FINAL REPORT – DSC465

Group Name : Visual Thinker Group

Dataset Title: Chicago Divvy Bike Sharing Data

Source: Chicago Data Portal – Divvy Trips Dataset

Team Members: Amar Sawant , Ankita Kshirsagar and Sultan Almazrouei

1. Introduction :

The Chicago Divvy Bike-Sharing dataset is a comprehensive collection of trip-level records from Chicago's official bike-share program, Divvy. Each entry in this dataset represents a single bike trip and contains detailed information such as start and end timestamps, station names and IDs, trip duration, bike type (classic or electric), and user type (member or casual rider). Covering data from 2013 to 2022, the dataset captures the evolution of bike-sharing usage across nearly a decade in one of the largest cities in the United States.

The data is publicly available through the Divvy System Data portal and is also accessible on Kaggle under the dataset "Chicago Divvy Bicycle Sharing Data", which consolidates multiple years of historical trip data. The full dataset includes millions of records, making it extremely rich but also computationally heavy to process. Earlier snapshots of the dataset contained more than 9 million rows and 23 columns, and as the system expanded, the total number of records grew significantly reaching tens of millions of trips city-wide. Due to this large size, handling the combined 2013–2022 data requires careful filtering, aggregation, or year-wise preprocessing, especially when using visualization tools like Tableau.

This dataset provides a unique opportunity to study real-world urban mobility patterns. It enables analysis of:

- Peak hours and seasonal trends in bike usage
- Most popular stations and routes across the city
- Differences in behavior between casual riders and annual members
- Growth trends in ridership over the years
- Adoption of electric bikes and changing user preferences

- Spatial usage patterns when combined with station geolocation data

Beyond basic trip analysis, the dataset can support deeper applications such as infrastructure planning, station rebalancing strategies, demand forecasting, sustainability studies, membership conversion insights, and understanding the relationship between bike usage and urban design.

Overall, the Chicago Divvy dataset is ideal for data visualization because it transforms massive, complex raw data into meaningful stories about how people navigate the city. With millions of real trips recorded over multiple years, this dataset offers valuable insights for city planners, researchers, businesses, and anyone interested in the future of sustainable urban transportation.

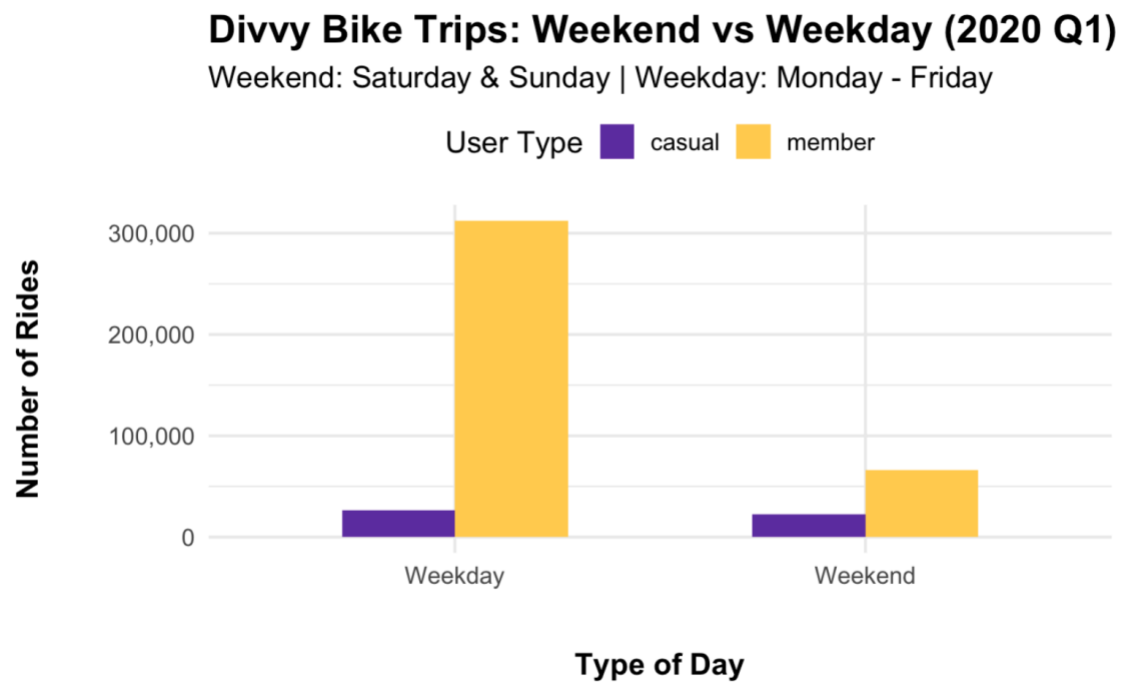
2. Dataset Analysis and details :

This dataset contains detailed trip records from the Divvy Bike Share program in Chicago, where each row represents a single bike trip with information such as start and end times, station locations and IDs, bike type, and user type (member or casual). Depending on the selected time period (monthly or yearly), the dataset can range from approximately 100,000 to over 1,000,000 trip records. It includes a mix of categorical variables (bike type, station names, user type), numerical fields (latitude and longitude), temporal elements (trip start and end timestamps), and geographical data that maps trips across the Chicago area, making it suitable for time-based, spatial, and behavioral analysis.

