Cyclistic\_Analysis\_20\_21

Kartik Soni

22/05/2021

## Executive Summary

* **Cyclistic’s finance analysts** have concluded that annual members are much more profitable than casual riders
* Cyclistic experiences **more rides booked by their annual members rather than casual riders**
* However, **Average duration per ride** for a casual rider is almost **three times** of mean of ride duration for annual member every day

## Business Task

Cyclistic marketing team needs to design marketing strategy to convert casual riders into annual members as it is believed that maximizing the number of annual members will be the key to future growth

## Description about data sources

* Data was collected from the given link <https://divvy-tripdata.s3.amazonaws.com/index.html> provided by the Coursera platform in order to prepare a case study for the capstone project of Google Data Analytics Professional Degree Certification Program.
* Downloaded the **past 12 months data (May/2020 - April/2021)**
* Data is reliable, original, comprehensive, cited and current.
* Each monthly dataset have **13 variables** (ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_station\_id, end\_station\_id, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual)
* Defined other columns for weekday, month, year and ride\_length (*ended\_at-started\_at*) to go ahead with the processing of the data.
* Converted data type of the column (start\_station\_id and end\_station\_id) **from Double to Character** to merge all the 12 months data.

## Cleaning and Manipulation of Data

* Dataframe includes a few hundred entries **when bikes were taken out of docks** and checked for quality by Divvy or **ride\_length was negative**.
* Removed rows with **Start\_station\_name == HUBBARD ST BIKE CHECKING (LBS-WH-TEST)" | “WATSON TESTING - DIVVY” | “hubbard\_test\_lws”**
* Created a new version of the dataframe (v2) since data is being removed.

r all\_trips\_v2 <- all\_trips[!(all\_trips$start\_station\_name == "HUBBARD ST BIKE CHECKING (LBS-WH-TEST)" | all\_trips$start\_station\_name == "WATSON TESTING - DIVVY" | all\_trips$start\_station\_name == "hubbard\_test\_lws" | all\_trips$ride\_length<0),] r all\_trips\_v2 <- all\_trips\_v2 %>% drop\_na(ride\_length)

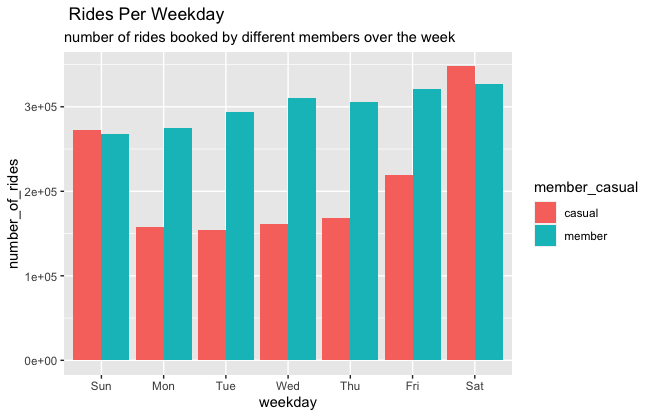
## Summary of Analysis

* **Casual riders average ride duration is way higher than the annual members**, same pattern being followed for median, max.
* Intensity of Casual riders increases at weekends which means **more of the casual riders prefer bicycles for leisure activities than making use of it to go to work** where annual members bookings don’t encounter much change over the week.
* **Docked Bike** is the most preferred bike by all riders including annual and casual.
* **Casual riders perform twice as round trips as compared to annual member** (start\_station\_name == end\_station\_name)
* Summer fests **(mainly in July and August)** in chicago leads to increase in demand of rides for both type of the members.

