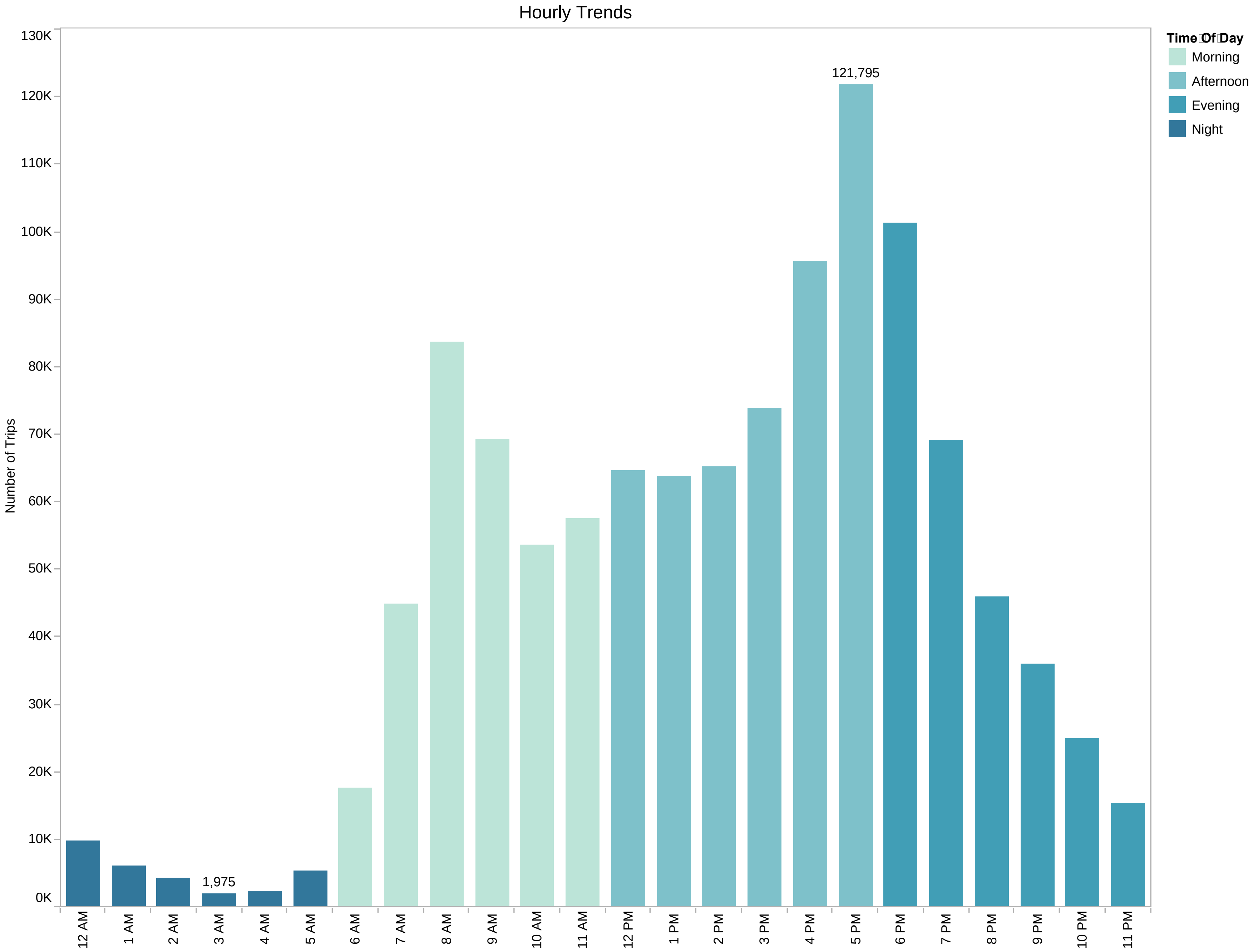
- Weekdays see the highest number of trips, driven primarily by member usage. Tuesday, Wednesday, and Thursday are the busiest days.

