Martin Thai

**Data Analytics Project: Cyclistic**

**All related code can be found in my Github:**

[**https://github.com/martintthai/data-analysis-projects/tree/main/cyclistic**](https://github.com/martintthai/data-analysis-projects/tree/main/cyclistic)

**\*Please visit my Tableau for a better viewing experience of the visualizations:**

[**https://public.tableau.com/views/CyclisticProjectDashboard/Dashboard1?:language=en-US&:display\_count=n&:origin=viz\_share\_link**](https://public.tableau.com/views/CyclisticProjectDashboard/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link)

**Introduction/Background:**

I’ll be doing my project on the data made publicly available by Cyclistic, a bike-share company based in Chicago. The company offers 3 types of pricing plans, single-ride passes, full-day passes, and annual memberships. In this project, I focus on two types of customer segments, casual riders and annual members. Cyclistic users that use single-ride and full-day passes will be referred to as casual riders, and users that have an annual membership will be referred to as members. My goal is to maximize the number of annual memberships as annual members are much more profitable than casual riders. In order to convert casual riders into annual members, I need to answer the question: **How do annual members and casual riders use Cyclistic bikes differently?**

**What I did in a nutshell:**

* Conducted an in-depth SQL analysis applying subqueries, joins and aggregations, and DML (data manipulation language) statements to differentiate the behaviors of annual members and casual users within a bike-share company, yielding valuable insights into the diverse ways they use the service
* Managed end-to-end data operations, including cleaning, processing, and analysis, incorporating factors such as ride duration, origin-destination patterns, and the days users typically use the bike-share service
* Utilized data-driven insights to create a dashboard analyzing key metrics in Tableau, recommending flexible membership options to accommodate diverse preferences and targeted marketing initiatives for specific user segments, with the goal of converting casual users into annual members to drive increased revenue

The data that I use is publicly available here: <https://divvy-tripdata.s3.amazonaws.com/index.html>

**Datasets:**

There are 3 csv files/datasets that I will be using:

* 04/2023 Data (Cyclistic data from April 2023)
* 05/2023 Data (Cyclistic data from May 2023)
* 06/2023 Data (Cyclistic data from June 2023)

These datasets contain historical bike trip data from quarter 2 of the year (April, May, June).

Columns include:

* ride\_id (for each ride)
* rideable\_type (type of bike used)
* started\_at (when the ride started)
* ended\_at (when the ride ended)
* start\_station\_name
* start\_station\_id
* end\_station\_name
* end\_station\_id
* start\_lat
* start\_lng
* end\_lat
* end\_lng
* member\_casual (if rider is a member or casual rider)

**Dataset statistics:**

-- There are 426,590 rows in april\_trips.

SELECT COUNT(\*)

FROM `cyclistic`.`april\_trips`

-- There are 604,827 rows in may\_trips.

SELECT COUNT(\*)

FROM `cyclistic`.`may\_trips`

-- There are 719,618 rows in june\_trips.

SELECT COUNT(\*)

FROM `cyclistic`.`june\_trips`

There are **1,751,035** total rows throughout the 3 datasets pre-cleaning and filtering.

**Processing and cleaning the data:**

**I combined the 3 datasets into 1 new table called cyclistic\_q2.**

-- Combine all 3 tables to get a complete Quarter 2 table.

CREATE TABLE cyclistic\_q2 (

SELECT \* FROM `cyclistic`.`april\_trips`

UNION ALL

SELECT \* FROM `cyclistic`.`may\_trips`

UNION ALL

SELECT \* FROM `cyclistic`.`june\_trips`

)

I then cleaned the new table, cyclistic\_q2, by removing null and blank values. In this table, there are no null values but there ARE blank values. Columns start\_station\_name, start\_station\_id, end\_station\_name, end\_station\_id, end\_lat, and end\_lng have missing values.

/\*

Remove null and blank values. Only return rows with no null or blank values.

There are no null values, only blank values. Columns start\_station\_name, start\_station\_id,

end\_station\_name, end\_station\_id, end\_lat, and end\_lng have missing values.

\*/

DELETE

FROM cyclistic\_q2

WHERE start\_station\_name = ''

OR start\_station\_id = ''

OR end\_station\_name = ''

OR end\_station\_id = ''

OR end\_lat = ''

OR end\_lng = ''

After removing missing values, we are left with **1,322,180 rows**.

