



Navigating Speedy Success

How do our annual members and casual riders use our platform differently

August 26, 2024

Overview

This case study aims to research how the different user types, casual vs annual, interact with the Cyclistic platform and compare the two. Data analysis will focus on total rides between users, ride duration, popular locations, bike types, and how the different users interact with the platform over time. Cyclistics finance analysts have concluded that annual members are significantly more profitable to the business than casual riders. Increasing annual members is the current primary business goal in order to provide the strongest position for future growth. Annual members generate more consistent revenue streams and demonstrate higher levels of customer loyalty, showing that they are of high value to the long term growth strategy. Therefore, understanding how casual users interact differently will allow us to target them appropriately with a strong marketing campaign to effectively increase total annual users.

Goals

1. **Understand how casual riders and annual members use Cyclistic bikes differently:** This process will occur by analysing the historical data provided by Cyclistic over a 12 months period. With intentions to discover trends and correlations that show how the different members interact with the bikes over the course of the day, week, month and year. Specific data such as ride duration, location and frequency will be utilised.
2. **Design a new marketing strategy:** using the insights gained from the data analytics a new marketing strategy will be developed to convert casual riders into annual members. Potential outcomes include targeted campaigns, personalised offers and membership incentives. The new marketing strategy will be measured for its success in the following

months utilising data on conversion rates from casual to annual memberships and retention rates ongoing.

Specifications

Cyclistic is a Chicago based bike share company that started in 2016. Since its creation the company has grown to a fleet of 5,824 bicycles that are geo-tracked and locked into a network of 692 different stations located across the city of Chicago. Members interact with the bikes by unlocking them at one station and return to any other station in the system at any time of day.

Our primary stakeholders are two fold:

1. Cyclistic Executive Team - responsible for reviewing and actioning the new marketing strategy
2. Lily Moreno, Director of Marketing - responsible for the development of campaigns and initiatives to promote the bike-share program.

Cyclistic's finance analysts have determined through their research that Annual Members provide greater profit opportunities to the business than the Casual riders. Accepting that whilst the pricing flexibility helps the business attract new members, converting existing Casual riders to Annual members is key to longer term growth. As such, the opportunity exists to create a new marketing plan that will target converting Casual Riders to Annual Members. With this clear goal in mind, first the data team needs a better understanding of how the casual and annual members differ, why casual members would buy a membership and how digital media could affect the marketing tactics chosen.

Milestones

1. Prepare historical data

The first step of this case study is to access, clean and organise the historical data provided by Motivate International Inc, that covers a 12 month period of bike usage. The data used was permitted for use under this licence agreement (<https://www.divvybikes.com/data-license-agreement>).

2. Process and analyse data

Once the data is cleaned, processing can begin including performing calculations to assist with further insights such as determining individual trip duration, identifying potential outliers, breaking down available information into logical/comparable columns. Once each dataset is processed it will be exported into Tableau for analysis and prepared for visualisation.

3. Design Visualisations for Effective Comparisons

Using the processed and combined datasets in Tableau, clear and insightful visualisations will be created to show trends, correlations and helpful comparisons between the user types. From this, a basic marketing strategy can be developed.

Methodology

For this case study, Excel and Tableau were the tools of choice. Excel was chosen for its robust cleaning capabilities of files of this size and Tableau chosen for its powerful visualisation tools. The process began by sourcing and downloading the 12 CSV files that covered the 12 month period from August, 2023 to July, 2024. These files were provided by Cyclistic via Motivate International Inc and are publicly available for download. The data has been provided under a specific licence, which permits its use for analysis but prohibits sharing the dataset on its own. Once the files were downloaded, they were individually opened in Excel where the following methodology was performed.

Data Cleaning and Preparation:

- Checked the ride_id column, which served as the unique identifier, to ensure there were no duplicate records. The rideable_type column was cleaned using the TRIM function to remove extra spaces and renamed to bike_type for clarity.

- The start_date_time and end_date_time columns were reformatted to dd/mm/yyyy hh:mm:ss format to standardise the timestamps. To aid a detailed analysis, additional columns were created by extracting specific components from start/end date/time columns such as Start_date, start_month, start_day and start_hour.
- A new column was created that calculated the ride duration in minutes and seconds, providing a clearer understanding of trip lengths.
- At this point, in depth cleaning of the trip length was required. Outliers were addressed by examining the trip duration in seconds. Rides shorter than 10 seconds were deemed as unreliable and therefore removed, these may have been recorded due to a system or user error. Additional cross-referencing of start_lat/lng and end_lat/lng data was completed. If both the start and end coordinates matched and the ride duration was under 30 seconds the data was removed as being potentially unreliable.
- Rides under 30 seconds that had different start and end coordinates were retained as they were deemed likely valid.

Handling Long Ride Durations:

- To address excessively long ride durations, an interquartile range (IQR) was used. Any rides above the upper bound were labelled as outliers. A new column, "is_extreme", was created, categorising each ride as either "extreme" or "normal". These values remained as they could have a significant impact on the overall analysis. The additional column allows for these "extreme" outliers to be excluded for analysis and comparison later if required.

Handling Missing Data:

- A significant amount of data was missing from the start_station_name/end_station_name and start_station_id/end_station_id columns. As string information is easier to interpret and provide more intuitive insights, the station_id columns were removed. Missing station names were populated by cross-referencing available latitude and longitude data. If the coordinates matched existing station names, the blanks were filled accordingly; otherwise, they were labelled as "unknown".

Final Adjustments:

- Lastly, the “member_casual” column was trimmed for extra spaces and renamed to “user_type” for consistency and clarity across the dataset.
- All unnecessary columns were deleted, leaving: ride_id, bike_type, start_date, start_month, start_day, start_hour, ride_durations_mins, start_station_name, end_station_name & user_type. This cleaned and streamlined dataset was then ready for analysis enabling a clear comparison of usage patterns between casual and annual users.
- The 12 files had the same process applied and at the end were combined into a single CSV using a code on MacOS Terminal. Then imported into Tableau for the visualisation process.

Data limitations:

Once data cleaning, processing and preparation was completed it became evident that additional data could have provided even deeper insights. The absence of a unique rider identifier was notable as it limited the ability to fully understand how each user type interacts with the platform over time. Knowing the number of non-repeat users would have opened up opportunities for targeted marketing to those who left the platform after a single use.

Additionally, the substantial amount of missing start and end location data could mean that certain popular locations were missed in the analysis. Moreover, the dataset did not provide information on bike availability at each station, leading to potential location bias for those stations with more bikes. For instance, it was discovered that the most popular location was *Streeter Dr & Grand Ave*, most likely due to its proximity to the popular tourist location, *Navy Pier*. This may have led to there being more bikes available consistently, providing a greater opportunity for users to begin their journey here. The missing location data may obscure important trends or key stations that could be targeted in the marketing campaign.

Weather data was also excluded from the dataset, weather patterns could have revealed seasonal trends and temperate data that influenced usage behaviours of different user types.