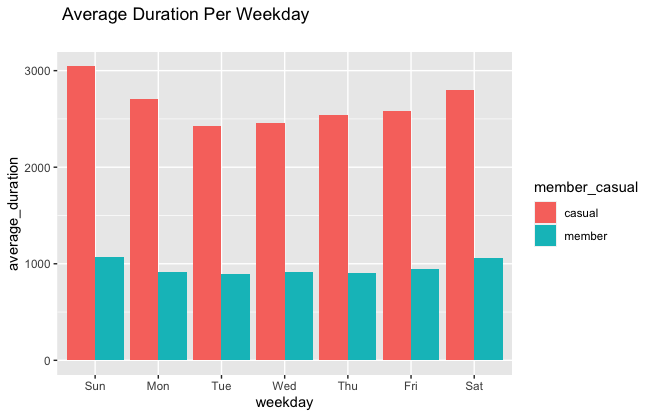
r all\_trips\_v2$day\_of\_week <- ordered(all\_trips\_v2$day\_of\_week, levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday")) r aggregate(all\_trips\_v2$ride\_length ~ all\_trips\_v2$member\_casual + all\_trips\_v2$day\_of\_week, FUN = mean) r table(all\_trips\_v2$month) table(all\_trips$rideable\_type) r round\_trips <- all\_trips\_v2 %>% select(c(start\_lat, start\_lng, start\_station\_name, end\_station\_name, ride\_length,member\_casual)) %>% filter(start\_station\_name == end\_station\_name)

## Supporting Visualizations and key Findings

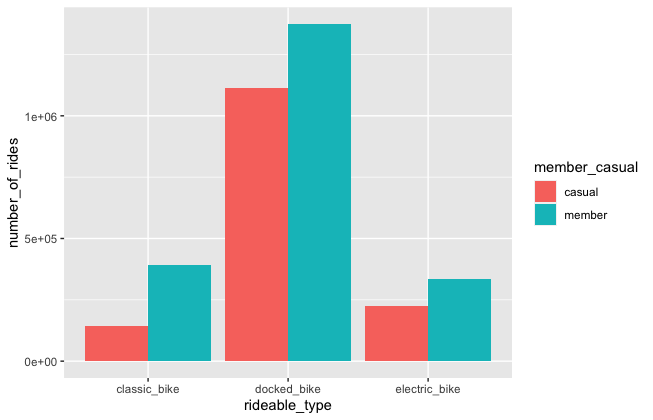
### **Weekend (Saturday and Sunday)** -> High turnup of casual riders as compared to week days



