Cyclist Case Study - Journal

Botond Koroknai

July 2025

1 Brief structure

Throughout my case study I will follow the recommended framework by the Google Data Analytics course, which consists of 6 steps:

1. Ask

2. Prepare

3. Process

4. Analyze

5. Share

6. Act

The data files I was using gave me a brief summarization about the trips that were taken in 2019/Q1, Q2, Q3 and Q4.

2 Background information about the company

Cyclistic is a bike-sharing service operating primarily in Chicago. It offers two user types: casual riders who pay per ride and annual subscribers who pay a yearly fee for unlimited access. The goal of this study is to identify usage patterns and develop strategies to convert casual users into annual subscribers, driving business growth.

3 Ask

Business statement:
The task is to analyze trip data to understand differences between casual users and annual subscribers, and find actionable insights to increase subscriber numbers.

Key stakeholders:

- Marketing team: aiming to design targeted campaigns to increase subscriptions.

- Product management: interested in understanding user behavior for service improvements.

- Data analytics team: responsible for data-driven insights and recommendations.

Guiding questions include:

- What motivates users to subscribe?

- How do usage patterns differ between casual and subscribed users?

- When and where should marketing efforts be focused?

4 Prepare

I used R within a cloud-based RStudio environment (Posit Cloud) to prepare the data. Two datasets covering 2019 were uploaded there.

Data organization and integrity:

- Checked data structure using the `str()` function to verify data types.

- Verified data credibility: As first-party, internally collected data, it is considered reliable and representative.

- Privacy concerns: The data is anonymized and does not include personally identifiable information.

- Licensing: Data usage follows the company's non-commercial data license agreement.

- Security: Access to the data is restricted to authorized users only.

- Accessibility: Data can be shared securely with trusted collaborators.

Data verification:

- Checked for missing values using `sum(is.na())` and found missing data primarily in `gender` and `birthyear` fields, likely due to optional user registration details.

- Verified accuracy by confirming correct data types and logical value ranges.

5 Process

Data cleaning steps included:

- Corrected data types with `as.datatype()` functions.

- Created new columns such as `trip_duration` and `Weekday`.

- Trimmed whitespace and removed duplicates.

- Sorted data by date for better organization.

- Removed invalid entries, such as trips with negative duration or trips longer than one week.

- Maintained data integrity by only working on data copies and documenting changes thoroughly.

6 Analyze

How should you organize the data to perform analysis on it?
My first step was to unify the structure for the 4 data files in order to perform the same analytical steps on them.

Has your data been properly formatted? I have made some minor formatting mistakes, like selecting faulty data types for some columns, but it was quickly corrected using Excel.

What surprises did you discover in the data?
When I first read that the business goal was to attract more people to be subscribers, my first thought was that there are way less subscribers than "casual" users. In reality there are way more subscribers than users, which made me dig deeper into the details in order to find underlying trends to attract more people to be a regular user of the service.

What trends or relationships did you find in the data?

- Both the number of "casual" users and subscribers are growing when the spring comes, however they peak at different times:

    - The peak of the subscribers come in Q2 (APR-JUN).
   - While the Costumers only reach their max in Q3 (JUL-SEP)

    My hypothesis behind this trend is that once the weather is getting warmer both categories start to rise. Once it is summer the "subscribers" mostly locals go away on holiday, while "Costumers" mostly tourist come to the City, that is why it reaches its maximum later.