**Analyzing the data:**

From the dataset, a couple of metrics that would be useful in answering the question: **How do annual members and casual riders use Cyclistic differently?** are

1. **The percentage of annual members and casual riders**
2. **Number of rides by bike type**
3. **Number of rides by bike type by user type (annual member vs. casual rider)**
4. **Average ride duration by members and casual riders**
5. **The days that members and casual riders prefer to use Cyclistic**
6. **Areas that members and casual riders tend to use Cyclistic in**

I added new columns to the table to make it easier to analyze:

* day\_of\_week (the day of the week that the ride started on)
* ride\_length (the duration of the ride)
* route (concatenated the station that the ride started at and the station that the ride ended at to see what areas the rider passed by)

ALTER TABLE cyclistic\_q2

ADD day\_of\_week VARCHAR(30) AS

(CASE WHEN DAYOFWEEK(started\_at) = 1 THEN 'Sunday'

WHEN DAYOFWEEK(started\_at) = 2 THEN 'Monday'

WHEN DAYOFWEEK(started\_at) = 3 THEN 'Tuesday'

WHEN DAYOFWEEK(started\_at) = 4 THEN 'Wednesday'

WHEN DAYOFWEEK(started\_at) = 5 THEN 'Thursday'

WHEN DAYOFWEEK(started\_at) = 6 THEN 'Friday'

WHEN DAYOFWEEK(started\_at) = 7 THEN 'Saturday'

END);

-- Add ride length to table.

ALTER TABLE cyclistic\_q2

ADD ride\_length INT;

UPDATE cyclistic\_q2

SET ride\_length = TIMESTAMPDIFF(MINUTE, started\_at, ended\_at);

-- Add route to table.

ALTER TABLE cyclistic\_q2

ADD route VARCHAR(150);

UPDATE cyclistic\_q2

SET route = CONCAT(start\_station\_name, ' to ', end\_station\_name);

**Metrics:**

1. **The percentage of annual members vs. casual riders**

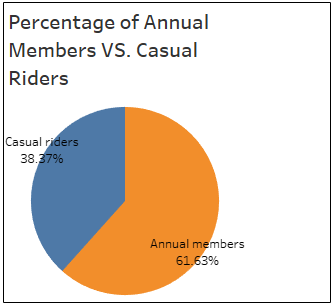
SELECT member\_casual, COUNT(\*) / 1322180 AS percentage\_user\_type

FROM cyclistic\_q2

GROUP BY member\_casual



* **61.63%** of Cyclistic rides in quarter 2 of 2023 were initiated by members.
* **38.37%** of Cyclistic rides in quarter 2 of 2023 were initiated by casual riders.



1. **Number of rides by bike type**

SELECT rideable\_type, COUNT(\*) AS total\_rides

FROM cyclistic\_q2

GROUP BY rideable\_type



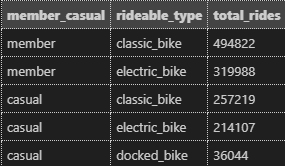
* Classic bikes were used the most, over electric and docked bikes.

1. **Number of rides by bike type by user type (annual member vs. casual rider)**

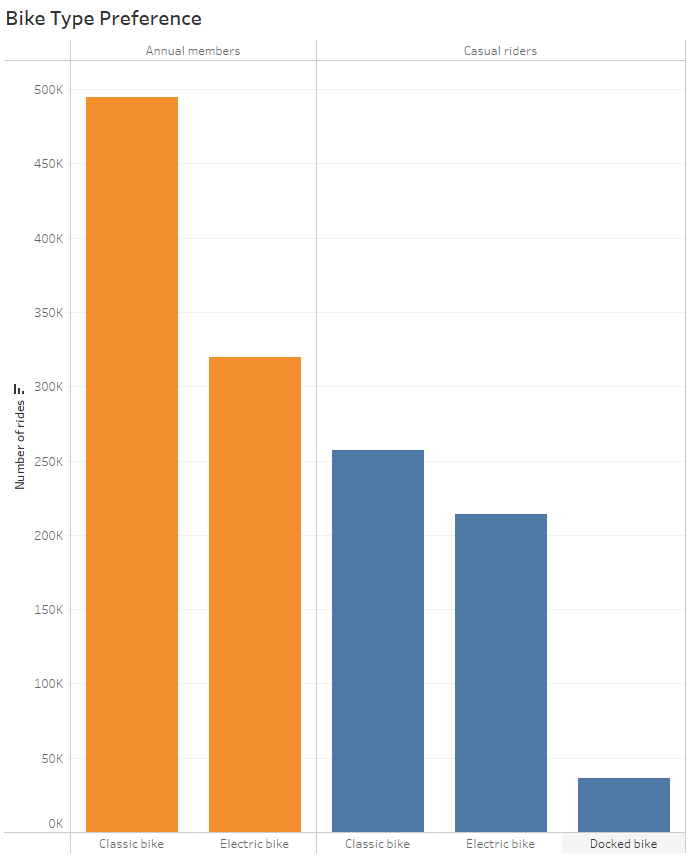
SELECT member\_casual, rideable\_type, COUNT(\*) AS total\_rides

