Nice Riding - How Weather Patterns Affect Minneapolis Cyclists

Introduction.

Minneapolis has been consistently recognized as one of the bike friendliest cities in the continental U.S, and as recently as 2015 was named the #1 bike friendly city in America (Vox). While much of that success has been down to changes in the city's infr astructure (adding bike lanes on major roads), and the already vast and overarching bike/trail system, the introduction of a bicycle ride -sharing program is a huge reason for that as well. It allows people of all ages, athletic ability, and income levels to enjoy biking around the city by providing affordable bike rentals and eliminate the need to invest in their own bike..

Here in the Twin Cities of Minneapolis & St. Paul, MN, we have the bike -sharing nonprofit company Nice Ride MN that was introduced in 2010. Customers can rent bikes at stations that are scattered throughout both cities. Customers can then bike around to their preferred location and return their bike at any other station, providing there's an empty dock for it. However, not all However , there isn't a year round usage of this system; Nice Ride MN shutters its doors once daylight savings time ends to avoid Minnesota's harsh winters.

As avid bikers during the spring/summer/fall months, we did notice that our own riding patterns dropped off once the days grew shorter and the temperatures began dropped below freezing. We decided to look at **if** and **how** ridership in the Twin Cities is affected by the weather patterns and changes throughout the year; does Nice Ride MN have any opportunity to see continual usage of its bikes if the weather goes below a certain point?

Related Work

Nice Ride MN provides public access to their historical data, so many people have done their own investigation into how usage has changed. Most projects focus on the system as a whole whereas our project concentrates on defining usage trends among existing docks.

Springboard Foundations of Data Science Capstone Project Final Report: Predicting Nice Ride MN Bikeshare Volume

http://www.rpubs.com/tonytusharjr/NiceRideMNFinal

The goal of this research is to combine Nice Ride data and weather data in order to predict overall ridership for the system as a whole. The author ultimately concludes that there are meaningful relationships between overall ridership and several factors, such as day of the week, precipitation and type of rider (member vs casual rider). This work differs from our research in that it assesses the Nice Ride system as a whole by clustering the docks into eight groups based on mean distance from other docks. There is no analysis related to individual stations.

Map Monday: Minneapolis Bicycling / Nice Ride Ridership Mashup

https://streets.mn/2015/09/28/map-monday-minneapolis-bicycling-nice-ride-ridership-mashup/#lightbox/2/

This project's goal was to combine information from a 2010 survey on Minneapolis bike ridership with geo-locations of Nice Ride docks in order to make recommendations on where Nice Ride could successfully expand its footprint. The similarity to our work is limited to the graphic representation of the Nice Ride docks and their relative size/usage. The central theme, however, is based on expanding the scope of docks while our research concentrates on defining usage trends among existing docks.

NACTO: If You Want Bike-Share to Succeed, Put Stations Close Together

https://usa.streetsblog.org/2015/04/29/nacto-if-you-want-bike-share-to-succeed-put-stations-close-together/

This 2015 article shows us that one of the most telling traits of a successful bike share system implementation at that time was the density of bike docks. The researchers looked at the number of riders per bike per day for various bike share systems in the United States and Europe and found a correlation between ridership and how closely the docking stations were placed. While very interesting and well-done, this differs from our research in that we are less focused on assessing the success of the Nice Ride system as a whole and more interested in the success or failures associated with each individual station.

Process

For our analysis, we started with basic Excel pivot tables and charts to get an overall summary of the data. This was just to do some initial discovery of the data. Following that, we took our datasets and developed a Python program utilizing a variety of libraries to help us visualize the data and draw conclusions, including **pandas** and **numpy** for the data aggregation and analysis, **matpoltlib & seaborn** for the data visualization, and **folium** for mapping the dock stations.

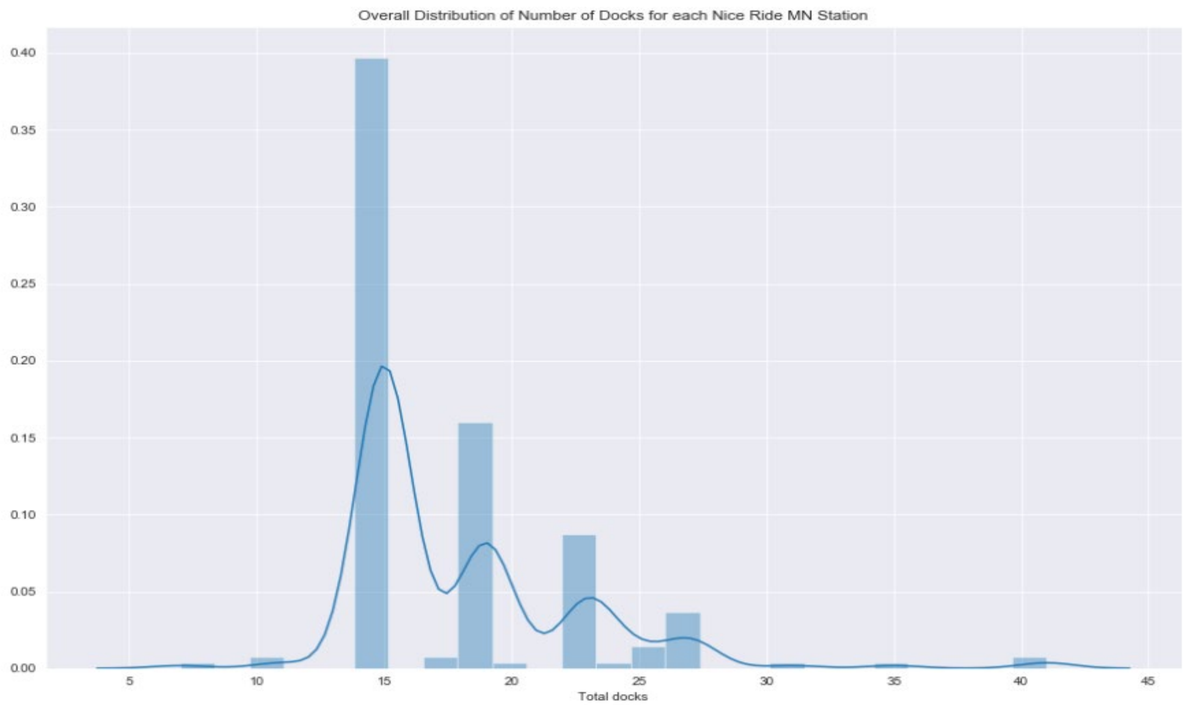
Results

Research Questions:

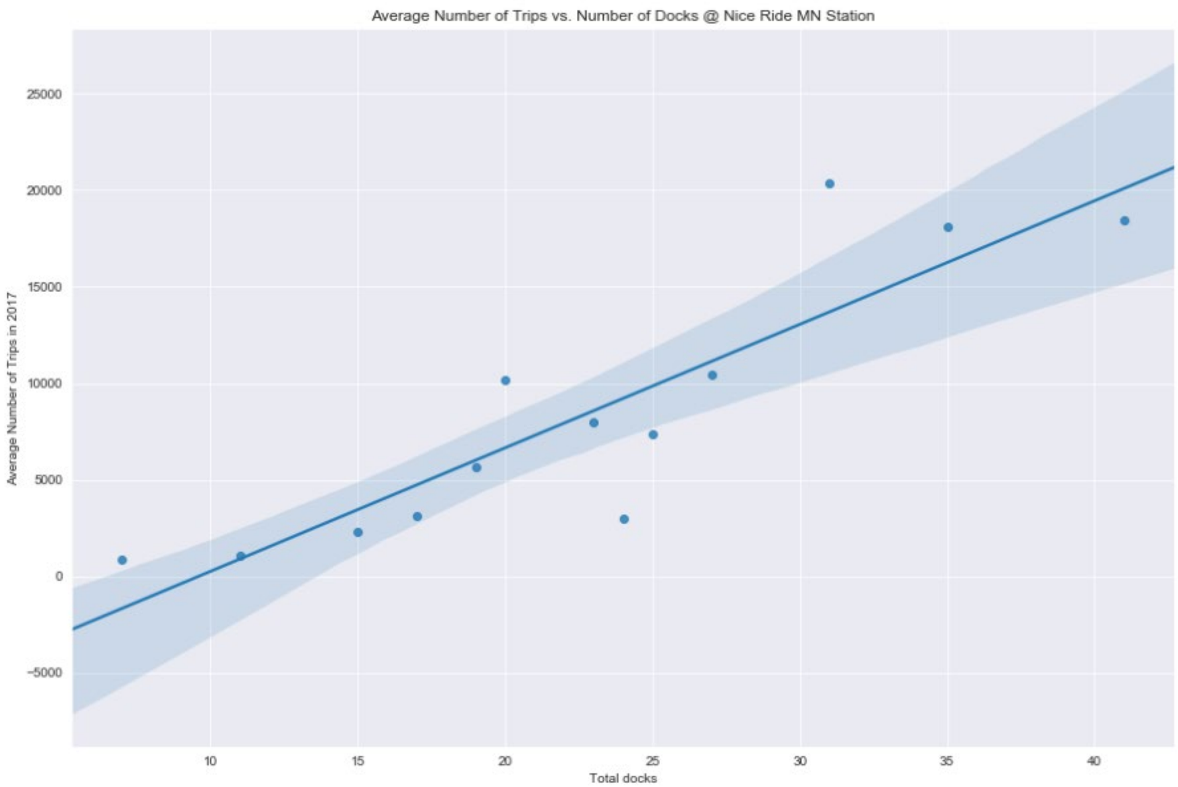
- What do we know about the docking locations and sizes?
- Does the number of docks at each station match the demand at those stations?
- Could the number of docks be more optimally distributed?

