Portfolio Project-Toronto Bike Share Service

EDA for Trip Patterns & User Behavior with R

Hello, I am Laks Malhotra, and I welcome you to my first ever end-to-end Data Analytics project. I recently completed my Google Data analytics Professional Certificate, more on that experience in <this>(link) article.

I submitted this project as a capstone project for my final submission, so anyone can use this as a reference for developing their projects.

Google Data analytics Professional Certificate provides a standard ASK, PREPARE, PROCESS, ANALYZE, SHARE and ACT project lifecycle which helps in understanding the workflow. For practical application, I’d recommend doing multiple projects using a different product or service domains, and then creating your portfolio containing tailored projects.

Improve your data analytics skillset by practicing the following projects:

* <Cyclistic Bike Share>(link)
* <Bellabeat Fitness App>(link)

(**Note:** You can review these practice projects by clicking on their more respective names and it will direct you to the articles)

Then, use the projects as templates for your projects.

Here, in this project, I used the ‘Cyclistic Bike Share’ project as reference material to develop a similar Data Analytics project for ‘Bike Share Toronto’. Let’s continue with our project by downloading the trip data in CSV format and saving the files in a proper location in your working environment.

Now, let’s check for ROCC <LINK> for the dataset. It’s a best practice to ensure data integrity and credible results. Inspect the data and data source against the given parameters:

* **Reliable:** The data is available open source at <https://ckan0.cf.opendata.inter.prod-toronto.ca/en/dataset/bike-share-toronto>(this Link) published by parking authority Toronto, hence is a reliable source.
* **Original:** This data is directly uploaded by the city of Toronto for the public.
* **Comprehensive:** Bike Share Toronto has a similar dataset if we compare it with our template project of Cyclistic bike share.
* **Current:** We will use data from January 2021 till May 2022 to understand the current trends and patterns.
* **Cited:**  Data utilized for this project is available at the given link <https://ckan0.cf.opendata.inter.prod-toronto.ca/en/dataset/bike-share-toronto>

Available for use under <open-government-license-toronto> license.

Now, let’s explore the dataset to understand the data types and records it contains.

Open January 2021 dataset to inspect the table:

View the first few records (rows) in the dataset to understand the attributes associated with each data point:

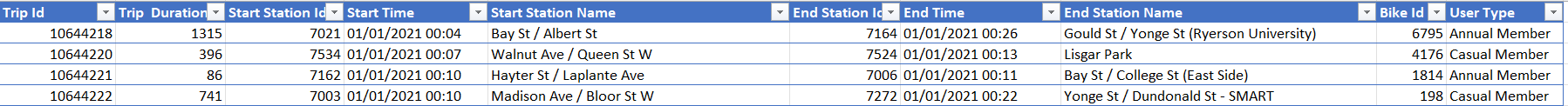


Figure 1. Records

Let’s view the columns and their datatype:

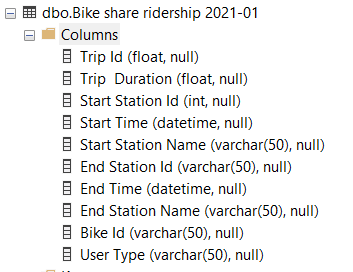


Figure 2. Columns detail

Now that we have ensured the integrity of our data and previewed our datasets, it’s time to upload the data into a workstation and identify some patterns and prominent trends based on the user type, the overall demand of the service, seasonal trends, source, and destination stations trends for our customers using Bike Share services in Toronto.

For a general overview, we will explore:

* Overall growth of the service
* Statistical summary of the Trips taken by the customers.
* Seasonal patterns of bike share service used by the customers.

For this project, we are using the R programming language in RStudio as our working environment.

We will use datasets from January 2021 to May 2022 for our analysis, I’ve done a separate article focusing on Data Cleaning, where my data transformation and cleaning process is explained in greater detail. Check out “**How to do Data Cleaning for your Beginners in R” <link>**

Also, you can find all the code and documents associated with this project on my **GitHub repository.**

**Data Uploading**

Firstly, import our library to the workstation and then upload the datasets on that platform. Load the datasets using the “**read\_csv()**” function and store the data in the respective data frames.