FROM cyclistic\_q2

GROUP BY member\_casual, rideable\_type



* Both members and casual riders preferred classic bikes.
* Members prefer classic bikes a lot more than casual riders. **Casual riders however used electric bikes proportionally more often than members.**

****

1. **Average ride duration by members and casual riders**

**SELECT member\_casual, AVG(ride\_length) AS avg\_ride\_length**

**FROM cyclistic\_q2**

**WHERE ride\_length > 0 AND ride\_length < 720 -- Filter out rides that were longer than 12 hours.**

**GROUP BY member\_casual**

****

* **Casual riders ride approximately 10 minutes longer than members.**

1. **The days that members and casual riders prefer to use Cyclistic**

**SELECT member\_casual, day\_of\_week, COUNT(\*) AS num\_rides**

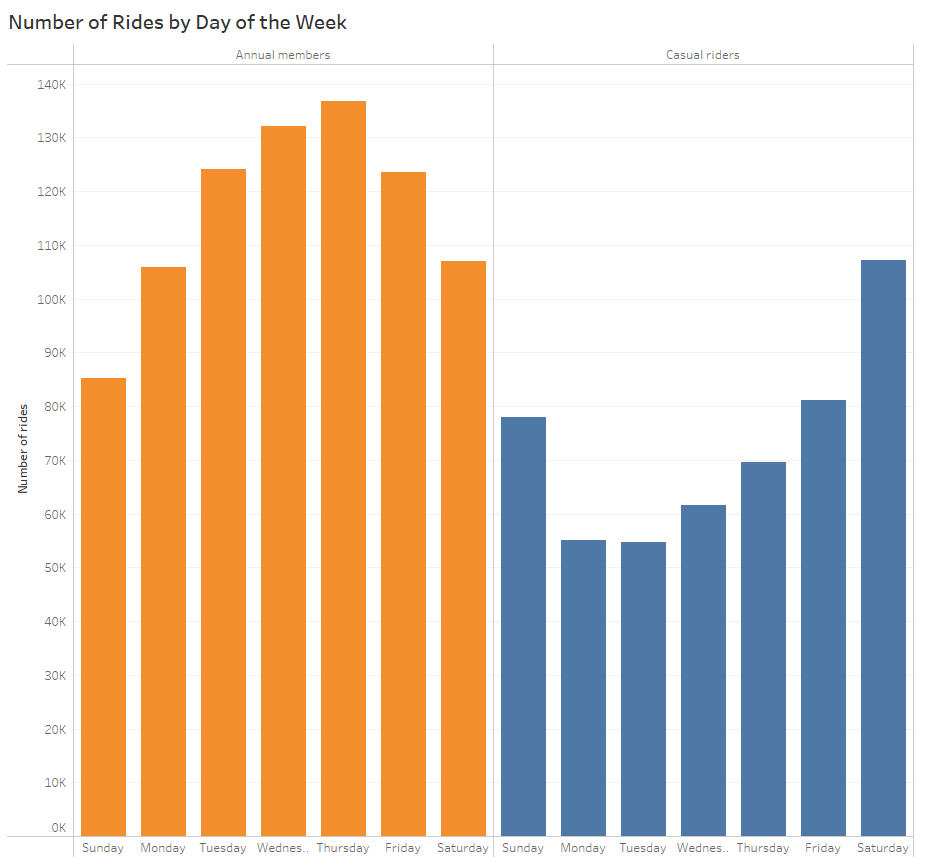
**FROM cyclistic\_q2**

**GROUP BY day\_of\_week, member\_casual**

**ORDER BY member\_casual, num\_rides DESC**

****

* Casual riders tend to use Cyclistic on the **weekend** more often.
* Annual members tend to use Cyclistic during the **weekdays** more often.



