**Capstone Final Project Summary**

**Name: Brittany Glover**

**Project Title: Chicago Divvy Bike 2022 Analysis – Divvying Up Divvy’s Users**

**Data Collection Source:** *The 2022 data for this project was collected directly from Divvy’s website. There was also data collected from the City of Chicago’s data portal. Links to both are listed below.*

* Divvy: <https://divvy-tripdata.s3.amazonaws.com/index.html>
* City of Chicago: https://data.cityofchicago.org/Transportation/Divvy-Bicycle-Stations/bbyy-e7gq

**Project Summary:**

The objective of this analysis is to compare the usage of Divvy’s services for member and casual consumers. With this analysis, my goal is to find where casual consumers stand with rental usage and how we could potentially convert those users into annual members for Divvy. The target audience is Divvy’s business development/marketing team. After completing the data analysis, the findings should show the Divvy team different trends for the casual consumers that could update their marketing strategies to help sway them from using Divvy for single rides/24 hours passes and using the system for an annual membership.

After pulling the monthly data sets from Divvy’s website, I brought each CSV file into Python to complete the data cleaning, formatting, and the initial exploratory analysis. The original size of the data contained 5,667,717 rows when considering January – December 2022 data. To complete the data cleaning, I checked for null values which resulted in 3,463,328 null values. I decided to remove those as I am trying to look at actual trends which require the dataset to be fully complete. After removing the null data, I removed special characters that were found in some of the columns, as well as updated the verbiage to be correctly uppercase/lowercase letters. The data types had to be updated to datetime for the ‘Started At’ and “Ended At” columns. I created two new columns to break out the date separately to make grouping of the analysis easier to complete. I also added an additional column to account for each trip length in seconds. I created another column to list the month of the trip, so the ride counts can be broken down per month with an aggregation calculation. A column was added for the starting hour of the trip which could help determine the time of day that’s busiest for rental rides. The last column created was to account for the day of the week of the trip; this column was added to analyze the busier days of the week for rentals.

After adding these additional columns, I now felt like I had all the information needed to get a better picture to complete my analysis. To further complete the data formatting, I sorted the “Trip Length” column values by smallest to largest and saw that there were negative trip values, I removed those rows and also removed rows that had trips less than 60 seconds as it was noted by Divvy that trips listed as less than 60 seconds, were most likely due to a docking issue and not an actual ride. I removed the rows related to “docked” bikes as I’m strictly looking for data that involves actual trip rentals from the two bicycle types, electric and classic. Lastly, I removed the rows that were related to temporary or test stations as I want to look at stations that are permanent and consistently used.

Once I had all the data cleaned and formatted in the way that I thought would give the best analysis, the total row count was 3,458,245. Due to the size of this cleaned dataset, I decided to take a look at the top 50 most frequented stations to compare the usage trends between the casual and member consumers so I could work with a more manageable sized dataset. After the query was done for the top 50 most frequented stations, my dataset was at a total of 985,282 rows. Once the Python analysis was completed, I brought the data into Tableau to create my visuals. Along with the data collected from Divvy’s website, I collected data from the City of Chicago’s data portal to pull in geographical data of all of the Divvy’s stations in the city. I used this data in Tableau to create a map of the top 50 more frequented stations that is used in my analysis of my dataset from Python.

After bringing my datasets into Tableau to start visualizing the information, I found a couple of outliers in the dataset. The majority of the top 50 most frequented stations, are either clustered into the downtown area or along the Lakeshore Drive in the northern part of the city. There are two outliers that are showing as popular stations in the Hyde Park/Woodlawn area. These two areas are further South when compared to the rest of the data. However, those two stations are near the University of Chicago so it could be popular due to students in the area. When comparing the totals for the trip count and the totals for the trip length, I used an average for each pane so it could be easily visible what the averages were for member vs. casual riders per month and bike type, as well as if each month fell either above or below the average.

When looking at the top 50 most frequented start stations, the analysis showed that there are a higher number of membership riders using Divvy’s service vs. casual riders. However, there isn’t a large variance in the total percentages between the two. Casual users account for 43% of the total rides, while membership users account for 57%. Members are more consistent with their rentals throughout the year and days of the week, whereas casual riders seem to be using the service more during the peak summer months and weekend days. When looking at trip lengths, casual riders have the higher average total in trip lengths for the busy ‘peak’ season and weekend days. When considering that causal riders tend to use the service more on the weekends and during the summer months, it can be assumed that the rides are used more for leisure purposes rather than work or commuting. Looking at rentals for the time of day, the busiest times for both members and casual riders seem to be between 4:00PM and 6:00PM which could infer that casual members are also using the bikes for work and not only leisure. Knowing this information, this could have potential to persuade casual riders into becoming an annual member.

After taking into consideration the full analysis of this dataset, I have the following recommendations:

1.**Notifications/email reminders** to casual riders after a rider uses a single/24-hour pass, there could be phone/email notifications sent that could convince them to switch to a membership based off of the long-term pricing vs. a single use price.

2. **Digital campaign** to appeal to more riders, highlighting the financial benefits of using the bike service vs. car rideshare service. The campaign could also note that Divvy bikes aren't just for tourist usage but could also be beneficial for work/leisure commuting for residents in the city.

3. **Divvy/Lyft membership** - put more of an emphasis on the Divvy & Lyft all access pass. If a casual rider is using Divvy and Lyft often, an all-access pass could allow them to enjoy the annual divvy membership benefits, free Lyft priority pickup upgrades, exclusive pricing on preferred Lyft rides, free roadside assistance for your own vehicle through Lyft, free rental car upgrades, and a free GrubHub+ membership (for $16.58/month billed annually)