Casual vs. Member Patterns:

- Members dominate weekday ridership for commuting, while casual users are more active on weekends compared to weekdays, likely for leisure or tourism purposes.

Operational Planning:

- Weekday operations should prioritize urban and commuter-heavy stations to support member demand. For weekends,

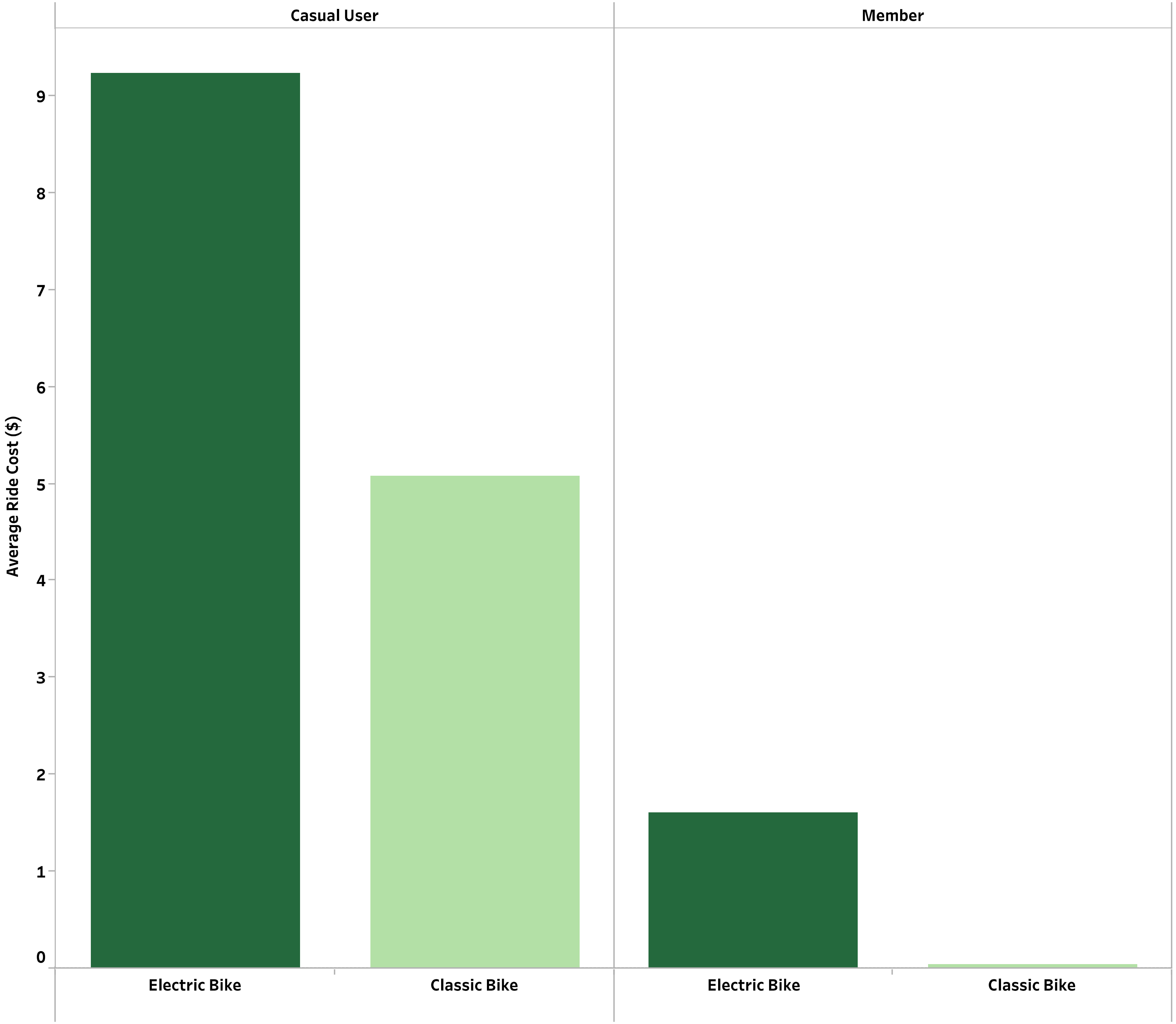


Key Observations

Peak Hours: High ridership during morning commute (7–9 AM) and evening commute (5–6 PM) reflects strong demand from commuters.

