

Bike Share Project - Analysis

Table of Contents

1. [Introduction](#)
2. [Project Goal](#)
3. [Description and Limitations of the Data](#)
4. [Analysis](#)
 - [Overall Differences](#)
 - [Overall Insights](#)
 - [Yearly Trends](#)
 - [Yearly Insights](#)
 - [Weekly Trends](#)
 - [Weekly Insights](#)
 - [Daily Trends](#)
 - [Daily Insights](#)
 - [Geographical Trends](#)
 - [Geographical Insights](#)
- [Key Insights](#)
- [Recommendations](#)

Introduction

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno - director of marketing - believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

Project Goal

The goal of this project is to **determine the differences between how casual users and members use the bikes** provided by this bike-sharing company, in order to come up with informed marketing strategies that **incentivize casual users to become full fledged members** of this bike sharing community.

To make the best decisions possible the company has gathered data from all rides made on the last 12 months, and made it available under this [license agreement](#). After collected, the data has already been cleaned, validated and verified. It has also been stripped beforehand from all personally identifiable information, and is in such a form that can be used effectively to make data-driven decisions regarding the purpose of this project.

Description and Limitations of the data

The data contains information for almost 6 million bike rides made between March 01 2022 and February 28 2023, which can be found [here](#). It has been cleaned and transformed and contains the following information:

- **trip_id**: a unique identifier for each ride made with one of the bikes.
- **bike_type**: the type of bike that was used. The types of bikes are electric, classic and docked.
- **start_time, end_time**: the time and date the ride started/ended. It is of datetime type.
- **start_name, end_name**: the name of the station where the trip started/ended.
- **start_id, end_id**: the code that identifies the station where the trip started/ended.
- **start_lat, start_lng, end_lat, end_lng**: the coordinates for the station where the trip started/ended. It has been rounded to 5 decimal places.
- **usertype**: the type of user, casual or member.
- **year, month, time_of_day, date**: extracted from the start_time field.
- **duration**: time elapsed between start_time and end_time, it is of timedelta type.
- **minutes**: the duration in minutes, rounded to 1 decimal place.
- **day_of_week**: the day of the week when the trip started.
- **distance**: the approximate distance that was traveled from the start to the end of the ride. Calculated using a flat surface approximation.
- **speed**: the average speed for the whole ride, disregarding completely the route that was taken (speed=distance/minutes).

There are some quirks and limitations to the data, these following are the main ones:

- The data doesn't include any information that could allow us to determine patterns concerning individuals. For example, there could be a small group of users that use one particular station a lot, parking and taking out bikes from it several times a day. There is no way for us to know from the available data if the popularity for that station is because of this small group or if it truly is representative of the population.
- Information about the state or country in which a user resides could also allow us to create better social media ads, targeted for specific communities, and also reducing the investment in it by avoiding regions that have next to none tourist flow towards Chicago. This information is not included in the data, though, so any kind of targeting that is more specific than just the user type is impossible without gathering additional information.
- The data shows approximately twice as many station names and id's as there should be, even after cleaning the data. Whether this is because each station has multiple names/ids, that they changed at some point or something else would need to be clarified with the company.

- Stations' coordinates vary within a small range of values; that range is not something to worry too much about, but it is definitely not fixed to a single value, as one would expect for a non-moving object. This introduces a lot more coordinates than there should be.

During the cleaning process there have been steps taken to decrease the effect of the last two issues listed above, although further work with the company to clarify things even further is still needed. The data has also been cleaned for outliers, concerning mainly to really long duration trips, spanning even for several days, and also for trips starting or ending really far away from the area covered by the stations. These steps were taken in order to avoid skewing the analysis results by non-representative individuals. This can only take us so far though, and there could be patterns and trends that might still be skewed because of individuals not being able to get identified by the data, as was mentioned above. The dataset is so large, though, having **5,601,735 recorded bike rides** after cleaning, that it would be almost impossible for a single individual or group of individuals to skew the results in a significant way.

Analysis

The data is already formatted in a useful way and has many useful features for analysis. With it we will go over several stages of analysis, going from overall differences to yearly, weekly, daily and geographical trends and traits that characterize casual users.

Overall differences

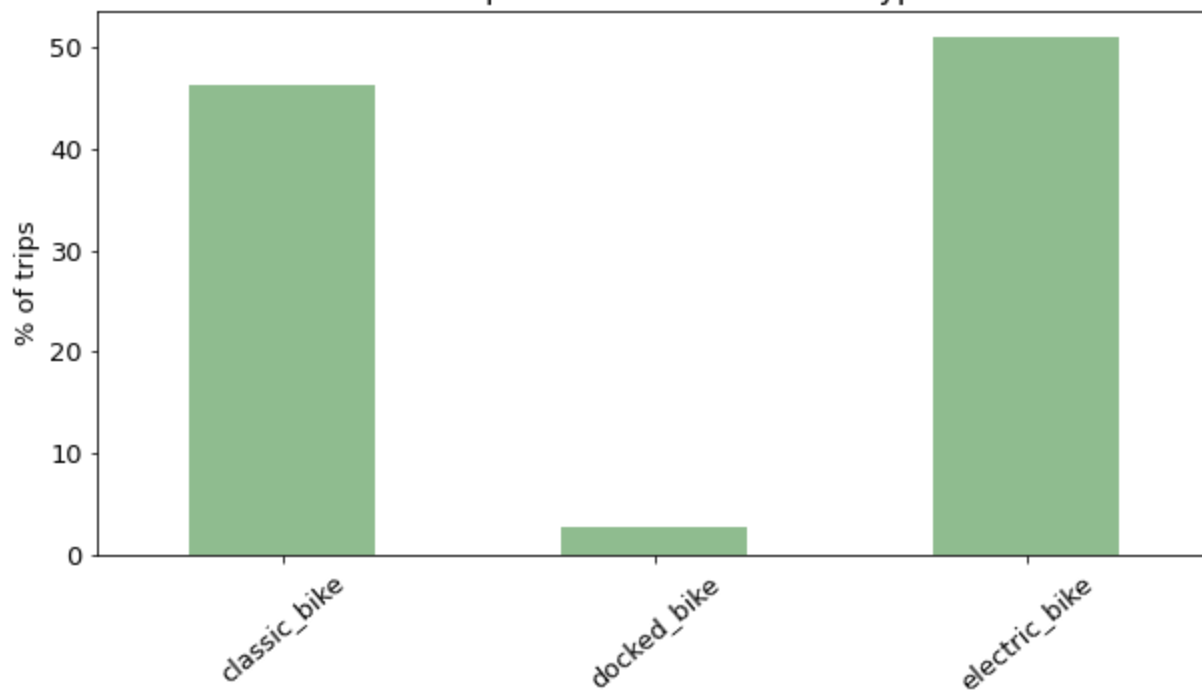
Here we look at the differences between members and casual users over a full year of bike usage, that is, with a yearly granularity.

To begin with, the next table shows that, although there are **more trips made by members** with an annual subscription, **casual users spend more time using the bikes**, overall. This indicates casual users have longer rides, at least as it regards to their duration.

	% of all trips	Total riding time (in years)
usertype		
casual	40.065526	77.223375
member	59.934474	75.420496

From the next chart we see that docked bikes are rarely used, while electric and classic bikes take almost 50% of the share each, with electric ones having a slight advantage.

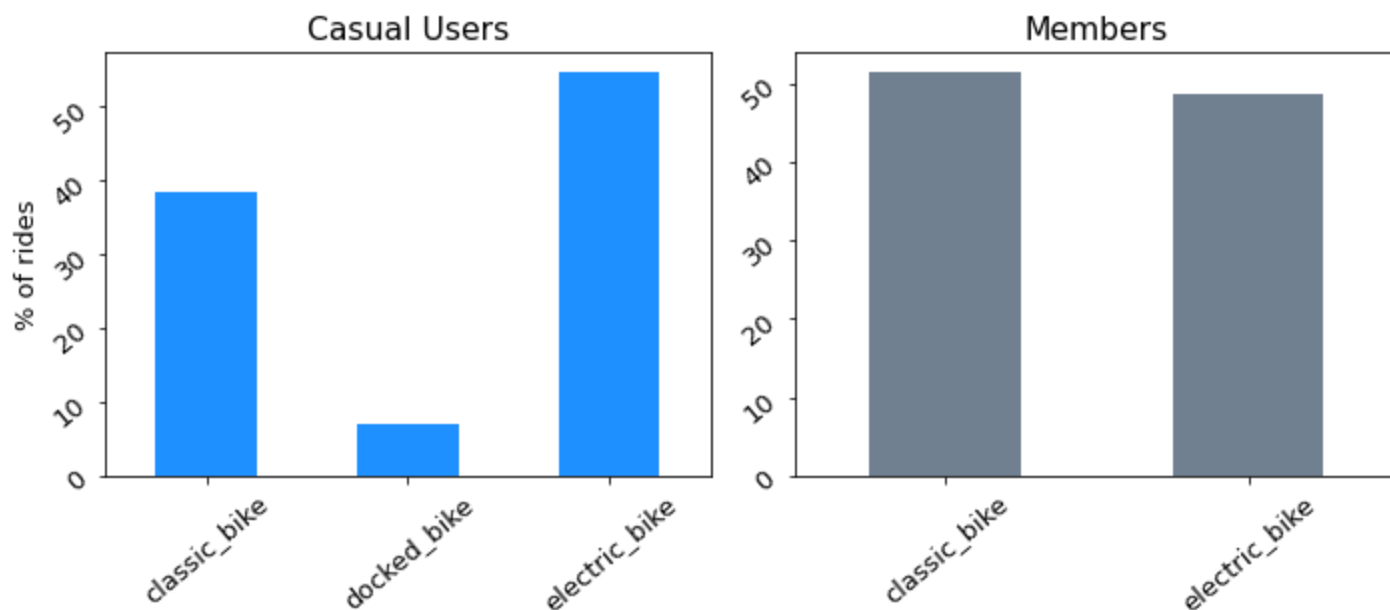
% of trips made with each bike type



when we split the trips into those made by members and by users, though, we find that **casual users generally prefer electric bikes**, while members are almost evenly split between classic and electric.

One thing worth noting is that **members didn't use a single docked bike** in the whole year.