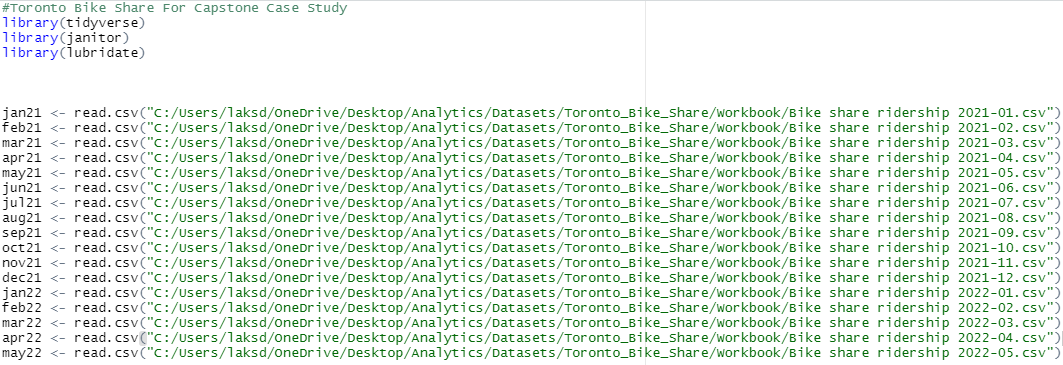


Figure 3. Data Uploading

**Data Preparation**

If we observe no error while uploading our data, then we should jump right into data cleaning and manipulation steps.

If you do encounter some trouble in the primary stages, don’t feel bad, it’s just a part of your learning process. For me, my first few attempts were painstaking, my learning curve required lots of trial and error, which is why I can probably help you with that, check this article “**How to avoid the mistakes that I made during data cleaning in R” <Link>.**

Getting back to our analysis; Let’s combine the multiple datasets into one **workbook**, this will streamline and ease any analytical process in the workflow.

We will use the **“** **rbind()**” function in R to create a compiled workbook.

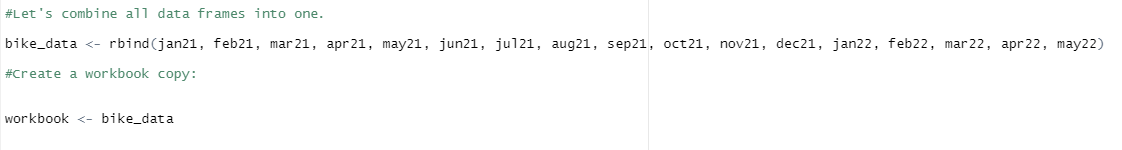


Figure 4. Compiling Workbook

Then preview the data types of the workbook columns by various functions in R

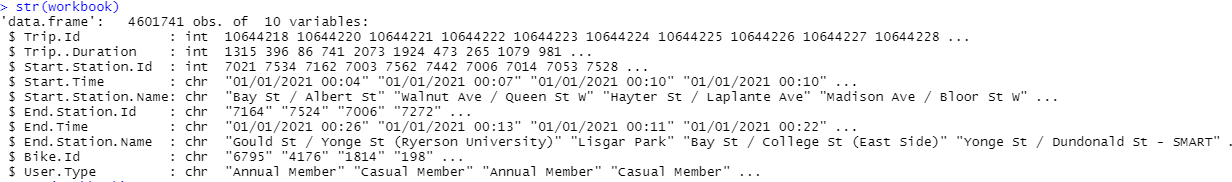


Figure 5. Str function in R

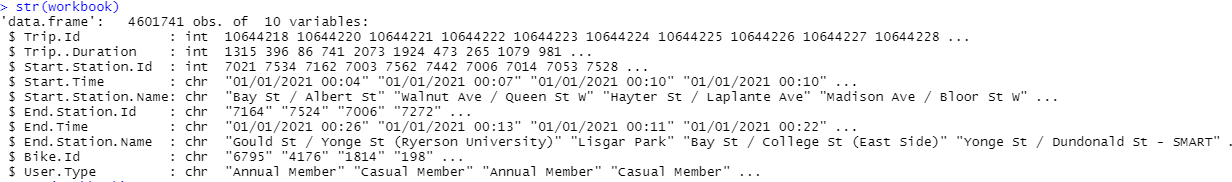


Figure 6. Glimpse function in R

**Data transformation and manipulation**

Let’s convert some relevant columns in their respective data types, to begin with, data transformation and manipulation.