3. Visualizations:

3.1 Visualization 1 : Grouped Bar Chart Comparing Weekday vs Weekend Trips by User Type

In this analysis, I first loaded the required R libraries such as dplyr and ggplot2, followed by importing the Divvy trip dataset. After loading the data, I created a new variable to categorize each trip as either a weekday or weekend based on the day of the week. Once this classification was added, I grouped the dataset by type of day and user type to calculate the total number of rides in each category. Finally, I used ggplot2 to generate a bar chart comparing weekday and weekend ridership for casual and member users, allowing me to visually analyze differences in usage patterns across the two user groups.

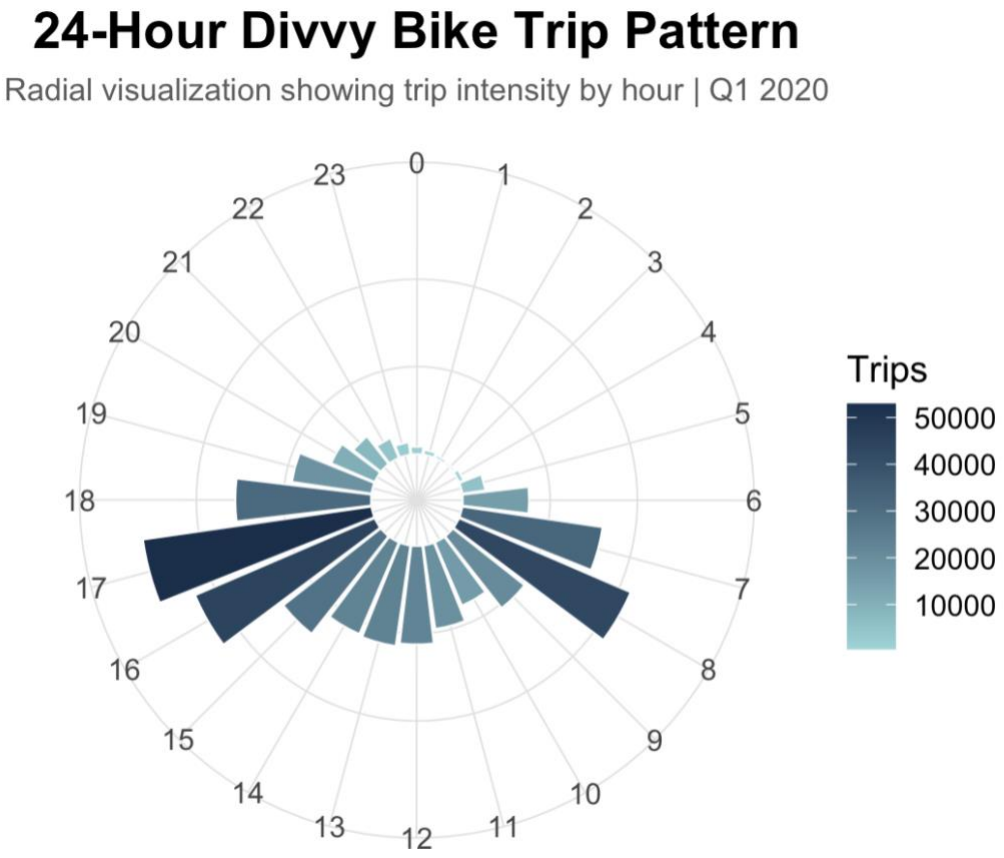


Observation

The graph shows a clear difference in Divvy bike usage between weekdays and weekends during Q1 of 2020. Member riders dominate weekday usage, contributing well over 300,000 trips, which suggests that members primarily use Divvy bikes for commuting or routine travel. In contrast, casual riders have significantly fewer weekday trips, indicating they are less likely to use bikes for daily commuting. On weekends, both member and casual rides decrease, but the drop is much larger for members. Casual usage remains relatively steady across weekdays and weekends, suggesting casual riders tend to use bikes more for leisure or occasional trips rather than structured daily routines. Overall, the data highlights strong weekday commuting patterns among members and more flexible, leisure-focused patterns among casual users.