Distribution of Rides by Bike Type

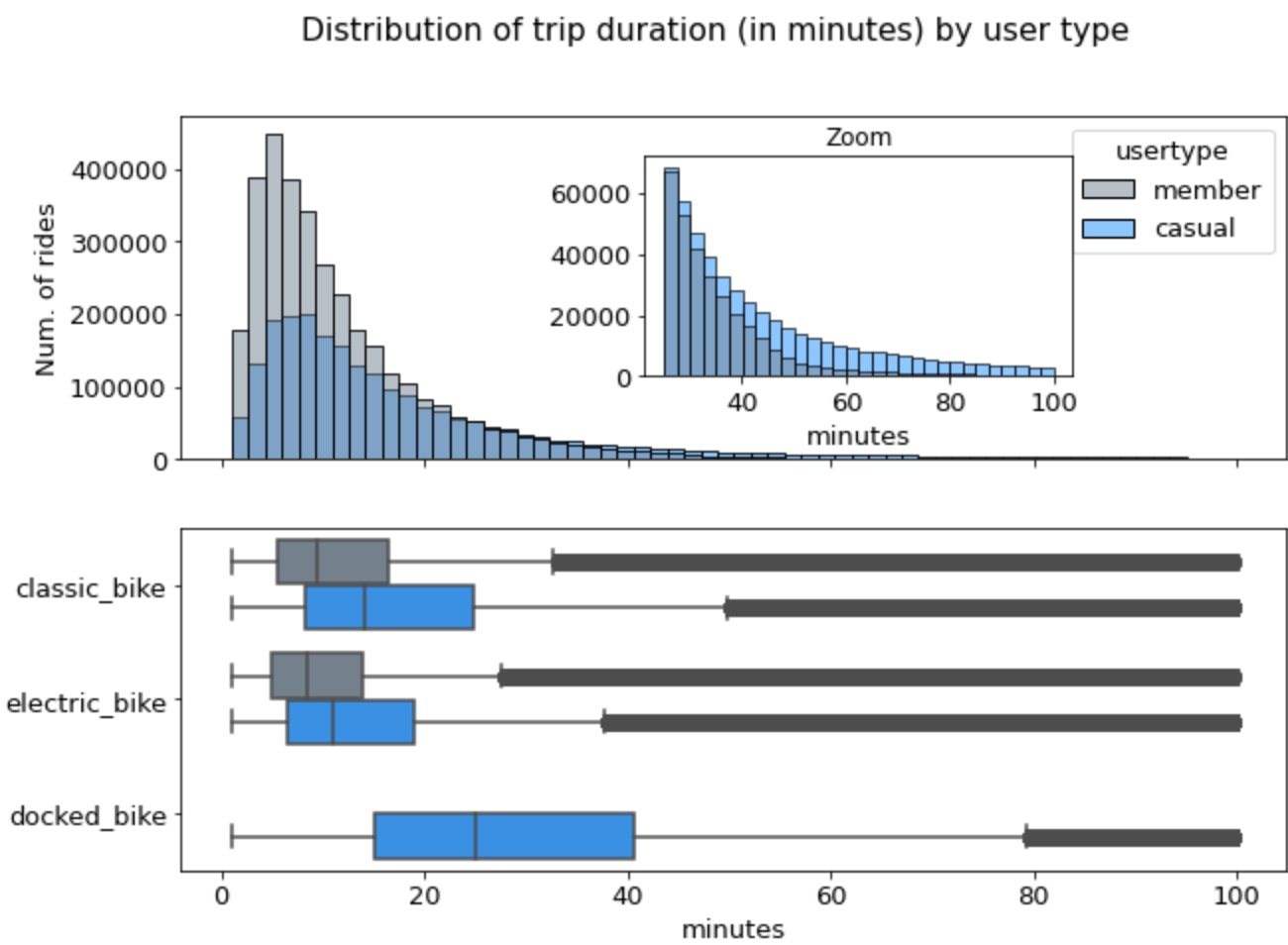


The next table shows some descriptive statistics for the trip duration.

Casual users go on rides with a longer duration than members. They have a higher standard deviation also, indicating there was a high number of casual users that had considerably longer lasting trips than average.

	mean	median	std
usertype			
casual	18.171739	12.8	16.500592

The following boxplot and histogram show that, indeed, **casual users tend to go on trips with a longer duration**. Most members keep their trips short, under 30 minutes. Beyond that duration the number of rides by casual users is greater than that of members, which is the reason why the mean gets dragged so much to the right from the median. Both casual's and member's distributions are skewed to the right, but casual's do so to a greater extent.



There is also a measurable difference between durations for different types of bikes. **Electric bikes** seem to be **used for efficient, short trips**, while **classic and - especially - docked bikes are used for longer, probably more recreational rides**.

For all types of bikes casual users have more lasting rides than members.

The next table shows the mean duration for trips made by casual users and members for each type of bike, and it reinforces some of the insights gathered from the last plot.

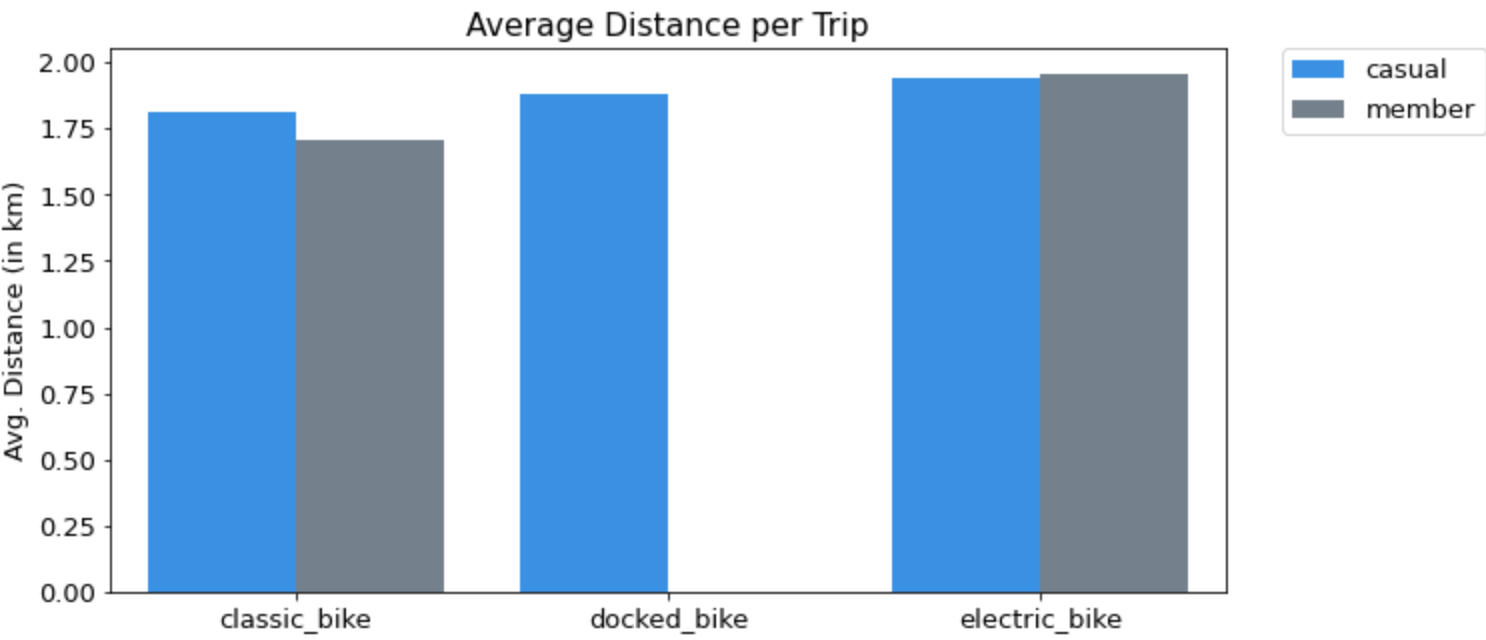
Although rides with classic bikes are longer in general, the difference is a lot more drastic for casual users.

This suggests that **members use the bikes mainly as a convenient method for transportation**, having shorter ride times with electric bikes simply because they are faster, while **casual users take out classic bikes to enjoy the ride**, taking their time to go from one place to another. They do so to such extent that even using electric bikes they still have longer lasting trips than members using classic bikes.

usertype	casual	member
bike_type		
classic_bike	19.849787	12.669245

docked_bike	30.945141	NaN
electric_bike	15.374943	11.011041

Another factor that plays a role here is the distance traveled. Perhaps casual users not only ride more slowly, but they make longer distance trips as well?. The chart below shows this is not at all the case. Year round average trip distances are almost exactly the same for all bike and user types: around 1.9 km.



A more detailed view of the distribution of trip's distance reveals that

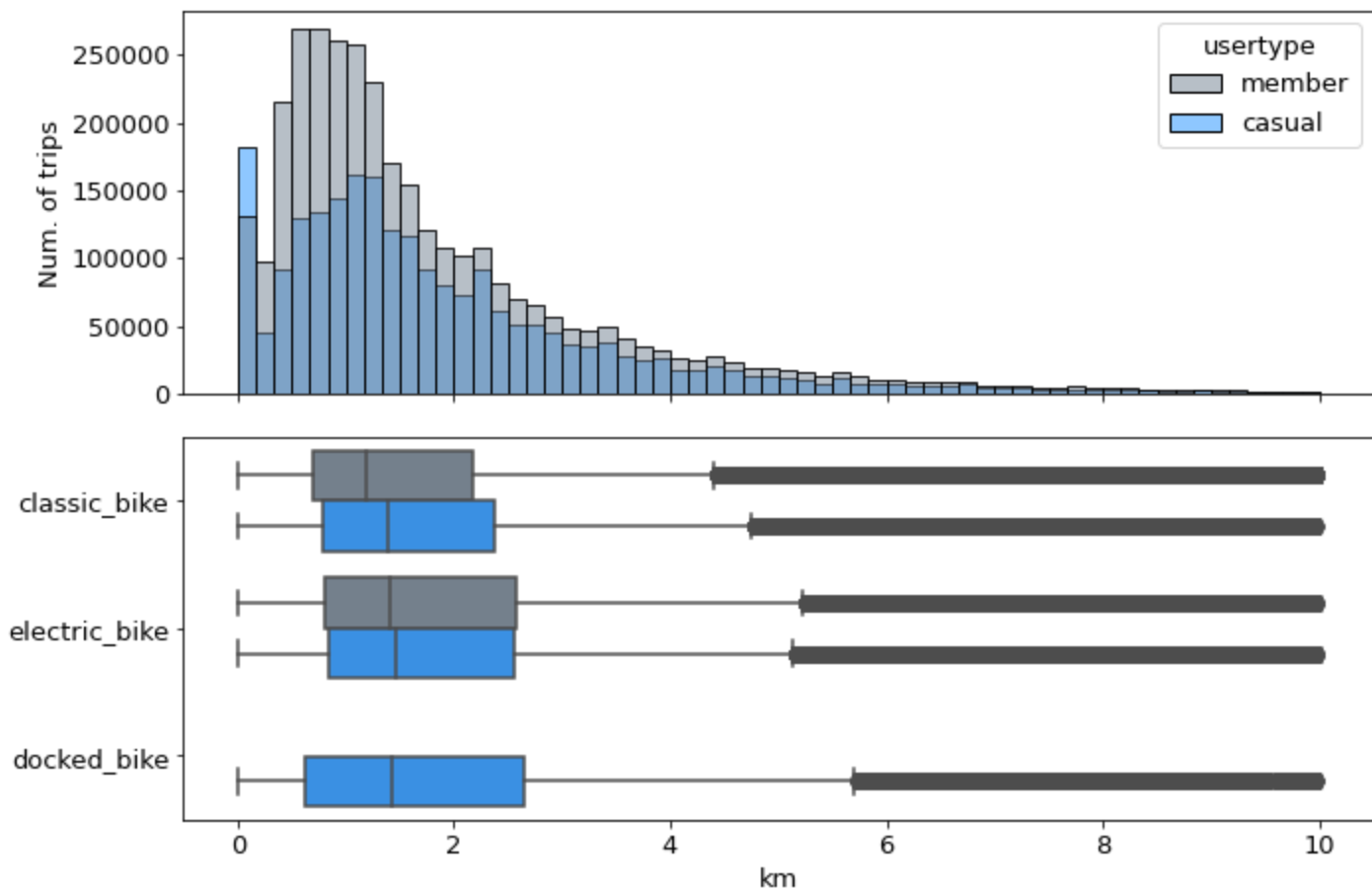
- Most **member's trips are very short**, between 0.5 and 2 km aprox.
- **Casual users** have a less pronounced peak for short trips; the **distances** traveled by them **are more spread out**.
- **Casual users most common trips** have close to **0 km** total distance traveled. This could either be false starts or round trips.

