Do snow, or rain, or heat or gloom of rush hour stay Toronto's Bike Share users?

The Bike Share program in Toronto is one way for citizens and tourists to get around the city without cars, taxis, or public transit. Bike stations are located throughout the city. You can pick up a bike at any station, ride it and drop it off at another station when you're done. The bikes are meant for short trips of less than 30 minutes.

Toronto weather varies a lot throughout the year. With warm humid summers and cold snowy winters. The Bike Share program is available all year round. But how does ridership change with Toronto's variable weather? Can you predict the number of rides based on weather conditions and other characteristics?

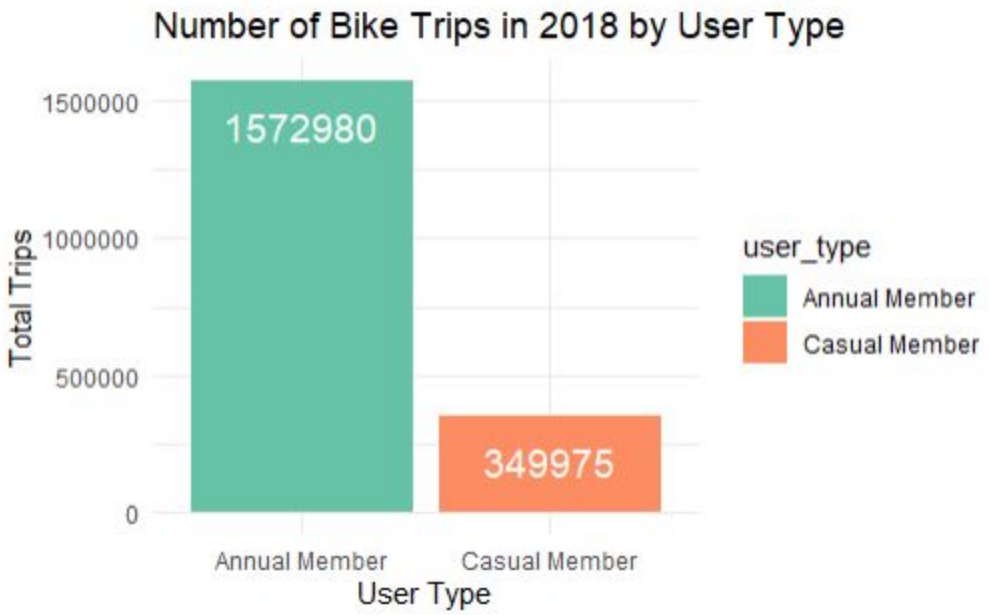
Data

The following analysis uses Bike Share Toronto usage statistics for 2018 from the city's Open Data Portal[1]. As well as data about daily weather statistics made available by Environment Canada[2].

Descriptive statistics

There are two types of Bike Share program users, annual members and casual members. Annual members pay a yearly fee to have unlimited use of the system. Casual members pay for short-term unlimited access for either 24 or 72 hours. In 2018, more than 80% of the trips taken in Toronto were by annual members.

[1] https://open.toronto.ca/dataset/bike-share-toronto-ridership-data/
[2] https://climatedata.ca/download/

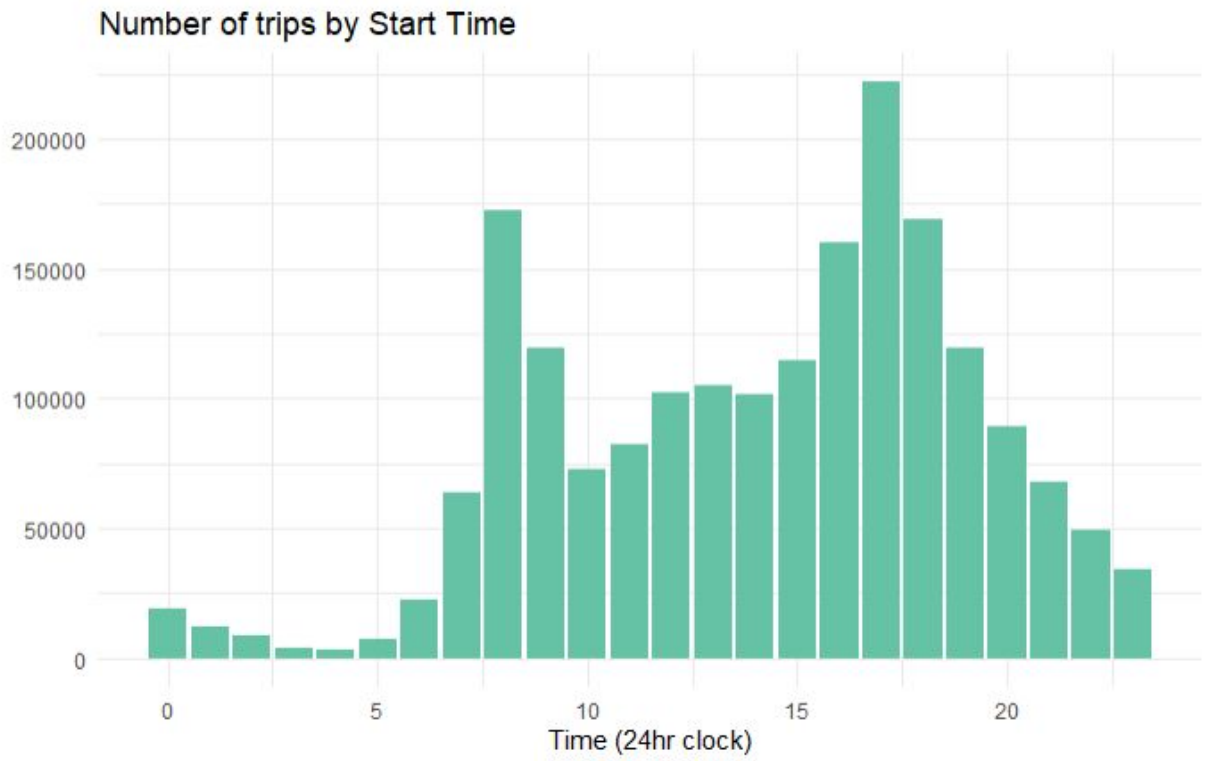


The bikes are meant for short trips and users get charged overage fees if their trips exceed 30 minutes. Most Bike Share users stick to the time limits. Only, 6% of all trips in 2018 were longer than 30 minutes and the average trip duration is 16min and 3 seconds. The shortest trips in the open data set were 60 seconds and the longest trip was more than 15 hours long.

