**INTRODUCTION**

**Divvy Bike Sharing Program in Chicago: A Detailed Analysis of 2017 Data (quarter 1 and 2)**

****

Bike sharing programs have emerged as a Sustainable, Healthy and Efficient alternative to cars in the dynamic urban mobility landscape. Among them is Divvy Bikes, a network of bicycles and docking points designed for both residents and visitors which can be found in Chicago. The report analyses the Divvy bike sharing programme in detail, with an emphasis on the 1st and 2nd quarters of 2017.

Two major datasets are at the heart of our analysis: **the trip records (Divvy\_Trips\_2017\_Q1.csv and Divvy\_Trips\_2017\_Q2.csv) :** The trip records provide an overall overview of each journey carried out by users of this program, which shall include trip duration, start and end times, station details as well as user types.

**the station information (Divvy\_Stations\_2017\_Q1Q2.csv) :** The station dataset, On the other hand, valuable information regarding geographic spread, capacity and online dates for bike sharing stations are available in a station dataset.

**Objective:**

* ***To understand user behaviour***: By analysing trip patterns, durations, and frequencies, we seek to gain a deeper understanding of how users interact with the bike-sharing system.
* ***To assess the operational efficiency***: By examining station capacities and usage, we aim to evaluate the efficiency of the bike distribution and docking system.
* ***To identify trends and growth opportunities***: By scrutinizing the data from multiple angles, we aim to uncover trends that could guide future expansion and operational strategies.

**DATA PRE-PROCESSING AND EXPLORATION**

***Ensuring Data Integrity and Usability***The foundation of any data analysis lies in meticulous data preprocessing and exploration. In this project, we embarked on a journey through the extensive datasets of Divvy Bikes, encompassing trip details and station information for the first half of 2017. This section outlines the critical steps undertaken to transform raw data into a clean, coherent format suitable for in-depth analysis.

**1. Data Loading and Initial Assessment**Our analysis began with the loading of two key datasets: trip records for the first (Q1) and second (Q2) quarters of 2017, and station information for the same period. Leveraging the powerful capabilities of Python and its libraries, we efficiently imported these datasets into our analytical environment.

**2. Data Merging and Concatenation**With two separate datasets for Q1 and Q2 trip records, our immediate task was to merge them into a single comprehensive frame. This step ensured consistency in analysis and enabled a holistic view of the data across the six-month period.**3. Handling Missing and Incomplete Data**Data quality checks revealed instances of missing and incomplete records. We meticulously handled such anomalies by removing records with missing start or end times or filling or replacing the genders and birthdate columns.

A black background with white text

Description automatically generated

**4. Data Type Conversions and Formatting**To facilitate accurate calculations and interpretations, we performed essential data type conversions. Date and time fields were formatted to appropriate datetime types, enabling temporal analyses like identifying peak usage hours. Additionally, numerical fields were confirmed for correct data types to ensure accurate mathematical operations.

**5. Enriching Data with Calculated Fields**Our exploration extended to the derivation of new variables from existing data. For instance, we calculated the day of the week (based on their start dates) and categorized trips. These enhancements provided deeper layers for analysis, allowing for more nuanced insights into user demographics and behavior patterns.

**6. Exploratory Data Analysis (EDA)**The EDA phase provided initial insights into the data through descriptive statistics and visualizations.

**DATA ANALYSIS AND VISUALIZATION**

* **Geospatial Analysis**: Using latitude and longitude data, we conducted a geospatial analysis to visualize the distribution of stations and their usage across Chicago. This helped in pinpointing areas with high demand and those needing additional infrastructure.

**A blue and orange dots

Description automatically generated**

*A colorful image of a person

Description automatically generated with medium confidenceHeatmap across Chicago map:*

* **Docking point capacity** - daily bike check-outs and check-ins at each Divvy station to assess whether the docking capacity (dp capacity) of the stations is sufficient to handle daily traffic. (False states the insufficiency of capacity)

A screenshot of a computer

Description automatically generated

* **Bike Utilization and Maintenance Needs:** Bike Usage Analysis: We tracked the usage patterns of individual bikes, identifying those with the top 50 highest usage and potential maintenance needs. This aspect of the analysis was vital for efficient maintenance scheduling and ensuring the longevity of the bike fleet.

A graph of blue and green bars

Description automatically generated

**A graph of a number of trips

Description automatically generatedStation Utilization**: Station Traffic: We assessed the flow of bikes in and out of stations, identifying the most and least popular stations. This analysis was crucial for understanding spatial patterns in bike usage and for potential station capacity optimization.

* **Revenue Generated:** The code snippet provided is a straightforward approach to estimate the revenue generated by Divvy Bikes from various user types, categorized as Subscribers, Single Ride users, and Day Pass customers. It begins by setting fixed prices for each user type: The revenue from each user type is then calculated by multiplying these rates with the respective counts of users or trips, derived from the dataset. Specifically, it counts the number of unique bikes used by Subscribers, the total number of trips by Single Ride users, and the total number of trips by Day Pass customers .

A screen shot of a computer

Description automatically generated

* **Peak Day in a week:** visualizes the distribution of Divvy bike trips across different days of the week to identify the busiest day . This visualization aids in understanding weekly patterns in bike usage, which is crucial for operational planning and resource allocation.