3.2 Visualization 2 : Polar Bar Chart / Clock-Style Visualization - 24-Hour Radial Trip Pattern Chart

Steps: In this part of the analysis, I first grouped the dataset by hour of the day and calculated the total number of trips for each hour. After preparing the summarized data, I used ggplot2 to create a bar chart and then transformed it into a polar coordinate system using coord_polar() to achieve a radial or clock-style visualization. I mapped trip counts to both bar length and color intensity so higher ridership hours appear longer and darker. Finally, I customized the labels, theme, and color scale to clearly show trip volume variations across the 24-hour cycle, resulting in an intuitive visualization of hourly Divvy bike usage.



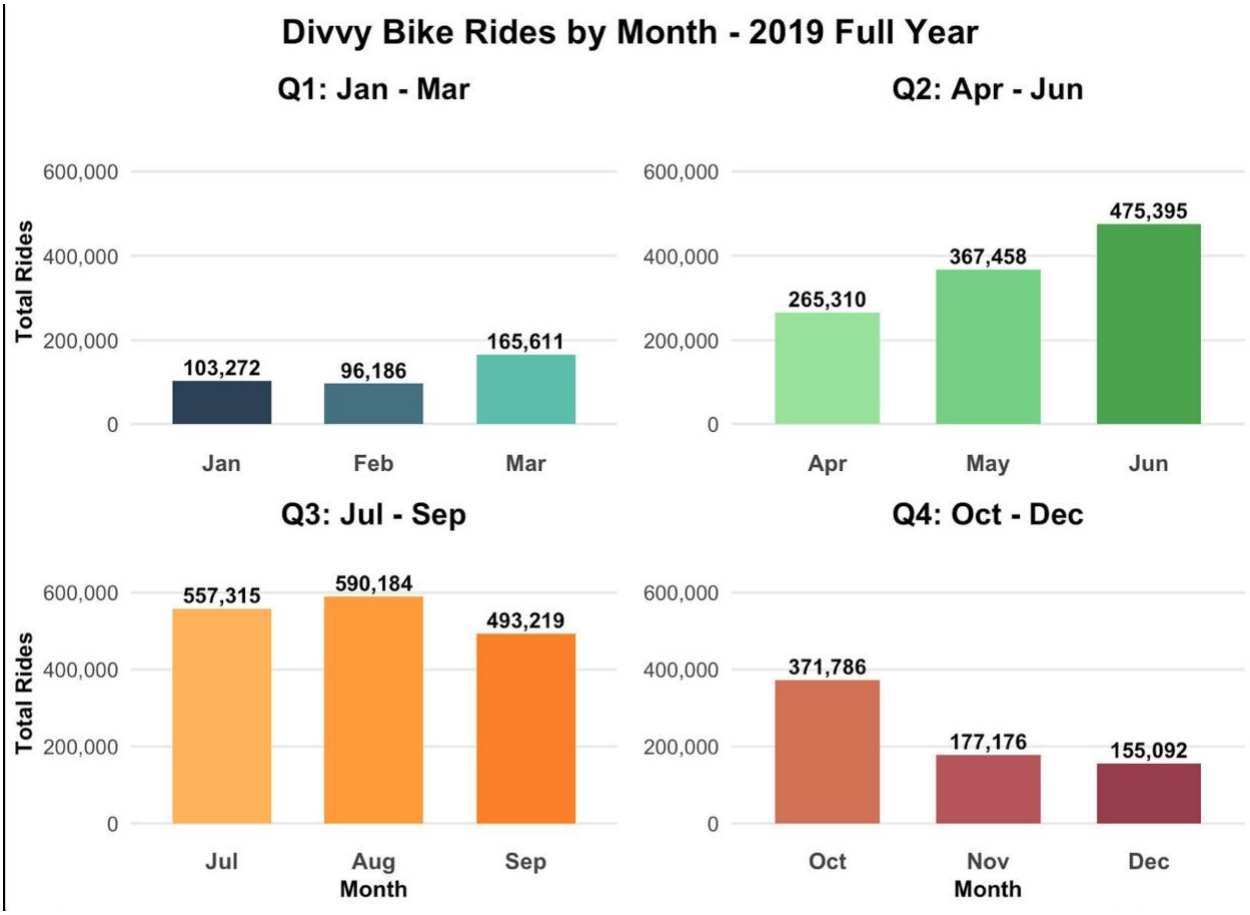
Observation

This radial bar chart visualizes Divvy bike trips across a 24-hour day for Q1 2020, revealing clear hourly usage patterns. The chart shows that bike activity is lowest during early morning hours (around 1 AM–5 AM), steadily increasing after 6 AM. There is a noticeable peak around 5 PM–6 PM, which aligns with typical evening commute hours, indicating that many users rely on Divvy bikes for returning home after work. A smaller but visible increase also appears around 8 AM–9 AM, reflecting morning commute activity. Midday hours maintain moderate trip levels, while late-night usage drops sharply. Overall, the pattern highlights a strong commuter-driven behavior, with the highest ridership occurring during evening rush hour.