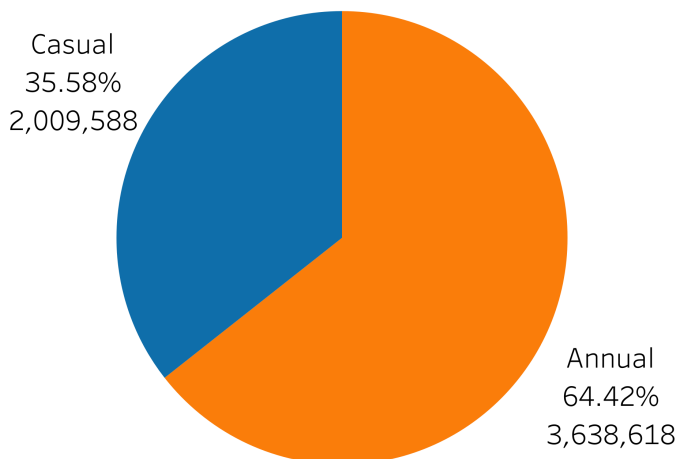
Additionally, the lack of demographic data, combined with the absence of unique rider IDs, restricts the ability to design targeted marketing campaigns for specific user groups, such as younger casual riders. Furthermore, there was no mention or data regarding previous promotional or advertising campaigns that could have led to spikes in casual membership usage. The dataset only covered a 12 month period, which limits the analysis to a short term perspective. A longer time could provide more comprehensive insights.

Despite these limitations, it can still be confidently determined that the findings are both insightful and actionable. The available data, while incomplete in certain areas, reveals clear trends and key insights that can inform the development of an effective marketing strategy.

Analysis & Key Findings

Tableau was utilised for its powerful visualisation tools and how it can convey data precisely and with clarity. Visualisations were created to compare the differences between annual users and casual users of Cyclistic over the 12 months period. It should be noted that the data taken for analysis covered the date range of August 2023-July 2024 and that when a 12 month period is shown it covers this period. Firstly, it is important to understand how the user base is split.

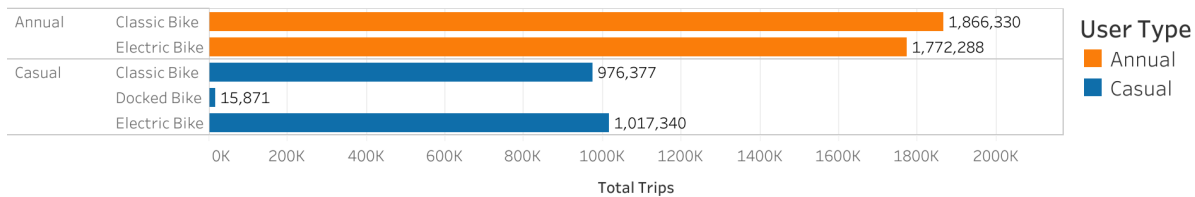
Cyclistic User Comparison



This chart shows us the breakdown of the Cyclistic user base between two user types; annual and casual users. It shows that Annual users take 64.42% of the rides in a 12 month period for a total of 3,638,618 trips. Whereas, casual users take just 35.58% of the total trips, as 2,009,588 trips. Initially this might infer that there is a greater volume of annual users taking more trips, however, it in fact shows a limitation to the data itself. All it shows us is that the majority of trips taken were done so by annual members, as there is no data for rider ID to link multiple rides to a single user.

From here bike type data was investigated to discover any potential trends.

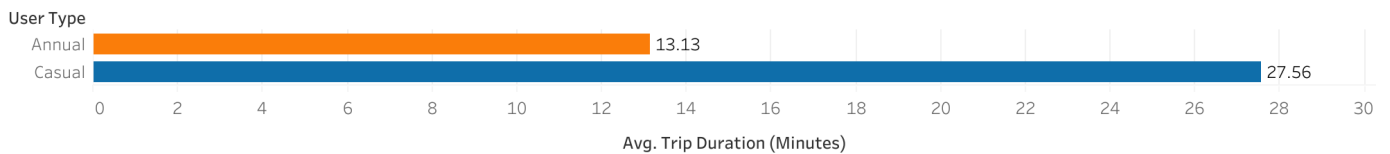
Total Trips per Bike Type by User Type



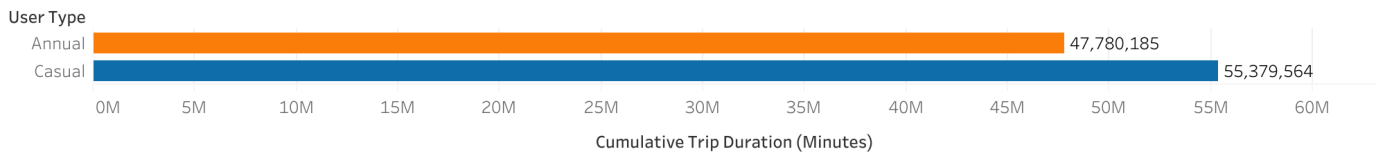
The chart above shows that annual users tended to use classic bikes slightly more than electric bikes, whereas casual users tended towards using electric bikes slightly more. At this stage more investigation would be recommended to understand the distribution of bike types geographically to learn if that may impact user decisions. This would allow better understanding for a marketing campaign to place certain bikes in certain locations that would improve user experience.

Next was to analyse trip durations, here there are two charts to show the average trip duration in minutes for each user type as well as the total minutes spent on a bike per user type over the 12 months period. 1

Average Trip Duration by User Type

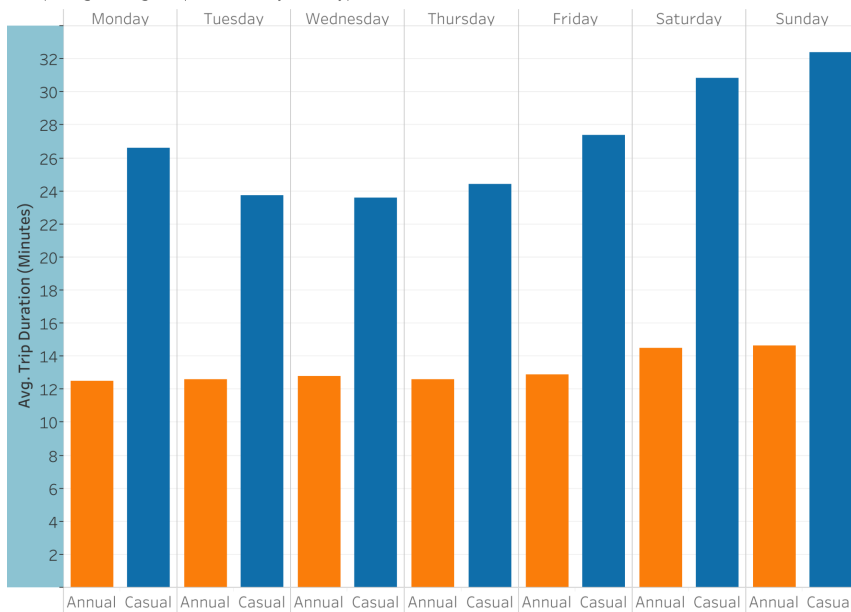


Cumulative Trip Duration by User Type



These two charts show that the average trip duration by casual users was significantly higher than that of the annual users, just over twice as long. Cumulative trip duration data shows us that, even though the Casual user makes up only 35.58% of total rides taken, the casual user type spent more time on any bike compared to the annual user. Once again, the limitation of not knowing how many unique riders make up this dataset limits us to making reliable inferences as to how this may impact any decisions. However, learning that casual users spend more time on the bikes per ride is noteworthy. While not necessarily informing marketing decisions, this does provide a good usage point of comparison between the two user types.