To begin with, our first order of business was to get a sense of what the Nice Ride system of docks looks like as a whole. The chart below breaks out the stations by the number of bike docks available (size) and shows us what percent of the Nice Ride system utilizes stations of that size. We see that almost 40% of the Nice Ride system uses stations that can accommodate 15 bikes.

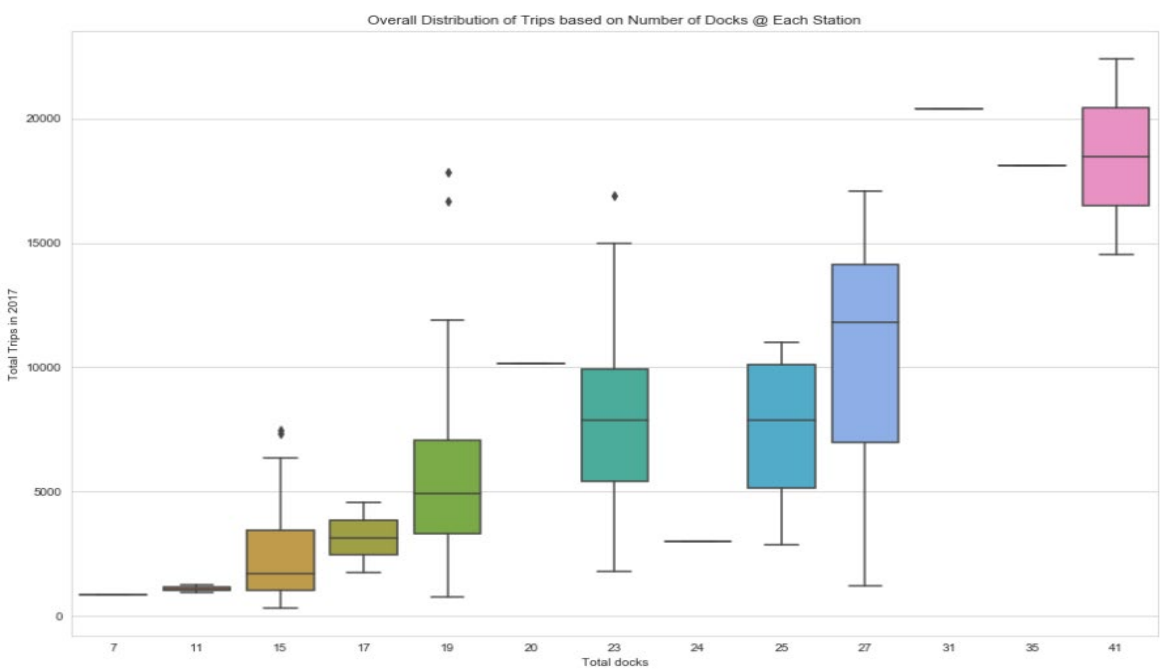
Plotting Distribution of Total # of Docks per Station



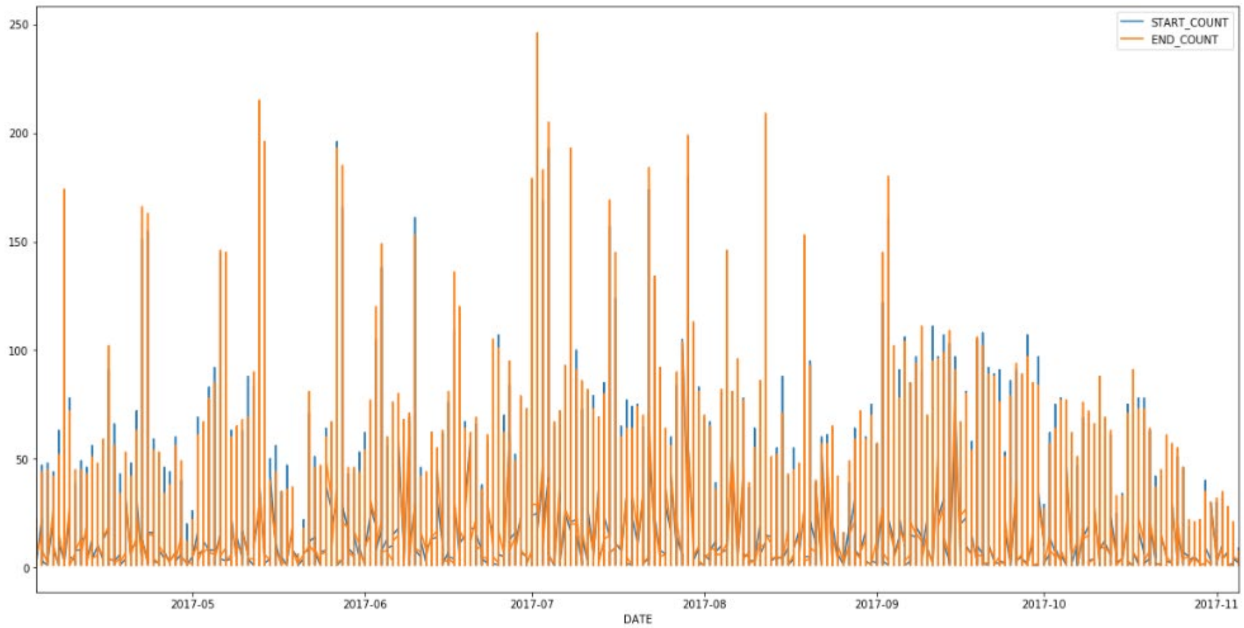
Measuring Number of Docks vs. Total Trips in 2017