3.3 Visualization 3 : Grouped Monthly Bar Charts (Quarterly Breakdown)

Steps:

To create this visualization, I first grouped the dataset by **month** and calculated the total number of rides for each month in 2019. After summarizing the data, I added a column to indicate each month's **quarter (Q1–Q4)**. Using **ggplot2**, I plotted monthly bar charts and then used facet_wrap() (or separate plots combined later) to display each quarter in its own panel. I applied distinct color palettes for each quarter to visually differentiate the seasons, and added data labels showing the exact monthly ride counts. Finally, I customized the axes, titles, and themes to clearly present the seasonal ridership patterns across the entire year.



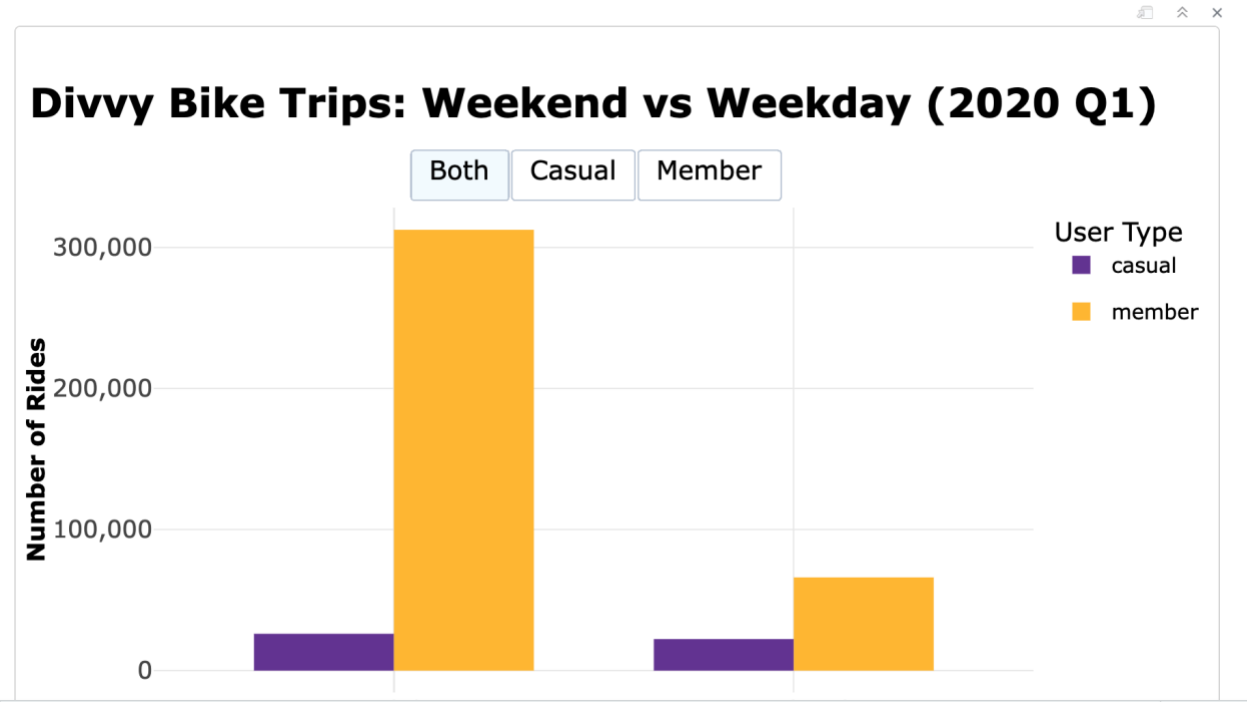
Observation

This visualization shows the total number of Divvy bike rides for each month of 2019, separated into four quarterly panels. The data clearly illustrates strong seasonal trends. Ridership is lowest in Q1 (Jan–Mar), with January and February showing the smallest counts due to winter weather, and a slight increase in March. Q2 (Apr–Jun) shows a sharp rise, indicating the start of the biking season as temperatures improve. Q3 (Jul–Sep) represents the peak biking period, with July and August crossing over 550,000 rides, making them the highest ridership months of the year. In Q4 (Oct–Dec), ridership drops again, with October still moderately high but November and December falling significantly as winter approaches. Overall, the visualization highlights a strong correlation between weather/season and bike usage, with clear peaks in summer and lows in winter.

