**Case Study 1 – Coursera**

The scenario I choose was working with a bike share company called Cyclistic. My main goal for this was to determine how annual members and casual riders use Cyclistic bikes differently. So, I took the previous 12 months of ride data and put them in a folder on my computer. Right away I noticed something, the size of the documents. The fall, spring and summer months were considerably larger than the winter months. I have reached one conclusion already! The winter months in Chicago…not ideal for riding bikes! The data I wanted to dig up became clear. How do ride types, duration and sheer numbers differ between riders seasonally.

After opening the first month (June 2020) of ride data, I quickly realized there was a lot going on. After sifting through some columns, I decided to make a new column called ride\_length. Where I took the end ride time and subtracted it by the start time to determine ride length. Next, I inserted a new column next to ride\_length, called day\_of\_week. I used the WEEKDAY function to calculate the day of week (1 being Sunday and 7 being Saturday). Honestly the numbers got very confusing to look at, so I created a new tab with the following table below:

|  |  |
| --- | --- |
| 1 | Sunday |
| 2 | Monday |
| 3 | Tuesday |
| 4 | Wednesday |
| 5 | Thursday |
| 6 | Friday |
| 7 | Saturday |

Using the chart above, I created a new column called day\_of\_week (yes, it’s the same name as the other column, but that’s all I could come up with). I created a VLOOKUP function and tied it to this table so I could convert all number weekdays into actual text!

Now that I can finally distinguish a Sunday from an average start time, its time to clean some data! First things first I deleted all latitude and longitude data, it didn’t seem necessary for the data I wanted to dig up. I added a filter to the top row and immediately started deleting any NULL values in ride\_length, start\_station, and end\_station columns. After cleaning the data on all 12 months I made a pivot table for each. They all included various elements to each rider type, number of rides, and on average how long those rides were. This information only made me want to dig deeper, I needed to find out the trends seasonally. It was time to open RStudio. In RStudio I used rbind to combine all data into the four seasons. After that, I exported this data back into XLS files. For each season I created a new Pivot Tables to calculate and visualize the total rides of each type and the average ride times.

I was able to find out that causal riders tend to use the service on the weekends and ride much longer than a member rider. Our member rider friends are more consistent in their riding, their popular days are during the week and average a ride time of under 20 minutes. Breaking the data down by ride also showed some interesting details. Classic bikes are the most popular in the summer months. During the spring, fall and winter electric bikes usage drastically increases (especially for casual riders).

A couple takeaways:

* The better the weather, the more people will ride and the longer they will ride for.
* The worse the weather, riders (especially casual) will shift into an electric bikes and ride for less time. In general, will ride less overall.

Overall, I am very pleased with my findings. This project was very fun to work on and made me think outside the box. We were originally told to combine all data into a full year view, I felt this wasn’t necessary. Leading me to compile and break the data down seasonally. Which gave an accurate representation of how casual and annual members ride differently.

* Kyle Burns