### Average ride duration for casual members is **almost three times** of annual members

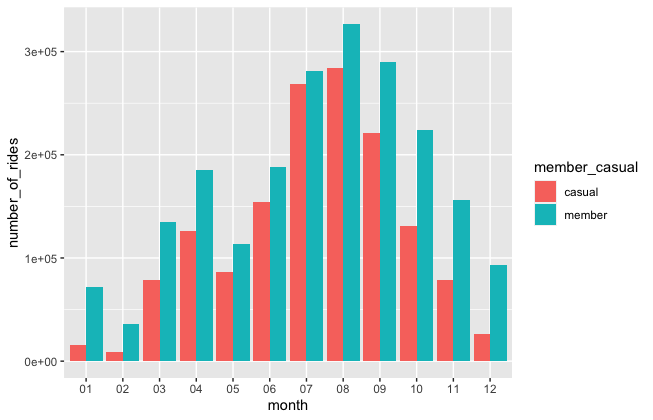


### **Docked Bike** most preferred one among the all

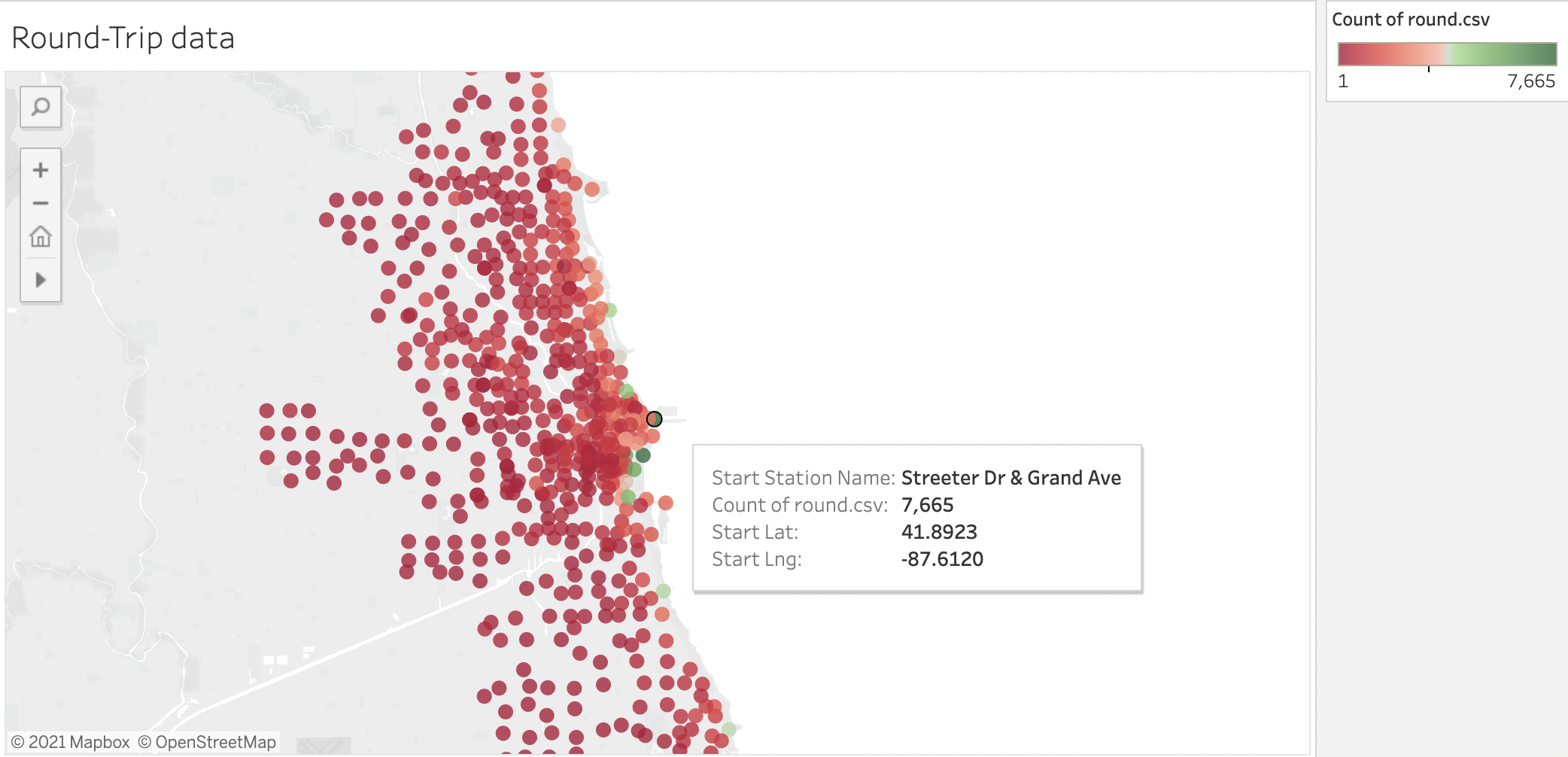


### **July and August**, these two months experience high demand in bikes over the region

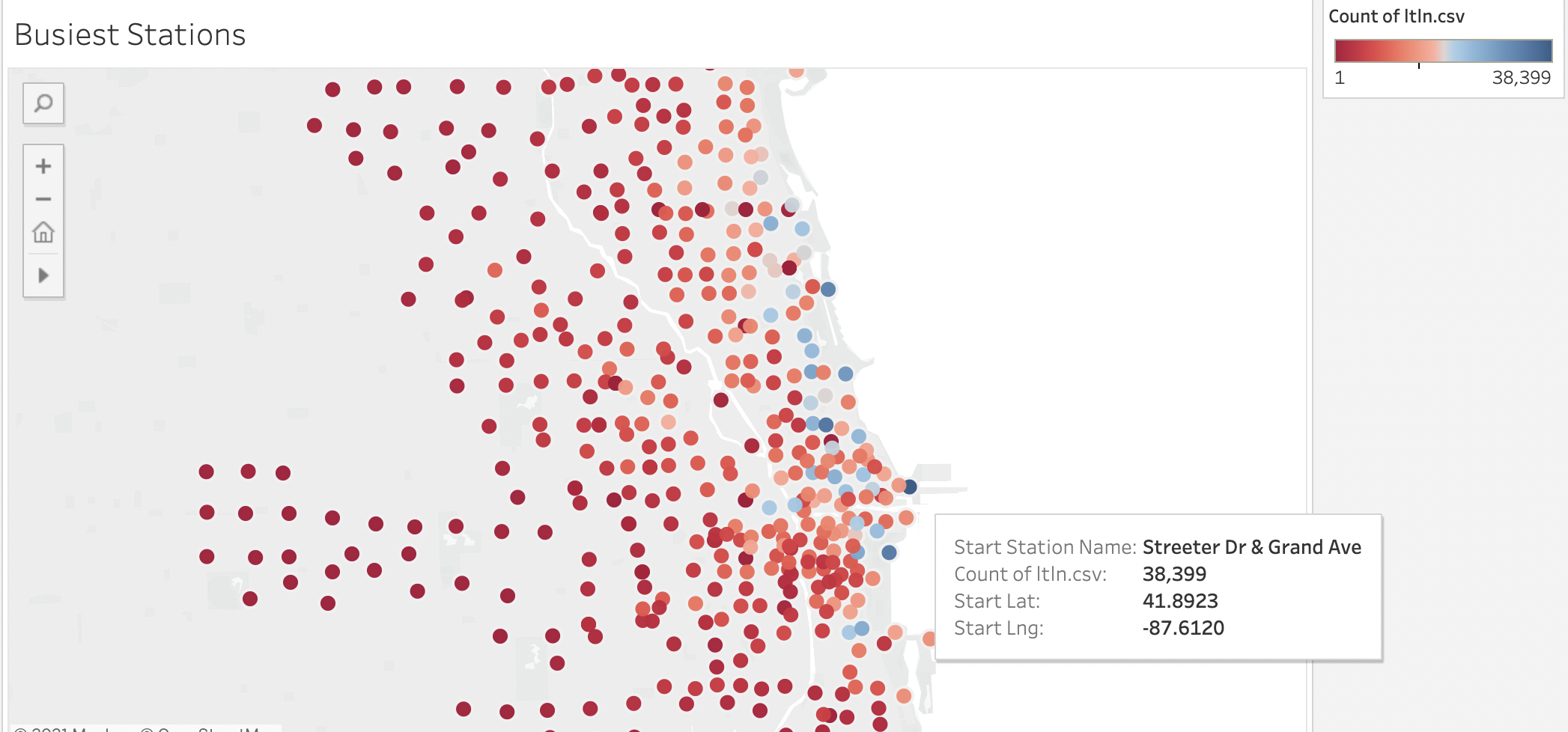
(may be because of summer fests)



### **Station with round trip entries**, for testing bikes at these stations on regular time intervals



### **Busiest Stations**, to keep a check on volume of bikes so that numbers satisfies the customer demands



## Recommendations

**Additional Data to be included** (Survey on happiness metric of annual members, are there any problems faced by casual riders, (if yes, what are those?) and also get to know the reason for not choosing annual membership program.

1. **Share Posts on different digital media platforms highlighting Cyclistic’s finance analyst data** on how annual members are much more profitable than casual riders.
2. **Highlight the benefits of using bicycles to travel on daily basis** both financially and physically to casual riders, this will help to make them think of these bikes not only for leisure activities but also while going for work.
3. Share average duration insights to casual riders through visuals and **state the problems faced by different customers, especially on longer rides**(higher ride\_length) and **respective solutions as well** (i.e. exchange bicycles at any other station, start ride without worrying about payment and free from the responsibility of taking care of bikes on your own)
4. **Examine the increase in demand for the months (July and August)**, team should see the particular reason behind it and work on the same.