3.4 Visualization 4 : Interactive D3 Bar Chart (Toggle Filter: Both / Casual / Member)

Steps:

To prepare the data for my D3 visualization, I first loaded the Divvy dataset into R and created a new variable to classify each trip as either weekday or weekend based on the day of the week. Then, I grouped the data by type of day and user type and calculated the total number of rides for each group using dplyr. After summarizing the counts, I cleaned and formatted the dataset to keep only the required columns - weekday/weekend, user type, and total rides. Finally, I exported this summarized dataset as a CSV file from R so it could be used directly in the D3.js script for creating the interactive bar chart.

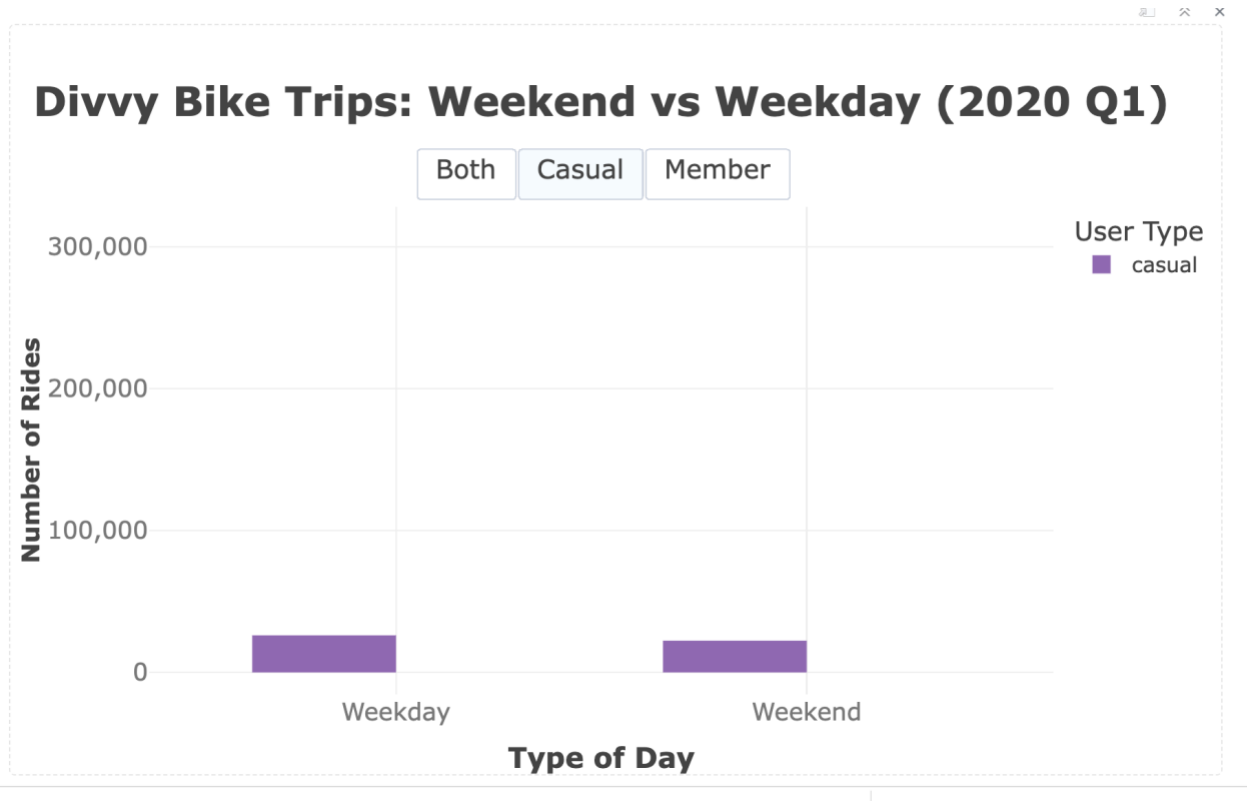


Observation

This interactive D3 visualization compares Divvy bike rides between weekdays and weekends in Q1 2020, with toggle buttons that allow switching between Both, Casual, and Member riders. When viewing Both, the chart clearly shows that member riders dominate weekday usage with over 300,000 rides, highlighting strong work-week commuting patterns. Weekend ridership is significantly lower for members but remains more balanced for casual riders, who use bikes more consistently for leisure. Switching to Casual shows much lower

ride counts overall, with casual riders maintaining similar but modest usage across weekdays and weekends. The Member view highlights a sharp contrast: high weekday rides and much smaller weekend activity. This interactivity helps emphasize behavioral differences between user groups and makes the comparison more intuitive.