Looking at the boxplots for each bike type we also notice that

- **Electric bikes** are more popular for slightly **longer routes**.
- **Classic bikes** are **used mostly for short rides**, with the distribution for casuals a little bit more skewed toward longer trips than that of members.
- **Docked bikes** are used pretty much for any distance, although as we've seen before they are rarely used.

	mean	median	std
usertype			
casual	1.886470	1.432115	1.633178
member	1.829475	1.291099	1.626537

Distribution of trips by distance travelled



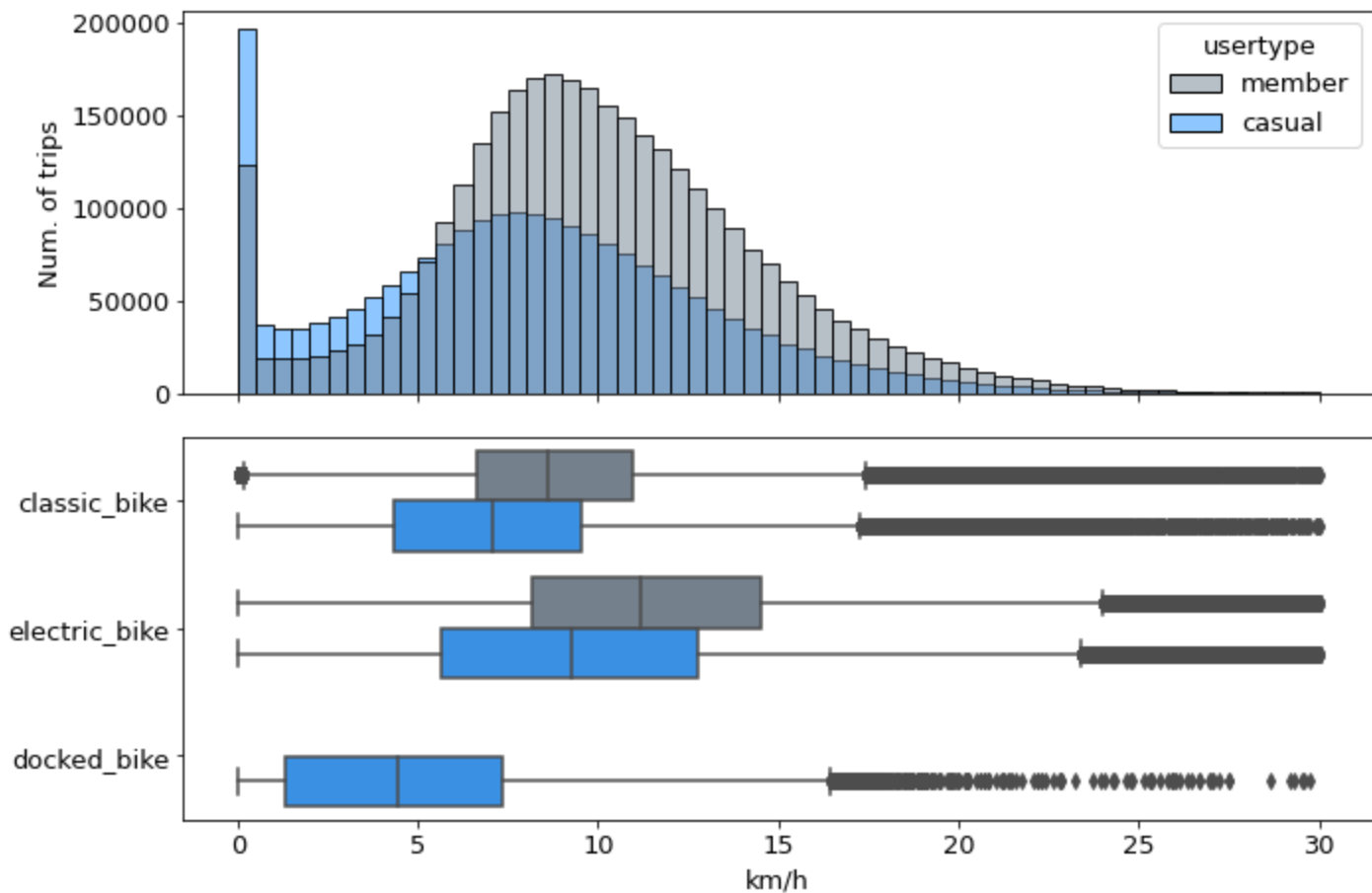
With these results about the duration and length of the trips it is fair to guess that the speed for casual users will be less than that of members, and electric bikes will be fastest, with docked ones being the slowest. The next table and plots show just that.

Both members and casuals reach **speeds up to ~ 18 km/h on classic bikes** and up to **~ 24 km/h with electric bikes**, and that is without counting the most extreme users. This goes to show that electric bikes are in fact exploited for their speed.

With this said, it is worth noting that the 25-75% portion of the boxplot for members using classic bikes is fully contained within the 25-75% range for casuals using electric bikes, which shows that casual users, even when using electric bikes, can go really slow.

	mean	median	std
usertype			
casual	8.045784	7.92	4.944675
member	9.941216	9.67	4.588424

Distribution of trips by speed



Insights for overall differences

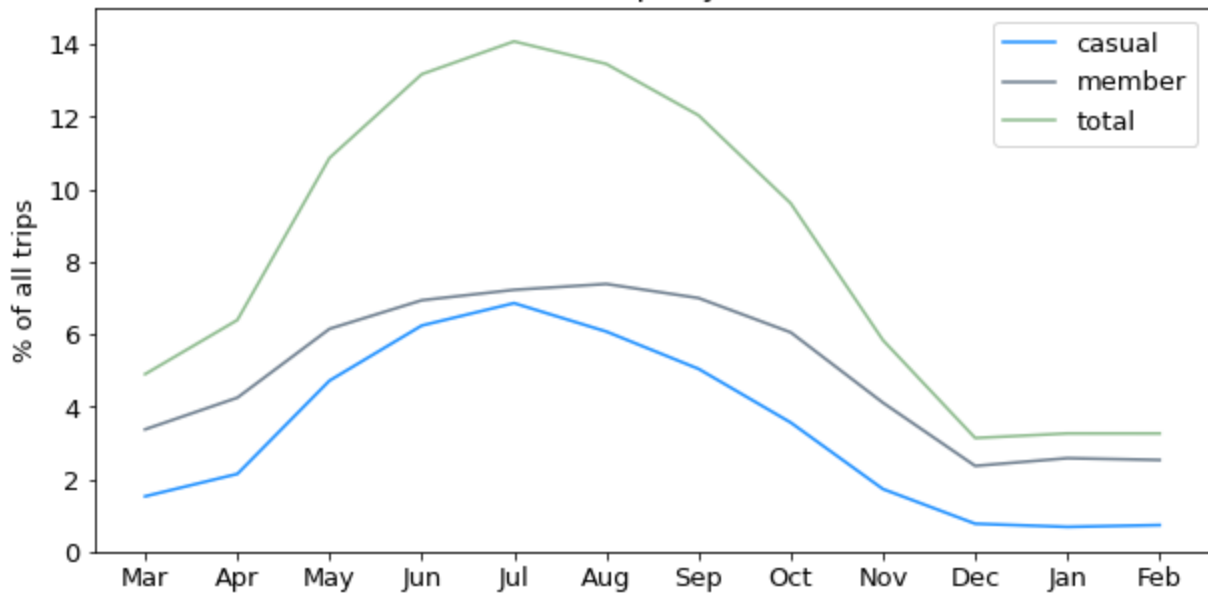
- **Members are faster** than casual users by almost 2 km/h on average.
- **Casual users use each bike for more time**, on average, than members. Around 6 minutes more per trip, to be precise.
- **Casual users prefer electric bikes**, members don't have a clear preference.
- **Electric bikes** are used for **faster rides**. Docked bikes are used for slow rides.
- **Electric bikes** are used for **longer distance rides**
- **Docked bikes are rarely used**, and exclusively by casual users.
- **Casual users** most common trip is **false starts and/or round trips**.

Yearly trends

The next plot shows the percentage of total for trips happening each month. The green line is the sum of the other two, and shows the total of trips for each month.

Both user types display similar trends, although the curve is more pronounced for casual users: colder months (November to march) see the least amount of rides. After April comes the number of trips steadily increases, reaching a maximum on July for casual users and august for members, and decreasing after that until reaching a valley around December and lasting for at least 3 months (up to the present date on which this analysis was made).

% of all trips by month

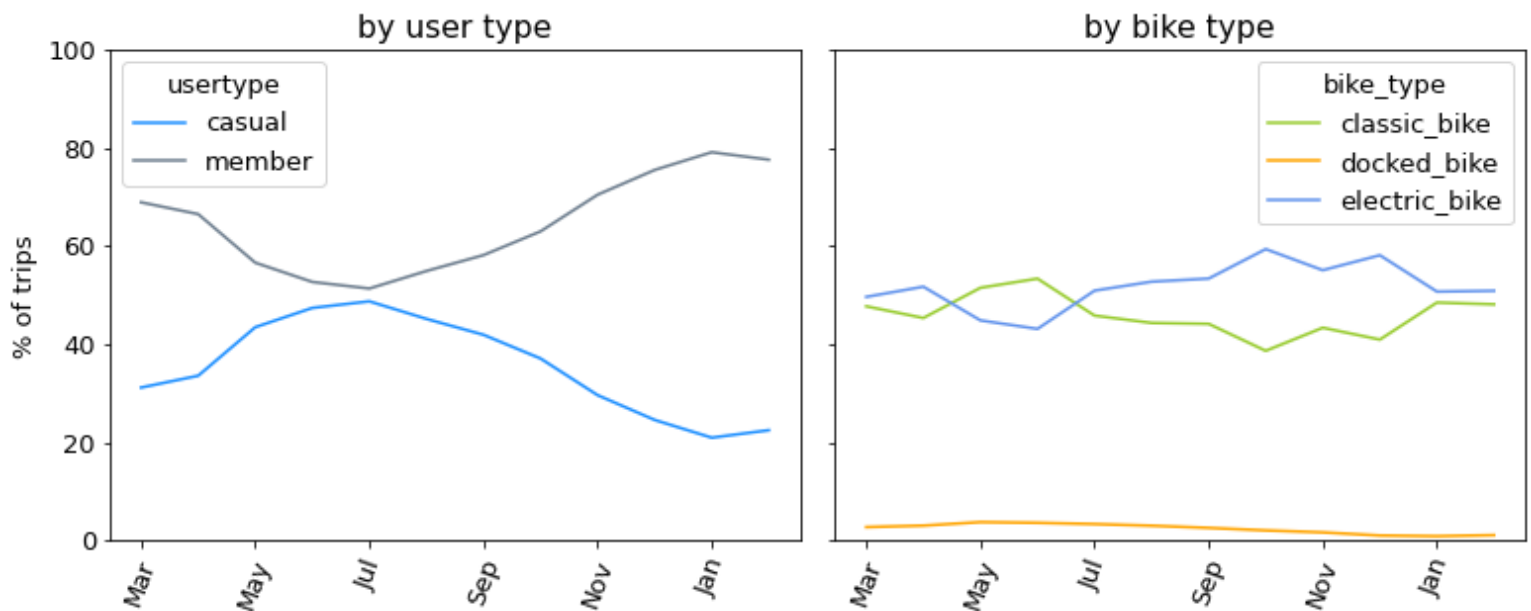


The next plot shows the percentage of trips made within each month by casual users vs members. It is clearly shown that, although the yearly trends are similar for both types of users in the sense that the monthly number of trips increases in warm months and decreases in cold ones, the share of those trips that are made by members vs casual users display opposite tendencies: casuals have only 30% of the share of monthly trips in March but they rise up to almost 50% in July, while members hold 70% of March trips and decrease to about 50% in July. The difference is even higher in January, with casuals holding only ~20% and members ~80%.