Causal riders exceed the 30 minute ride limit much more often than annual members. Of the 120,439 trips that resulted in an overage fee, 73% were for casual riders.

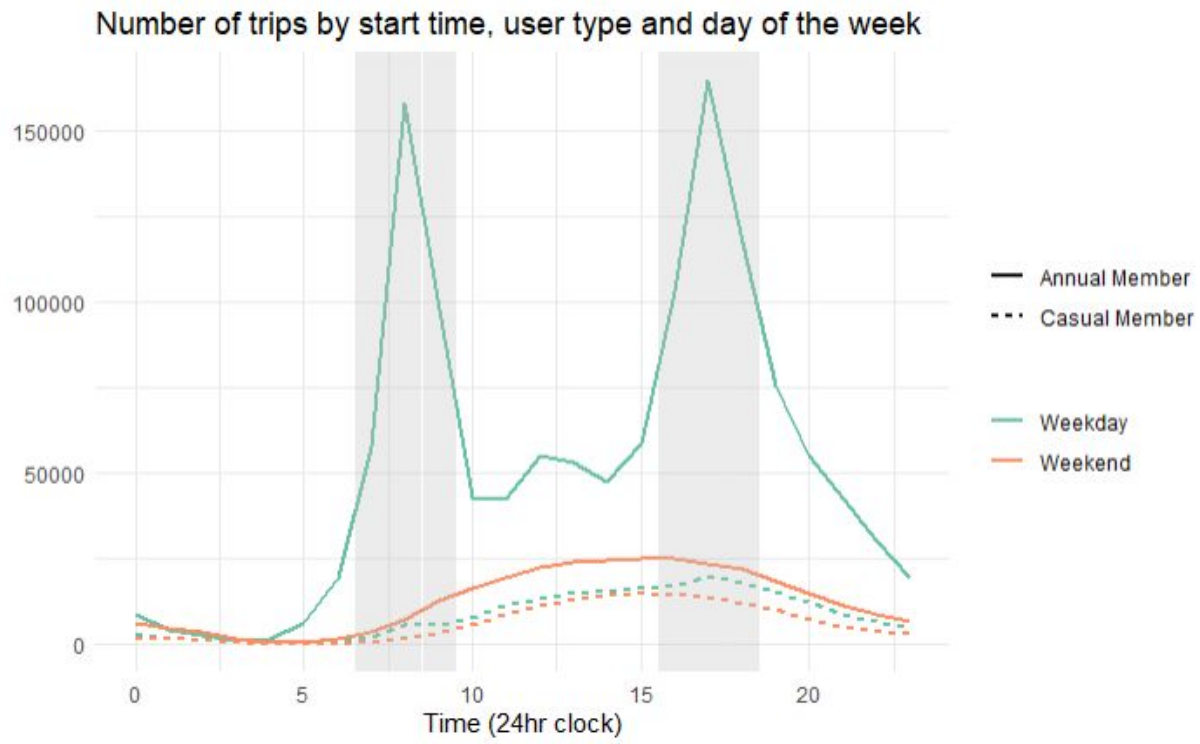
	30 min or less	More than 30 min
Annual Member	1,540,333	32,647
Casual Member	262,183	87,792

The figure below shows the number of trips grouped by the hour they started. The number varies through the day. There are very few bike trips between midnight and 6am. Ridership peaks during the daily commute, from 8-9 am and from 4-6 pm.