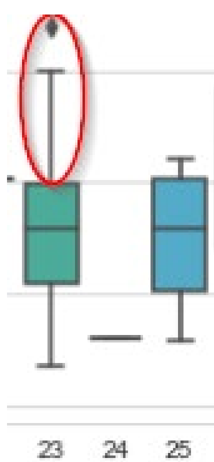
Total Trips in 2017 vs. Total Docks



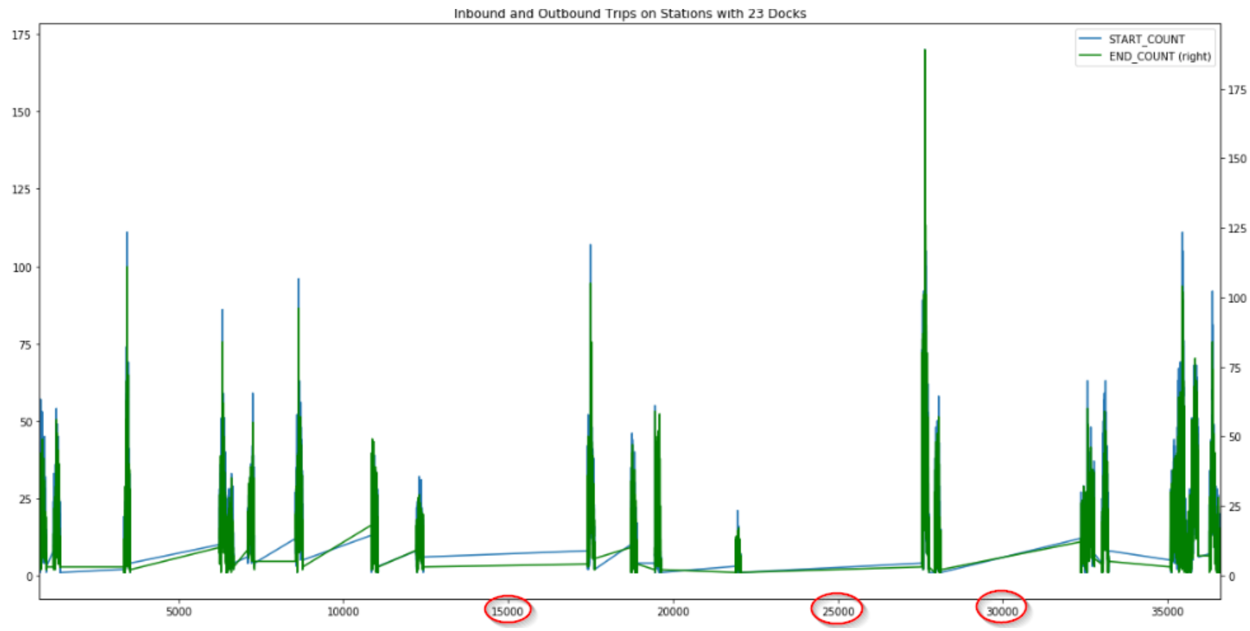
There are a number of charts that indicate the proper distribution of stations and bikes, especially on the high demand zones. The chart above indicates a linear relationship between trips and docks by number of dock stations. The company is addressing the demand in certain areas by increasing the number of docks and bikes available. The relation is positive so we can observe a uniform increase to the right. For the most part, Nice Ride distributes bikes and docks based demand per stations with no major oversights. The chart below shows the line blue, most of the year for every station popping over the orange line.



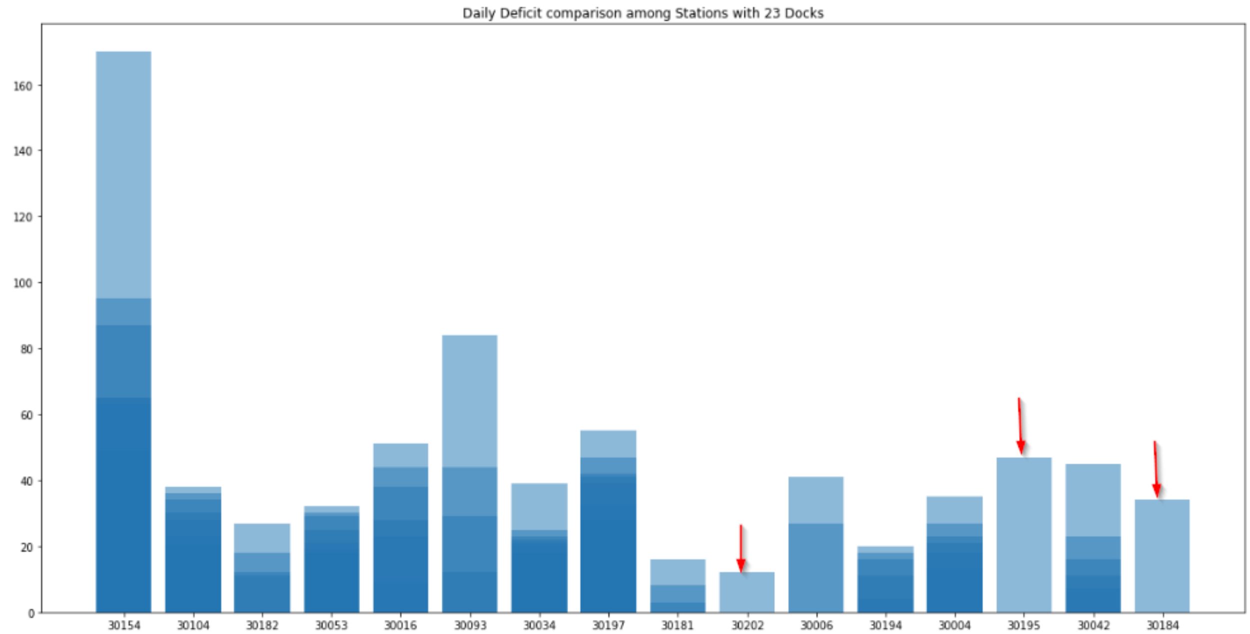
However, if we go back to our boxplot, we can also observe that the range between 23 to 25 doesn't show a significant increase, thus the mean is the same. Nevertheless, stations with 23 docks, shows us some outliers that can be further analyzed.



In the chart below, we identified Stations 3xxx as being one candidate to be analyzed, based on the peaks of green lines that can be seen after the mark 31xx.



When examining the top 100 daily deficits on bike stations with 23 Docks, we got three stations that clearly are isolated events as they had this issue o n just one day.

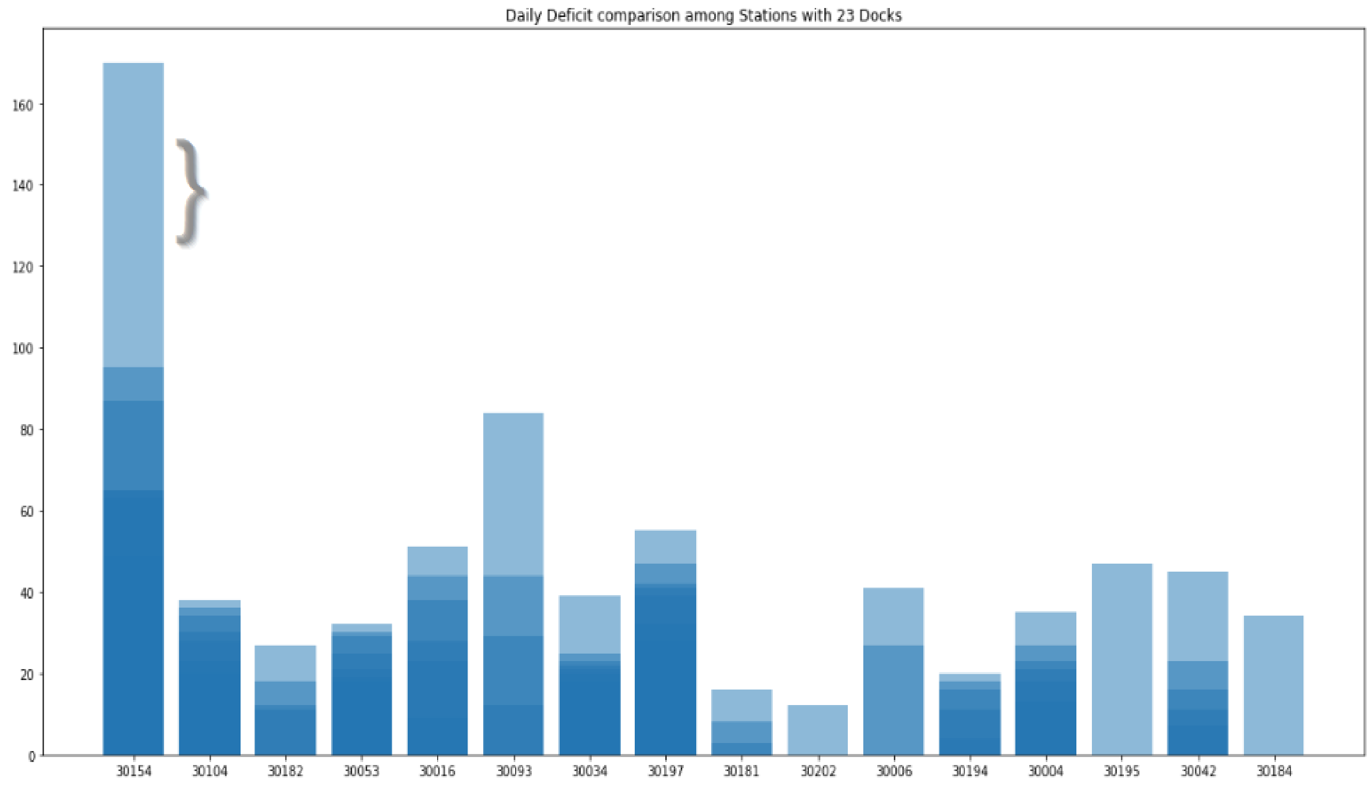


	STATION_NUMBER	DATE	START_SIZE	START_COUNT	END_SIZE	END_COUNT	DIFF
STATION_NUMBER							
30184	30184	2017-07-28	23	34	23	46	-12

