



Google's Data Analytics Certification

Capstone Project

Cyclistic Membership Analysis

Nicolas Urban

June 19, 2025

Introduction

Background Information

In this scenario from the course, I am a junior data analyst on the marketing analyst team at a fictitious company called Cyclistic. In this project, due to the company's main revenue stream being subscriptions, I am expected to develop a marketing strategy to convert casual users of our application into yearly subscribers using data insights and visualizations.

Purpose

Increase revenue by developing a marketing strategy that will target the casual user market segment and convert them to subscribers.

Outline

- A. Outlined the project process and objective based on the purpose of the project and possible issues I may face

Issues

- B. Limited Data Set
 - a. Only the first quarter of two years' worth of data was given to analyze, which limits the ability to identify long-term trends for casual users.
- C. Narrow Market Segment
 - a. Casual users are a small minority compared to the overall user demographic, with limited data on only ride durations and application usage.

Questions

- D. How do annual members and casual riders use Cyclistic bikes differently?
- E. Why would casual riders buy Cyclistic annual memberships?
- F. How can Cyclistic use digital media to influence casual riders to become members?

Data

I. Data Variables

A. *trip_id*

1. Unique ID for each ride

B. *start_time*

1. Start date and time

C. *end_time*

1. Ending date and time

D. *start_station_id*

1. Start stations' unique ID

E. *start_station_name*

1. Start stations' street names

F. *end_station_id*

1. End stations' unique ID

G. *end_station_name*

1. End stations' street names

H. *usertype*

1. Usertype: Subscriber (paid membership) or Casual (free member)

II. Data Cleaning

A. Deleted Irrelevant Variables

1. Deleted the following columns:

a) *bikeid*

- (1) Due to this variable existing in only one dataset and containing some inconsistencies, this variable was eliminated.

b) *tripduration*

- (1) Due to formatting differences and missing values for some rows, this variable was eliminated and will be replaced.

c) *gender* and *birthyear*

(1) In order to keep personal user data private, these variables were eliminated from the dataset.

d) *rideable_type*

(1) Irrelevant variable

e) *start_lat*, *start_lng*, *end_lat*, and *end_lng*

(1) Due to these variables only existing in one dataset and there not being sufficient mapping data, these variables were eliminated from the dataset.

B. Added Variables

1. Created *ride_length* column that calculates the duration of a ride from *start_time* and *end_time* in HH:MM:SS time format
2. Created *day_of_week* column that calculates the day of the week the ride started on using *start_time* with values 1-7 (Sun=1, Mon=2, etc)

C. Trim white spaces

D. Made columns in both datasets consistent

E. Reformatted necessary columns and sorted sheets by *started_at*, then *trip_id*

II. Data Calculation and Visualization

A. The software applications used were: Google Sheets, Microsoft Excel, and BigQuery.

B. Documentation:

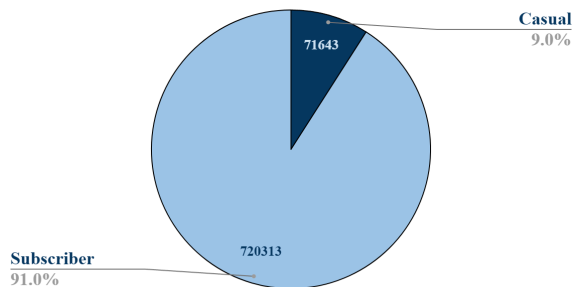
1. [Calculation Documentation](#)