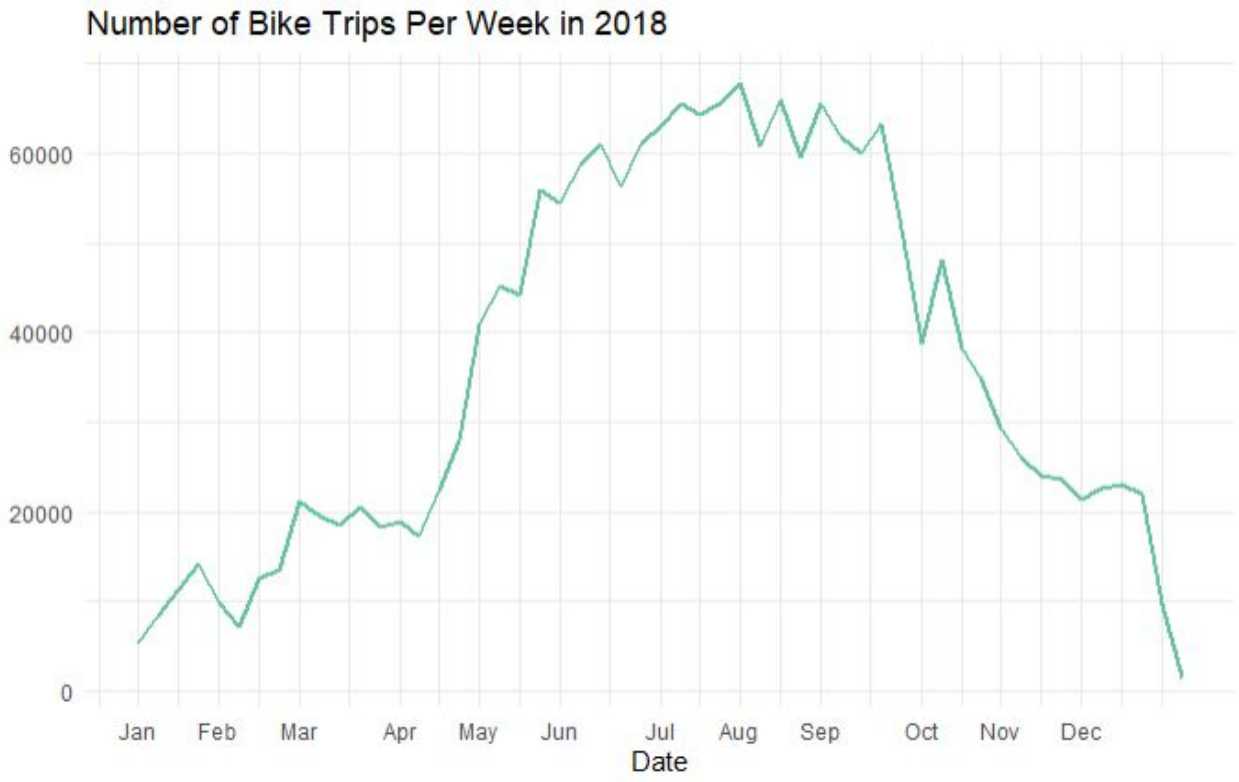


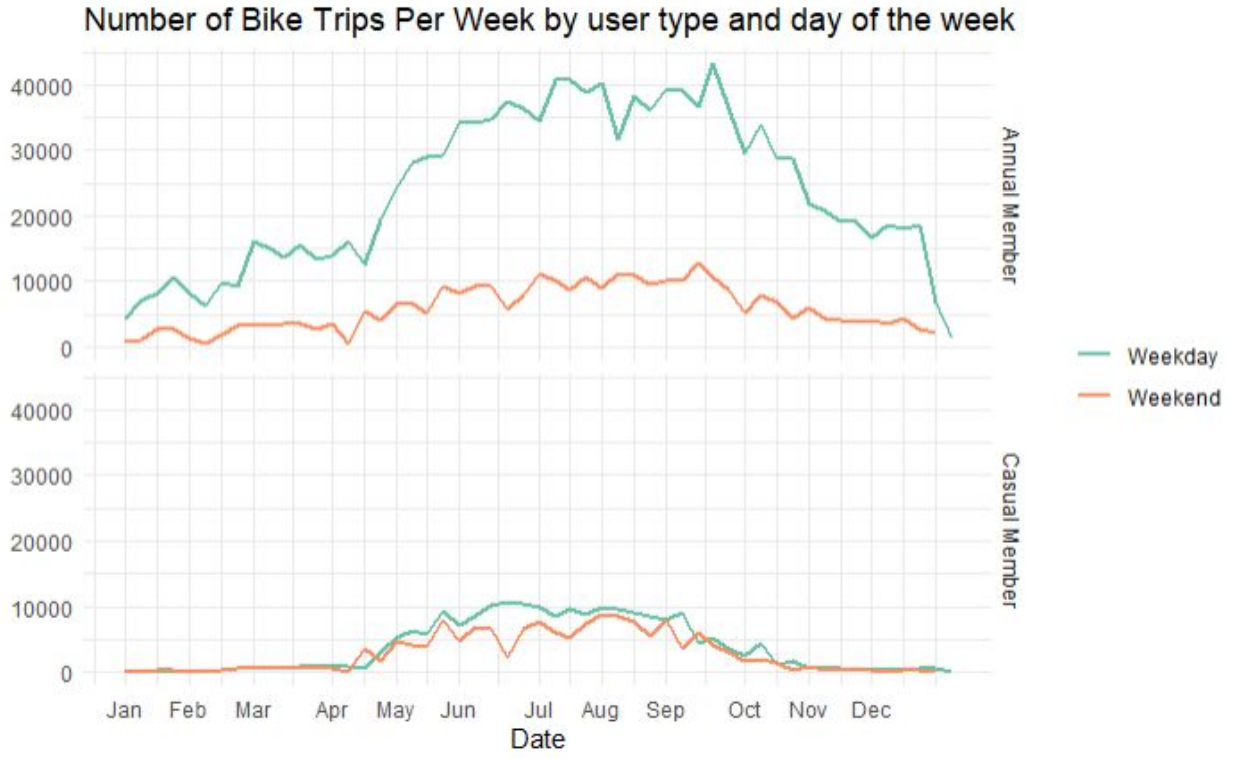
This pattern makes the most sense for annual members during weekday travel. The peaks occur during the morning and evening rush hour (highlighted in grey below). Many of the annual members must be commuters who use the service to get to and from work.

The pattern is different for casual members and annual members on weekends. First, there are fewer trips on weekends except very early in the morning, from 1-2 am. Second, the number of trips rises gradually from about 9am to 4pm and decreases again through the evening. Third, the pattern for casual users is similar to the weekend annual members. This could be because casual users may be visitors to the city and their schedule of sightseeing may be similar to a Torontonian's weekend plans.



The number of trips varies through the year, with more trips when the weather is warmer (May to October). This trend doesn't vary by day of the week or user type. Everyone seems to prefer biking in the spring and summer.