A graph with blue line

Description automatically generated

* **User Behavior Analysis**: Types of Users: We distinguished between different user types and analysed their usage patterns. This analysis provided insights into the varying preferences and demands of different user segments.A comparison of a green and blue bar chart

  Description automatically generated
* **Temporal Trends**: Seasonal and Hourly Trends: Our analysis extended to uncovering how usage varied by time of day and across different months. We observed notable variations, with peak usage during specific hours and seasonal fluctuations that aligned with Chicago’s weather patterns.

*A graph of blue green and blue bars

Description automatically generatedPeak Usage Hour in a Day:*

*Bike Usage by Season:*

*A graph with numbers and a bar

Description automatically generated*

* **Bike Route Patterns:** an analysis to identify the busiest routes in the Divvy Bikes dataset. This visualization provides a clear view of the most popular routes in the bike-sharing network, highlighting areas with the highest traffic and potential focal points for resource allocation or infrastructure improvements.**A graph of a number of colored bars

  Description automatically generated with medium confidence**
* **Bike traffic imbalance:** an analysis is conducted to identify Divvy bike stations experiencing a significant imbalance in bike traffic, specifically those with more bikes checked out than checked in. Stations with a net change less than zero indicate more check-outs than check-ins. The analysis further focuses on stations with a substantial deficit, defined as those with a net loss of more than 100 bikes. The resulting data is then visualized in a bar chart, highlighting stations with the most significant deficits in bike availability.

A screenshot of a graph

Description automatically generated

* **Correlation between User Demographics and Usage Patterns** : Gender Dynamics By integrating user demographic data we explored correlations between gender and bike usage. This analysis shed light on the demographic segments most actively using the bike-sharing system. (only for Subscribers , Customers are Not Provided)

A pie chart with text

Description automatically generated

* **Approximate Distance covered analysis :** the distance of each trip in the Divvy Bikes dataset using the Haversine formula, which determines the great-circle distance between two points on a sphere given their longitudes and latitudes. The process begins by defining the Haversine function to calculate distances distance in kilometers for each trip based on these coordinates.

**A screenshot of a computer

Description automatically generated**

* **Predictive Insights for Enhanced Resource Allocation :** In an effort to optimize the allocation of bicycles and better understand user demand patterns at individual stations, a time series analysis was conducted focusing on Station 50 of the Divvy bike-sharing network. This analysis done here is not accurate as there is no considerations of internal and external factors but this is critical in predicting future bike demand, thus aiding in efficient resource management and improving user experience.

**A graph with blue lines and orange lines

Description automatically generated**

**SUGGESTIONS AND CONCLUSIONS**

Through above visualization figures, I have noticed several interesting phenomena.

* Most usage is around Chicago Union Station and nearby. There are some stations rarely used.
* Heat maps to visually represent high and low usage areas. This can help in strategic planning for expanding or reducing station numbers in certain areas.
* 81,909 stations are having insufficient dp capacity in a day . where in-case there is a high possibility of getting traffic in check-in of bikes .
* station 35 is the most used station and can consider for marketing strategies and resource allocation for high bike demands.
* Tuesday is the peak day and 5pm is the peak hour.
* Sunday is the least used day . where the maintenance services can be performed.
* People use sharing bicycles more in spring than summer then winter (more frequently usage and longer trip duration)
* People use sharing bicycle more frequently in weekdays than weekends, but the average trip duration is longer in weekends than weekdays.
* Top most frequently used routes can be used to inform decisions on where to focus maintenance efforts and possibly expand bike lanes.
* Columbus Dr & Randolph St is the highest turn overed station where there is more bikes checked-out than check-in . there might be a need for more bikes or docking space.
* From station 66 to station 171 is the approximately highest distance of 3.54kms .
* Actual Bike demand is always high as compared to the Forecast bike demand.

# References

1. Verma, R. (2023b) 'Exploratory data analysis of Chicago Divvy Bicycle Sharing,' *Medium*, 16 January. https://python.plainenglish.io/exploratory-data-analysis-of-chicago-divvy-bicycle-sharing-c441f1036113.
2. Ghaniwala, B. (no date) *Bicycling to level the field: A study of Divvy in Chicago*. https://ecommons.luc.edu/luc\_theses/4433/?utm\_source=ecommons.luc.edu%2Fluc\_theses%2F4433&utm\_medium=PDF&utm\_campaign=PDFCoverPages.
3. Faghih-Imani, A. and Eluru, N. (2015) 'Analysing bicycle-sharing system user destination choice preferences: Chicago’s Divvy system,' *Journal of Transport Geography*, 44, pp. 53–64. https://doi.org/10.1016/j.jtrangeo.2015.03.005.
4. Jaber, A. and Csonka, B. (2023) 'Investigating the temporal differences among bike-sharing users through comparative analysis based on count, time series, and data mining models,' *Alexandria Engineering Journal*, 77, pp. 1–13. https://doi.org/10.1016/j.aej.2023.06.087.
5. Zhou, X. (2015) 'Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago,' *PLOS ONE*, 10(10), p. e0137922. https://doi.org/10.1371/journal.pone.0137922.