1. **Areas that members and casual riders tend to use Cyclistic in**

**SELECT member\_casual, route, COUNT(\*) AS num\_rides**

**FROM cyclistic\_q2**

**WHERE member\_casual = 'member'**

**GROUP BY route**

**ORDER BY num\_rides DESC**

**LIMIT 10;**

**SELECT member\_casual, route, COUNT(\*) AS num\_rides**

**FROM cyclistic\_q2**

**WHERE member\_casual = 'casual'**

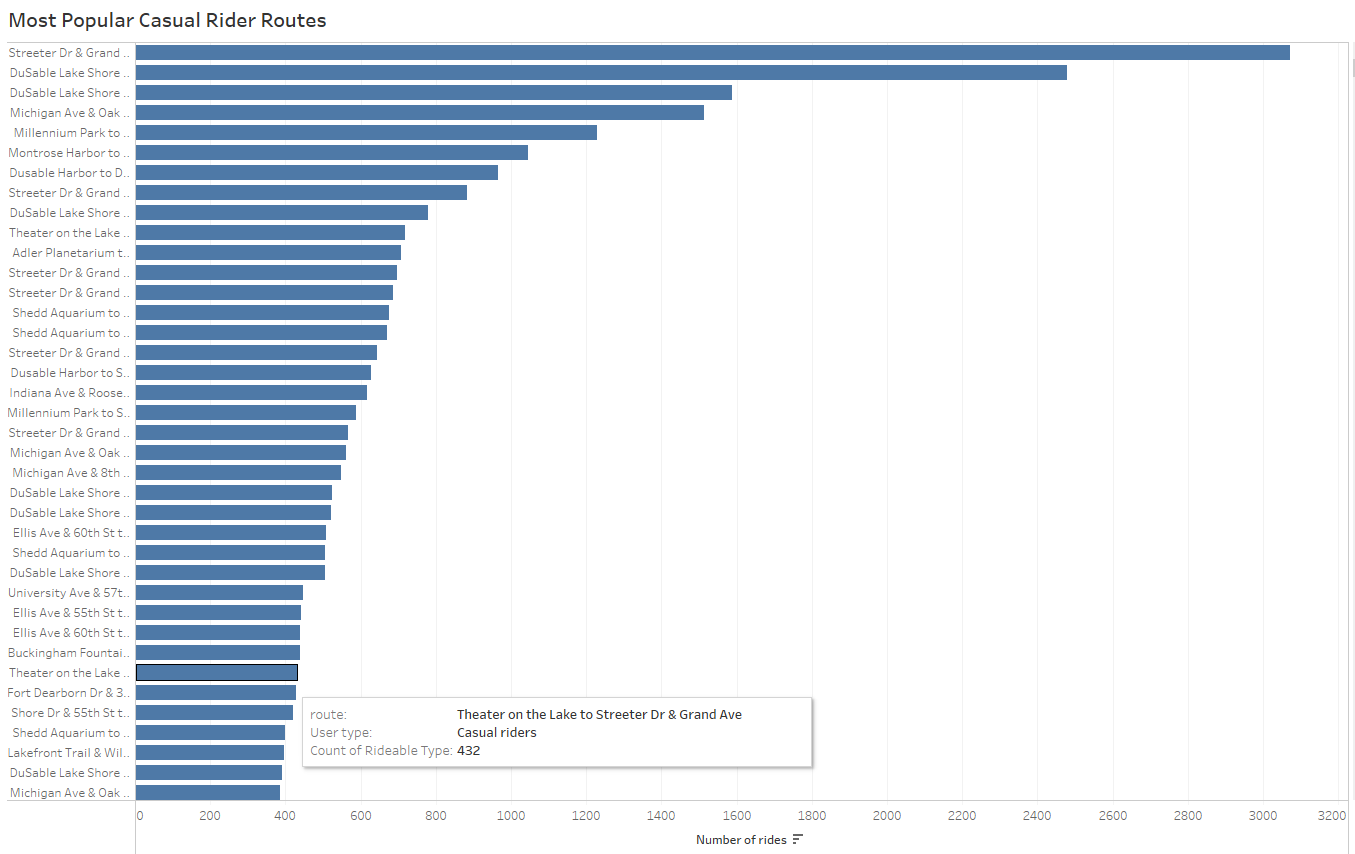