	STATION_NUMBER	DATE	START_SIZE	START_COUNT	END_SIZE	END_COUNT	DIFF
STATION_NUMBER							
30202	30202	2017-05-12	23	12	23	27	-15

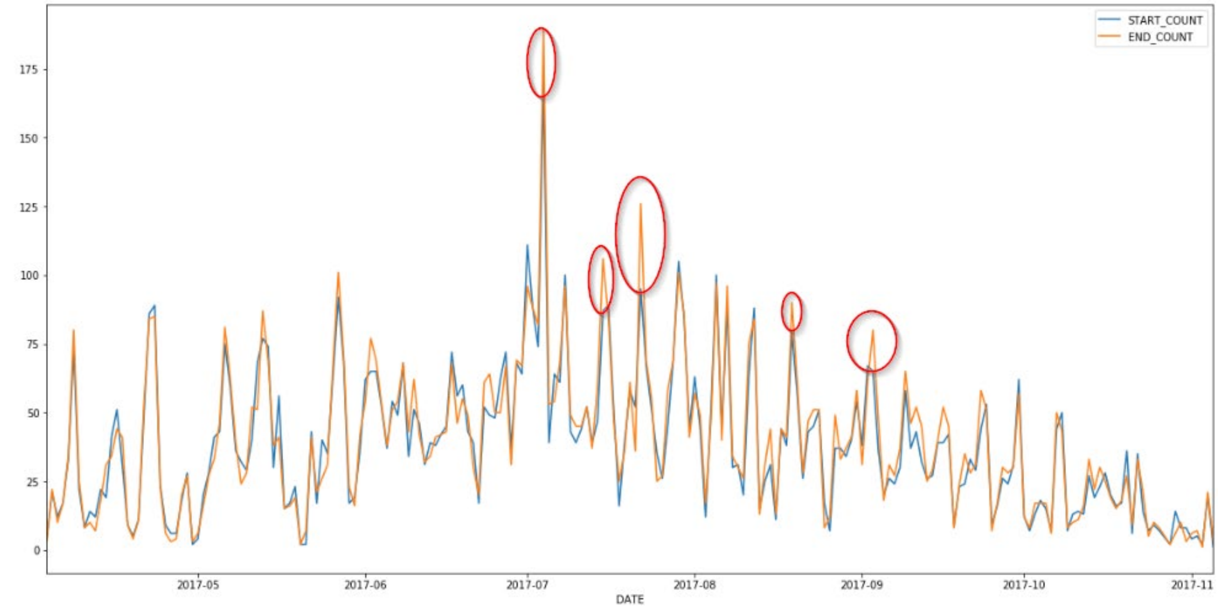
	STATION_NUMBER	DATE	START_SIZE	START_COUNT	END_SIZE	END_COUNT	DIFF
STATION_NUMBER							
30195	30195	2017-10-23	23	47	23	61	-14

An isolated incident of deficit for these three stations is contrary to what we see at Station 30154, which clearly shows a consistent pattern. Please note that the gradient of Blue is determined by how many days its start count was a deficit. More Gradients mean more days.

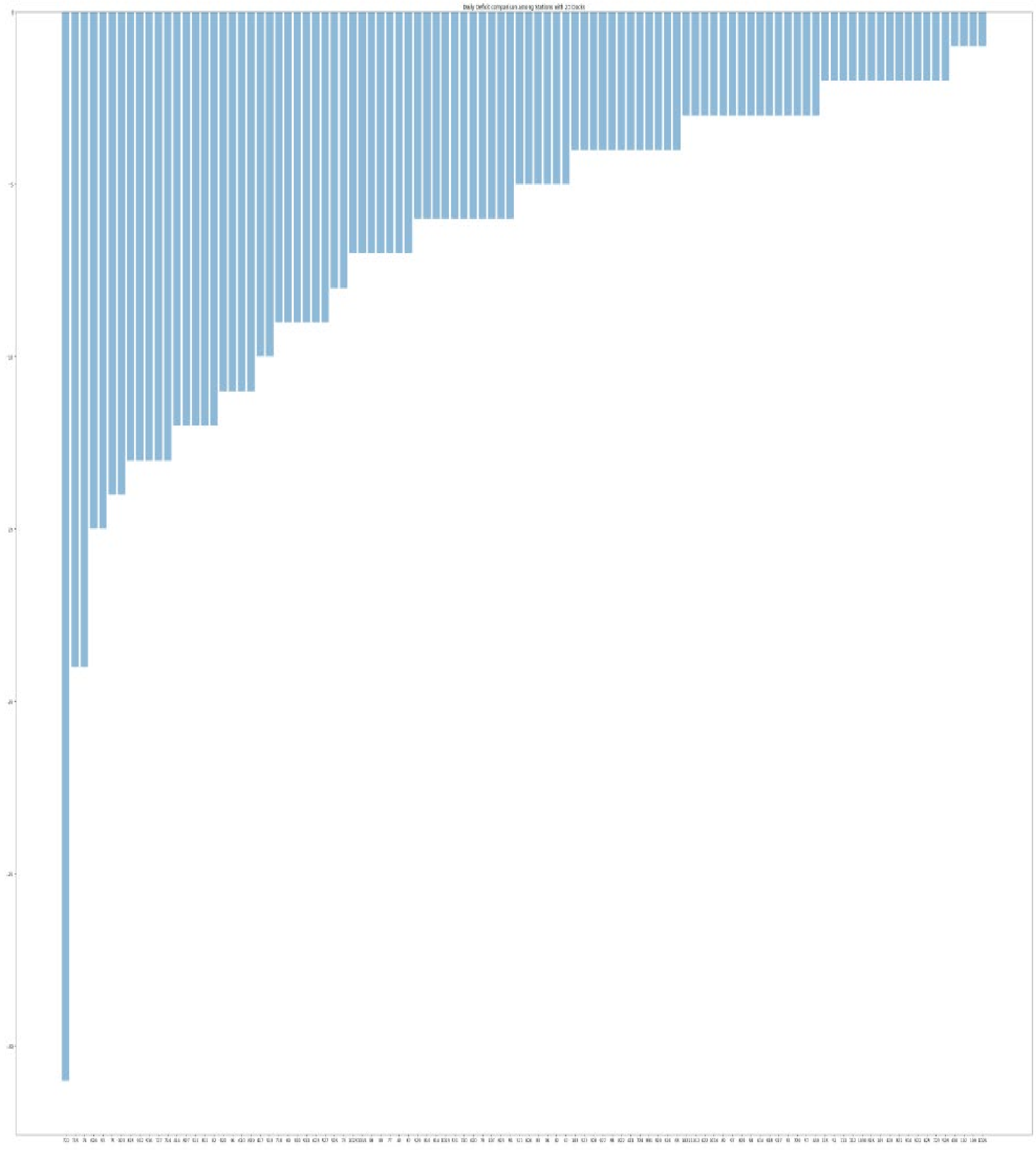


	STATION_NUMBER	DATE	START_SIZE	START_COUNT	END_SIZE	END_COUNT	DIFF
STATION_NUMBER							
30154	30154	2017-07-22	23	95	23	126	-31
30154	30154	2017-07-15	23	87	23	106	-19
30154	30154	2017-07-04	23	170	23	189	-19
30154	30154	2017-06-24	23	49	23	64	-15
30154	30154	2017-09-03	23	65	23	80	-15
30154	30154	2017-07-05	23	39	23	53	-14
30154	30154	2017-09-23	23	44	23	58	-14
30154	30154	2017-08-15	23	31	23	44	-13
30154	30154	2017-09-12	23	32	23	45	-13
30154	30154	2017-09-16	23	39	23	52	-13
30154	30154	2017-07-27	23	45	23	58	-13
30154	30154	2017-07-14	23	47	23	60	-13
30154	30154	2017-04-14	23	19	23	31	-12
30154	30154	2017-08-27	23	37	23	49	-12
30154	30154	2017-05-11	23	40	23	52	-12
30154	30154	2017-08-11	23	63	23	75	-12
30154	30154	2017-06-02	23	65	23	77	-12