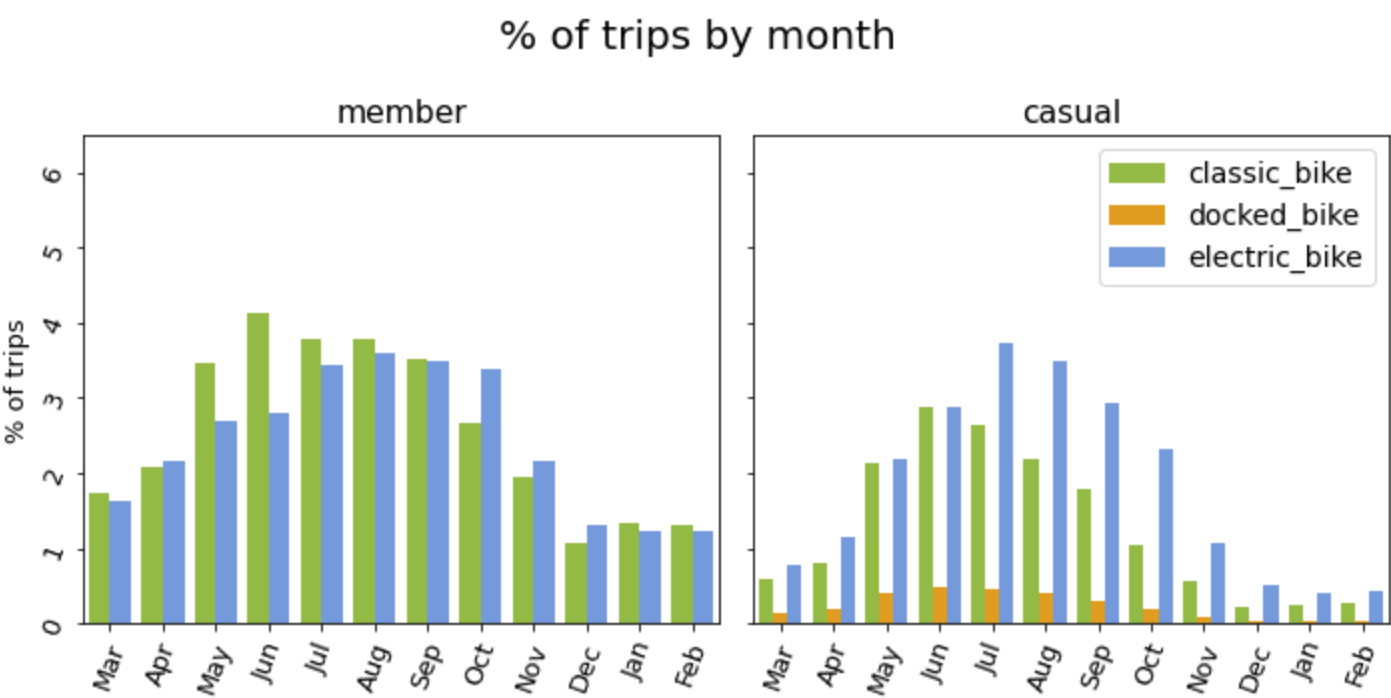
As for the bike types, they seem to change their popularity along the year as well. Docked bikes are the least popular and stay pretty much the same all year. Classic and electric bikes, on the other hand, have exactly opposite trends.

Electric bikes are most popular almost all year long, especially from July to January. Classic bikes only take over for a brief period during springtime, in May and June. Perhaps this has to do with the weather being nicer for a ride in these months while, when it's cold outside, people want to get to their destination ASAP, which is more achievable using electric bikes. This doesn't explain why classic bikes start making a comeback in January, although the drop in tourism after the holidays might have something to do with this.

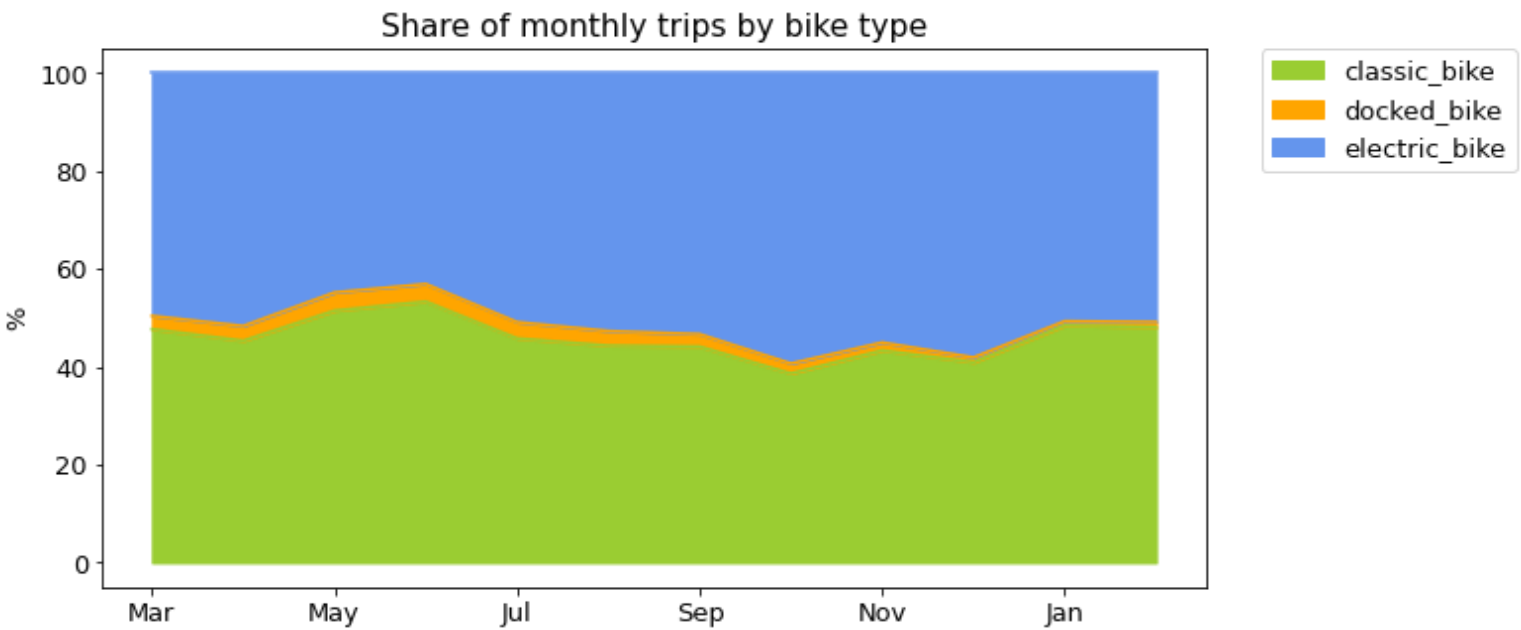
% of monthly trips



Below we can see the % of all year's trips each type of user made each month, for every type of bike. The overall trend for both types of users is the same but they have an important difference. For members classic bikes are the most popular in the spring; after July they start decreasing in popularity while electric bikes rise through the ranks, becoming the more popular ones from October to January. On the other hand, for casual users classic bikes are never on 1st place, they are even with electric ones on may and June, but other than that electric bikes win by a wide margin, especially from July to November.



<Figure size 720x576 with 0 Axes>



To test the previous hypothesis that users go on longer rides when it is warmer outside, the table below shows the average duration for trips per month. We see that, indeed, **in warmer months the trip duration is highest**, while it is **lowest on colder months**. This applies both to casual users and to members as well, although again we see that the curve for casuals is a lot steeper.

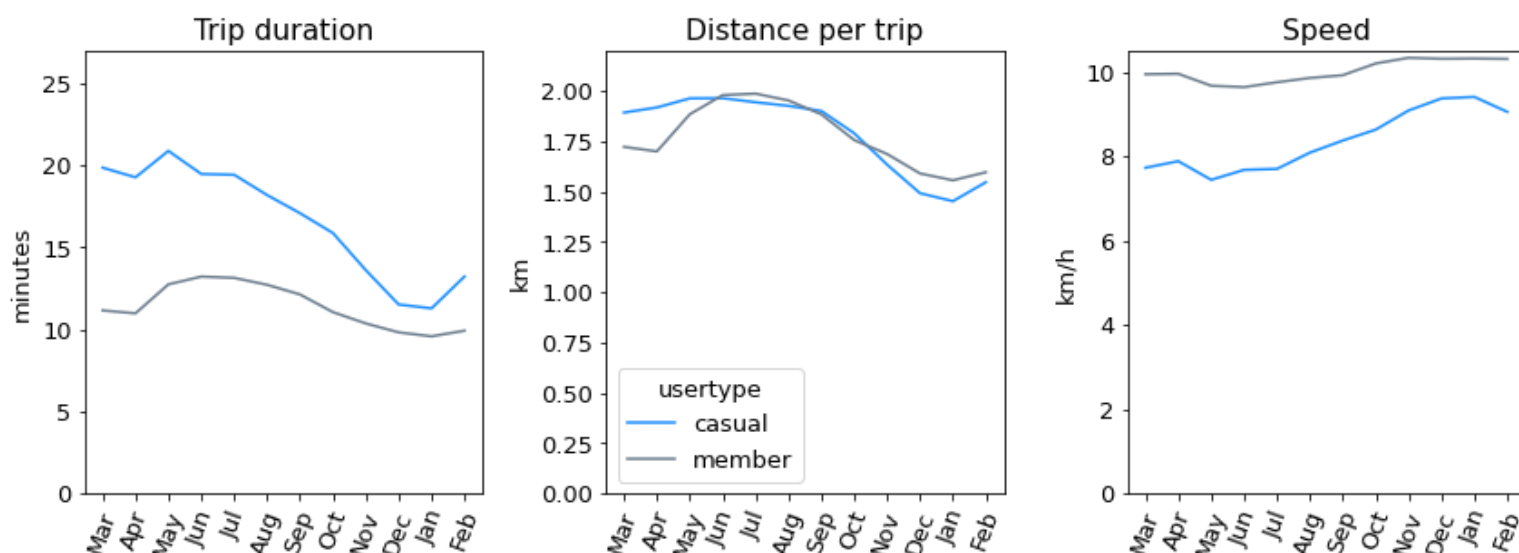
This effect could happen either because on warm months trips are longer in distance, because users ride more slowly, or a combination of both.

The second plot shows that both types of users do in fact go on longer distance rides when it is warmer. More importantly, they **both go on rides of almost the same length all year** except in march, April and to a lesser extent on may.

But this is not the only reason why spring and summer rides last longer. The last plot shows that **casual users are consistently slower than members**, although **both of them slow down in warmer months**. This seems to further verify that casual users tend to enjoy the ride and take their time going from one place to another, while members use it more as a convenient method of transportation. All users make haste when its cold outside, though.

So the longer average duration for trips in warm seasons has two causes: 1. on those seasons trips are longer in distance, and 2. users are slower while using the bikes when its warmer. We can say it is causation in this case since $\text{speed} = \text{distance} / \text{duration}$, so $\text{duration} = \text{distance} / \text{speed}$. If distance is larger and speed is smaller the duration will be larger on both accounts. As for the cause of these effects being the weather, this is only a hypothesis.

Monthly Averages



One other factor that probably has some influence in the fact that members are faster overall is that they are very likely to be fitter from using the bikes regularly as one of their methods of transportation.