**GROUP BY route**

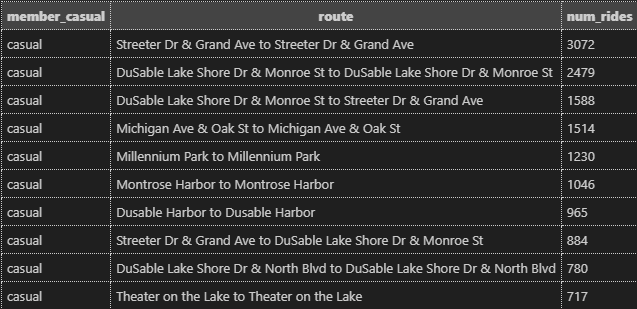
**ORDER BY num\_rides DESC**

**LIMIT 10;**

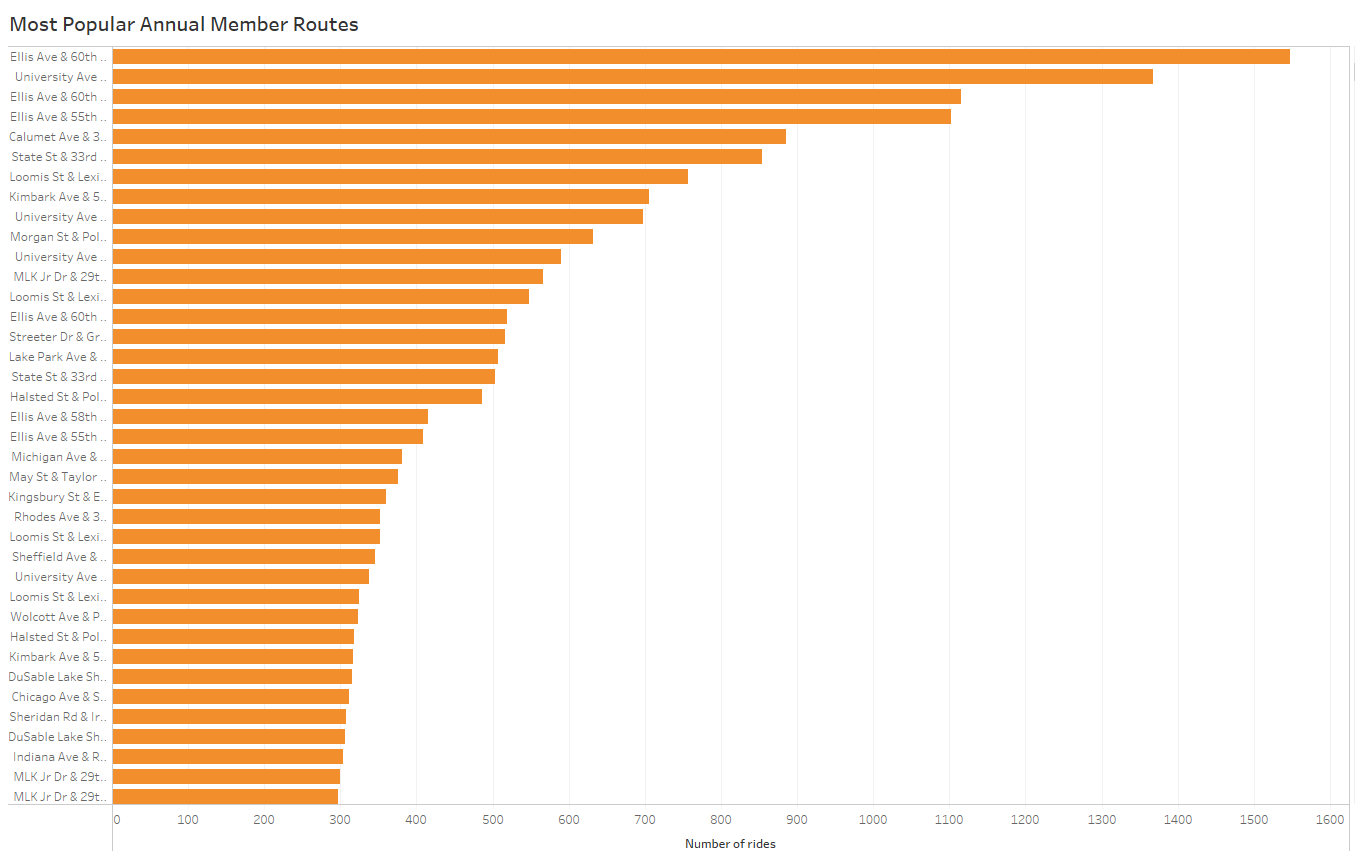
****

* The routes that members tend to take are located near the Hyde Park neighborhood on the South Side of Chicago, which is home to the University of Chicago. There is a mix of **residential, academic, and cultural institutions** nearby.
* Most member routes go through **residential and institutional** spaces.