Number of Bike Trips Per Week in 2018
60000
40000
20000
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Date

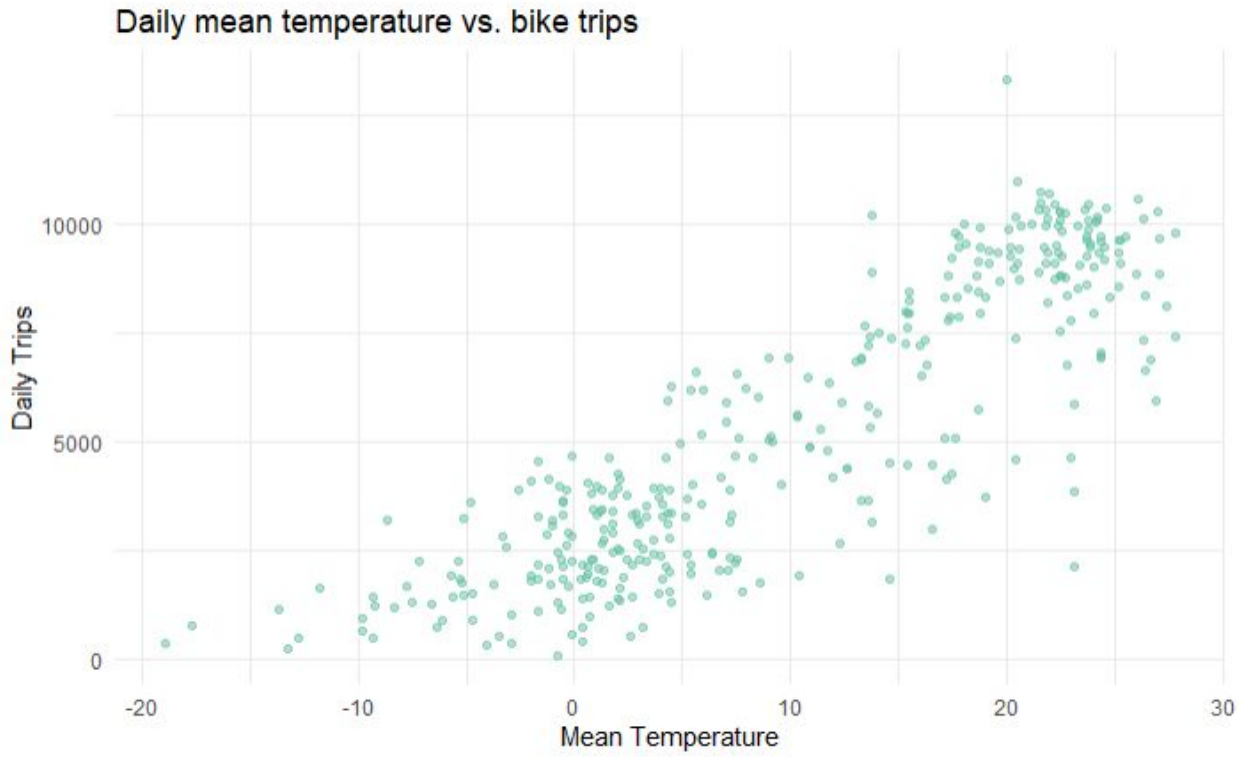

Number of Bike Trips Per Week by user type and day of the week
40000
30000
20000
10000
0
Annual Member
Casual Member
Weekday
Weekend
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Date

Weather and Bike Share Use

Let's dig deeper into the impact of weather on ridership. I used the daily weather statistics for the City of Toronto for each day from January 1st to December 31st 2018. I focused on three measures: Mean daily temperature, total precipitation and amount of snow on the ground. Some records were missing for each, so I imputed missing values using kNN.

Temperature

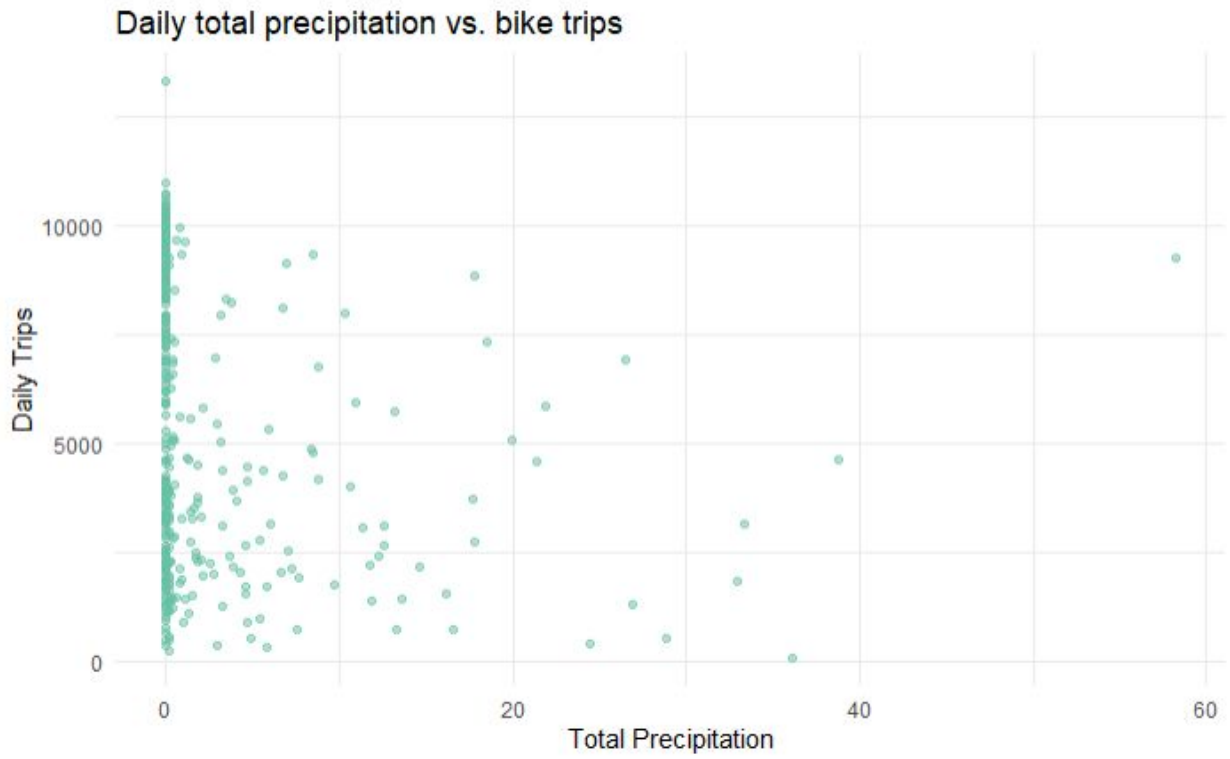
There is a strong positive correlation between mean temperature and the number of daily trips (r=0.87). As it gets warmer, more people use Bike Share. But the relationship isn't perfectly linear. The number of daily trips rises more between 10 and 20 degrees than it does between -10 and 0 degrees. Also, the number of trips seems to plateau at the extremes. For example, when the mean temperature is 20 degrees or higher there are about 10,000 trips.