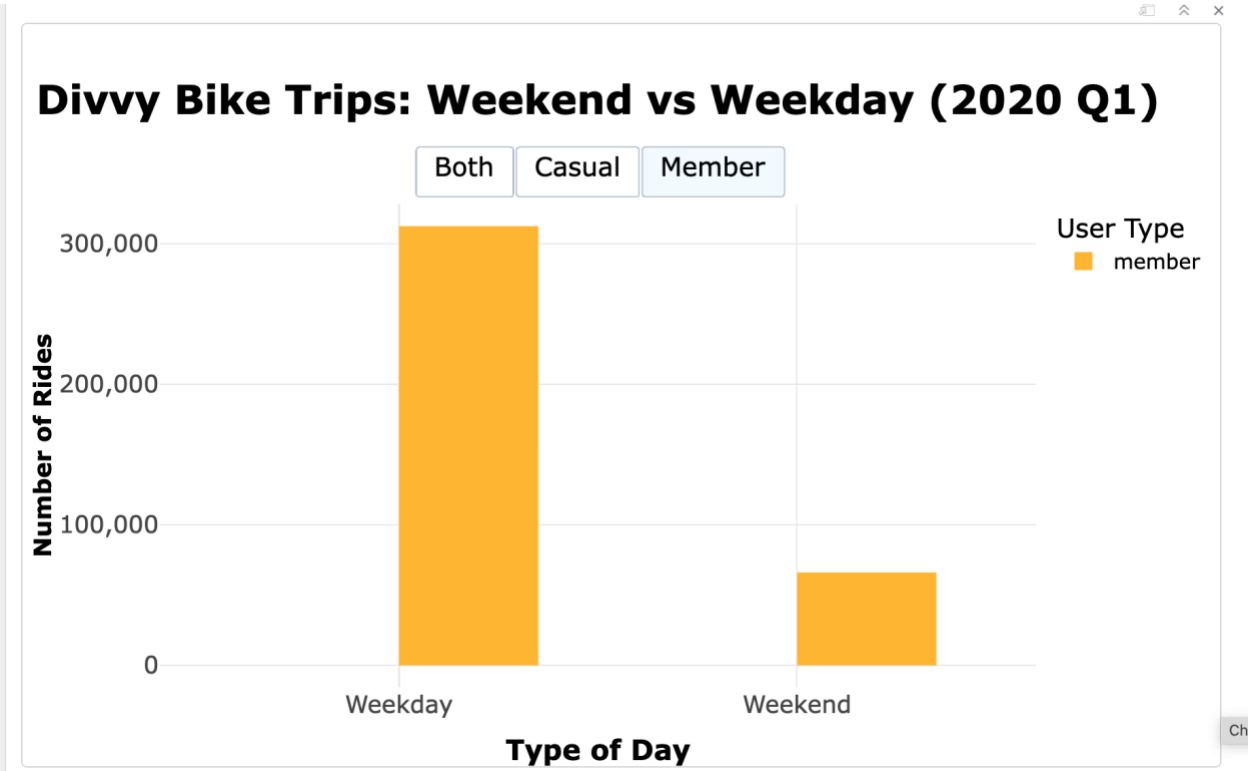
User Type: Casual



Observation :

In the casual-only view of the Weekend vs Weekday chart for Q1 2020, the number of rides remains relatively low and consistent across both weekdays and weekends. Casual riders show no strong preference for either type of day, suggesting that their bike usage is more leisure-oriented and less tied to daily commuting patterns. The similar ride counts across days indicate flexible, occasional usage rather than routine trips.

User Type: Member



Observation

In the member-only view of the Weekend vs Weekday chart for Q1 2020, member riders show a very strong preference for weekday usage. Weekday trips exceed **300,000 rides**, which is significantly higher than weekend usage, reflecting consistent commuter behavior. On weekends, member ridership drops sharply to less than half of weekday levels, indicating that members primarily use Divvy bikes for routine weekday travel rather than recreational weekend trips.

4. Conclusion

Across all visualizations, the Divvy bike-sharing data reveals strong and consistent patterns in how riders use the system throughout the year. The monthly and quarterly charts show that ridership is highly seasonal, climbing sharply in the warmer months (April–September) and reaching its peak in July and August, while winter months experience the lowest usage due to weather conditions. The 24-hour radial chart highlights clear commuting behavior, with noticeable peaks during morning and especially evening rush hours. The weekday-vs-weekend visualizations further emphasize the difference between rider groups: **member riders** primarily use Divvy bikes on **weekdays**, reflecting routine commuting patterns, whereas **casual riders** maintain lower but more balanced usage across all days, suggesting leisure-oriented trips. The interactive D3 charts reinforce these insights by allowing a focused comparison of casual, member, and combined usage patterns. Overall, the visualizations together demonstrate that Divvy bike usage in Chicago is strongly influenced by season, time of day, and rider type, with members driving weekday commuter traffic and casual riders contributing more to flexible, recreational travel.

