# Divvy Bikes Insights and Recommendations Report

## Executive Summary

In the ever-evolving landscape of urban mobility, Divvy Bikes stands as a beacon, providing an accessible and eco-friendly solution to Chicagoans. This report provides an in-depth analysis of Divvy Bikes’ operational data for the first two quarters of 2017. Through a conscientious examination of trip and station-wise data, along with demographic, geographic and operational insights, we aim to discover and show key trends and issues in the business operations, offering recommendations for strategic and growth-oriented improvements.

**Table of Contents**

## 1. Introduction 2. Data Overview 3. Data Cleaning and Preparation 4. Exploratory Data Analysis

1. Trip Duration Analysis
2. Station Wise Analysis
3. Quarter Wise Analysis
4. Bike Wise Analysis
5. City Wise Analysis

## 5. Operational Issues

1. Capacity Discrepancy
2. Battery Health and Maintenance
3. Geographical Disparities

## 6. Future Opportunities

## 7. Conclusion

## 1. Introduction

In the current urban environment, the demand for sustainable and convenient transportation solutions has surged. Divvy Bikes offers a simple, affordable, and enjoyable means of exploring Chicago. Just hop on a classic bike, grab an e-bike or scooter — and be on your way. This report delves into the operational data of Divvy Bikes, unravelling trends and providing actionable insights for improved performance.

## 2. Data Overview

The dataset comprises trip-wise and station-wise data for the first two quarters of 2017. The trip data includes information such as trip ID, timestamps, trip duration, customer type, birth year, gender, and station details. Meanwhile, station-wise data encompasses station coordinates, capacity, name, ID, and city. The commonality between both datasets lies in their shared attribute, specifically pertaining to the station names and IDs.

## 3. Data Cleaning and Preparation

Prior to analysis, the data underwent thorough cleaning. The initial data preparation involved concatenating the quarterly datasets, introducing a ‘Quarter’ column as a secondary key to denote the respective quarter. Subsequently, the station-wise data was integrated into the combined dataset through a left join, in order to provide us with comprehensive details about both the pickup and drop-off (“from” and “to”) stations for each entry in the amalgamated dataset of the two quarters. Subsequently, upon inspecting the final prepared dataset for missing values, we identified and addressed multiple instances of empty values in the gender and birth year columns. To mitigate bias associated with removing these values, we opted to replace gender null values with ‘Prefer not to say,’ acknowledging the sensitivity of personal information. Concurrently, birth year null values were marked as ‘Null’ due to inconsistencies in the observed data; notably, the minimum birth year recorded as 1899, an implausible age of 124 years.

## 4. Exploratory Data Analysis

### 4.1 Trip Duration Analysis

The trip duration, initially provided in seconds, was converted to minutes for better readability. Examination exposed a left-skewed distribution marked by notable outliers. Adopted a logarithmic scale for x-axis visualisation for improved clarity. The analysis underscores an average trip duration of around 16 minutes (15.66 precisely) for all users within this timeframe, with over 75% of values encapsulated within one standard deviation of the distribution, proving the skewness of this distribution. **Before Logarithmic Transformation**

A graph of a distribution of tripuration

Description automatically generated

## After Logarithmic Transformation

A graph of a distribution of tripuration

Description automatically generated

## Further breakdown and grouping by Quarters, Gender, and User Type

A graph of a number of different colored lines

Description automatically generated with medium confidence

A graph of two people

Description automatically generated

A graph of a person with a number of numbers

Description automatically generated with medium confidence

Subsequent dissections by quarters, gender, and user type revealed fascinating patterns. Notably, there are distinct variations in trip duration across different quarters and among user types and genders. It is evident that Subscribed Males exhibit the highest bike usage, while day pass customers opting not to disclose their gender surprisingly boast a higher average trip duration than any other group. Additionally, the average usage in Q2 surpasses that of Q1, indicative of business growth during this period.

A graph with blue and orange bars

Description automatically generated A graph of different colored bars

Description automatically generated

Having discerned the usage patterns within the customer base, our focus shifts towards a comprehensive examination of the data from diverse perspectives and aggregations. The objective is to scrutinize the intricacies of the business operations, identifying latent gaps or challenges and formulating strategic solutions.