Precipitation

There is a weak negative relationship between total precipitation and bike trips (r = -0.18). The data cluster on the left because for 209 days of the year there is zero precipitation. When there is more than 25cm of rain there tend to be fewer bike trips.

But there are also many exceptions, like the day with 58cm of rain and 9256 bike trips.

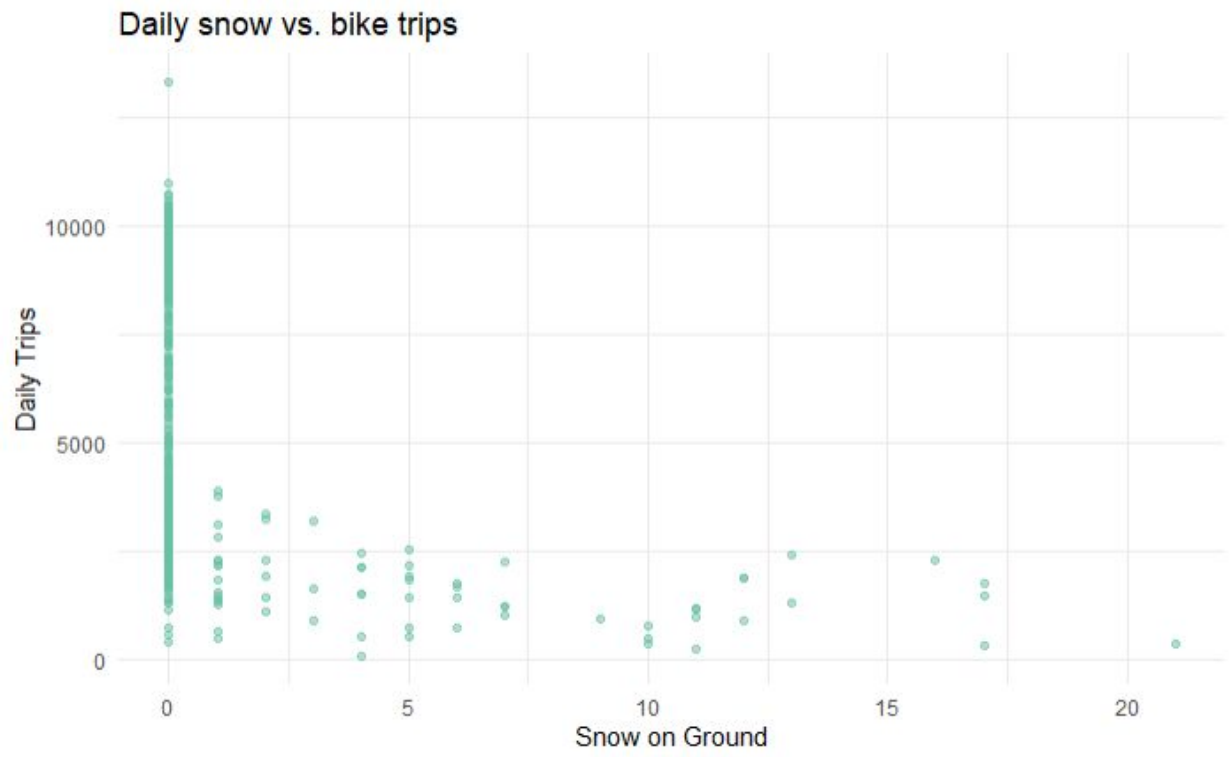


The weak relationship may be a limitation of the data. More detailed information about when it rains through the day may be a better predictor of bike trips. Chris McCray used hourly precipitation data in his analysis of Bike Share trips in Montreal[3]. He found that the number of bike trips was impacted differently by a brief shower versus sustained rain throughout the day.

Snow on the ground

There is a moderate negative relationship between the amount of snow on the ground (in centimeters) and the number of bike trips (ρ = -0.57). Again most of the data clusters on the left, because most days of the year there is no snow. When there is snow, there are fewer trips overall. And as the amount of snow builds up there are fewer and fewer trips.

[3] https://web.meteo.mcgill.ca/cmccray/weather-bike-traffic-montreal/



Predicting the number of bike trips

The next step is to build a model to see if the number of Bike Share trips can be predicted by the weather, day of the week and time of year.

I was influenced by the work of Todd Schneider who built a model to predict the number of bike share trips in New York City based on the weather[4]. Schneider used a non-linear least-squares algorithm for two reasons. First, the dependent variable, number of bike trips, is always positive and a linear regression model can't guarantee a positive number. Second, the relationship between bike trips and weather, particularly temperature, are not linear.

I'm facing the same issues with the Toronto bike and weather data. So, I used the same nlsLM() function from the minpack.lm package as Schneider to implement the Levenberg-Marquardt algorithm. You need to specify the form of the model and the start values up front. I used the following model:

---

4 https://toddwschneider.com/posts/a-tale-of-twenty-two-million-citi-bikes-analyzing-the-nyc-bike-share-system/#citibike-weather

$$d_{trips} = Baseline(d) + Weather(d) \quad (1)$$

$$Baseline(d) = e^{\beta Weekday * dWeekday} + \beta Season * dSeason \quad (2)$$

$$Weather(d) = \beta_{Weather} * 1/(1+e^{(-(WeatherFactor(d)-\beta Centre)/\beta Width)} \quad (3)$$

$$WeatherFactor(d) = d_{Mean\ Temperature} + \beta_{Precip} * \log(1+d_{Precip}) + \beta_{Snow} * d_{Snow} \quad (4)$$

This is a slightly modified version of Scheider's model. I introduced a season variable and removed a constant and a correction factor that he used to fit data from New York City.