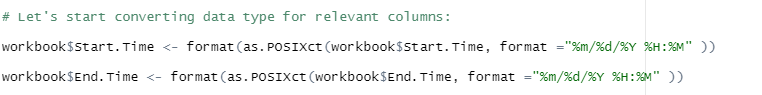


Figure 7. converting Date Time columns

This **Datetime** conversion saves us lots of trouble by standardizing the data types for the columns at the very beginning of the analysis.

Now, let’s add relevant columns to the workbook by adding Ride Duration, day, month, year, and day\_of\_week columns



Figure 8. Adding relevant columns

This **Date Time** columns extraction trick adds an extra layer to our analysis. Check this article, “**How to improve your analysis insights by extracting DATE TIME Columns?**” <link> for more practical applications.

Now, let’s check for any missing values, negative values, and other discrepancies with our dataset.

Finish it off with a new “**workbook cleaned”**, a better version of the workbook for the analysis process.

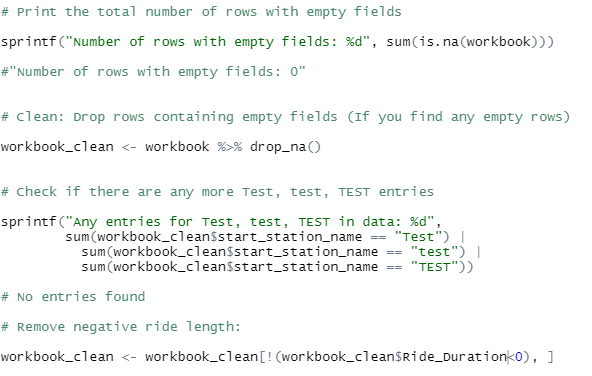


Figure 9. Workbook cleaned

Now that we have completed our ***Data preparation*** stage by transforming and manipulating the data; Let’s get our hands with some Data Analysis!

**Data Analysis**

**What are Bike Share Toronto’s current growth Trends?**

Since January 2021, our analysis shows that Bike Share Toronto had **3,575,182** customers in **2021** and is showing growth trending with **1,026,559** customers till **May 2022.**

I have used this given code to visualize the trend:

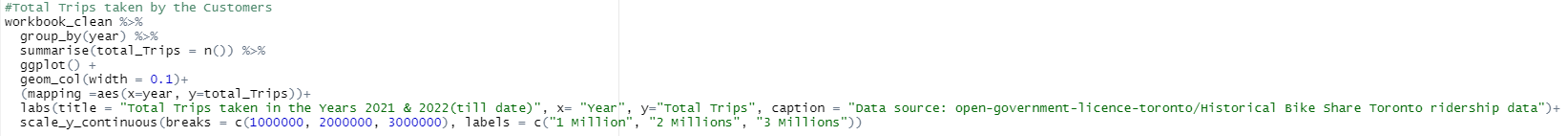


Figure 10. Total Trips code

Here’s the Visual from the code:

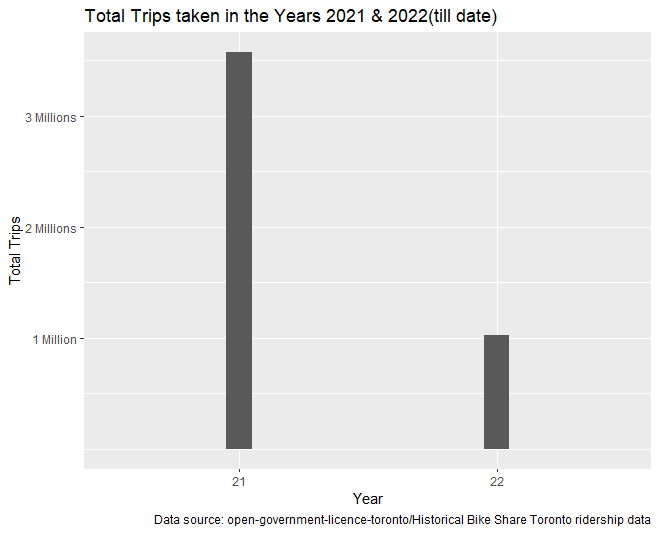


Figure 11. Total Trips Visual

***What are User type Trends & Patterns for Bike Share Toronto?***

**Total Trips Taken**

Bike Share Toronto has two distinctive categories for their customers, annual members with the memberships and casual riders who purchase daily passes. Our analysis shows that in the given time period **more** than **2.59 million** annual members and **2.003 million** casual members have used this Bike Share service.