Yearly insights

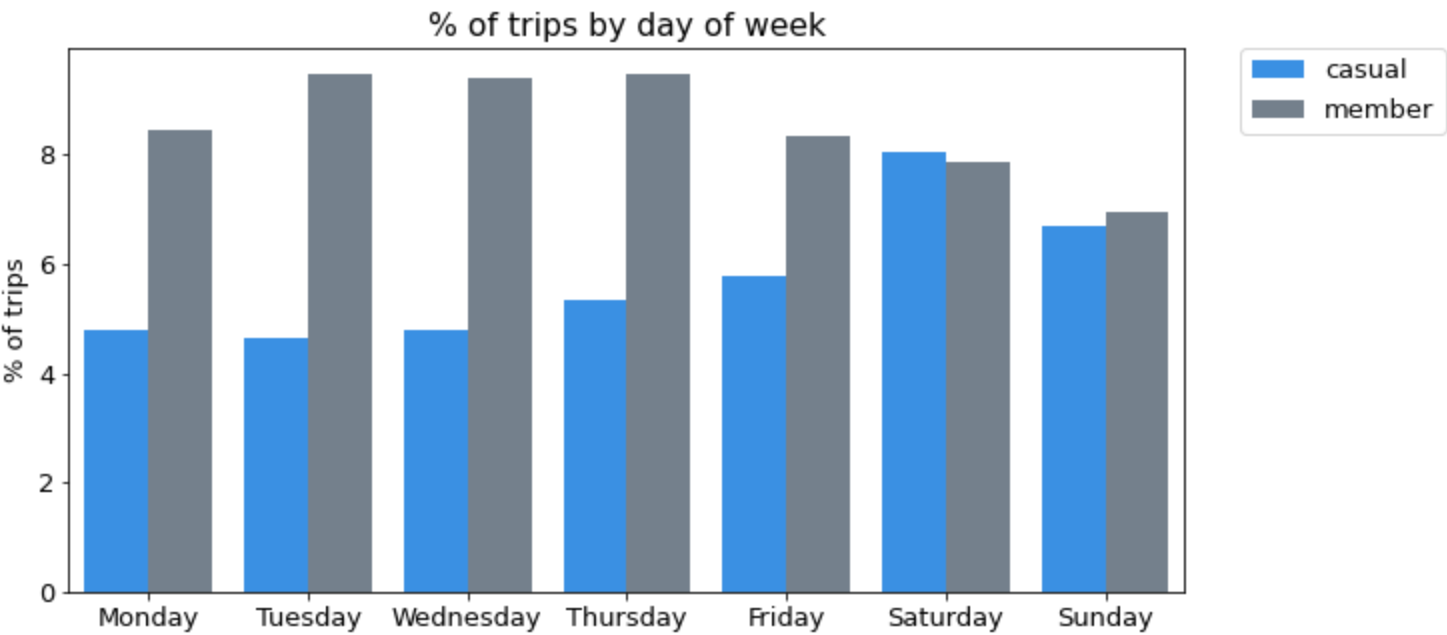
- There is a **high demand for electric bikes from casual users from May to October** (approx).
- **April to November** are the **most active months** overall.
- **November to March** have a really **low rate of casual users**.
- **Casual users are slower** than members.
- Trip*distances are longer on warm seasons.
- Speed is higher in cold seasons

Put as a story, we have seen that trips follow a certain trend over the course of a year; on March there not that many daily trips, they are short in duration, made at higher speeds and mostly by members. As the spring and summer come along casual users start using the bikes more and more, users in general tend to go a little bit slower and taking slightly longer trips, which makes the time they use each bike a little bit longer. People also start using less electric bikes and more classic bikes. As the summer fades out and the cold starts to settle, casual users gradually stop using the bikes, the rides grow shorter and faster in both speed and duration, and electric bikes become the most popular once more.

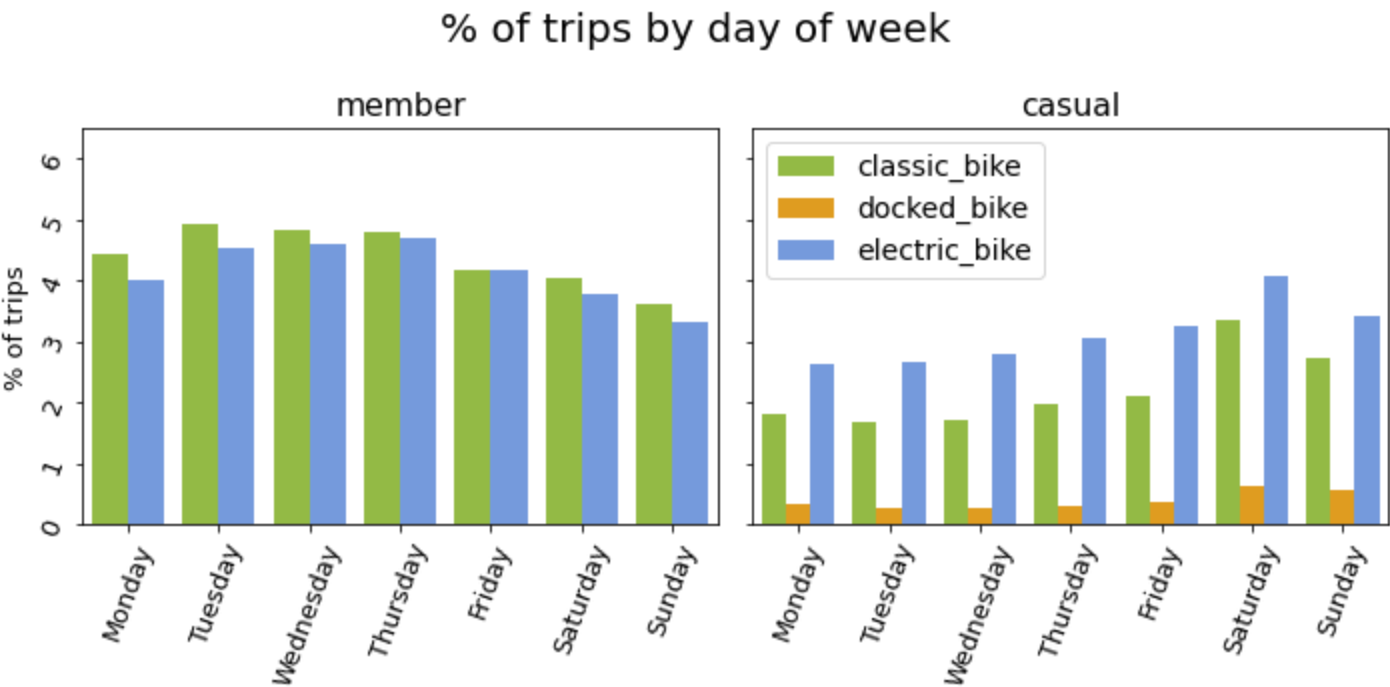
Weekly trends

In this section we will now focus on weekly trends.

The first thing we notice is that on a weekly basis members and casuals display different trends. Members use the bikes mostly during weekdays, while **casuals tend to use them most on weekends**, especially on Saturdays.



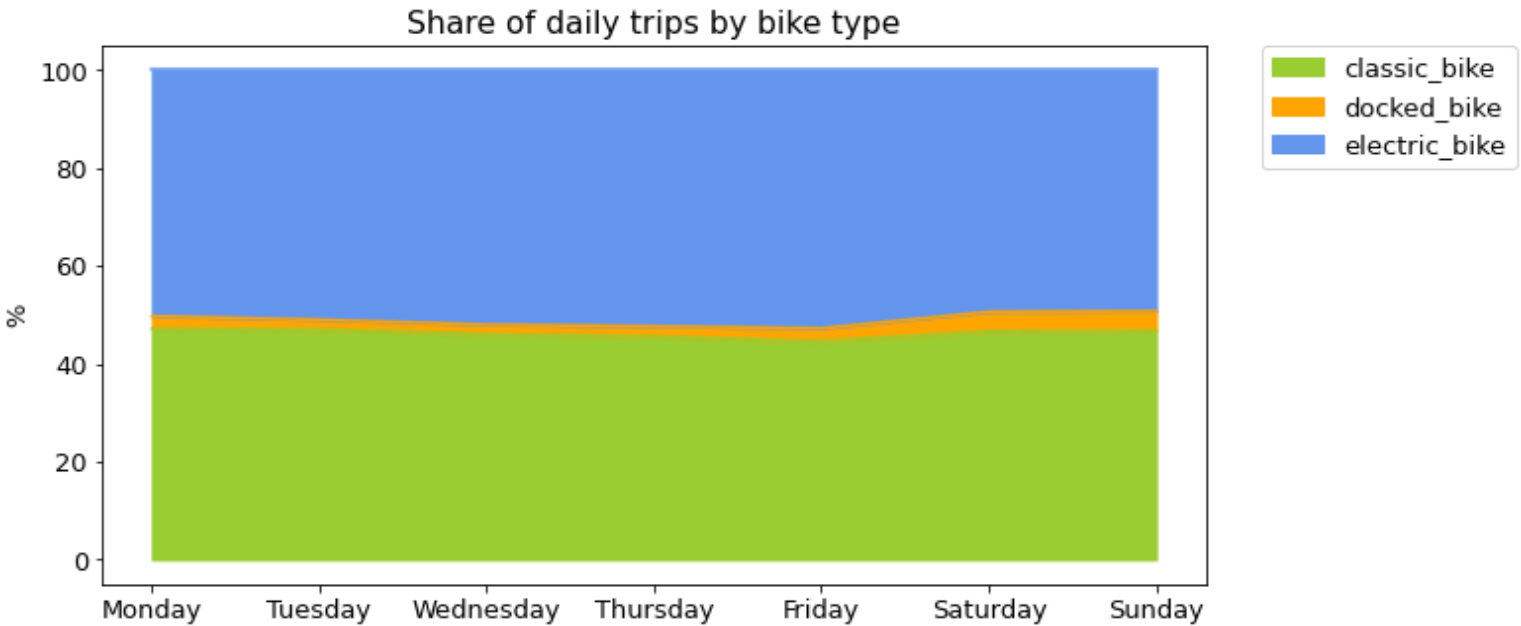
A closer look at the distribution of trips along the week, showing the distributions for each user type by bike type, shows that the trend is the same for all bike types within each user type, although the scale at which each bike type is represented for casual users is very different since they use electric > classic > docked.



The area chart below proves that the trends are almost exactly the same for all bike types (within each user type), since the ratios of classic, docked and electric bikes is pretty much the same throughout the week. If the trends were different the ratios of classic to electric to docked would change, and the proportions for each color shown in the plot would vary.

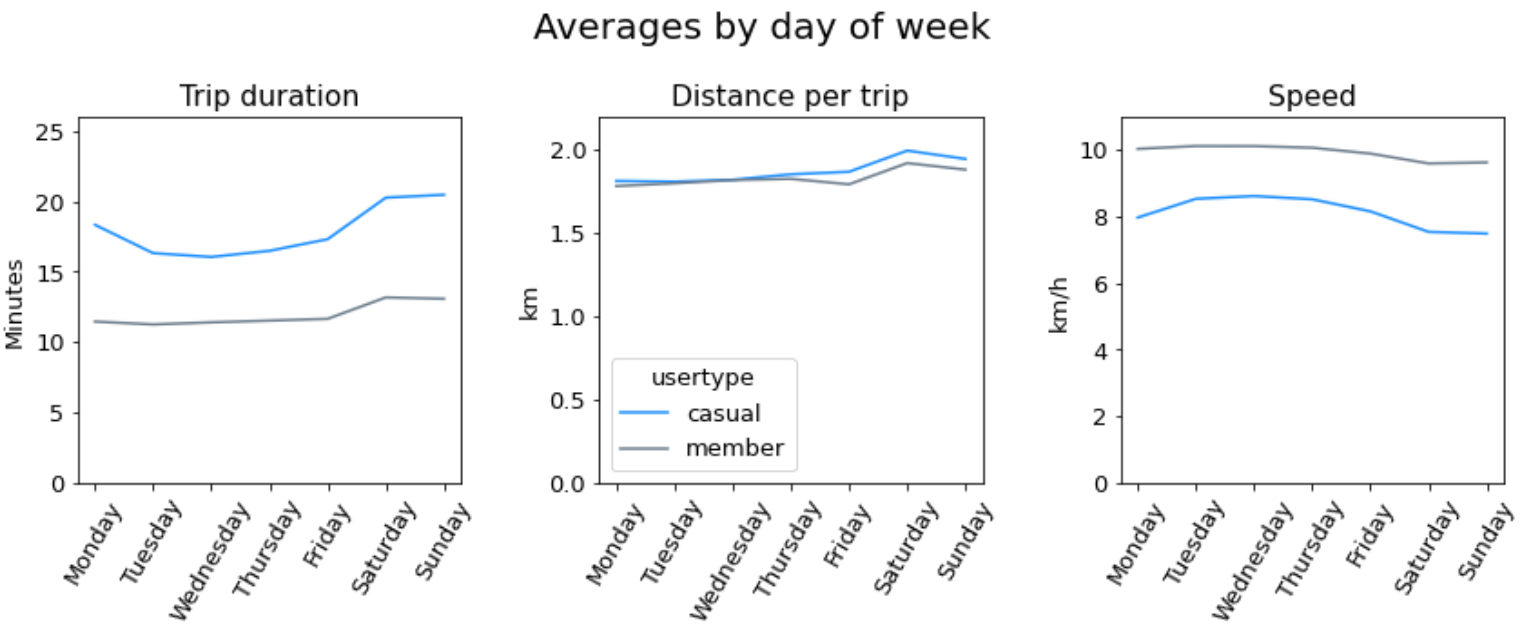
A close look reveals that docked bikes have a bit of a steeper curve, making their ratio higher on weekends, but the difference is so tiny that it doesn't really make a noticeable difference.