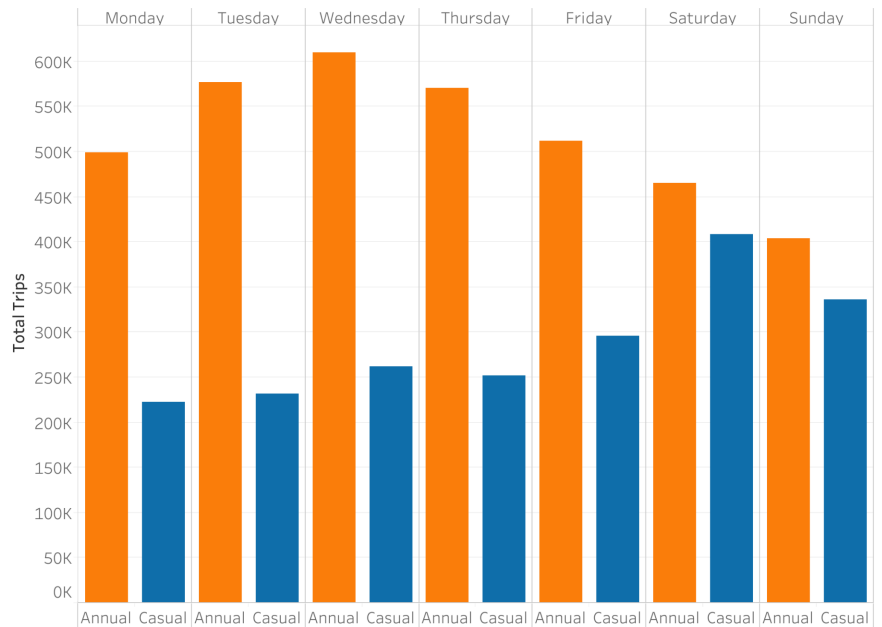
Comparing Average Trip Duration by User Type



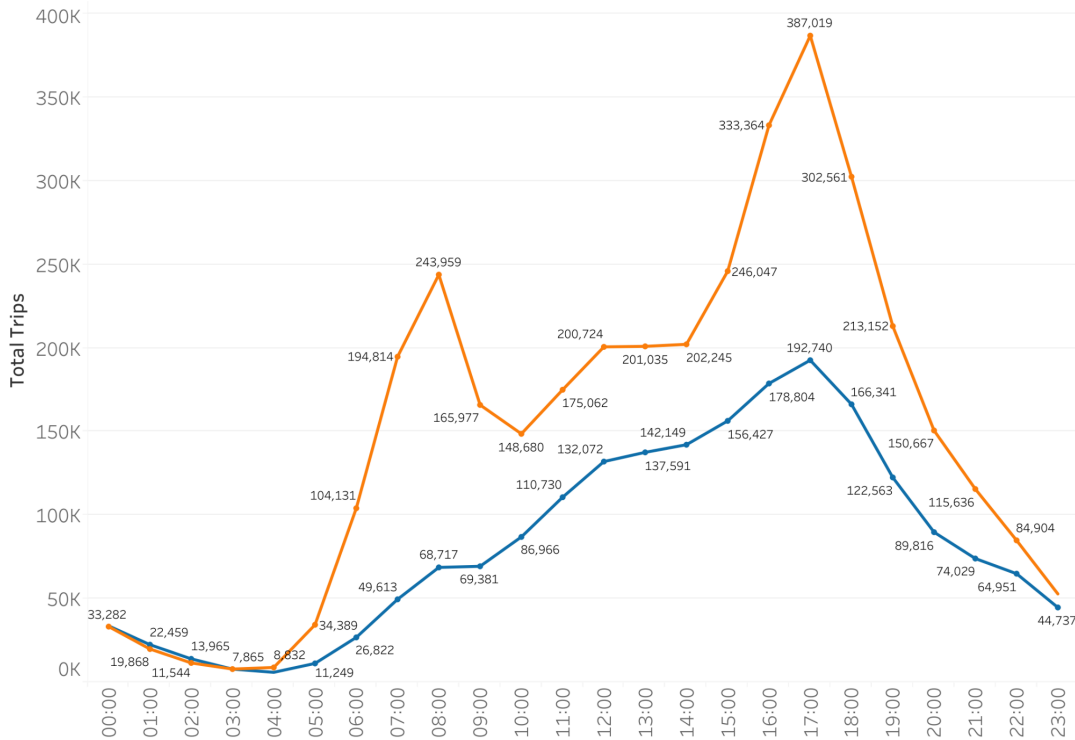
This chart further shows that average trip duration for casual users is substantially higher, especially on the weekends. This is very valuable, and shows a key potential time frame for a marketing campaign.

Next we look at total trips taken per user over the course of the week by user type. It can be seen that annual users far outweigh casual users from Monday-Friday, peaking on Wednesday before declining towards Sunday. Whereas casual users steadily grow reaching peak usage on Saturdays.