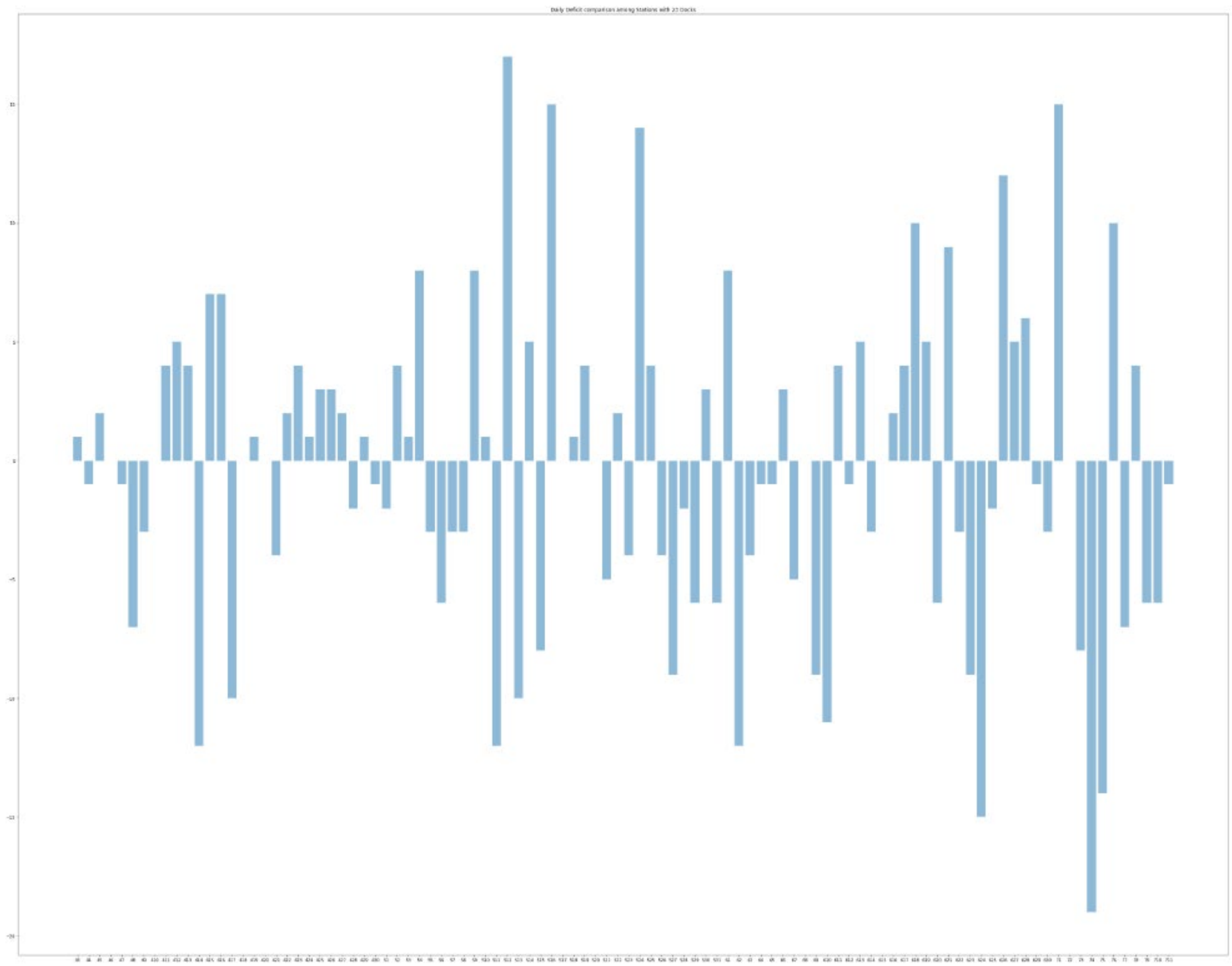
By taking a look at the daily distribution of trips for Station Number 30154 we find several peaks of deficit. No surprise to see July here. We have established that July is the most demanding month.



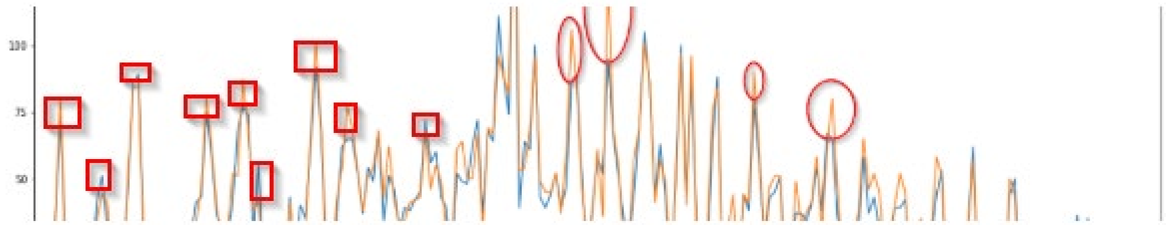
The deficits of Station 30154 increase exponentially and it may imply that this station suffers from high demand more than Nice Ri de was even expecting.

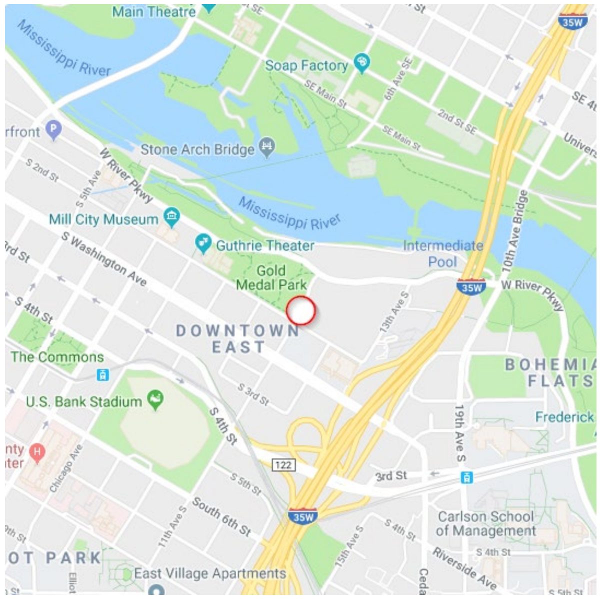


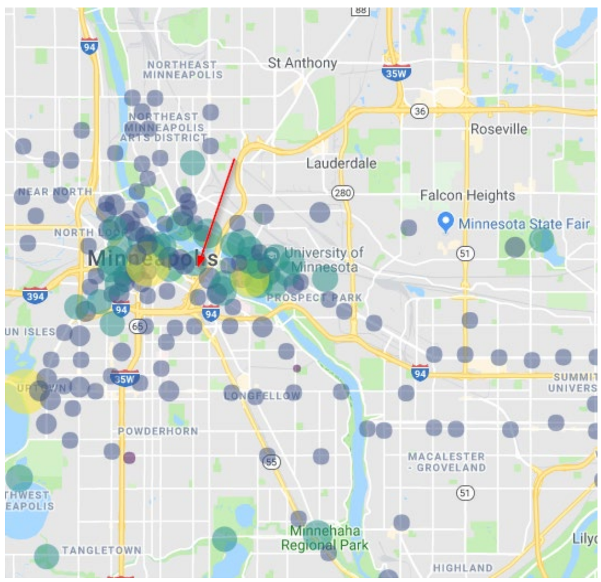
Here is a look at the Daily Deficit. We can clearly see peaks at any day regardless of the month. This is not just a July problem.



Bear in mind that the difference between Daily Deficits and Daily Distribution of trips from above is that Daily Deficits basically plots just the subtraction between Inbound and Outbound trips, so the scale is smaller but the proportion is larger. The values go from -100 to 100. The plot above is representing each rectangle below.