The $\mathbf{d}$ variables are the known values from the data set for each day of the year. The $\boldsymbol{\beta}$ variables are parameters, whose optimal values are identified using the Levenberg-Marquardt algorithm.

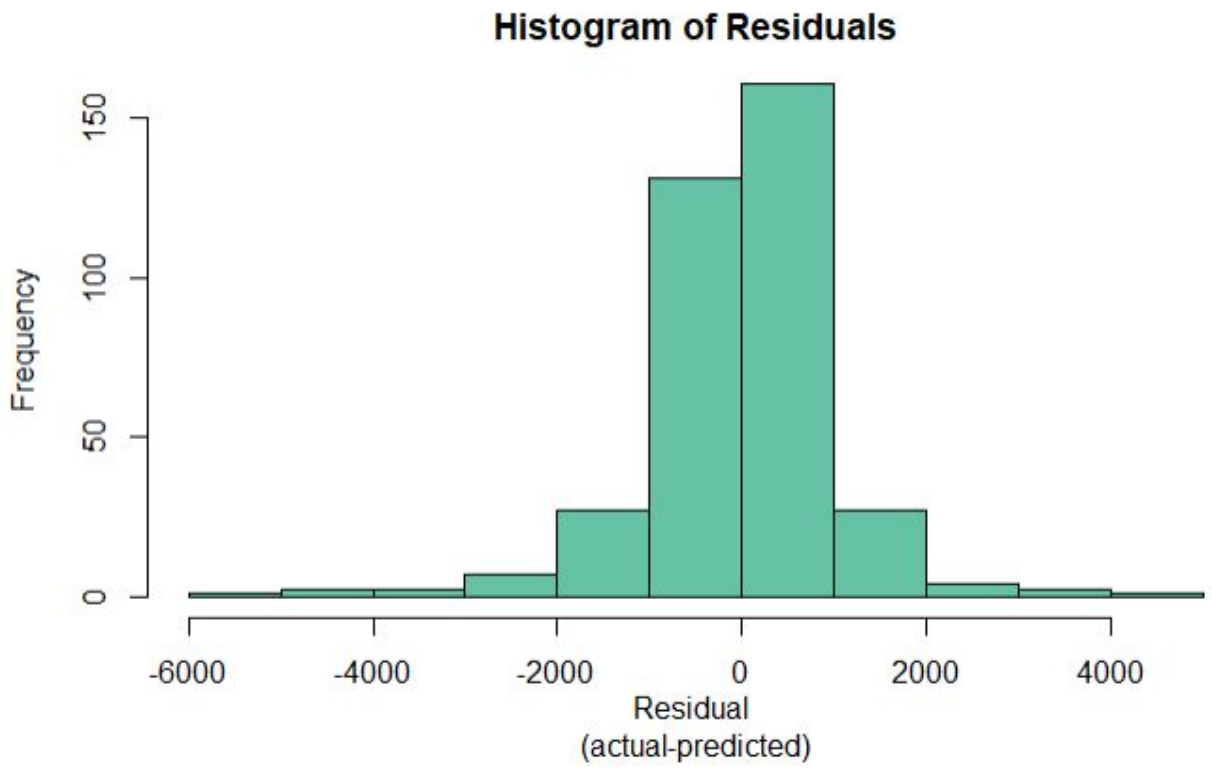
There are two main components of the model. The first component is the baseline which has an adjustment for day of the week and the season. As we saw earlier, the number of trips vary through the week and through the year. By using an exponent, it ensures that the outcome is a positive value.

The second component is the weather and it uses the s-curve function. Which reflects the relationship between mean temperature and trips that was "s" shaped.. The three variables used in the curve were mean temperature, total precipitation and amount of snow on the ground. A log transformation was applied to total precipitation because it was strongly, positively skewed.

The model seems to converge on reasonable parameters. And all the parameters included in the model seem to contribute significantly to the outcome.