Total Trips per Day of Week by User Type

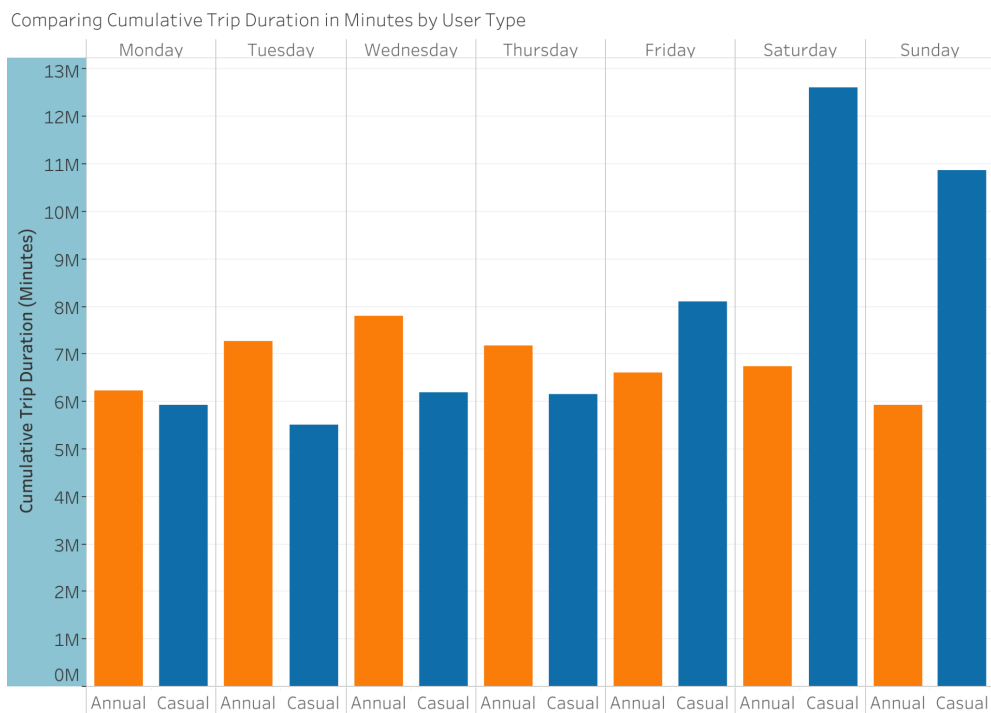


Comparing Total Trips taken by User Type per Hour of Day

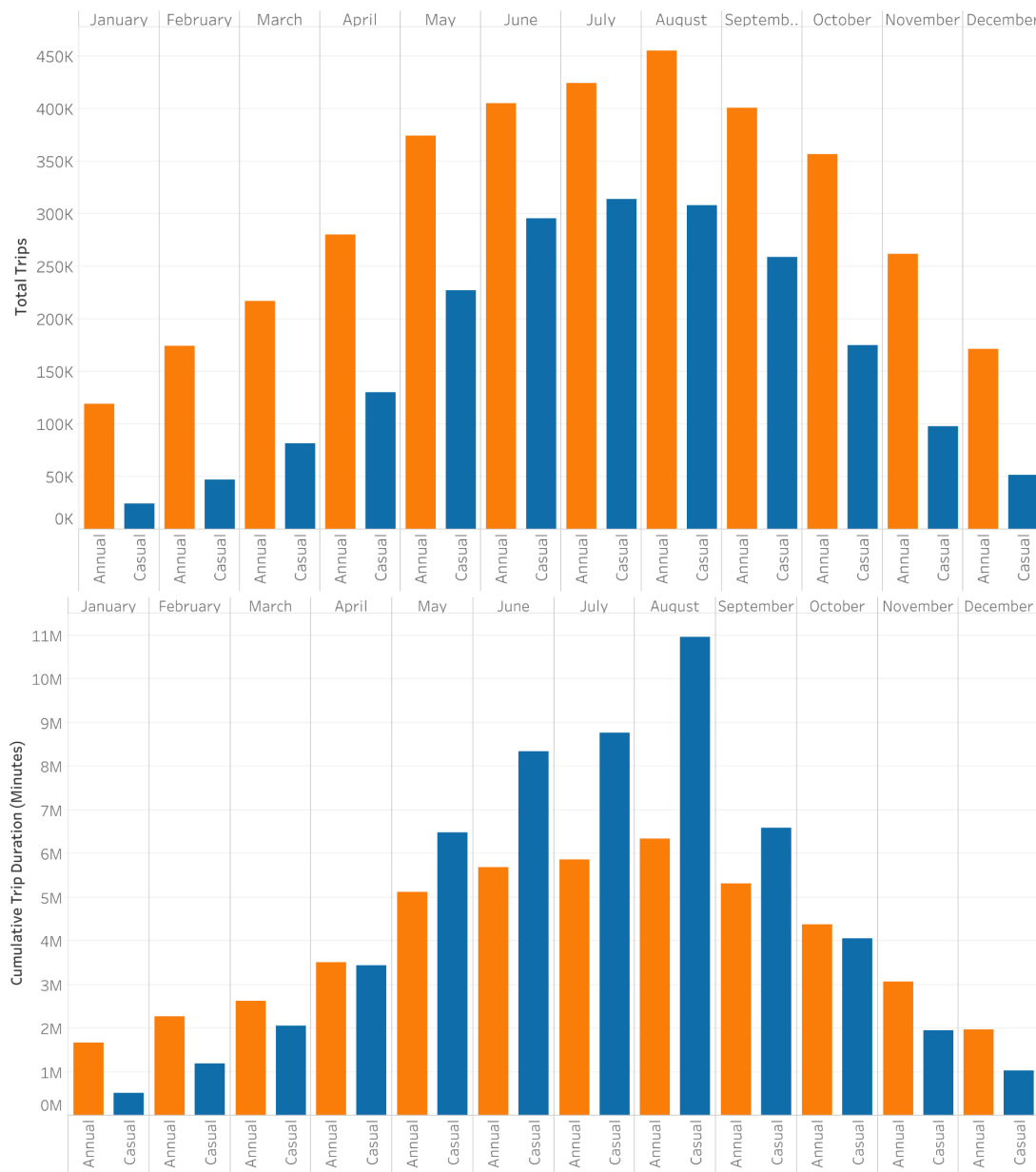


A hypothesis can be drawn from these charts that states annual member users use the platform considerably more for commuting, as usage peaks Monday-Friday and at typical commuting hours.

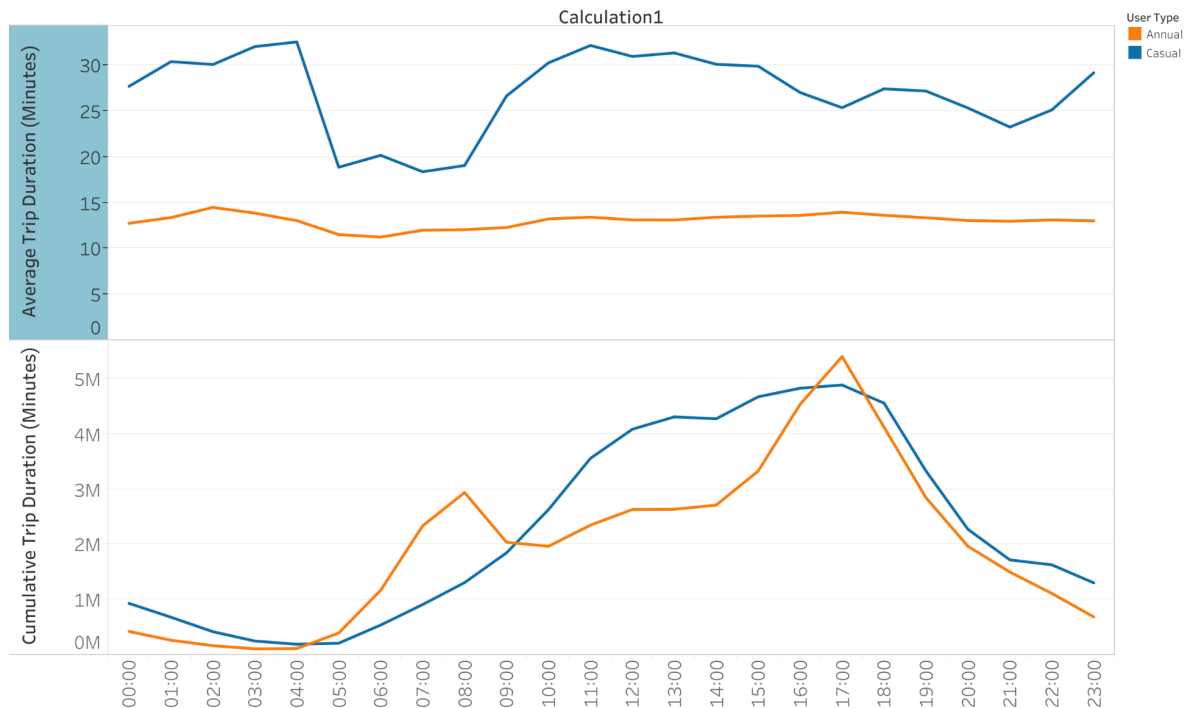
Annual users peak between 8am and 9am in the morning and around 5pm in the evening. Defining peak hours as between 7-9am and 4-6pm, rides between these hours take up 44.73% of total rides by annual users, compared to 36.11% by casual users. This information is important as it provides more context to how these two user types differ. This suggests that an effective marketing campaign could consider targeting advertising at public transport hubs and prime transport routes specifically between these hours, which could help engage new users. However, If we wish to specifically target existing casual users we can see that it would be more valuable to target them in the afternoon hours, and most notably during weekends for maximum exposure, as this is when casual usage peaks. Weekend usage by casual users in terms of cumulative trip duration is substantially higher on the weekends, as shown by the following chart. This could mean that, in app offers/advertising could be most effective on the weekends, especially as these days have the most amount of casual users too.



Understanding how the users interact with Cyclistic over the course of the year is important, it helps provide context as to why annual users are particularly beneficial. The chart below shows very low casual usage during the winter months. Weather conditions could be a distinct factor in affecting usage at this time, as well as reduced daylight hours. Casual users may be less willing to ride in winter conditions or low light. Overall trips peak in summer months for both users. This time period would provide greatest exposure to the largest volume of casual users, especially leading into the quieter months. The hypothesis could further be backed for the following chart which shows that casual users spend less time on the bikes during winter than annual users but substantially more time in cumulative minutes during summer.