### 4.2 Station Wise Analysis

Initiating our exploration from the standpoint of station-wise data, we bifurcate our analysis into two pivotal components: the pickup stations (termed as "from" stations) and the drop-off stations (referred to as "to" stations). Our investigation reveals that 'Streeter Dr & Grand Ave' emerges as the station with the highest number of trips both commencing and concluding, signifying its paramount popularity. This prominence can be attributed to its strategic location near Jane Addams Memorial Park, Navy Pier, and the harbour, making it a frequented spot by tourists. Despite this, the station with the maximum bike storage capacity is 'Field Museum,' introducing a potential quandary. A comparative analysis showcases 'Streeter Dr & Grand Ave' recording over 7 times more trips than 'Field Museum,' underscoring a noteworthy problem.

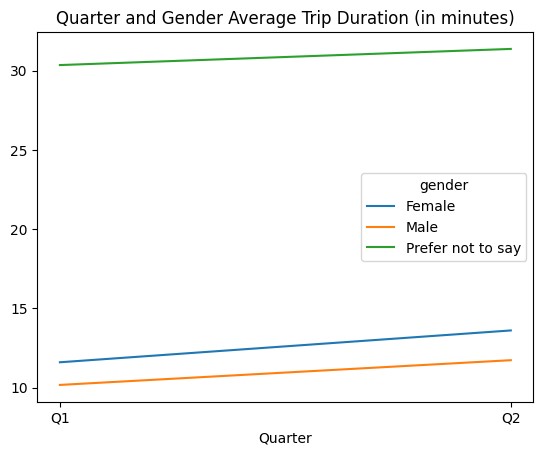
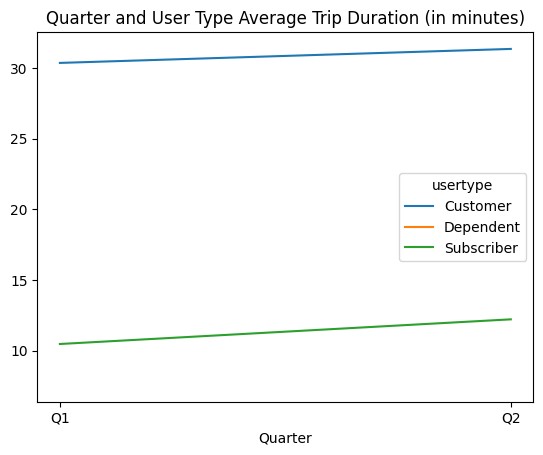
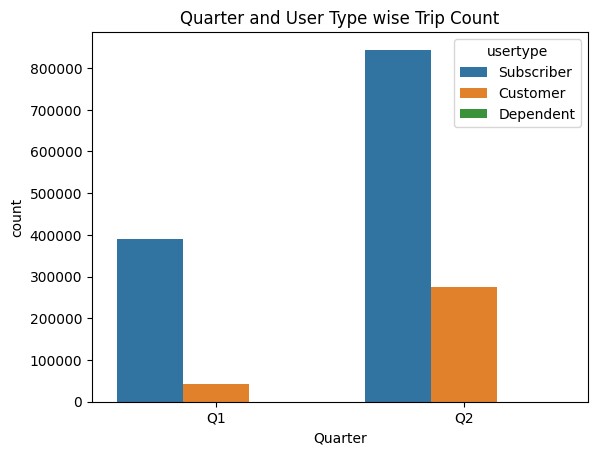
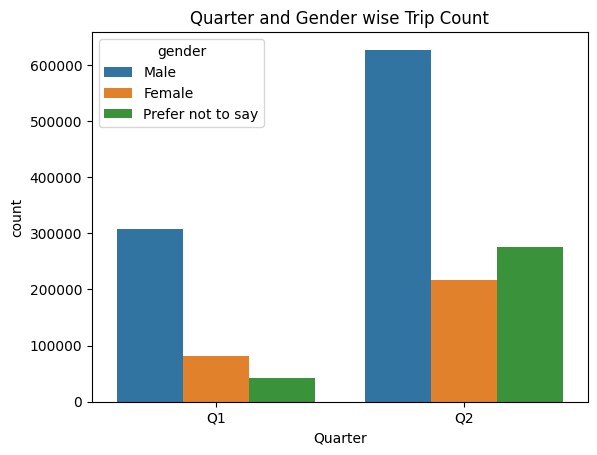
Our analysis distinctly shows the imperative need to enhance the capacity of the most frequented station while concurrently optimizing the capacity of less-utilized ones. This strategic adjustment will help improve cost-effectiveness and also bolster revenue generation.

Further, upon diving deep into the examination of maximum trip durations for both pickup and drop-off stations, we again identify noteworthy instances. 'Cicero Ave & Lake St' in the western part of Chicago and 'Wabash Ave & 87th St' in the southern sector stand out due to their considerable distance, exceeding 9 miles, from the most frequently used station (‘Streeter Dr & Grand Ave’). This geographical insight sheds light on specific areas with extended trip durations, providing valuable data for operational and strategic location planning considerations. Number of stations should be increased alongside with optimisation in the capacities of existing stations to again ensure cost-effectiveness and encompass revenue growth.