5. Future Work

Several opportunities exist to expand this analysis and gain deeper insights:

- Include Weather Data: Integrating temperature, precipitation, and wind data could help explain daily and seasonal fluctuations in ridership.
- Incorporate Geospatial Analysis: Mapping routes, identifying the busiest corridors, and visualizing station density can reveal spatial mobility patterns.
- Predictive Modeling: Building machine learning models to forecast ridership based on season, weather, and time of day could support operational planning.
- Bike Rebalancing Optimization: Using trip flow data to analyze where bikes accumulate versus where shortages occur can help improve fleet distribution.

6. Individual report

6.1 Amar Sawant - Polar Bar Chart - 24-Hour Radial Trip Pattern Chart

The 24-Hour Radial Trip Pattern chart provides a detailed visualization of Divvy bike usage across an entire day for Q1 2020, capturing how ridership varies from hour to hour. Unlike standard bar charts, this radial design mirrors the shape and flow of a clock, making it intuitive for understanding daily cycling rhythms. The chart reveals that ridership is extremely low during the late-night and early-morning hours between 1 AM and 5 AM, then begins to rise steadily after 6 AM as people start their day. A smaller peak appears around 8–9 AM, reflecting morning commute activity, while the highest and most prominent peak occurs around 5–6 PM, indicating the evening commute rush. Midday hours show moderate levels of usage, representing a blend of casual riders, errands, and shorter non-work trips.

During the creation of this visualization, I faced several challenges, particularly in transforming the standard hourly bar chart into a clean, interpretable radial format. First, I had to ensure that the trip timestamps were correctly processed and converted into hourly groups without losing data due to time formatting issues. Once the hourly counts were accurately summarized, I experimented with multiple chart types-including line graphs and traditional bar charts-but found that these formats did not clearly capture the repetitive 24-hour cycle or highlight peak commute times in an intuitive way. I then decided to use a polar coordinate system in R through coord_polar(), which required additional adjustments to spacing, bar alignment, and color intensity so the chart remained readable. Another challenge was controlling overlap and ensuring the colors and labels did not clutter the circular layout. After several iterations, the radial bar chart emerged as the most effective and visually engaging way to show the natural ebb and flow of Divvy ridership across a full day

Overall Conclusion

The 24-Hour Radial Trip Pattern chart clearly illustrates how Divvy bike usage varies across the day and highlights strong hourly riding behaviors. The radial design makes it easy to see that ridership remains very low during late-night and early-morning hours (1 AM–5 AM), begins to rise around 6 AM, and reaches its highest levels during the typical **evening commute window around 5–6 PM**. A smaller but noticeable increase also appears in the morning around 8–9 AM, consistent with workday travel patterns. The circular layout effectively emphasizes the cyclical nature of daily movement, visually mirroring the rhythm of a 24-hour clock. Overall, this chart demonstrates that Divvy bike usage in Chicago is strongly shaped by commuter schedules, with clear peaks during traditional rush hours and minimal activity during nighttime.

6.2 Ankita Kshirsagar - D3 Interactive Weekend vs Weekday Visualization (2020 Q1)

This D3.js interactive visualization presents a comparison of Divvy bike usage on weekdays versus weekends for Q1 2020, with the added ability to filter the data by rider type: Both, Casual, and Member. When "Both" is selected, the visualization highlights the stark contrast between member and casual riders, with members dominating weekday ridership and showing a noticeable decline on weekends. The "Casual" view reveals much lower but relatively even usage across weekdays and weekends, supporting the idea that casual riders primarily use the service for leisure or non-routine trips. Meanwhile, the "Member" view shows a significant spike during weekdays and a large drop during weekends, reflecting strong commuter behavior typical of regular Divvy subscribers. By providing these three perspectives in a single chart, the D3 visualization helps users interactively explore usage differences that would be difficult to observe in a static plot. Setting up the interactive buttons in D3 was another challenge, as each button needed to trigger a re-rendering of the bars without refreshing the entire page.

Why I Chose D3 : I chose D3 for this visualization because it provides the interactivity and dynamic updates that static R or Python charts cannot offer. The toggle buttons allow users to instantly filter the data and compare rider types more easily. Smooth transitions and automatic axis updates make the visualization more engaging and intuitive. D3 is also ideal for web-based sharing, making it the best choice for creating a flexible, interactive, and user-friendly chart.