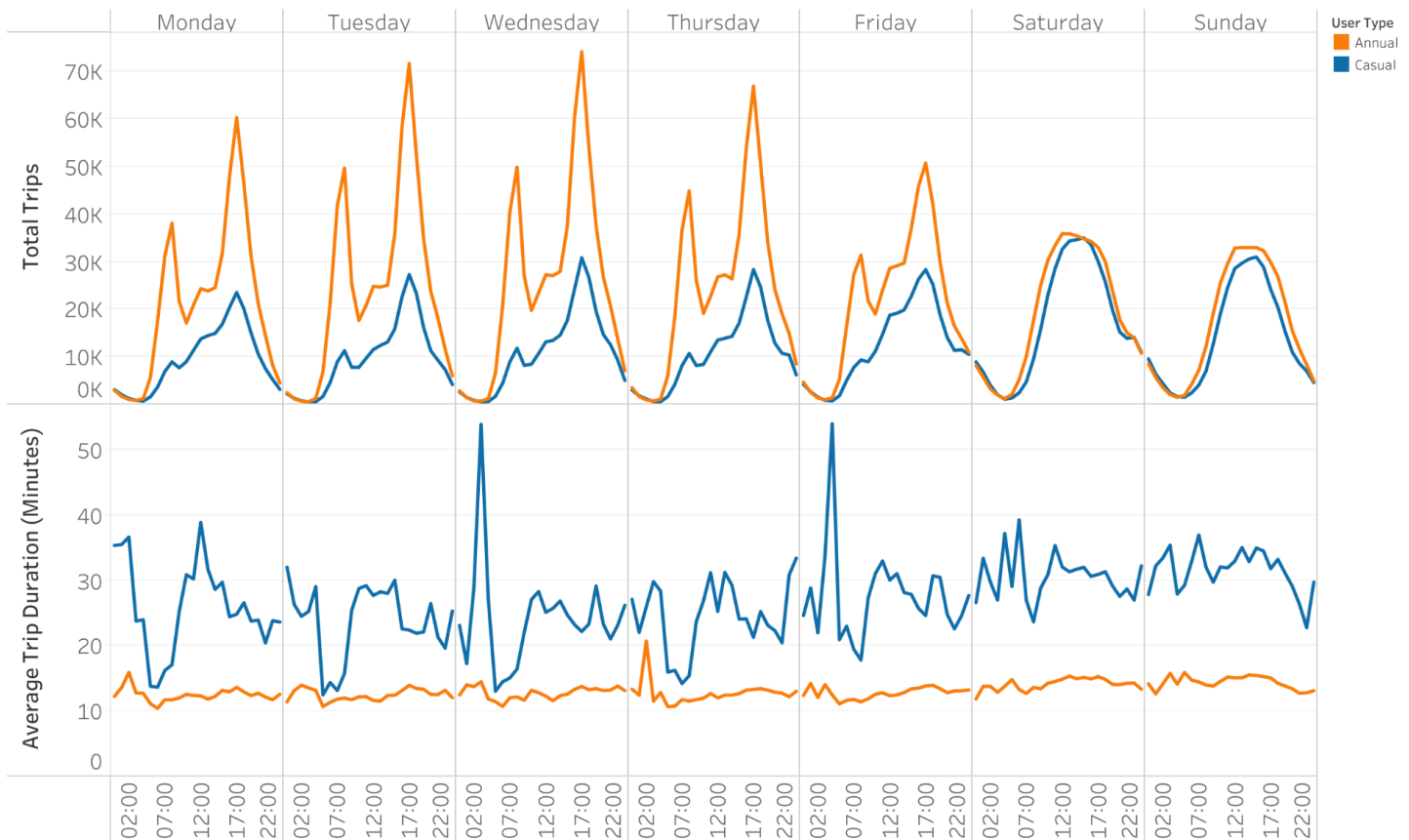
Comparing Average Trip Duration and Cumulative Trip Duration (Minutes) by Hour of Day



The first chart showing average trip duration across the hours of the day displays that annual users have consistent trip lengths, and that the average doesn't vary significantly across the day. This could be for a number of reasons but it would suggest that annual users have regular trips with the same start and end location. By comparison, casual users' average trip duration varies substantially throughout the day, dropping by almost a third during morning 'rush hour', and varying throughout the rest of the day. This could imply that casual users are more spontaneous with using the bikes, and not necessarily for consistent and regular trips. Once again, the limitation of knowing more data about regular casual users, learning unique users would be greatly beneficial to better understand this data.

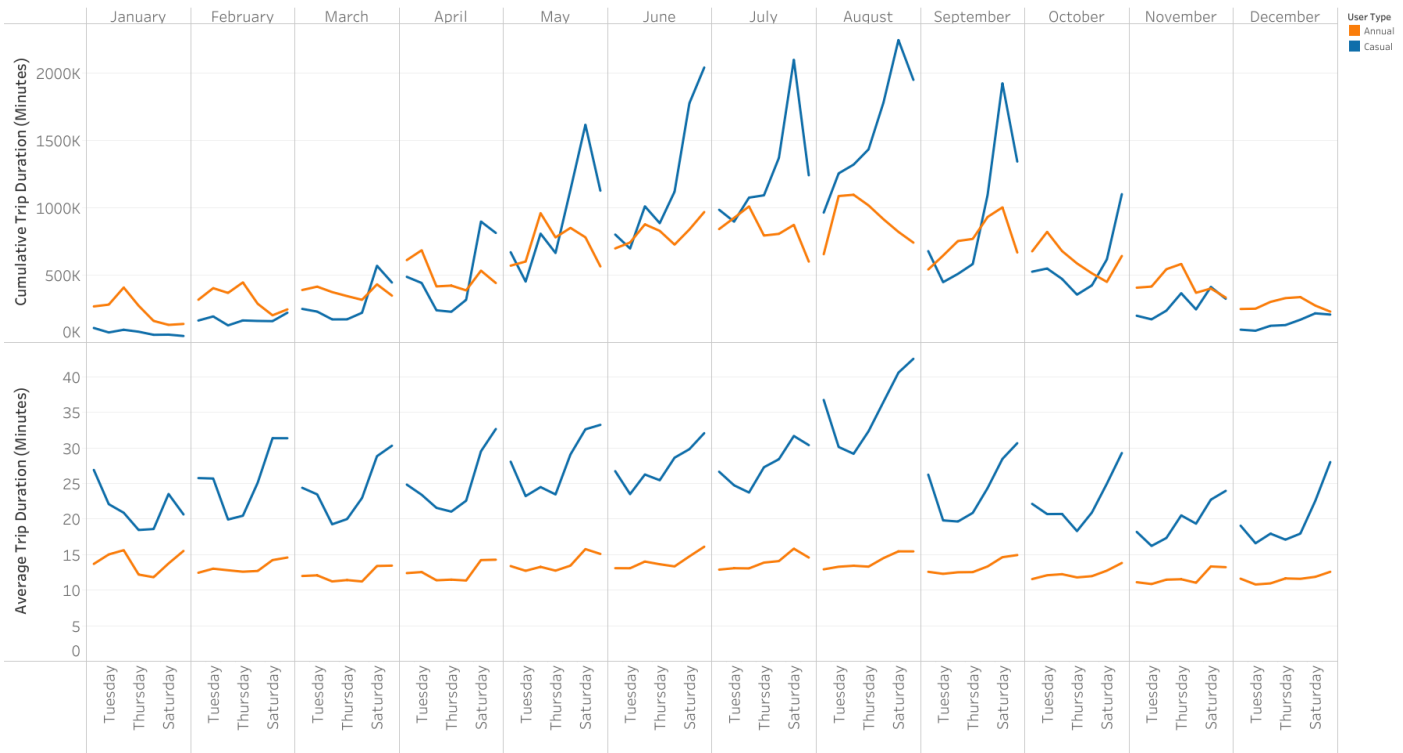
From here we start looking at trends by multiple factors of date/time. Such as how the users compare over hours of the day, depending on the day of the week both in terms of total trips taken and average trip duration. Then how the average ride duration by day of the week changes by months of the year.