### 4.3 Quarter Wise Analysis

Subsequently, we delve into a comprehensive time-series data analysis, focusing on quarterly trends. This approach enables a nuanced examination of various user groups over time.

The findings reaffirm previous observations, presenting a more accessible depiction. Notably, Q2 exhibits a heightened number of trips compared to Q1, with male subscriber trips registering the highest count. Furthermore, an examination of average trip durations reveals that Day pass customers, opting not to disclose their gender, exhibit the longest average trip duration. The discernible trend indicates an increase in trip duration averages from Q1 to Q2, attributed to a surge in usage volume.



### 4.4 Bike Wise Analysis

The pivotal aspect of Divvy Bikes' success lies in the meticulous management of bikes, encompassing their condition, maintenance, and battery health. Consequently, we transition to a granular analysis based on bike IDs.

Ensuring the seamless operation of the business mandates a commitment to the quality and accessibility of bikes within their designated area. This necessitates a proactive approach, encompassing regular inspections, battery replacements, and timely bike substitutions. The bikes that have accrued substantial total trip durations in these two quarters warrant vigilant scrutiny. To uphold their condition, Divvy Bikes should implement routine battery replacements for these bikes. Moreover, after reaching a predetermined threshold of total trip duration, the company should consider retiring these bikes.

Similarly, for bikes with a high count of trips, even if of shorter durations, the company should maintain a vigilant watch. The elevated number of trips signifies heightened utilization, suggesting a greater impact on wear and tear. Despite shorter individual trip durations, these bikes should undergo regular assessments to identify damages and ensure their continued usability. Two such examples of bikes are with IDs 5880 and 2565.

### 4.5 City Wise Analysis

After scrutinizing various business facets, one crucial realm awaits exploration: the location and geographical analysis. Divvy Bikes spans three cities—Chicago, Evanston, and Oak Park. Chicago, as the largest, operates as the hub with nearly 50 times more stations than the other two combined. Unsurprisingly, Chicago dominates Divvy Bikes' customer base. The eastern region, notably the most frequented station, 'Streeter Dr & Grand Ave,' fosters a seamless biking experience. However, stations in the west exhibit lower usage, presenting untapped potential. Despite existing customers, the dispersed stations pose challenges. This scenario implies significant expansion opportunities, particularly in Evanston and Oak Park. Addressing spatial distribution and enhancing accessibility could propel Divvy Bikes towards broader market capture.

(Geo Heat Map attached in the Python File for reference)

## 5. Operational Issues and Recommendations

### 5.1 Capacity Discrepancy

The mismatch between station usage and capacity poses a significant operational challenge.

Recommendations include reallocating bikes based on demand and expanding capacity at high-traffic stations.

### 5.2 Battery Health and Maintenance

High-usage bikes, such as those with IDs 5880 and 2565, should undergo regular battery replacements to maintain service quality. Establishing a preventive maintenance schedule is crucial for bike longevity and customer satisfaction. Can also launch bike rating schemes in the application for users to rate their experience on the basis of bike. The CSAT score can be used to conduct immediate proactive checks on bike quality.

### 5.3 Geographic Disparities

While the east of Chicago dominates in usage, the west exhibits potential for growth. To address this, strategically placing new stations in underutilized areas inside and outside Chicago (including Evanston and Oak park) can foster balanced accessibility and increased usage.

## 6. Future Opportunities

### 6.1 Gross Profit Analysis

Upon obtaining customer ID data, conduct a thorough gross profit analysis to understand revenue patterns and optimize membership plans. This would boost customer acquisition and help consumer base growth.

### 6.2 Demographic Segregation

Enhance data collection techniques and hence the quality to enable demographic analysis based on age, put limits on the birth year input which cannot be greater than the maximum age limit present in the specific dynamic scenario. Tailor station placement and capacity to the preferences of different age groups.

### 6.3 Customer ID Analysis

With the inclusion of customer ID data, strategize customized offers and promotions. Shorten the customer journey to yearly subscriptions by offering personalized incentives.

## 7. Conclusion

In conclusion, this report encapsulates a comprehensive analysis of Divvy Bikes’ operational data. By unravelling usage patterns, station dynamics, and geographic nuances, we’ve identified operational challenges and provided actionable recommendations. Embracing these insights can propel Divvy Bikes towards enhanced efficiency, customer satisfaction, and sustainable growth.