Overall Conclusion

Overall, the D3 interactive Weekend vs Weekday visualization clearly highlights how differently casual and member riders use Divvy bikes during Q1 2020. Member riders show a strong weekday dependency on the system, with significantly higher ridership connected to daily commuting, while their weekend trips drop sharply. Casual riders, in contrast, show much lower but steady usage across both weekdays and weekends, indicating that their riding patterns are more leisure-based and not tied to work routines. The interactive filters Both, Casual, and Member allow the viewer to isolate each rider category and interpret patterns without overwhelming the chart with too much information at once. This level of interactivity provides a more intuitive and engaging understanding of usage behavior than a static plot could. Through this visualization, it becomes clear that rider type plays a major role in shaping Divvy usage trends, with members driving weekday commuter traffic and casual riders showing a more flexible, recreational pattern.

6.3 Sultan Almazrouei - Grouped Bar Chart Comparing Weekday vs Weekend Trips

This graph visualizes Divvy bike usage during weekdays and weekends for Q1 2020, comparing how casual and member riders differ in their riding patterns. I created a grouped bar chart because it clearly shows the contrast between the two user types across two categories of days. The chart reveals that member riders take significantly more trips on weekdays - over 300,000 rides indicating strong commuter behavior, while their weekend rides drop sharply, showing less leisure use. Casual riders, on the other hand, maintain relatively consistent and much lower ride counts on both weekdays and weekends, suggesting more flexible, recreational usage. While working on this graph, I faced challenges such as categorizing each trip correctly into weekday or weekend, handling missing or inconsistent date formats, and ensuring that the summarized data for each user type was accurate before plotting. I also had to decide the best chart type initially considering line and stacked bar charts but ultimately chose a grouped bar chart because it best highlights differences between user types and is easier to interpret visually. Plotting this graph helped me clearly understand how travel behavior varies by rider type, making it an important visualization for analyzing Divvy usage patterns.

Additionally, in my initial version of this graph, I plotted all individual days (Monday to Friday separately, and Saturday–Sunday separately), but the result looked cluttered, repetitive, and harder to interpret. It also introduced unnecessary detail when the goal was simply to compare weekday versus weekend behavior. After reviewing the visualization, I decided to simplify it by grouping days into just two clear categories Weekday (Mon–Fri) and Weekend (Sat–Sun). This modification made the graph much cleaner, more readable, and more aligned with the analytical question I wanted to answer. The simpler version effectively highlights the core pattern without overwhelming the viewer.

Overall Conclusion

Overall, this graph shows a strong divide between member and casual riders in how they use Divvy bikes during weekdays and weekends. Member riders clearly dominate weekday ridership, reflecting consistent commuting behavior tied to the traditional workweek, while their weekend usage drops significantly. Casual riders, in contrast, maintain low but steady usage across both weekdays and weekends, indicating that their trips are more leisure-focused and not driven by daily routines.

6.4 Combine group contribution: Grouped Monthly Bar Charts (Quarterly Breakdown)

After each group member generated their own set of visualizations for the Divvy dataset, we collectively reviewed all the graphs to identify common trends and meaningful patterns. Through this collaborative analysis, we realized that comparing monthly ride counts across all four quarters would provide a clear and comprehensive understanding of Divvy's seasonal ridership behavior. Each member contributed insights from their individual graphs some focused on hourly trends, others on weekday vs weekend usage, and others on rider-type comparisons which helped us recognize the importance of summarizing ridership by month. Based on these combined observations, we decided to create this grouped monthly comparison chart for the full year of 2019. This visualization allowed us to present a unified conclusion about how weather, seasons, and commuter habits influence ridership patterns, making it a key result of our group collaboration.

7. Team Work :

Throughout this project, our team collaborated closely to analyze the Divvy bike-sharing dataset and create meaningful visualizations. Each team member worked independently at first, exploring different aspects of the data such as hourly patterns, weekday-weekend trends, rider-type behaviors, and seasonal changes. After generating our individual graphs, we met to discuss our findings, compare insights, and identify overlapping patterns. This collaborative review helped us refine our approach and decide which comparisons would provide the most value in explaining Divvy's ridership trends. By combining everyone's strengths-data cleaning, R visualization skills, D3 interactivity, interpretation, and design we produced a coherent set of visualizations that highlight different perspectives of the dataset. Our teamwork ensured that the final results were comprehensive, accurate, and visually clear.

8. Conclusion:

In conclusion, our analysis of the Divvy bike-sharing data revealed clear and consistent patterns driven by season, time of day, and rider type. Ridership peaked during warmer months and dropped significantly in winter, showing strong seasonal influence. Member riders showed pronounced weekday commuting behavior, while casual riders displayed more flexible, leisure-based patterns. To reach these insights, we used **both R and Tableau for exploratory data analysis**, which allowed us to validate our findings across multiple

tools and ensure accuracy. R helped in detailed data manipulation and custom visualizations, while Tableau provided quick, interactive exploration and pattern identification. Together with additional D3 visualizations, these tools helped transform a large and complex dataset into clear, meaningful insights about Chicago's urban mobility patterns.