<Figure size 720x576 with 0 Axes>



As for the duration, distance and speed of the trips, again we notice that member's are shorter in duration, higher in speed and almost the same in distance.

They both display exactly the same weekly trends, though: for both types of users **rides on weekends are a little bit longer in duration and in distance, while speed goes down a little**. This is probably because people are, in general, in less of a hurry on weekends. Maybe they also go in more recreational rides on weekends or visit their friends and family, traveling greater distances than usual.



Weekly Insights

- Casual users use bikes more often on weekends.
- Rides are slower, last longer and go further on weekends.
- Members use bikes more often on weekdays.
- Trends are the same for all bike types for each user type.

Daily trends

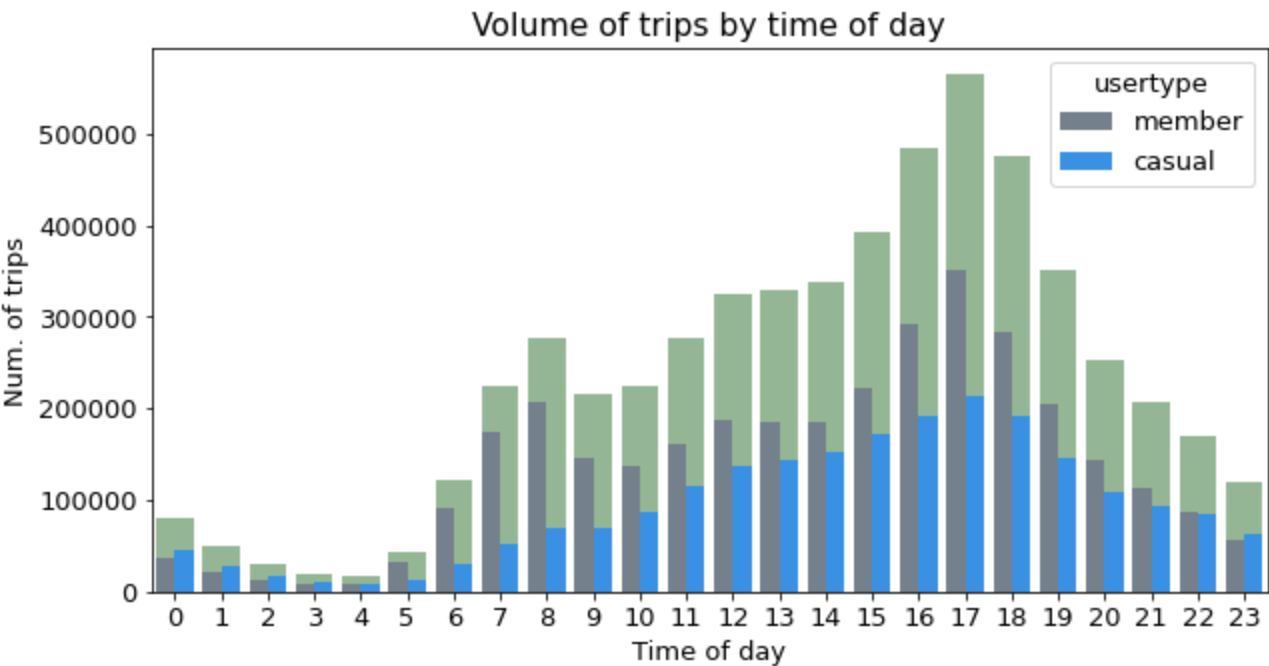
It is time to look at how rides change during the day, on an hourly basis.

In the next barplot the green shaded bars represent the total number of trips started within each of the 24 hours, that is the sum of the casual and member trips.

We see that the daily pattern is very similar for both types of users, although the peaks around 8am and 5pm are a lot more noticeable for members, very likely because of commutes to and from work. Very few rides happen after 11pm and before 6 am, and the **busiest hours are 4pm to 7pm**.

An interesting thing is that from 11pm to 4am there are more rides made by casual users than by members. At all other times the number of rides made by members is considerably higher than that of casual users.

Text(0.5, 1.0, 'Volume of trips by time of day')

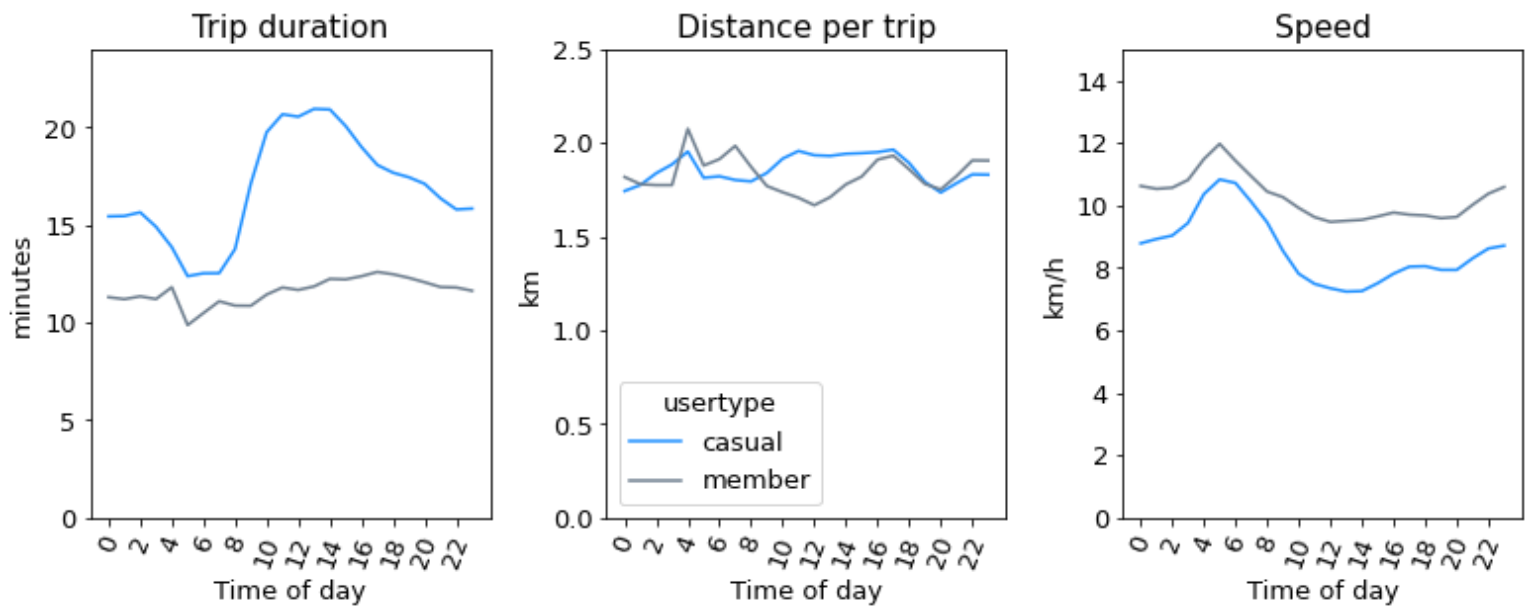


the next plots show how the duration of the trips, their distance and speed change throughout the day.

The duration of the trips is really low in the early morning, from 5am to 7am there is a minimum of 10-12 minute trips that coincides with a speed peak of 10 to 12 km/h. The slowest and most time consuming trips, on the other hand, happen between 9am and 6pm. This could have something to do with the higher traffic of all kinds (cars, bikes and pedestrians as well) at those hours.

The distances are all over the place during the day, although there is a clear drop in distance traveled around 9pm and a small peak at 4am. Why this is, I don't have a clue.

Averages by time of day

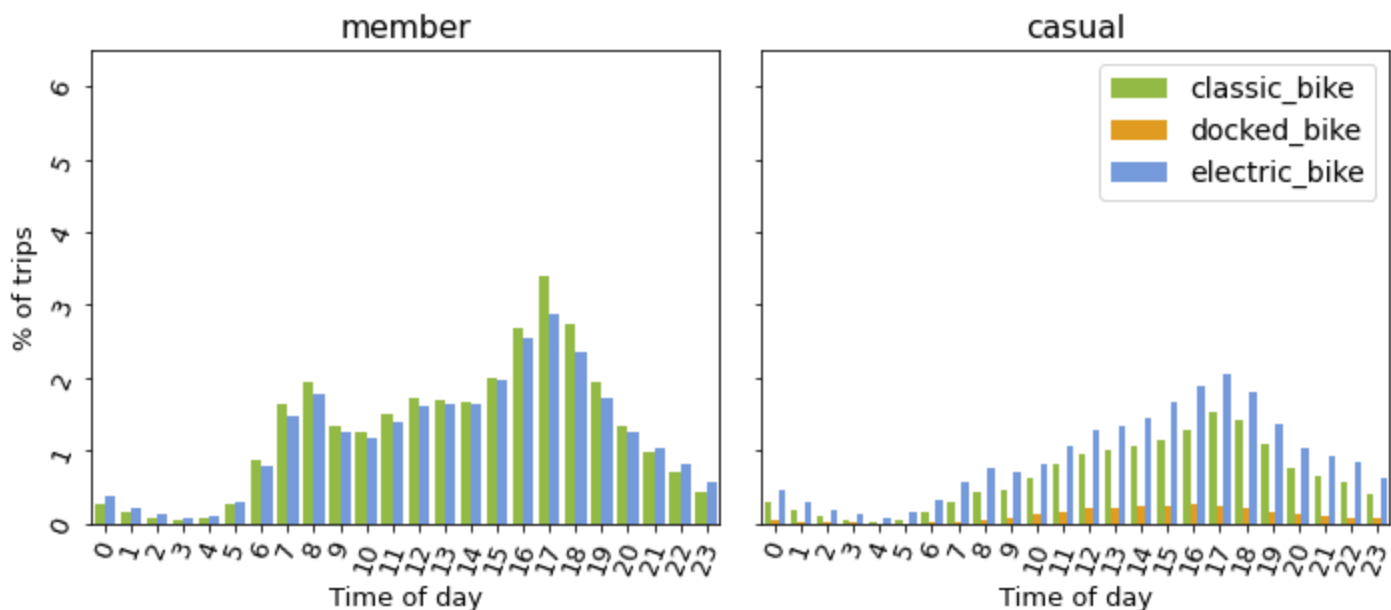


The following plots show the percentage of all trips made within each hour, for each bike and user type. At a first glance the behavior for members seems more complex, having that second peak around 8am, most likely caused by people commuting to work. After a closer look we see the casuals curve has this same behavior, although it is a lot less pronounced on the 8am peak.