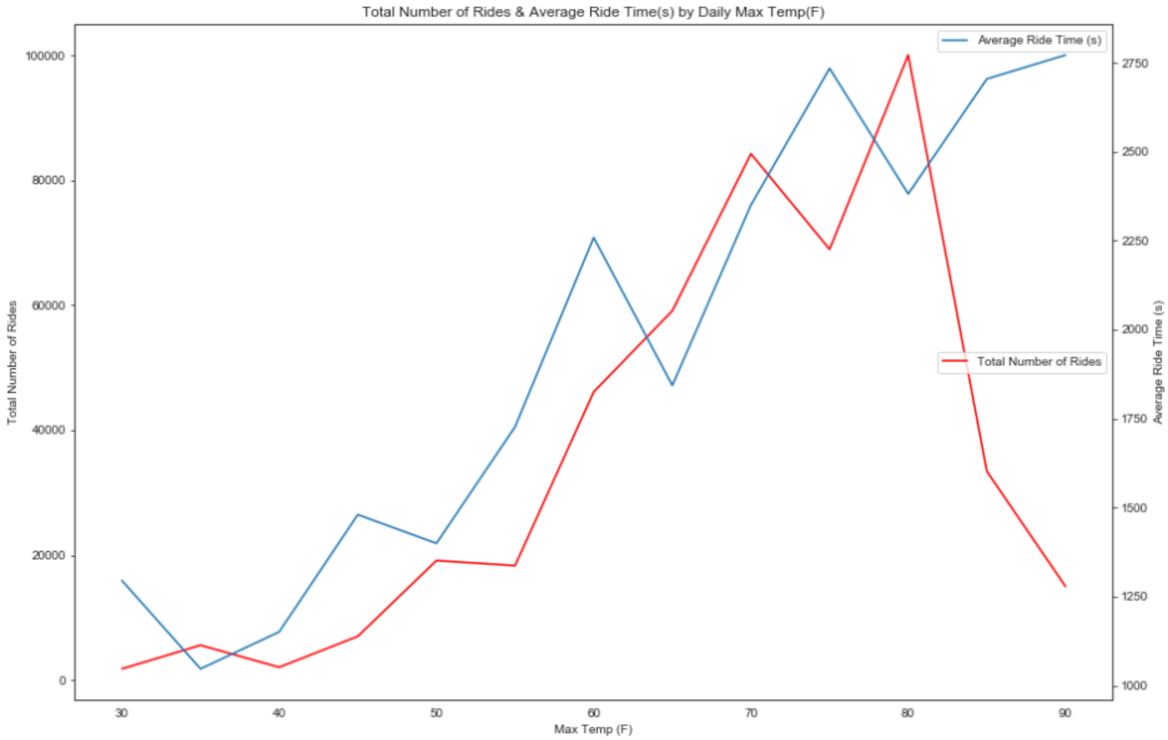


Location is critical here, as it determines the demand and traffic for every station. The Station 30154 is strategically located almost at the core of one of the two Nice Ride Hubs (St. Paul and Minneapolis). Nice Ride can easily address this by bringing some of the bikes from the outskirts of the Twin Cities into this station.

One of the main characteristics of the zone in which the station is established is that it is surrounded by Apartment Complexes. This quickly triggers an alert that we may need to assess population density as part of a later analysis in order to increase accuracy in our conclusions

Research Question : Does riding activity depend on weather?

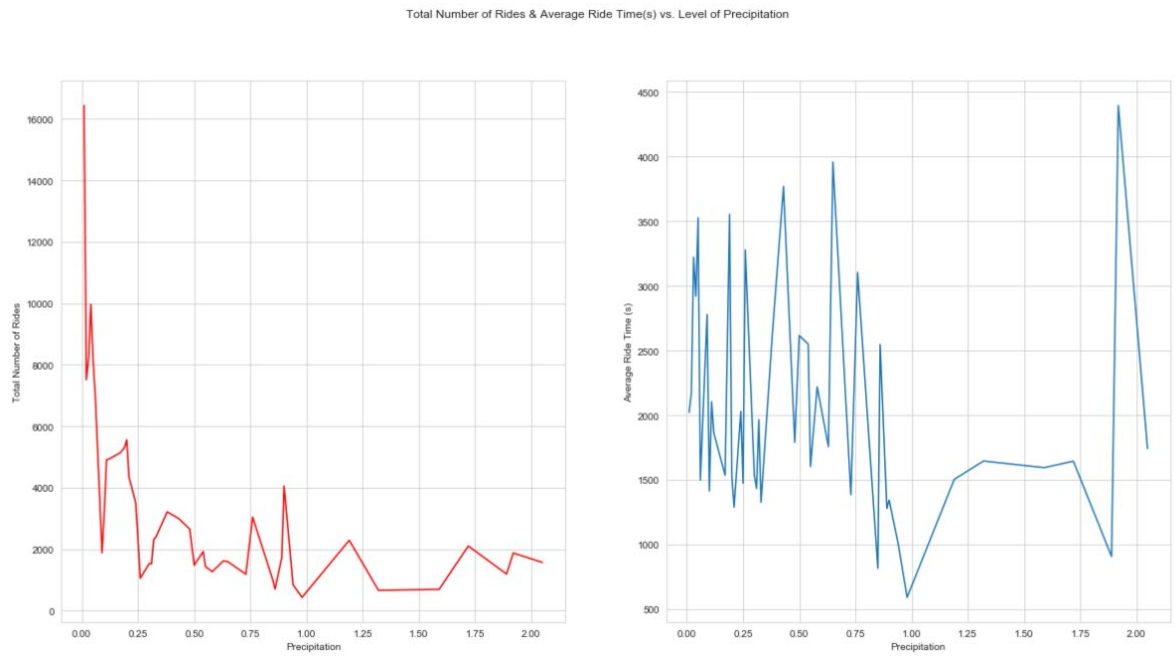
Total A verage Time and Number of Rides by Day Max Temp :



Overall, the changes in Nice Ride bike usage correlates with changes in Daily Max Temperature. As the daily temperature rises, more riders choose to use Nice Ride bikes and the average trip time increases. However, some exceptions to this trend can be seen at the extreme ends of the temperature scale.

First, we note that ridership count decreases on very hot days where the max temperatures exceeds 80 degrees. This drop is so pronounced that by the time max temperatures reach 90 degrees, ridership count dro ps to a similar level as days where the max temperature is only in the 50s. The other notable exception appears when the temperature drops below 35 degrees. Average ride time increases unexpectedly once the daily temperature drops to 35 degrees or lower. O ne might speculate that while bikes are away from the docks at these low temperatures, the time spent actually riding them may be lower than the time they spend parked outside a warm location. Qualitative research may be helpful in exploring this further.

Total Number of Rides & Average Ride Time by Precipitation Levels



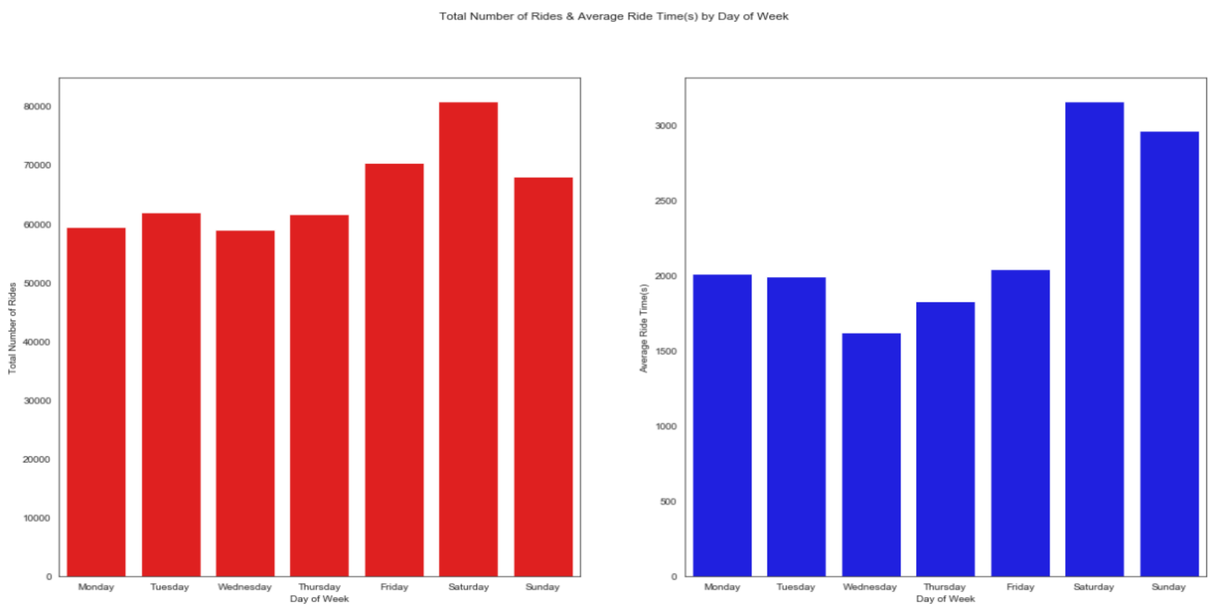
Number of Riders and Avg Ride Time have a different relationship with precipitation. As expected, ridership generally decreases as the level of precipitation increases. That said, Avg Ride Time does not appear to correlate as nicely.

We see unexpected spikes in Avg Ride Time throughout all periods of precipitation. This may be similar to the trend we see in increased Avg Ride Time when the temperature drops below 35 degrees in that rider s may keep their bikes away from the docks for longer, yet spend less time riding in favor of shelter from the elements. In other words, do riders park somewhere dry and wait for the rain to pass? More qualitative research may be useful if this is an area of interest.