Midday Usage: Moderate activity from 11 AM to 3 PM, likely driven by casual users or errands.

Low Demand: Minimal ridership from 12 AM to 5 AM, presenting opportunities for bike redistribution and maintenance.



Key Observations

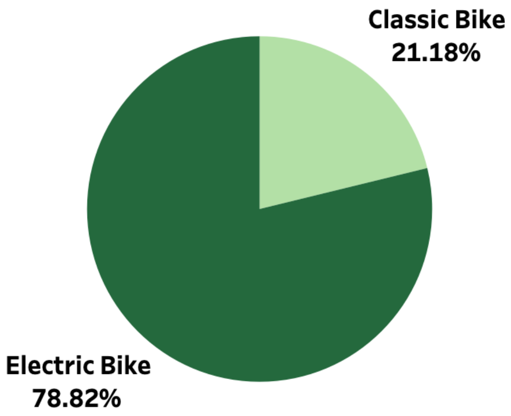
Electric Bikes Generate Higher Revenue:

- Casual users pay ~$9 per ride for electric bikes, significantly more than members, who pay ~$2 per ride.
- Electric bikes are the primary revenue drivers due to their higher costs and popularity.

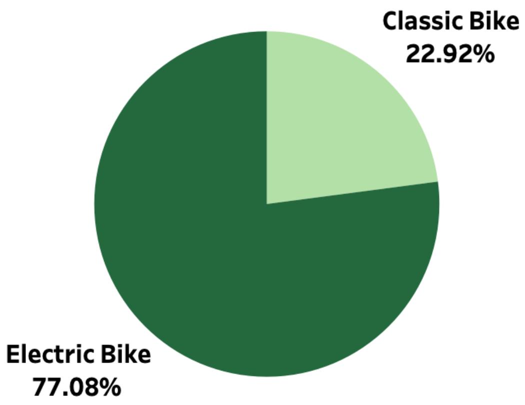
Revenue Dependency on Casual Users:

- Casual users contribute significantly more revenue per ride, making their engagement essential, especially for electric bike usage.

Casual User



Member



Key Observations

Electric Bikes are Preferred:

Casual Users: 78.82% of rides are on electric bikes, highlighting their preference for convenience and speed.

Members: 77.08% of rides are electric bikes, showing a similar trend, likely for commuting efficiency.

Summary of Key Findings

Seasonal Trends:

- Ridership peaks in summer and fall (July, October) and dips in winter (January).

Patterns by Day and Hour:

- Weekdays: Members dominate during peak commute hours (8 AM, 5–6 PM).
- Weekends: Casual users lead, primarily for leisure at recreational hotspots.

User Behavior and Bike Preferences:

- Electric bikes are preferred across all user types for convenience and efficiency.

Revenue Drivers:

- Casual users generate higher per-ride revenue, especially on electric bikes.
- Members offer consistent weekday demand but lower revenue per trip due to memberships.

Strategic Takeaway

Seasonal Optimization:	Launch winter promotions (discounted rides, membership deals). Use off-season for bike servicing and fleet rebalancing.
Operational Alignment:	Prioritize bike availability at commuter-heavy stations (weekdays). Shift resources to recreational/tourist areas (weekends). Promote midday discounts and use overnight hours for maintenance.
Engage Riders:	Offer electric bike discounts to casual users. Highlight membership savings to convert casual users. Introduce weekend perks to boost member activity.
Expand and Maintain Fleet:	Add more electric bikes in high-demand areas. Ensure proactive maintenance for reliability.
Infrastructure and Pricing:	Expand docking stations in high-demand areas. Implement dynamic pricing for off-peak hours.