Comparing User Types by Hour of Days and Day of the Week



The hourly trend of total trips reveals interesting data, particularly for the weekends. With Monday-Friday having a great contrast in how the user types compare, the weekends by comparison are almost identical. Average trip duration on the other hand shows a distinct spike on Wednesday and Friday early in the morning between 3-4AM. As this is a substantial spike in data it would be recommended that this be verified for reliability.

Comparing User Types over Days/Month of the Year



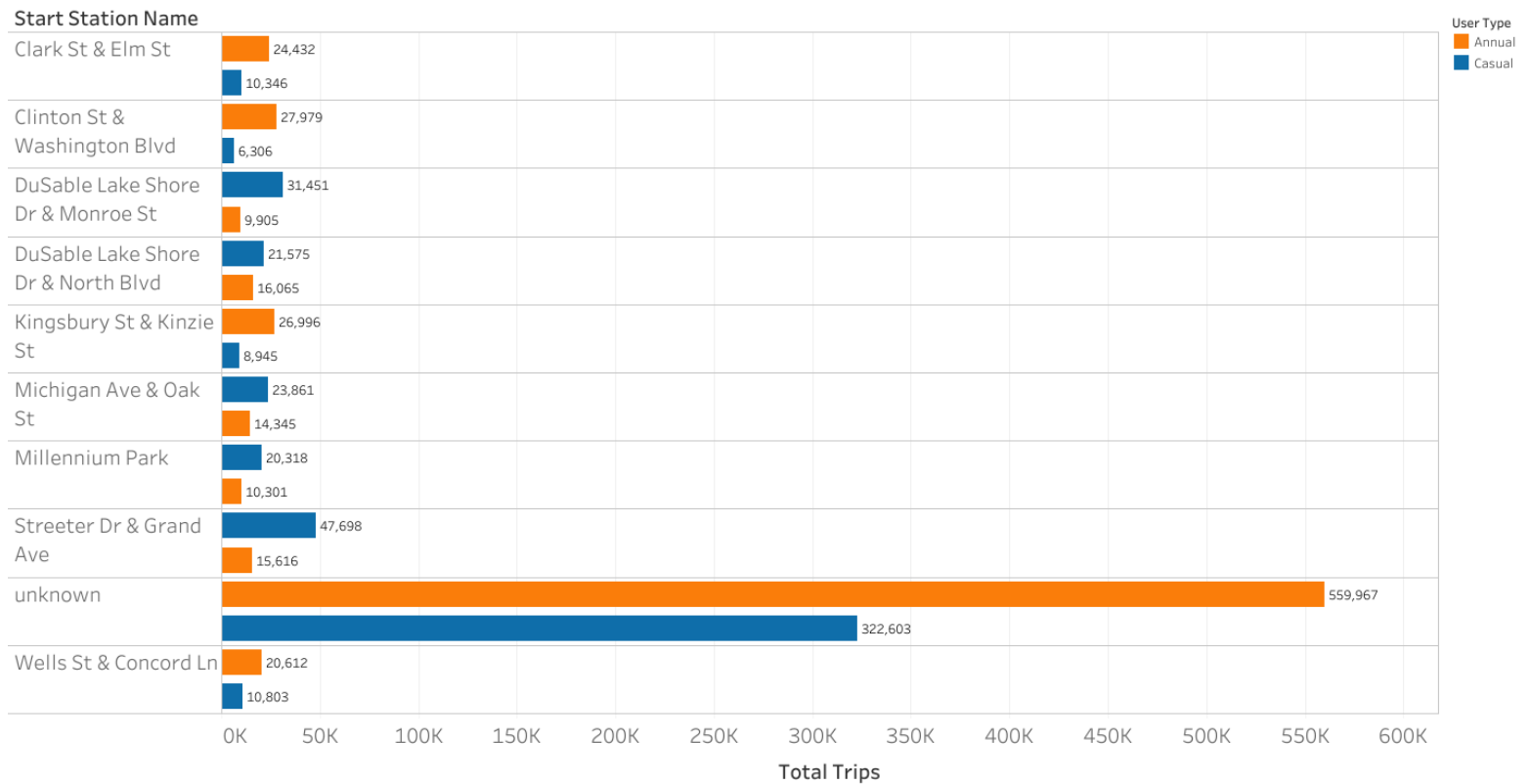
Trend of total rides broken down by the days of the week for each month of the year.

Following early trends that casual users accumulate more time spent on bikes than annual users, it can be seen clearly here, especially during the summer months from May through September. Casual users interact with the bikes for significantly greater lengths of time both in terms of average trip duration and total cumulative trip durations. This all while having 31.83% less rides taken than annual users.

Finally, location data was compared to determine where to target any potential advertising campaign. Top locations for both the start and end of the trips were compared between the user types.

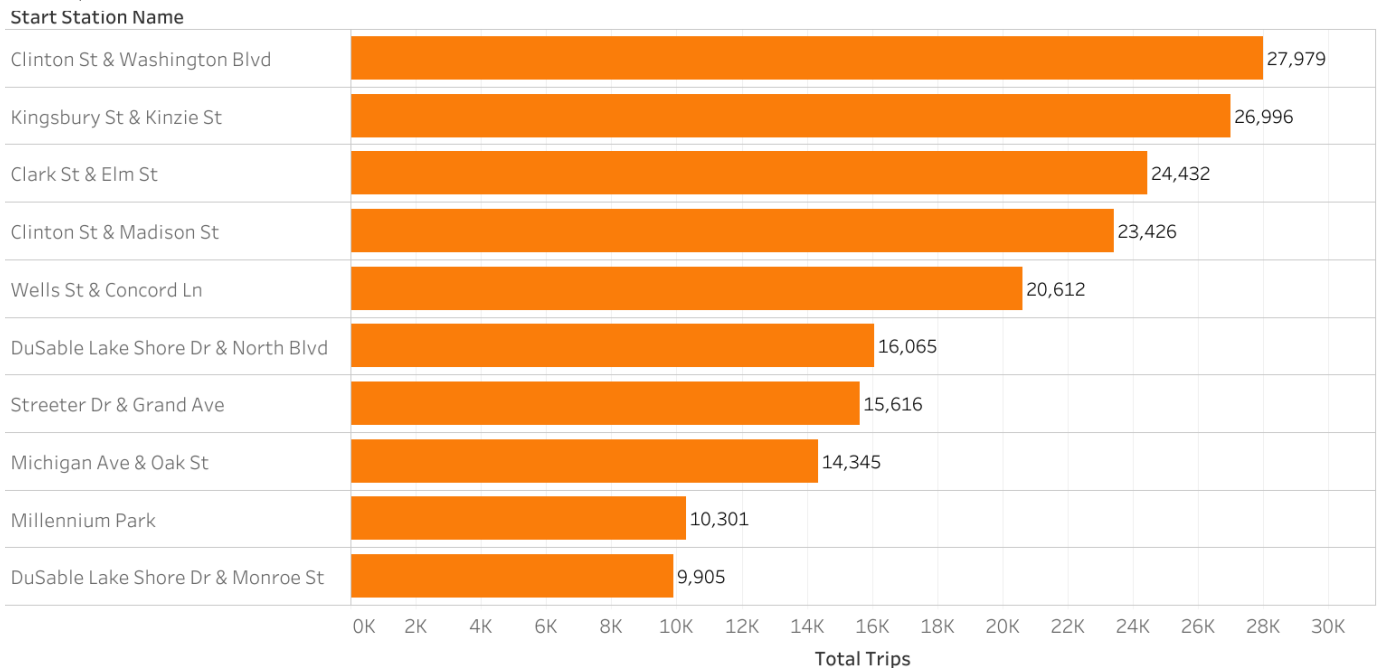
Unfortunately, this chart highlights a weakness in the data, that of missing data for start and end locations here shown as “unknown” for station name. However, once the unknown data was removed, reasonable insights can be determined that are valuable to a marketing campaign. .