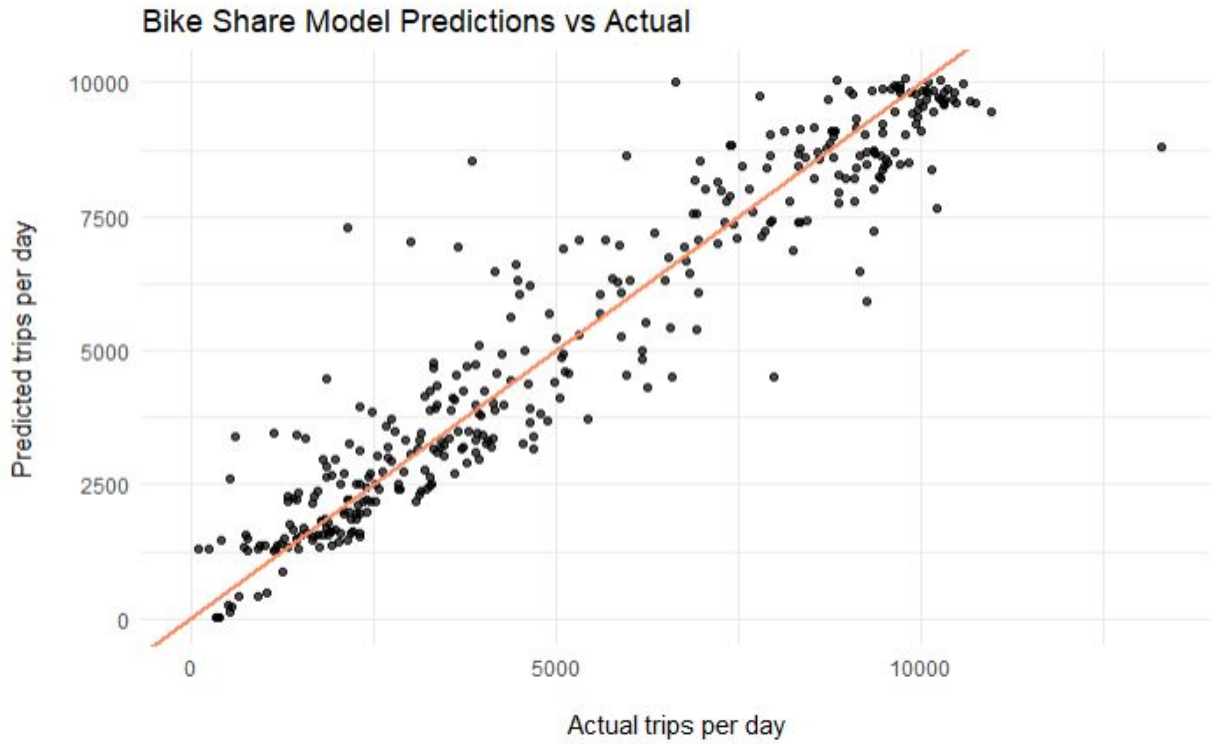
```
Formula: bike_trips ~ exp(b_weekday * WDay) + (b_season * Season) + b_weather *
    scurve(MEAN_TEMPERATURE + b_precip * log(1 + TOTAL_PRECIPITATION) +
        b_snow * SNOW_ON_GROUND, weather_scurve_center, weather_scurve_width)

Parameters:
                       Estimate Std. Error t value Pr(>|t|)
b_weekday               7.12883    0.08845  80.600  < 2e-16 ***
b_season              121.87962   17.32722   7.034 1.03e-11 ***
b_weather            7382.59044  424.90175  17.375  < 2e-16 ***
b_precip               -4.61118    0.35895 -12.846  < 2e-16 ***
b_snow                 -1.00966    0.40885  -2.470    0.014 *
weather_scurve_center   7.98056    0.85320   9.354  < 2e-16 ***
weather_scurve_width    6.69644    0.57898  11.566  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1033 on 358 degrees of freedom

Number of iterations to convergence: 15
Achieved convergence tolerance: 1.49e-08
```

In terms of goodness of fit, the model's root mean square error (RMSE) was 1,023 and the residuals were fairly normally distributed.



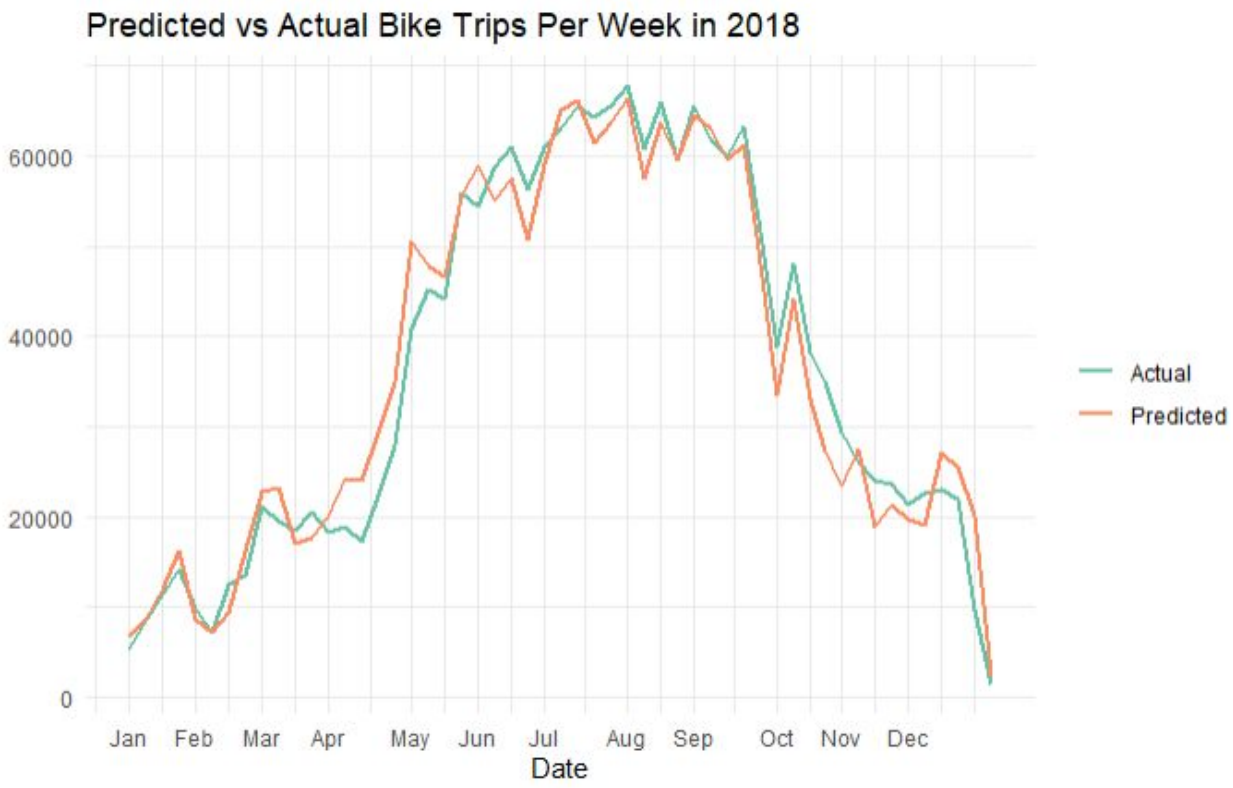
The scatterplot below compares the actual number of trips per day to the predicted trips per day. The residuals seem to have less variability on days with the highest and lowest number of trips. .