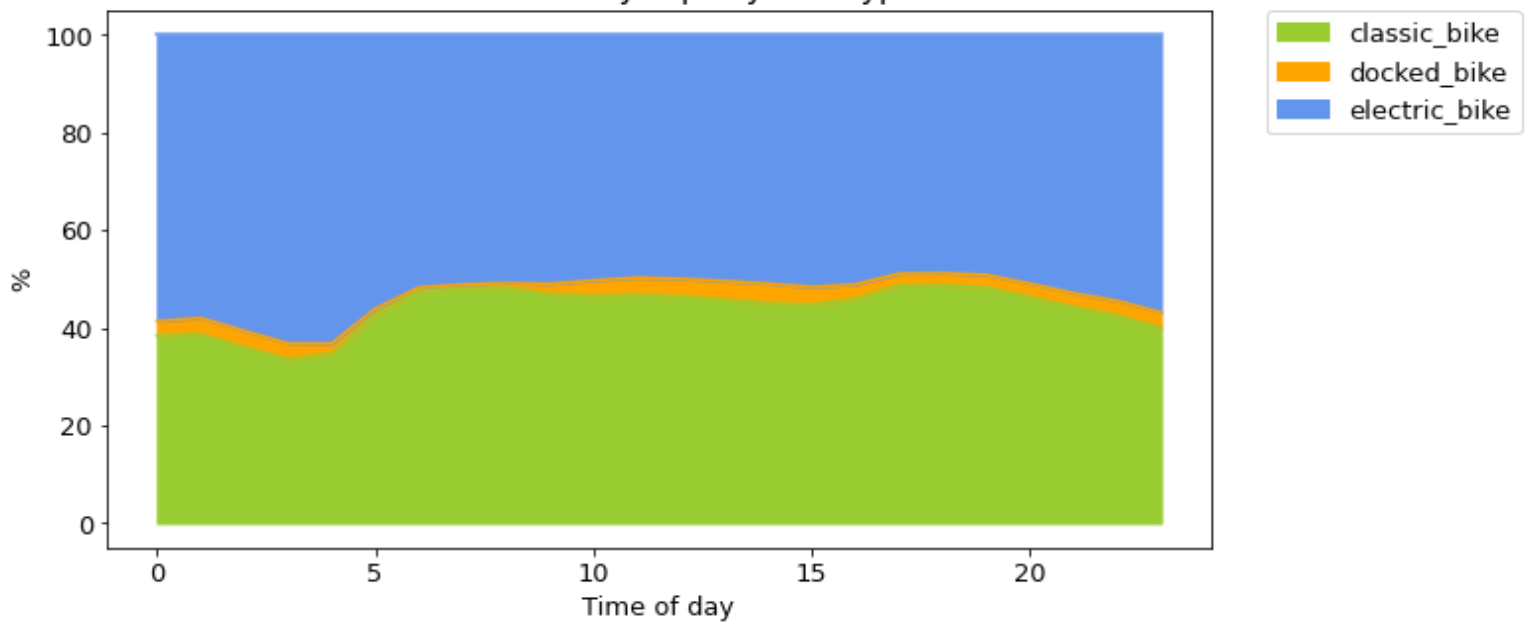
Other than that, the ratios of classic to docked to electric bikes is pretty much the same for both kinds of users throughout the day, except for the night time, from 9pm to 4am or so, when electric bikes take some of the share from classic bikes.

This partially explains why the speeds are faster during the night, since that is when - in proportion to all bikes made within the hour - most bikes are made with electric bikes. The speed spike at 6am shown above, though, falls outside this range, which is interesting.

% of trips by hour



Share of hourly trips by bike type



Daily Insights

- There are **two daily spikes, around 8am and 5pm**, most likely related to commutes to and from work, respectively.
- **Most rides happen between 10am and 9pm**
- Electric bikes are more popular at night.
- Casual users are slowest around midday.

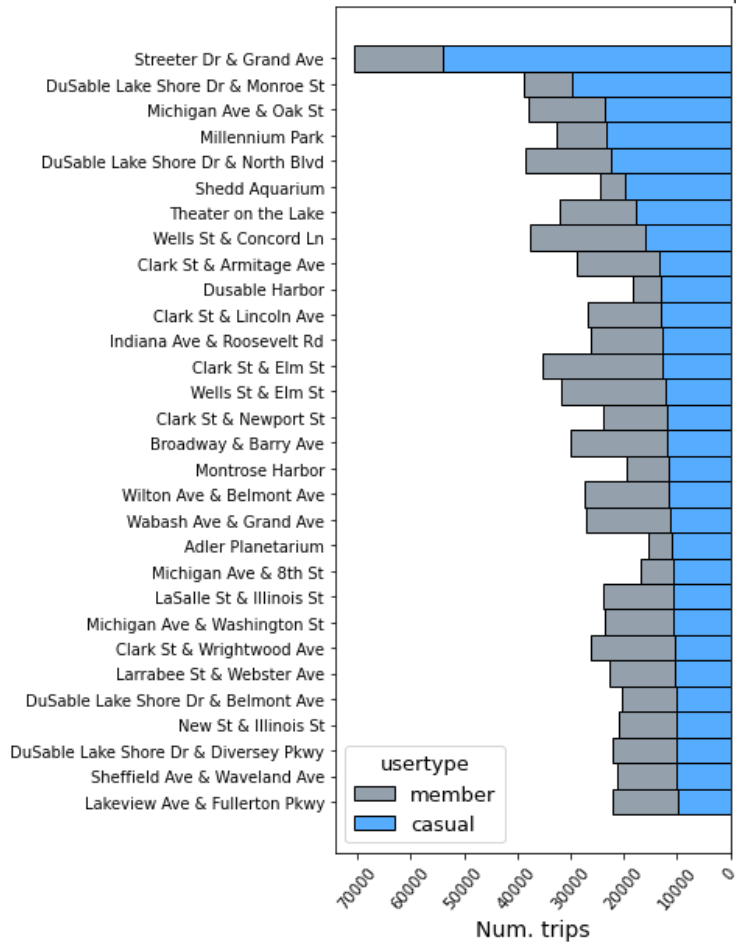
Geographical trends

Finally we will look at some of the geographical trends.

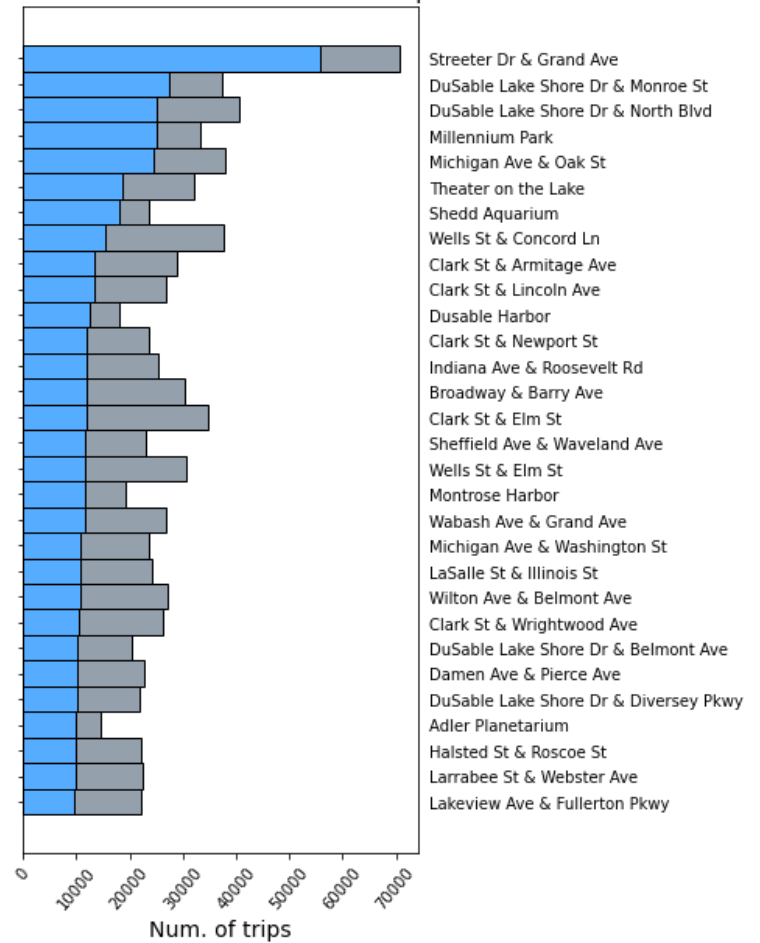
The following funnel plot shows the most popular stations among casual users. The blue bars show how many trips made by casual users started/ended at that station and the gray bars show the same thing but for members. It is ordered by decreasing number of round trips by casual users. From it is evident that there are some stations which are a lot more popular among casual users, as compared to members. Many of these **stations are located near museums, parks, and other recreational or tourist-oriented establishments**, which partially explains why most of the rides started or ended there are made by casual users, who in many occasions are tourists.

So, for example, to mention a few stations, Streeter Dr & Grand Ave, the most popular station overall but especially among casual users, is located within Addams memorial park, next to the navy pier, the children's museum, and many other attractions. Other stations that are mostly frequented by casual users are the Millenium Park, Shedd Aquarium, the Theater on the Lake, the Addler Planetarium, Dusable Harbor and DuSable Lake shore Dr & Monroe St., which is at the entrance to the yacht club.