Number of users in each Quarter
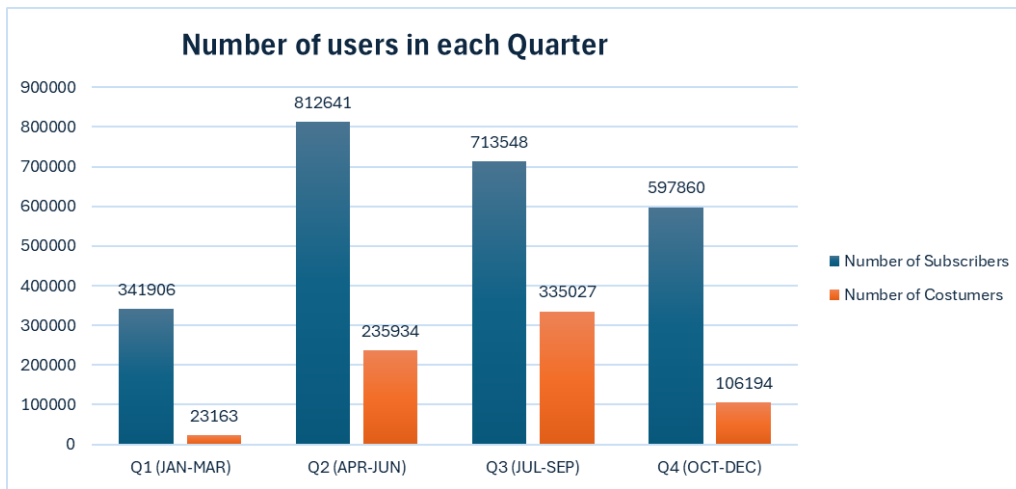
- By counting the number of trips for each day we can clearly see the differences between user types:
  - For the subscribers the majority of trips happens on weekdays, which indicates they are using the service for their everyday life like going to work/school.

COUNT OF TRIPS FOR EACH DAY - SUBSCRIBERS
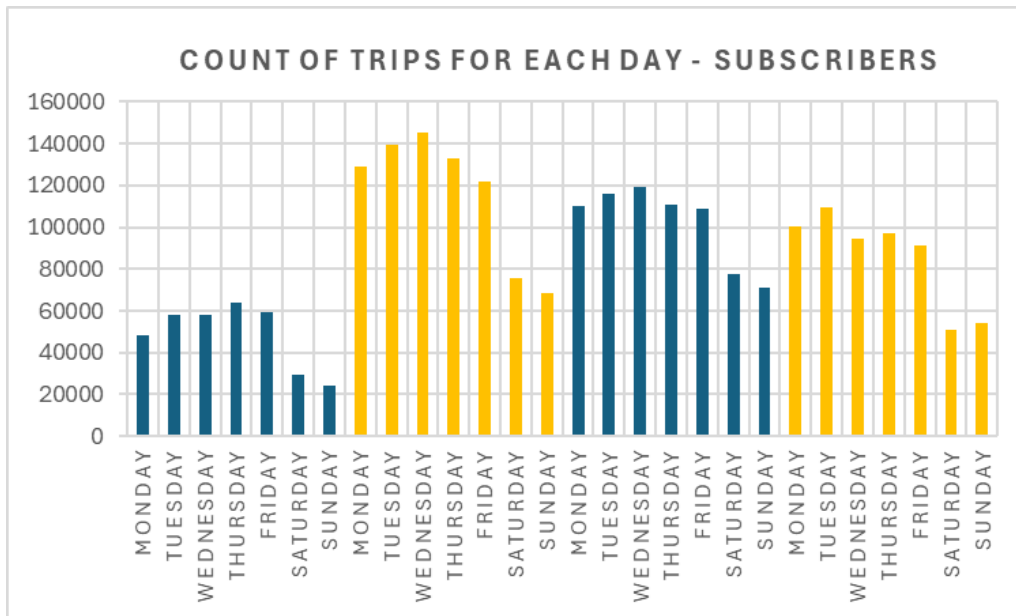
  - On the other hand we can clearly see that most of the trips for the "Costumers" happen in Q2 and Q3 with peaks on Saturday and Sunday which I think happens due to the extra "residential" use who are not subscribers.