There were some outlier values visible in the residuals plot.that could have impacted the model performance. For example, the maximum number of trips was on June 20, 2018, with 13,303 trips. Bike share offered free trips on Wednesdays for the month of June in 2018. And June 20th was one of the free trip days[5]. Free Wednesdays seems to be a semi-regular promotion, so perhaps adding a correction factor for the promotions could improve the models accuracy.



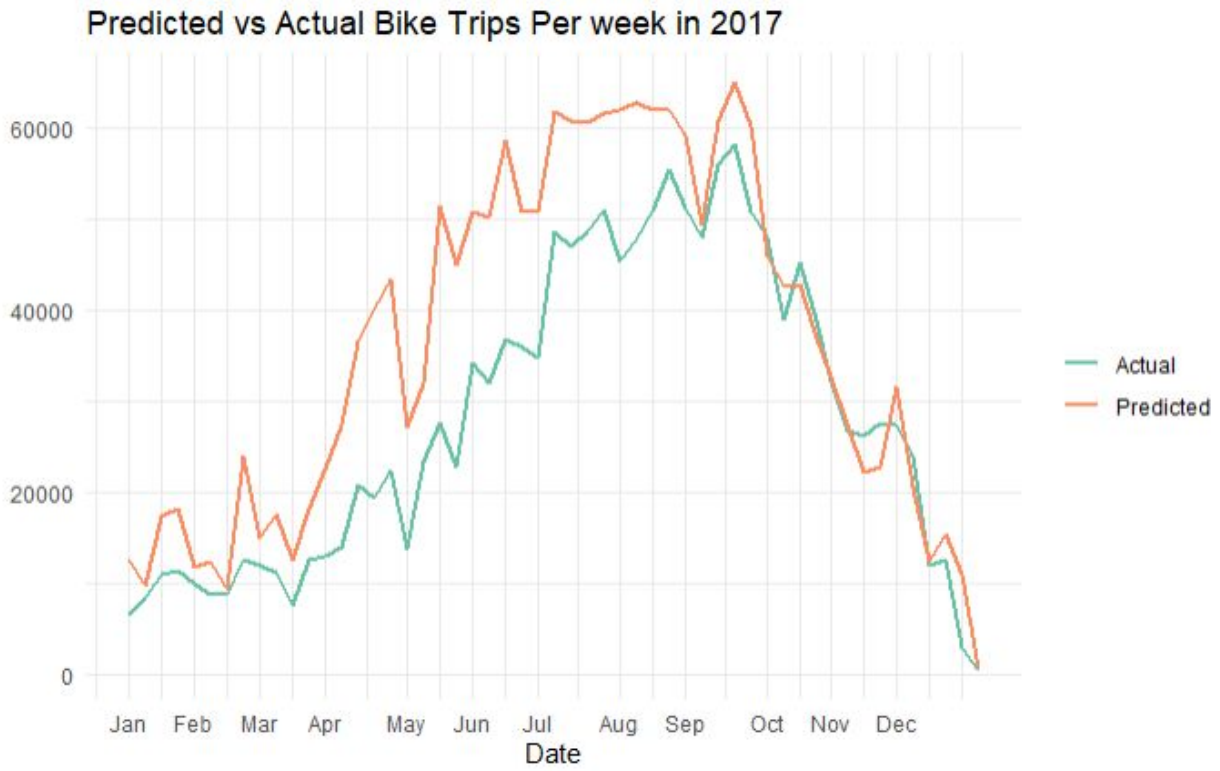
Another way to look at the model performance is over time. The graph below shows the actual number of trips per week compared to the predicted number of trips per week. The model over predicts a little in the off-peak season (e.g., December and April) and underpredicts during peak season (e.g., June, July & August).

[5] https://bikesharetoronto.com/news/june-2018-at-bike-share-toronto/



A better test would be to use the model to predict ridership for other years. Data for 2018 and 2019 are not yet available on Toronto's Open Data Portal, so I tested the model against Bike Share's 2017 data. The model predicted the annual ridership trend fairly well. But it over-predicts the number of trips per week from January to September. This is likely because Bike Share increasing their system capacity[6] in August 2017 and has reported that ridership in 2018 increased by 30% over 2017[7]. This is a limitation of using only one year of data to develop a model for a growing system.

[6] https://bikesharetoronto.com/news/bike-share-toronto-august-2017-expansion/
[7] https://bikesharetoronto.com/news/2018-highlights/



Schneider handles a similar capacity change in NY by using an expansion parameter in the model baseline. To improve the Toronto model, I could introduce an annual expansion parameter. Bike Share Toronto has had station, bike and user growth from 2015 to 2019[8]. And as Toronto opens up, post COVID-19, the Bike Share system may see further user increases as people seek out socially distanced transit options.

Conclusions

The number of Bike Share trips in Toronto follow predictable patterns for different types of users and times of year. Similar to Montreal and New York, Bike Share Toronto is used most often by locals to get to and from work. Trips peak during rush hour during the work week and on weekends the system is used less.

The weather impacts bike share use. Ridership peaks during spring and summer months. There are about ⅓ as many users in March, April and November (i.e., late winter and fall). The system is used the least in the December, January and February, which are the coldest and snowiest months of the year and a time when a lot of people are on vacation from work.

[8] https://en.wikipedia.org/wiki/Bike_Share_Toronto

Mean temperature has the strongest relationship with ridership. And while the amount of rain and snow is related to the number of trips, the relationship may be stronger if the precipitation data were more detailed (e.g., hourly).

The model developed to predict the number of trips based on day of the week, season and weather did a reasonable job of predicting ridership in the last few months of 2017 and throughout 2018. I think the model could be improved by adding in statutory holidays. Those are weekdays when ridership will be more like a weekend. I also think the annual growth of the system and regular promotions (e.g., Free Wednesdays) needs to be included as well to more accurately predict the number of trips.