* Routes that casual riders tend to take are near downtown Chicago and a large body of water, specifically Lake Michigan (which is a lakefront).
* The area that these routes pass include more **tourist-like attractions/sceneries**, such as Lake Michigan, Millennium Park, Frank Park, “The Bean”, and the Loop (which is a central business district of the city that is a hub for businesses, theaters, museums, and parks).
* The casual rider routes pass through **lively and vibrant areas**, a stark contrast from the areas that the member routes pass by.



**Key takeaways from Cyclistic Quarter 2 2023 Ride Data:**

1. Of the 1,322,180 rows of ride data that we looked at, 61.63% of those rides were initiated by annual members and 38.37% were initiated by casual riders. **Members take a majority of quarter 2’s rides, meaning optimizing profits means maximizing memberships.**
2. Classic bikes were used the most in general, as both annual members and casual riders preferred classic bikes. **However, casual riders did use electric bikes proportionally a lot more often than members, so there may be a market for electric bikes within casual riders.**
3. Casual riders ride for longer than annual members (approximately 10 minutes longer). **Casual riders seem to be taking their time with the service, using Cyclistic for leisure, therefore taking longer rides than annual members, who seem to be using the service to get from point A to point B (to get to school/get groceries/meet up with friends/etc).**
4. Casual riders tend to use Cyclistic on the weekend more often and annual members tend to use Cyclistic on the weekdays more often.
5. **Annual members tend to take routes that are located near residential and academic/institutional spaces** while **casual riders tend to take routes that are located near recreational areas and Lake Michigan, a popular lakefront with attractions nearby.** This is evident in the fact that annual members tend to use Cyclistic on the weekdays (when schools and institutions are usually open) and casual riders tend to use Cyclistic on the weekend (when they have free time to explore the city).

**Recommendations to convert casual riders into annual members:**

1. **Add longer-term ride packages, such as monthly memberships that mimic some features of an annual membership**

* Casual riders tend to use Cyclistic near recreational areas with tourist attractions. Once they have seen everything that areas such as downtown Chicago, Lake Michigan, and Millennium Park have to offer, their need to frequent those areas diminishes. This reduced frequency may be a barrier for converting them into annual members.
* Adding a monthly or even weekly ride package would **offer the flexibility for casual riders to use the service freely without having to pay an annual fee upfront.** This is not only cost-effective for casual riders, but also **encourages regular use**, which may even convert them from monthly/weekly users to annual members if they enjoy the service.

1. **Targeted marketing for students**

* Because casual riders tend to use the service nearby academic institutions such as the University of Chicago and Illinois Institute of Technology, they are most likely students, we should target our marketing and advertising towards them.
* Launch social media campaigns that highlight the convenience, cost savings, and environmental benefits of using Cyclistic.
* Recruit student ambassadors/brand representatives who can promote the service on campus through word-of-mouth.

1. **Targeted marketing in downtown Chicago/Harborside areas**

* Because casual riders are more likely to use electric bikes than annual members, advertise electric bikes as a smooth and pleasant way to explore the city..
* Introduce seasonal promotions to capture the attention of casual riders during peak times. This could be offering special discounts in the summer when outdoor activities are more popular.
* Run digital ads across the most popular routes among casual riders. This can be either through billboards or sponsorships with local businesses. This way, people on these scenic routes are aware of the service and can be convinced to turn to Cyclistic for their exploration of the city.

**Tableau Dashboard Link:**

<https://public.tableau.com/views/CyclisticProjectDashboard/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link>

**Extra:**

Thank you for getting this far! I hope you enjoyed my project as I had a lot of fun doing it!

Below I have linked different ways you can contact me if you want to reach out and chat or simply just view my portfolio. Thanks again!

LinkedIn: <https://www.linkedin.com/in/martin-thai-0a3955223/>

GitHub: <https://github.com/martintthai>