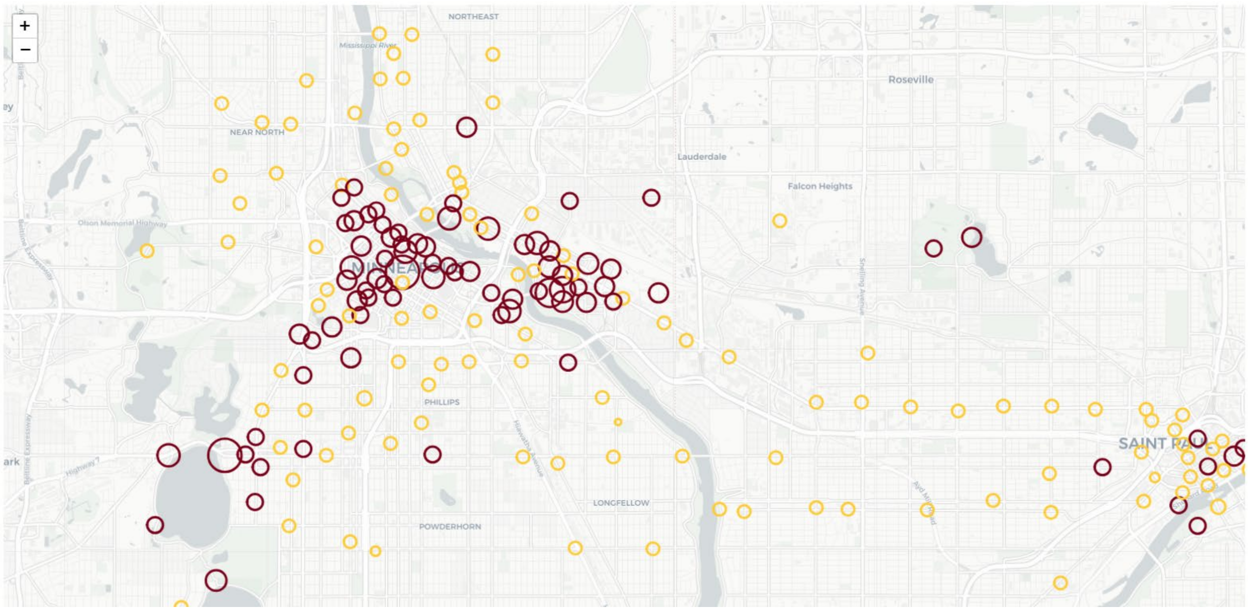
Research Questions :

- How does ridership activity change based on other factors such as day of the week, membership status, location and time of day?

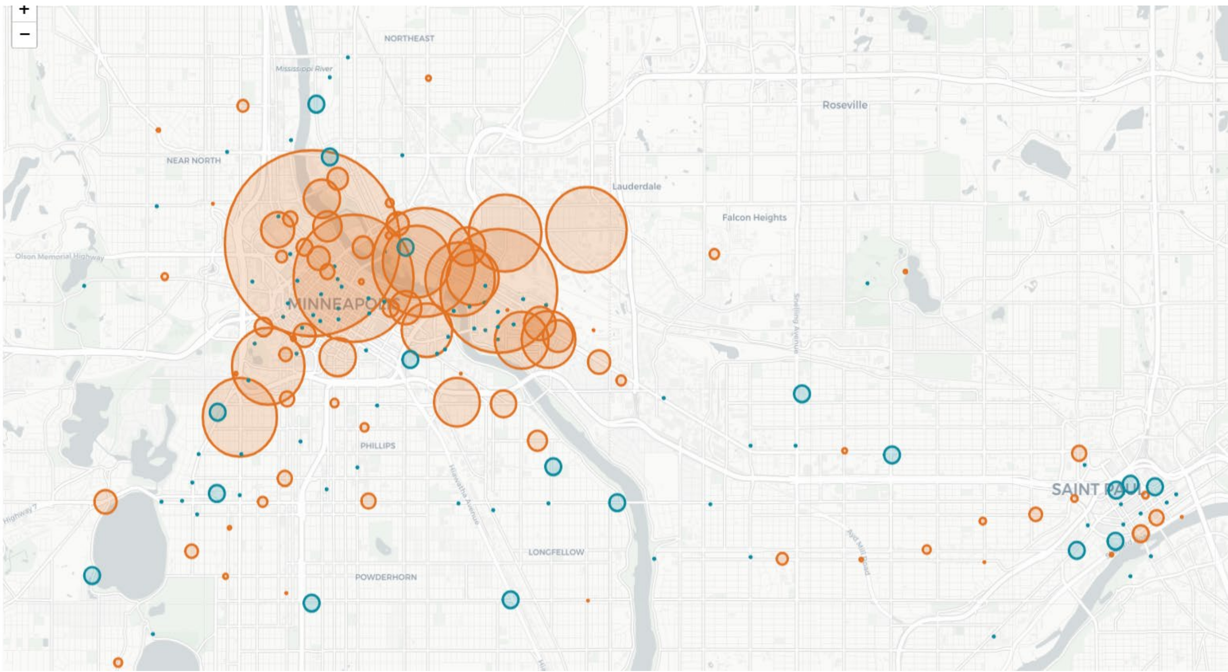
Total Number of Rides & Average Ride Time by Day of Week



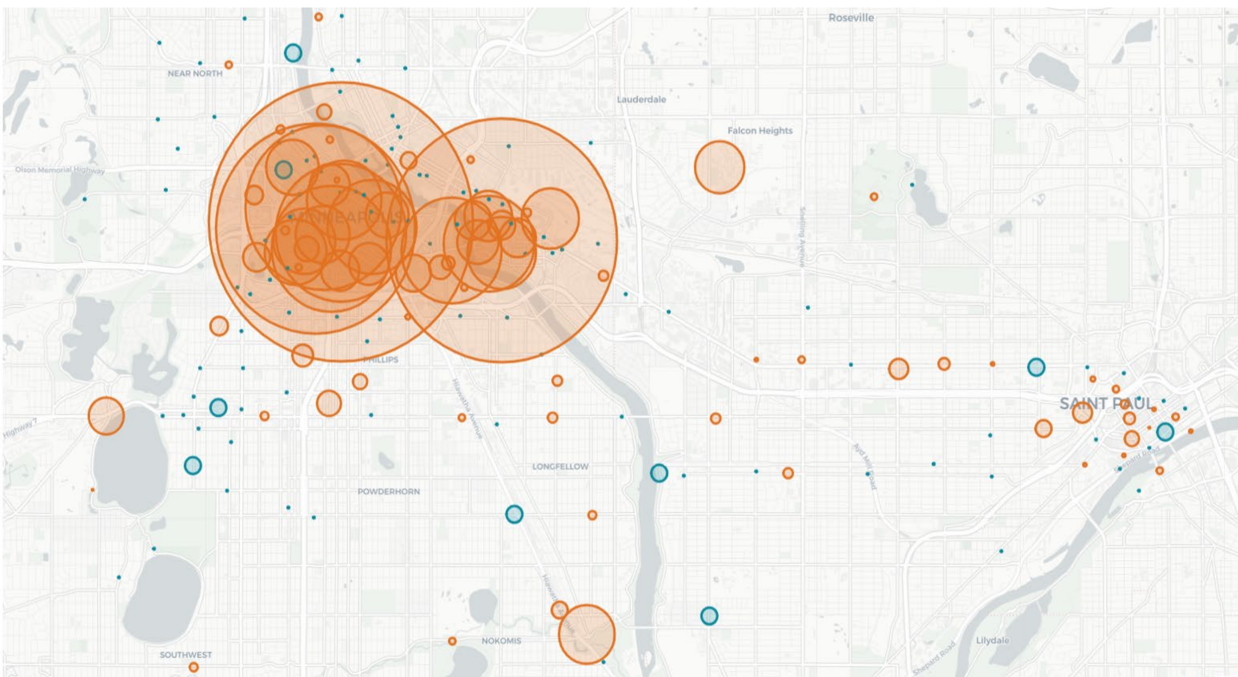
Overall Distribution of Nice Ride Locations in Twin Cities