COUNT OF TRIPS FOR EACH DAY - CUSTOMERS
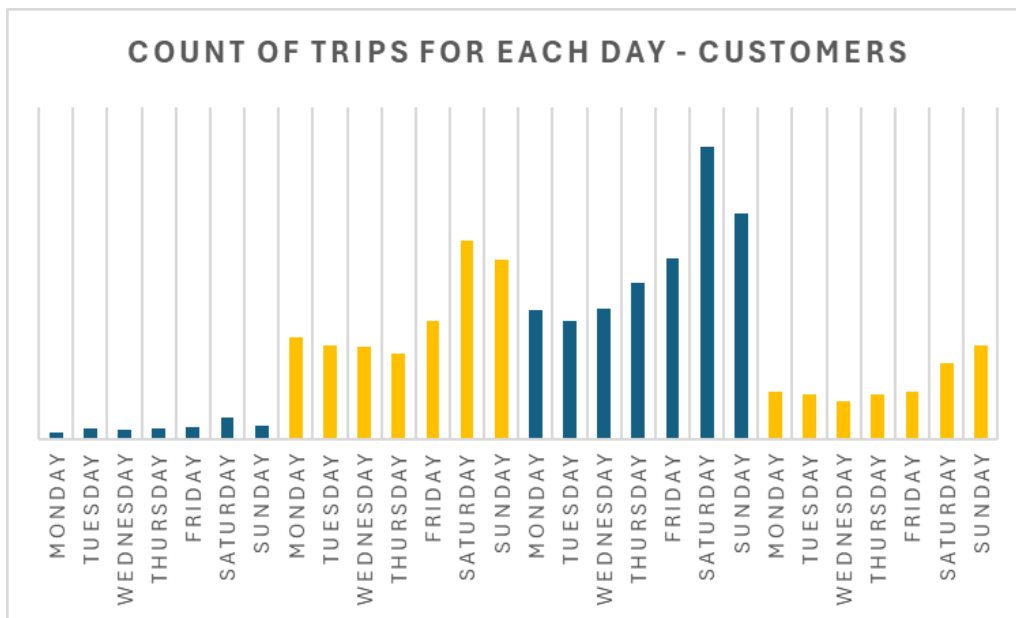
- What I consider also important is the geographical location of the trips: In case of costumers most trips was from / to the most popular attractions around Chicago

Start station	Count of trips
Streeter Dr & Grand Ave	22229
Lake Shore Dr & Monroe St	14445
Lake Shore Dr & North Blvd	9740
Michigan Ave & Oak St	9524
Millennium Park	8526
Shedd Aquarium	6949
Theater on the Lake	6908
Dusable Harbor	4746
Michigan Ave & Washington St	4548
Adler Planetarium	4278
Michigan Ave & 8th St	3847
Montrose Harbor	3703
Indiana Ave & Roosevelt Rd	3166

How will these insights help answer your business questions?

The insights reveal clear behavioral patterns that can inform strategies to convert casual users into subscribers. For example:

- **Seasonal trends** show that casual users peak in summer, suggesting that targeted marketing and subscription incentives during Q2–Q3 could be effective.

- **Day-of-week usage** indicates that subscribers mostly ride on weekdays, likely for commuting, while casual users prefer weekends. This implies that subscribers are typically locals, while casual users may include tourists or occasional riders.

- **Popular start stations** for casual users are located near major attractions and leisure areas. This highlights an opportunity to place targeted promotions or subscription ads at these high-traffic, tourist-friendly stations.

These insights help bridge the gap between casual and regular use by identifying when, where, and how casual users engage with the service and suggest touch points where we can introduce compelling reasons to subscribe.

7 Share

Summary of findings:
Two distinct user groups emerge:

- Subscribers are mostly locals using bikes for routine weekday trips, peaking in spring.

- Casual users are largely tourists or occasional riders who peak during summer weekends around key attractions.

These insights highlight opportunities to convert casual users by focusing marketing efforts seasonally (Q3) and geographically (tourist hotspots). **Audience and communication:**
As the primary audience includes marketing and technical teams. Effective communication includes:

- Clear visualizations and charts.

- Concise executive summaries and bullet points.

- Reproducible analysis scripts for transparency.

- Data-backed recommendations.

8 Act

Final conclusion:
Deploy targeted subscription advertising and promotions at high-traffic tourist locations (e.g., Millennium Park, Shedd Aquarium) during summer weekends to capitalize on peak casual ridership. **Next steps:**

- Develop seasonal marketing campaigns offering subscription incentives during Q3.

- Introduce location-based promotions at popular start stations.

- Monitor impact on subscriber growth and adjust strategy accordingly.