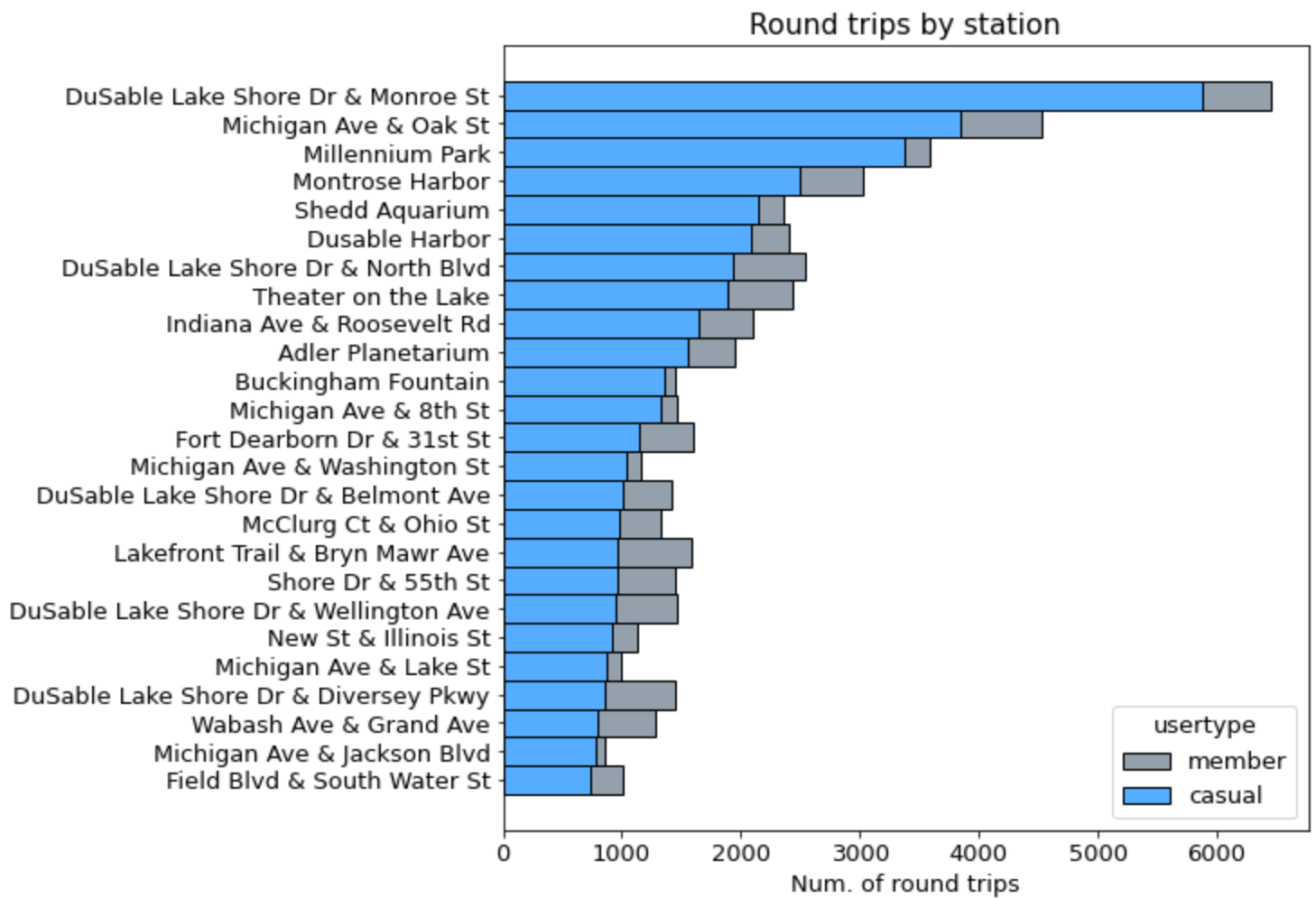
Most common stations to start a trip



Most common stations to end a trip



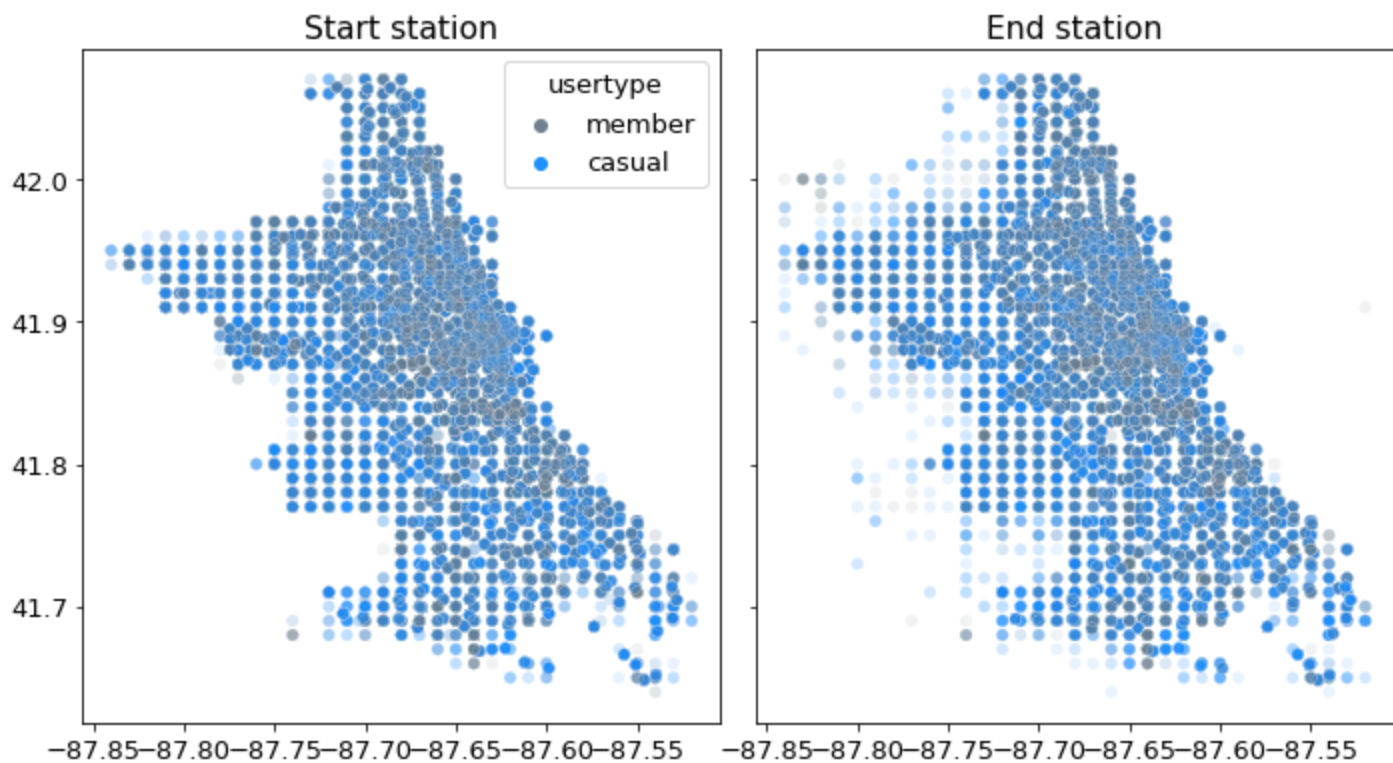
Earlier, from the distribution of distances traveled per trip, we hypothesized that casual users made a lot of round trips, that is, trips starting and ending at the same station, making their most common distance to be 0. The following shows the 25 most common stations where round trips happen. Many of the stations above are here again, and it is clear that **most of the round trips made at these stations are made by casual users.**



The maps below show that **the most busy region for both start and end stations is in the downtown Chicago area**, and in general all areas near shore, on the East side of the city. All areas seem to be almost equally frequented by casual users and members.

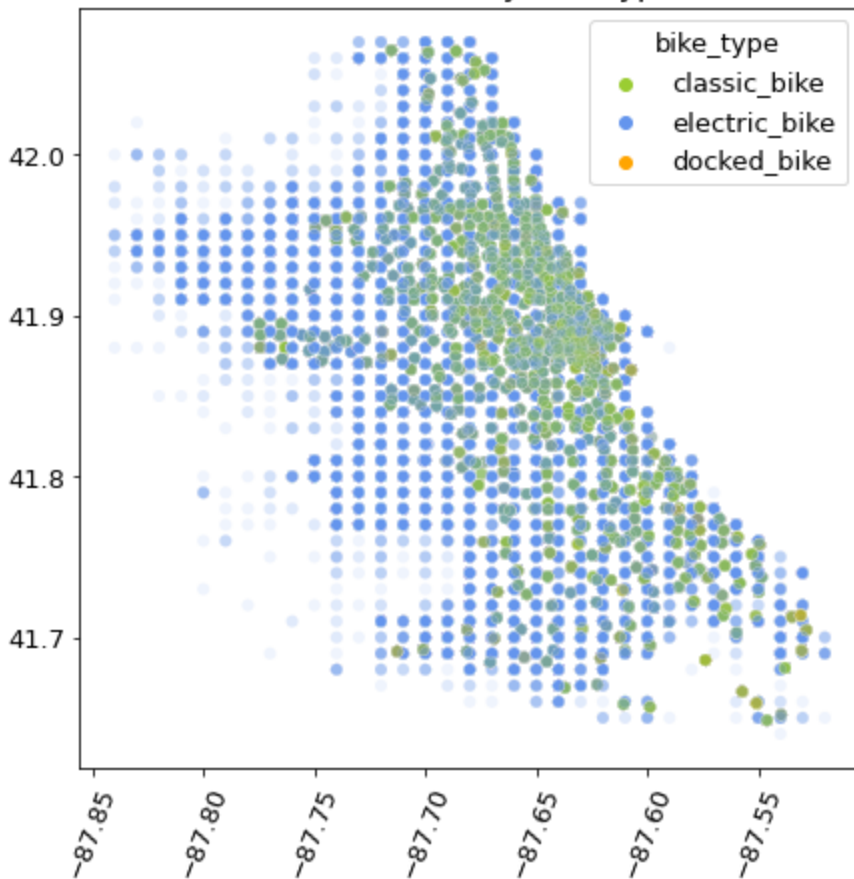
It is worth noting that, while the start location map has a more or less well defined border, determined by the locations of the stations, the map for the end location doesn't have a well defined border on the West side. There are plenty of trips that apparently ended in places where no trips start at; most of these trips seem to have been made by casual users. This means that casual users like going out West, out of the official area covered by the stations.

Geographical distribution of bike trips



The next map shows the end location for a sample of the data. Rides made with classic and docked bikes are contained within the most urban areas of the city, which is closer to the coast. Trips that go further West are almost always made with electric bikes, which correspond to some of the outliers noticed above. From this we can deduce that **many casual users take out electric bikes to go West, ending their trips at locations where there are no stations**. According to the company's website electric bikes can be left locked on poles, regular bike racks, etc, so the bikes can be recovered by the company and taken back into the city, leaving us with many trips ending in those areas but only a few starting there.

End locations by bike type



Geographical Insights

- Statscatterplotst used by casual users are **near touristic areas and establishments**.
- **Casual users** tend to make **round trips** a lot more than members.
- There is a fair amount of **trips ending in the North-West region** of the city, where there are **no stations** to start trips at.
- Most trips ending West are made by casual users on electric bikes.

Key Insights

Throughout this analysis there were a few differences that popped out between members and casual users. Here is a list of the most relevant qualities that characterize casual users:

- They prefer electric bikes, making **54.7% of the last 12 months trips with electric bikes**, 38.4% with classic ones and only 6.90% with docked bikes.
- They tend to go on **more lasting trips**, with an average of 18 minutes compared to the 12 minutes of members. From March to July trips are even longer, going as high as 22 minute per trip, on average. During this period the average distance per trips grows larger too.
- **81.09% of their trips happen between May and October** (inclusive), with only 18.91% happening in the other half of the year.
- Many of the **most frequented stations** by casual users are **located next to museums, piers, parks and other touristic attractions**. Most of these stations are not so frequented by members.
- There are **more rides** made by casual users **on weekends** than on weekdays, in contrast to members who are the other way around.

- Casual users use the bikes to make round trips more often than members, ending the ride at the same station where it began. **6.13% of casual user's rides are round trips**, compared to 2.13% of member's rides. Most of these rides start and end in the same touristic spots mentioned above.
- The busiest time of day for casual users is 5pm, same as for members. The demand for bikes from all users peaks from 4 to 7pm, making it the time most likely to leave users without available bikes. This is also when trips are longest in duration.

Recommendations

1. **Launch a semi-annual membership.** Casual users make more than 80% of their trips within only 6 months, and barely use them the other half of the year. Paying only for the seasons they might be tempted to use the bikes could be a good incentive for them to purchase a membership. This could push some of the current members out of their current plans, though, so the pricing for the different kinds of memberships should be thought of carefully to keep as many current members in their annual membership.
2. **Launch a spring or summer sale** a few months in advance, for any people expecting to travel into the city and planning their future vacation, as well as for people living within the city who are just waiting for the weather to be better.
3. **Launch a marketing campaign showcasing customers in iconic places** within the city using the bikes, showing all the most important landmarks, museums, parks, etc, can be easily reached using our bikes.
4. **Augment ride time/reduce extra time cost.** Casual users tend to go on longer rides; if the bikes are charged by the minute after a certain amount of time is reached in a single ride (as currently is), casual users won't commit since that system doesn't fit their needs. Pushing that free time a little bit upwards and/or lowering the price each minute is charged for after that could help some of those casual users decide to go for the membership after all. This could even be compensated by rising the annual fee at the same time, although not by a large amount since members who take short and efficient trips could reconsider their subscription.
5. **Change the rate for weekend rides.** The most popular days for casual users to go on a bike ride are weekends, which are the least popular days for members. Having the possibility to ride more at ease during those days could incentivize casual users to use the bikes more and maybe even consider purchasing a membership.
6. **Install more stations to the West**, specifically in the North-West regions. There seems to be considerable interest in those areas, judging by the many rides going there made by casual users on electric bikes.