Most Popular Start Station - User Comparison

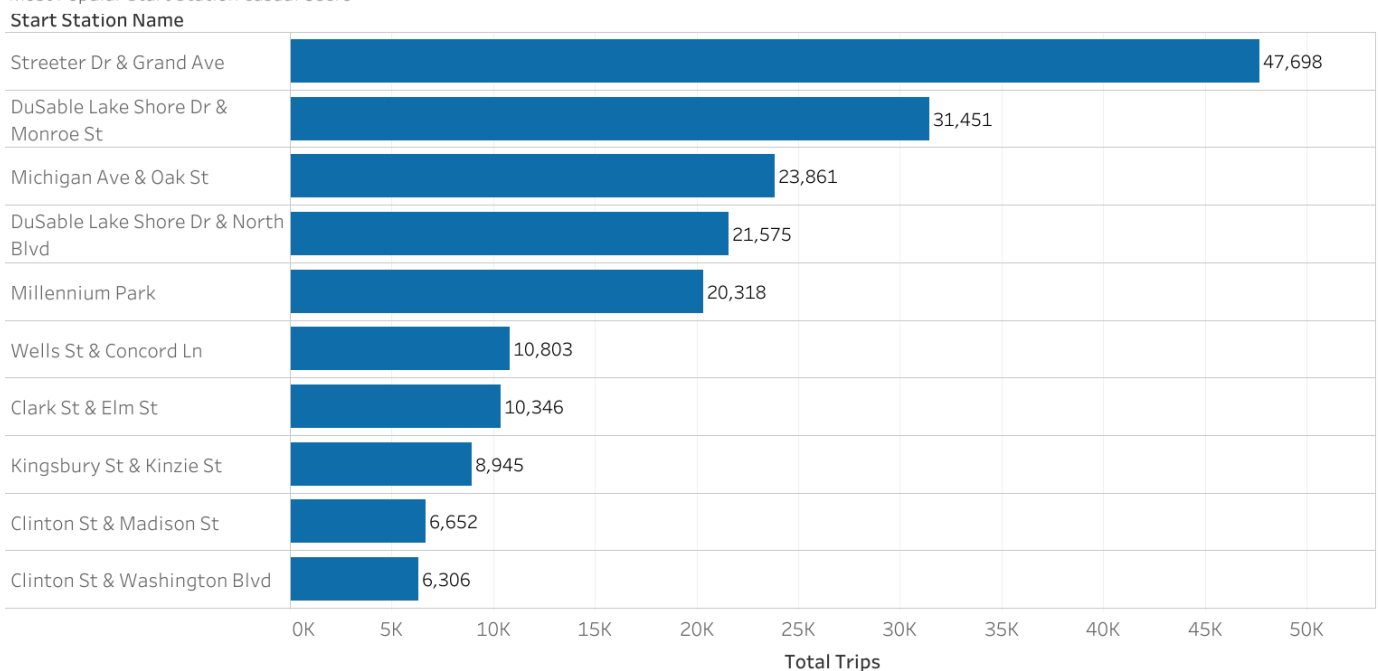


With the “unknown” station category removed and the two charts separated it is possible to identify the most popular known stations for each user type. Identifying that they do not share the most popular starting locations. Focusing on only Casual users and the top starting stations identifies locations that can be targeted for marketing campaigns. Such as billboards, posters, and location based in app advertising.