Analysis Results Summary

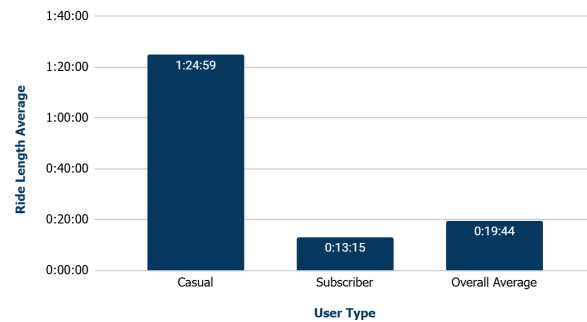
General Calculations

Total Rides by User Type

Total Rides: 791,956

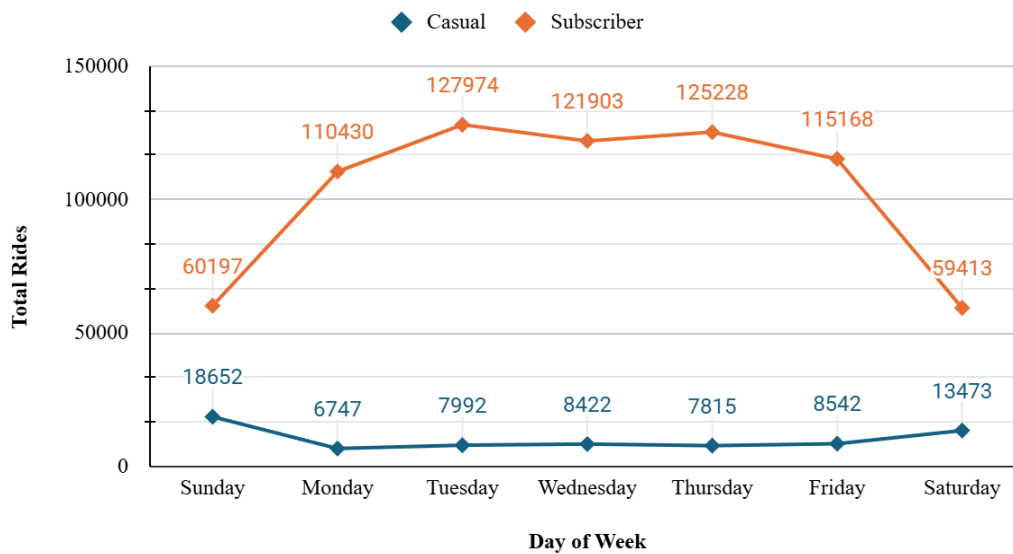


Ride Length Average Based on User Type



Despite making up only 9% of the total rides, casual members have a significantly longer average ride length in comparison to subscribers and the overall average. Casual users are a minority in the user demographic, with unique usage details.

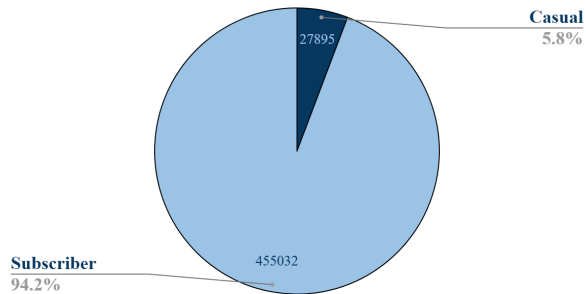
Total Rides By User Type



Although there is a decline in subscribers' usage during the weekends, there are more total rides on Saturday or Sunday than on any individual weekday.

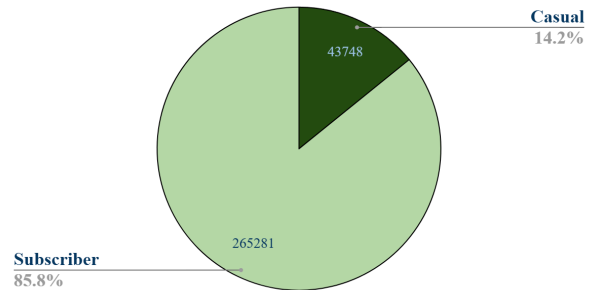
Total Winter Rides by User Type

Total Winter Rides: 482,927



Total Spring Rides By User Type

Total Spring Rides: 309,029



Casual users make up more total rides and a larger percentage of total rides in Spring than they do in Winter. The current assumption is due to the weather; however, due to the lack of weather data and seasonal data, seasonal analysis was inconclusive.

Conclusion

1. Subscription-only services
 - a. Design a new service/product line targeting long-distance cyclists
 - i. Create and market a new feature that makes tracking long distances easier
 - b. Design a new service that incentivises frequent use of the app (like a reward system from other apps), targeting occasional users, especially on weekends, since there is an increase in casual user usage during the weekends.
2. Collect more data
 - a. Lack too data that spans a long period of time, making long-term analysis and seasonal analysis difficult to conduct

Works Cited

- Google. 2019. Divvy_Trips_2019_Q1 [Data set]. Coursera.
<https://www.coursera.org/learn/google-data-analytics-capstone>
- Google. 2020. Divvy_Trips_2020_Q1 [Data set]. Coursera.
<https://www.coursera.org/learn/google-data-analytics-capstone>