Here’s the code and visual showing the exact scenario:

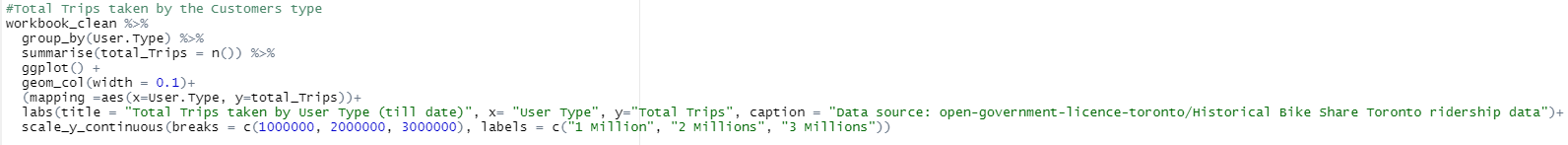


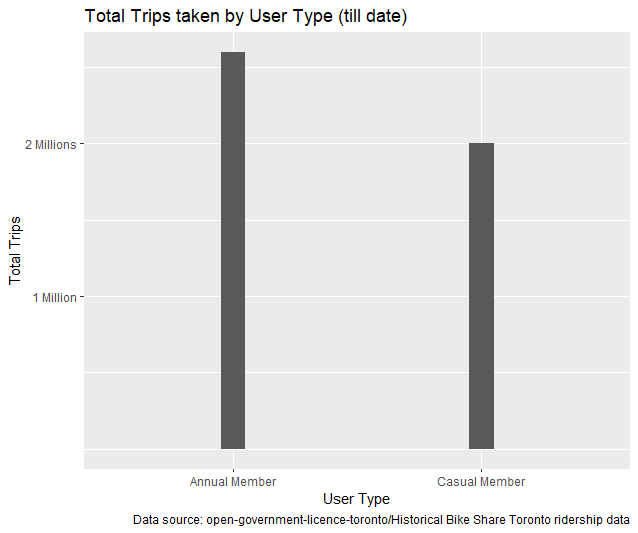
Figure 12. Total trips user type code

Figure 13. Total trips user type visual

**Average Ride Length**

Our analysis shows that Bike Share Toronto’s **annual members** use the service only for **12.6 mins on average** while **casual riders** have generally longer trips with an **average 22.7 mins** of trip durations.

Again, the code and visuals are as follows:

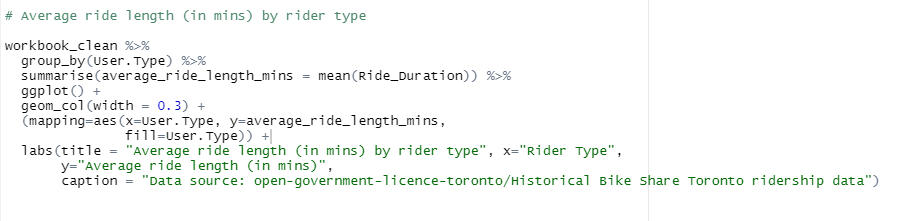


Figure 14. Average Ride Duration by User Type Code

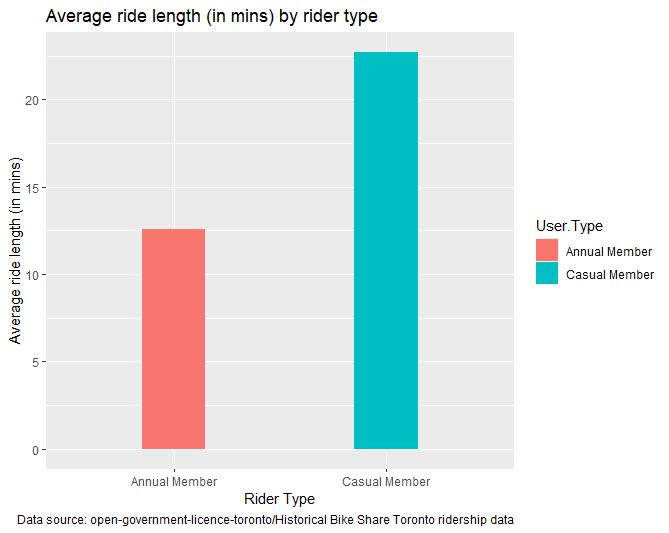


Figure 15. Average Ride Duration by User Type Visual

**Number of Trips made by Customers – Breakdown the user type on the weekly basis**

We can breakdown the daily bike service use by the customer and find out how **annual members** behave different from **casual members**. For this use used the following code:

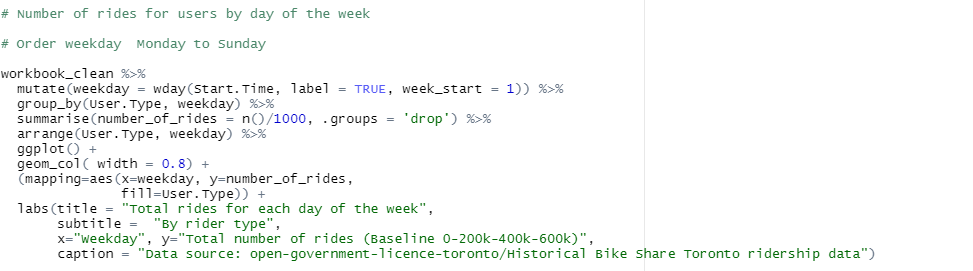


Figure 16. Breakdown total trips- daily use by user type code

**Note:** Always reorder the days like “Monday to Sunday” or any other logic for better aesthetics and understandability.

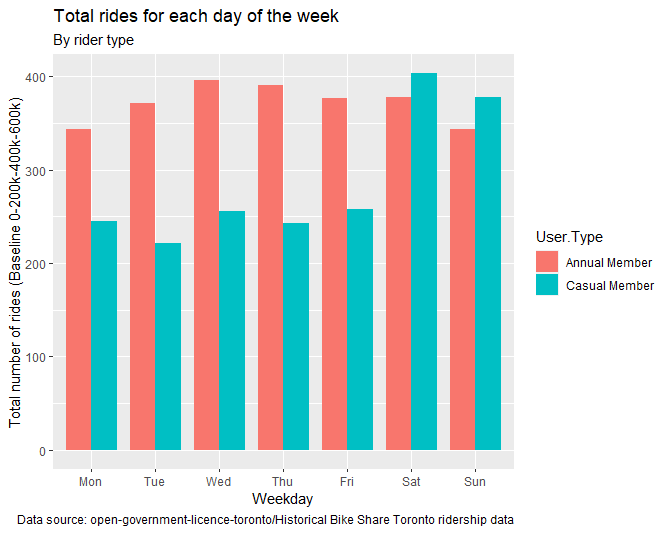


Figure 17. Breakdown total trips- daily use by user type Visual

We can see that all through the week **annual membership owners** are reliable customers providing consistent business and revenue to the parking authority of the Toronto city, with **minimum** **343k** and **maximum 396k** trips throughout the week.

While, on the other hand, **casual** users of the services show consistent but marginally less interest in the services then **annual** members on weekdays **(i.e., Monday to Friday)**, with the total number of trips between a **minimum of 222k** and a **maximum of 256k** trips.

But, on **weekends**, **casual riders** shine, where they counted **403k** rides, and **annual members** counted **378k** rides on **Saturday,** and on **Sunday, 378K Casual riders** took a bike ride as compared to **343k Annual members.**

That was a good insight to be put under consideration.

**Average Trip Duration of Customers – Breakdown the user type on the weekly basis**

Our analysis shows that the **annual member** rides a **minimum average of 12 mins** and a **maximum average of 13.6 mins** throughout the week, whereas the **casual member** rides comparatively longer than the annual membership, with a **minimum average of 13.6 mins** and a **maximum average of 26.0 mins**.