Most Popular Start Station Annual Users

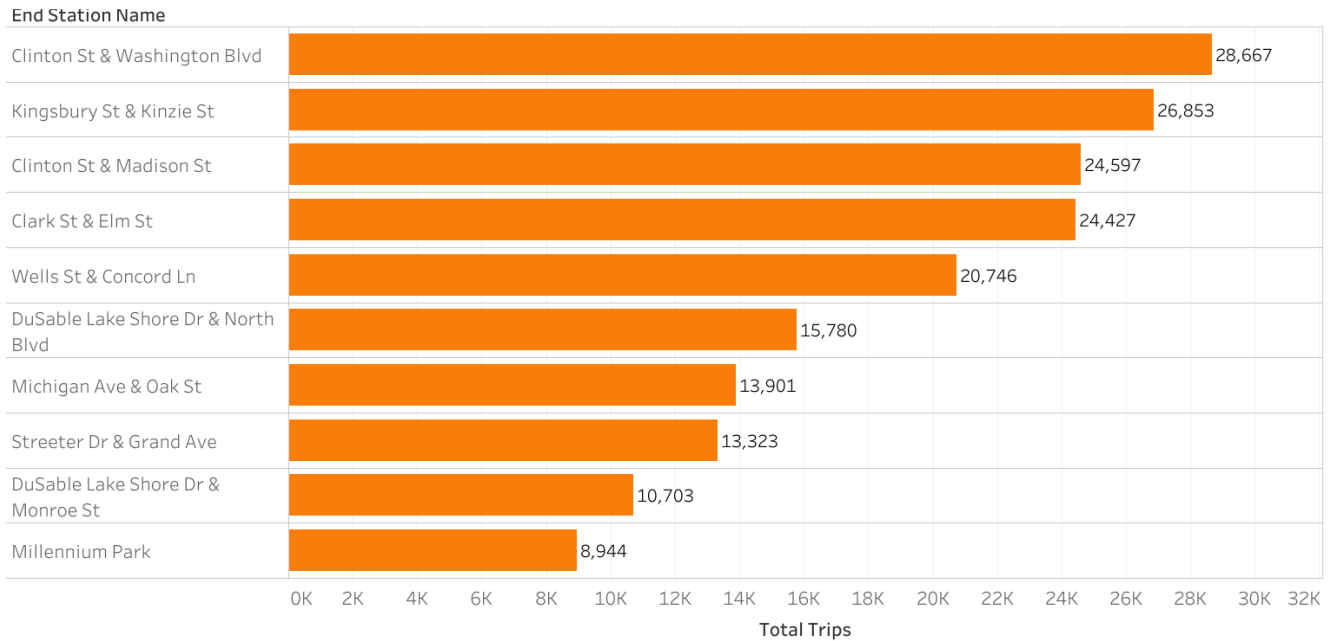


Most Popular Start Station Casual Users

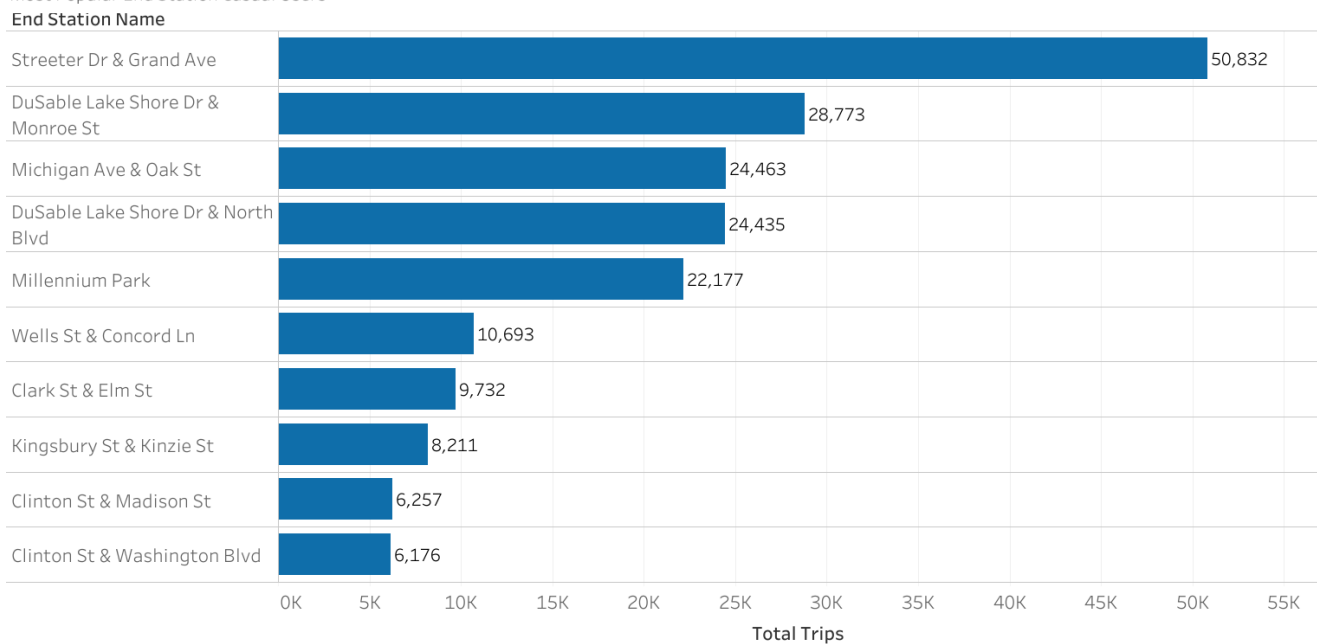


Comparing the end location data provides the same resources for the marketing team, and the top ten end station locations. For the same reasons as the start station comparison, the most popular end station locations are also very different. Therefore allowing us to target specifically just the casual users.

Most Popular End Station for Annual Users



Most Popular End Station Casual Users



Key Findings Summary

During the analysis phase and production of visualisations we were able to determine a variety of key insights that will lead to actionable recommendations. The following was learnt and provided context to the user base:

- Annual users accounted for 64.42% of total rides, whilst casual users made up 35.58%, highlighting a potential 2 million rides for conversion opportunities.
- Casual users spend, on average, twice as much time per trip compared to annual users on the bike. Despite casual users taking 45% fewer trips in total, they spent 16% more time cumulatively on their trips. Highlighting a high engagement rate with casual users on a per trip basis, here an opportunity to capitalise on the longer trips exists.
- Casual users favour electric bikes slightly, whilst annual users tend towards classic bikes.
- Further analysis of bike types and locations is recommended, as increasing electric bike availability may help drive casual to annual conversions.

The following provided actionable insights:

- **Daily trends** varied significantly between the two user types, casual users' peak usage occurred on weekends, particularly Saturdays. Suggesting weekend targeting of marketing campaigns towards casual users would be most effective.
- Annual users' **hourly usage** peaked during the weekdays, most notable on Wednesdays, especially during 'peak commuting hours' of 7-9am and 4-6pm. 44.73% of total trips taken by Annual users were taken during these hours, compared to just 36.11% for casual users. Suggesting that a potential avenue for targeting a marketing campaign during busy public transport hubs during these times would be effective in converting casual users and additionally creating new users.
- Analysing the **cumulative trip duration** further reinforced the notion that casual users spend more time on the bikes per trip and as a whole over the weekends. Providing further emphasis that this is where the marketing campaign should be focused to convert casual users to annual members.
- Analysing **monthly trends** helps narrow down the busiest period of the year to the summer months, from June to August. Once again, casual users' cumulative time spent

on the bikes peaks during these months, and again especially on Saturdays. Supporting the idea and allowing us to focus marketing on weekends during these months.

- Looking at **geographical trends**, the most popular start and end stations revealed, once unknown stations were removed and the two user types split, a contrast in the most popular stations for each user type. For casual users, *Streeter Dr & Grand Ave*, was the most popular station seeing 50%/76.64% higher casual usage than the second most popular; *DuSable LAke Shore Dr & Monroe St*, for starting and ending location respectively. The most popular stations for each user type contrasted in such that the most popular location for annual users was just the 10th most popular location for casual users. This informs the marketing campaign as to where resources should be focused the most. While some location data was missing, the available data still revealed clear trends and valuable insights for actionable results. Further research into weather patterns could provide additional context for season change that may affect usage, particularly among casual users.

Actionable Recommendations

As mentioned there is a significant opportunity to convert casual users to annual members with over 2 million trips taken annually by casual users. The analysis revealed a potential opening for appealing to new users as we discovered that Annual members had peak usage at typical commuting hours. A marketing campaign in hotspot public transport hubs utilising physical advertisements at this time could help bring in new users.

When it comes to increasing Annual members there are three key areas to target.

- **Weekend Campaigns:** Focus on casual users on Saturdays. Use in-app promotions or weekend-exclusive offers to encourage conversions to annual memberships.
- **Summer Promotions:** Given the high engagement during the summer months, offer seasonal deals to lock in annual memberships with benefits throughout the year. This could include a discount on annual memberships during the busiest months (June-August).
- **Commuter Campaigns:** Target busy public transport hubs during weekday commuting hours (7-9am/4-6pm) with physical ads, in-app offers, and location triggered social media campaigns aimed at converting casual users and attracting new customers.

Measuring Recommendations

Moving forward, should these recommendations be actioned they can all be tracked.

- For internet based advertisement the users that follow these offers can be tracked using cookies and noted in an ongoing data analysis of how effective this campaign was.
- Tracking user behaviour for in-app promotions will further strengthen the evaluation of the marketing campaign.
- For physical advertisements, it is possible to use discount codes and measuring how many of these were used during signup could offer a way of tracking this method.

To measure these as a whole, the conversion rates from casual to annual members will be tracked on a monthly basis. Including specific breakdowns of the above, allowing us to track the most effective campaigns for future reference.

Conclusion

The analysis of Cyclistic's user base provided clear insights into how marketing efforts can increase annual memberships. By focusing on key times, locations and user preferences, Cyclistic can effectively target casual users and convert them into annual members. Moving forward, targeted campaigns in transport hubs during commuting hours to pick up new users and weekend offers/promotions during summer months for casual users for the best opportunities for conversion. The data demonstrates that focusing on summer months and leveraging commuter habits will maximise user conversion and increased annual memberships. With ongoing tracking and adaptation of these strategies, Cyclistic is positioned to grow its annual member base and long term profitability.