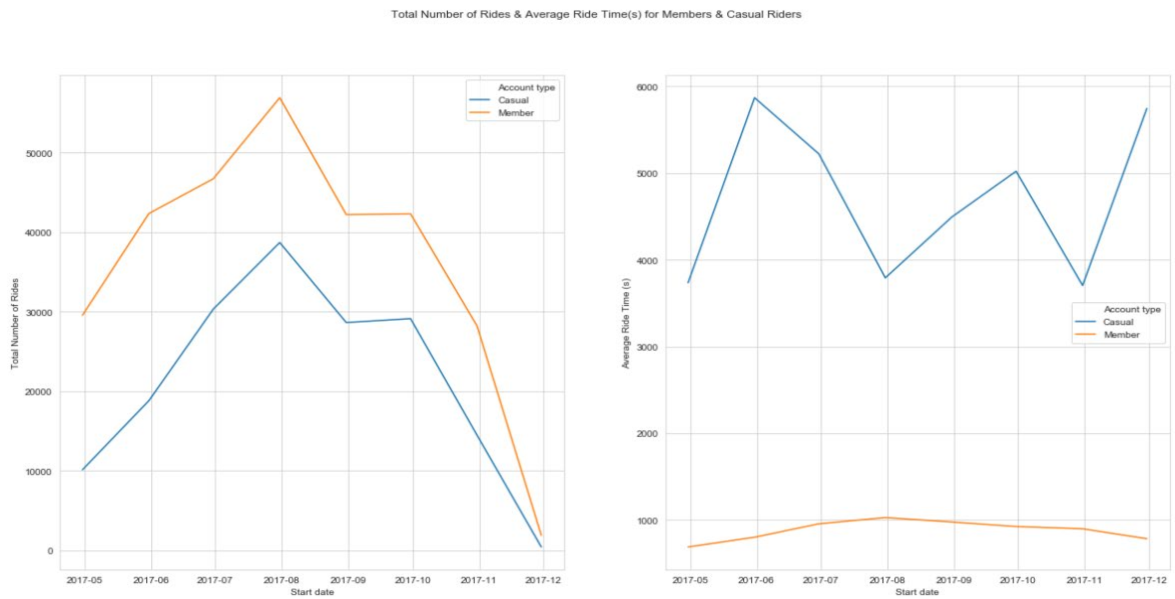


Observed patterns for 9AM ridership



Observed patterns for 5PM ridership





Conclusions

Does the number of docks at each station match the demand at those stations? Could the number of docks be more optimally distributed?

Is there a relationship between riding activity and weather?

Generally speaking, there is reliable evidence of a somewhat expected relationship between weather and ridership trends. The descriptive statistics show that fewer people use Nice Ride bikes when the weather is rainy and/or cold. There were, however, a few exceptions that may warrant further exploration. Why doesn't the average ride time decrease on days when it is raining? Also, why does the average ride time actually increase when the temperature falls below 35 degrees? Both of these trends may benefit from additional qualitative research that focuses on learning more about how the bikes are used when they leave the docks. We may also want to consider a more focused and granular assessment of usage during times when precipitation is actually present. Our current framework relies on daily precipitation data and we shouldn't assume that it was raining for the entire day. Perhaps we may see more useful trends in average ride time when viewing through that lens,

Ultimately, the relationship between ridership and weather can be leveraged to ensure the appropriate supply of bikes is present at high-demand areas when the weather forecast calls for warm, clear days. Likewise, costs related to shifting bikes to high-demand locations can be cut back when the forecast calls for extended rain or cold. These small advantages will help to maximize customer satisfaction, increase revenue and reduce cost s.

At which stations would it be optimal to place ads for Nice Ride membership?

The other interesting thing was the maps @ 9AM and 5PM, along with the members vs casuals average length of trips... this really points to members more than likely using Nice Ride to commute to/from work and home, and thus needing to get to a certain place at a certain time. And that it's really really heavily populated in downtown MN and the U of M campus. So probably concentrating more bikes there would make sense.

Informal Reflection

We only have data from 2017. A better analysis would include time dimensionality, although we have multiple days, we only have one sample of a day in our data set, by adding multiple years, we would be able to compare similar days across years and evaluate the tendencies.

Our work related to precipitation was completed using a daily interval. That analysis may be more telling if precipitation could be reviewed on an hourly basis and paired with ridership metrics during those same hours. We may not be capturing the true relationship between usage during times of inclement weather in the current framework.

We don't have population density information, this could benefit us to assess how much demand can we expect and it would also make our analysis more accurate in terms of Bike Distribution.

One of our discussion points centered around whether to assign the weather data to the start ("Start date") or the end ("End date") of the ride; after awhile we realized that if a user grabbed a bike, regardless of the weather that day, they obviously needed to get somewhere. The logical choice at that point is to use the start date and make the observations based on what the rider was thinking when they first undocked a bike.

When building a fully fledged out analysis in an interactive notebook, one of the key pieces for readability and making sense of what is happening in the background is updating column names so that they make sense, not only to the programmer or programmers involved in a project, but also to a user who might be looking at this for the first time. General slack attention and laziness column naming affects the code because references to old/ get lost and you end up with the wrong data

Another key piece of learning, as the analysis grew more and more complex, and different pieces were being investigated at different times, was the importance of slicing the data for a specific subset needed for each data visualization/graph. This was really critical when doing aggregations across time, or by a specific group, or by specific temperature ranges; building it piece by piece made it much easier to paint a cohesive picture.

When we started the analysis, we did some basic pivot tables and charts using Excel, just to make sure that there weren't any strange anomalies and to see if we could make sense of the data. The biggest coding challenges came from aggregating the results correctly, based on our initial observations using Excel; sometimes we could not get a one-to-one match, based on the different aggregation rules being used by Excel vs pandas, and sometimes led to frustration in getting the data to line up to what we wanted to present.

Another tremendous challenge came with the time series analyses. Although we only had about eight months worth of data to observe, converting our date fields to date time objects/indices was quite a painful experience, and had to be done multiple times because of the differing formats in each data set. We noticed that the built in pandas to_datetime functionality ran *incredibly* slow when a format was not specified, even for about 460,000 rows of data; that was a huge lesson learned. Wrangling date/time information is one of the biggest challenges for any programmer, and for future work on this project, code optimization will be to figure a format to do this better. Also, remembering that time of day when the hour is in military time is not the same for morning hours as it is for afternoon hours (so 5PM is actually 17 and not 5!)

And reiterating a lesson learned above, rememberin g to use ***unique*** variable names when creating copies or new versions of an existing dataframe, since the notebook can be many lines of code, if you forget a variable name or reuse it, it can give you bad or unexpected results! In this regard, print stateme nts are hugely important. Placing a print statement after every change to a column name/type, and double checking to make sure the code is working as expected will save you headaches when revisiting the code after a few days.

Appendix

Datasets :

Nice Ride Minnesota. (2018, November). *System Data*. Retrieved from https://www.niceridemn.com/system-data

Nice_Ride_2017_Station_Locations.csv contains data related to each individual dock station, including the number of bike docks, the latitudinal and longitudinal coordinates (helpful for locating them on a map), and the name of each station. Each row contains data for one specific station, and there is a total of 202 overall.

Nice_ride_trip_history_2017 _season.csv contains data for each *individual* bike trip in 2017; this includes account type of renter (member/non-member), duration of the trip, start/end stations, and start & end times/dates. Each row contains data for one trip/rental, and there is a total of 460,718 trips in the dataset.

WeatherDailyMinneapolis2017.csv contains data for the daily weather patterns for Minneapolis/St. Paul; this includes the daily high and low temperature (Fahrenheit), and level of precipitation (mm). Each row contains data for one day in 2017, and there are 365 rows overall.

References

Belluz, Julie. (2015, December 3). *Why Minneapolis was voted the most bike-friendly city in America*. Retrieved from https://www.vox.com/2015/12/3/9843562/minneapolis-bike-friendly

Anzilotti, Eillie. (2018, September 17). *Minneapolis would like to cure your dockless bike-share skepticism*. Retreived from https://www.fastcompany.com/90236724/minneapolis-would-like-to-cure-your-dockless-bikeshare-skepticism

Lindeke, Bill. (2015, September 28). *Map Monday: Minneapolis Bicycling / Nice Ride Ridership Mashup*. Retrieved from https://streets.mn/2015/09/28/map-monday-minneapolis-bicycling-nice-ride-ridership-mashup/#lightbox/2/

Schmitt, Angie. (2015, April 29). *NACTO: If You Want Bike Share to Succeed, Put Stations Close Together*. Retrieved from https://usa.streetsblog.org/2015/04/29/nacto-if-you-want-bike-share-to-succeed-put-stations-close-together/