Here, is the code and visual associated with this:

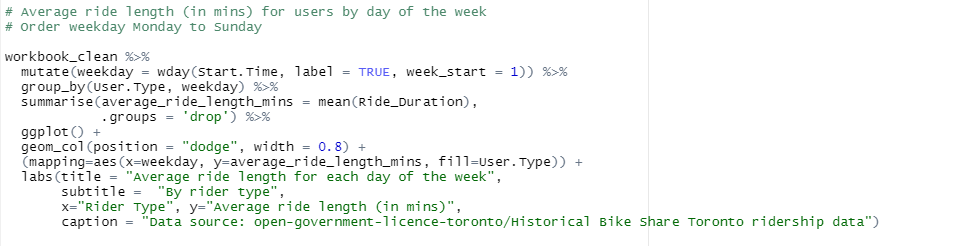


Figure 18. Average ride length for each day of the week code

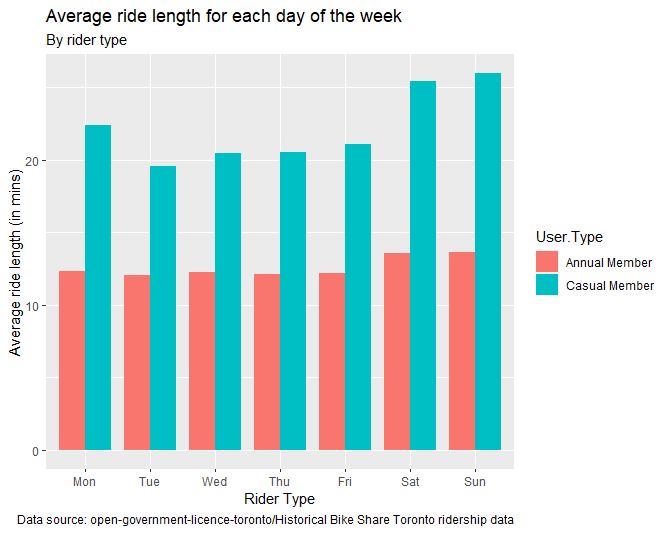


Figure 19Average ride length for each day of the week visual.

***Seasonal patterns of bike share service used by the customers.***

Let’s answer the following questions to better understand the seasonal trends.

**Total Trips per month by user type**

Here’s code and visuals:

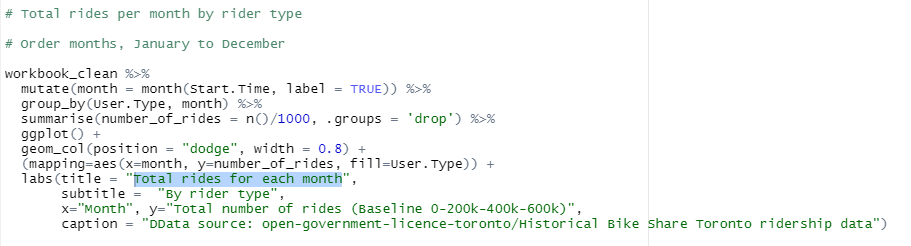


Figure 20. Total rides for each month code

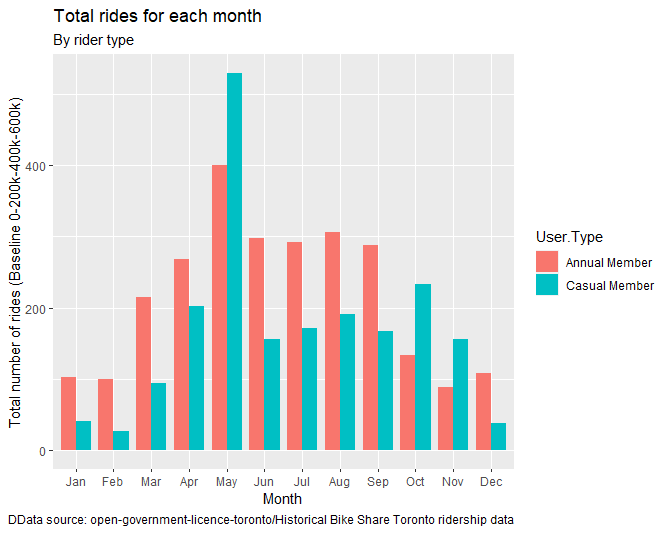


Figure 21. Total rides for each month visual

**Average Ride Length monthly**

Here’s the code and visual:

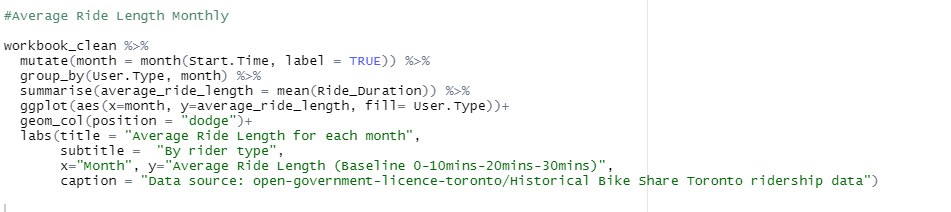


Figure 22. Average Ride Length for each month code

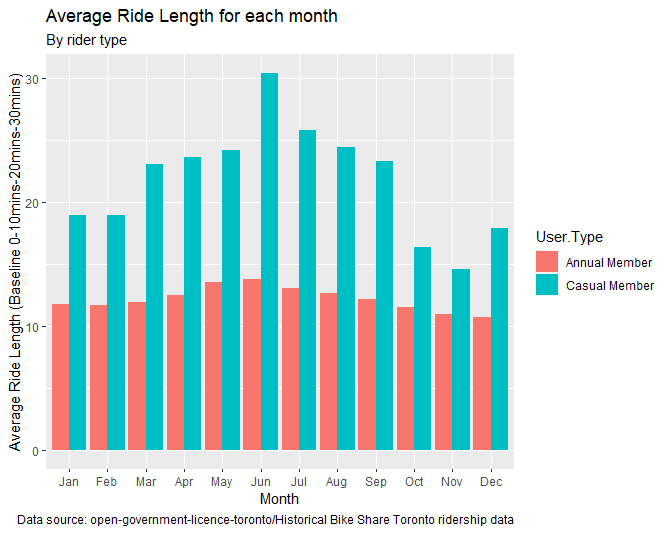


Figure 23. Average Ride Length for each month visual

It’s evident from figure 21., that in the month of **May** Bike Share Toronto served even more than **550k** **casual** bike riders who are probably enjoying the beautiful weather of late spring and early summer.

On the other hand, figure 21., also highlights that **annual membership** holders extensively use the Bike Share service all throughout the early spring till late summers. Hence, this time period is the most lucrative for the business as both **annual members** and **casual riders** make use of the Bike share service in Toronto City during this time.

**Conclusion**

In this report, we explored Bike Share Toronto’s ridership data for patterns and trends. These patterns can help us generate many useful insights, such as:

* Overall, Early Spring to late summer is the most preferred time of the year when people of Toronto and tourists love to ride bikes around the city.
* Annual members extensively use the service throughout the week, but casual riders peak during the weekend.
* Casual riders take longer rides than annual members.
* Casual riders take more trips than annual members.

**Recommendations & Concerns**

If we had some more information associated with the “type of bike” available, just like in Cyclistic Bike Share data then we could also add that factor to our project.

Also, if we had more details on the demographics of the customers like age or gender, it could help us generate more targeted insight.

Sky is the Limit!

When it comes to asking more questions to get more insight from the data, there’s no limit.

For example:

* We can generate insights on the **Most popular start and end Stations**, which **will** help us plan expansion strategy, and it can help launch a successful marketing campaign by increasing activities around more famous locations
* We can generate insights based on **Weather Patterns**, by using weather data available online in the public domain, and it can help understand customers’ best preferred time of the day when they take a bike share service.

(This project is enough for beginners, but as we improve our Data Analytics skillsets, we can build more insightful recommendations on top of this one.)

**Final Thoughts**

Don’t forget to save your work in standard format (MS.Doc file, R Markdown HTML Document, Jupyter Notebook, or SQL Script.)

Always remember that reporting is an integral part of your analytical Project. Also, it will be the final document that your potential stakeholder or client will receive from you, so make sure it contains everything that you need to communicate to your stakeholders.

I ensure that I have submitted the following documents:

* Project Report

MS. Word file. File, R Markdown HTML Document, Jupyter Notebook, or SQL Script.

* Changelog Document

This will be your timeline document highlighting your steps and solutions.

* Static/Interactive Dashboard (if required)
* Read Me text Document

Highlighting WHAT, WHY, WHO, WHEN, and HOW for our Data Analytics Project.

Finally, when we have saved all these documents into one folder or a repository online, then our project can ultimately be closed/finished.

***CONGRATULATIONS!***

Thank you for being kind and patient enough to be with me throughout this article, I hope you liked it and I wish you all the best for your future endeavors!

Keeping Learning!

Yours truly,

Laks Malhotra