Google Analytics Certificate Case Study: How does a Bike-Share Navigate Speedy Success?

By Kareena K.

Step 1: Ask

What is the problem you are trying to solve?

How do annual members and casual riders use Cyclistic bikes differently?

Step 2: Prepare

- Where is the data located?
- How is the data organized?
- How does this data help you answer your question?
- Are there any problems with the data?

ROCC

Reliable

Original

Comprehensive

Current

Cited

Step 3: Process

- What tools are you choosing?
- What steps have you taken to ensure that your data is clean?

```r
#Data Load In :
cyclist <- read.csv("C:/Users/Kareena/Desktop/Personal Project/Casestudy_cyclisticOct22/Data/Divvy_Trips_2020_Q1.csv")

#Taking a look at data
head(cyclist)  #13 variables, 426887 observations

# Checking if there are any NA values

sum(is.na(cyclist))
```

```
[1] 3
```

	ride_id <chr>	rideable_type <chr>	started_at <chr>
1	EACB19130B0CDA4A	docked_bike	2020-01-21 20:06:59
2	8FED874C809DC021	docked_bike	2020-01-30 14:22:39
3	789F3C21E472CA96	docked_bike	2020-01-09 19:29:26
4	C9A388DAC6ABF313	docked_bike	2020-01-06 16:17:07
5	943BC3CBECCFD662	docked_bike	2020-01-30 08:37:16
6	6D9C8A6938165C11	docked_bike	2020-01-10 12:33:05

6 rows | 1-4 of 13 columns

Step 4: Analyze

- How should you organize your data to perform analysis on it?
- What trends or relationships did you find in the data?
- How will these insights help answer your business questions?

Transforming Data: Dates (Month, Day, Year)

```
#separate(cyclist,start_date,c("Start_Year","Start_Month","Start_Day"),sep = " ")
#separate(cyclist,end_date,c("End_Year","End_Month","End_Day"),sep = "-")

#Separates the start date year, month, day
cyclist2 <- separate(cyclist,start_date,  c("start_year","start_month","start_day"))
#Separates the End date year, month, day
cyclist3 <- separate(cyclist,end_date,  c("end_year","end_month","end_day"))
```

Transforming Data: Time From 24 hour to 12 hour

```
#Data Transformation:
cyclist$start_date <- as.Date(cyclist$started_at)
#Convert the time from 24-hour to a 12-hour time
cyclist$start_time <- format(as.POSIXct(cyclist$started_at) ,format = "%H:%M:%S")
cyclist$start_time <- format(as.POSIXct(cyclist$start_time ,format = "%H:%M:%S"),format
="%I:%M:%S %p")

cyclist$end_date <- as.Date(cyclist$ended_at)

#Convert the time from 24-hour to a 12-hour time
cyclist$end_time <- format(as.POSIXct(cyclist$ended_at),format = "%H:%M:%S")

cyclist$end_time <- format(as.POSIXct(cyclist$end_time,format = "%H:%M:%S"),format
="%I:%M:%S %p")
```

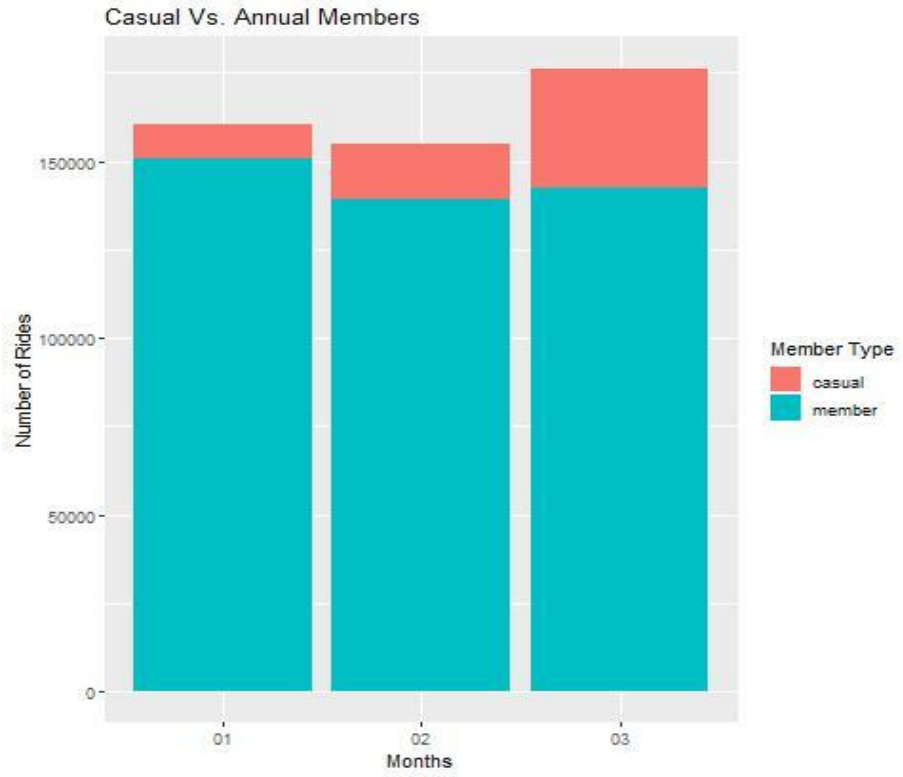
Transforming Distances (coordinates to miles)

```
casual_mem<- casual_mem %>% rowwise %>% mutate(distance =
distm(x = c(start_lng,start_lat),y=c(end_lng,end_lat),fun
= distHaversine)[,1]/1609)
```

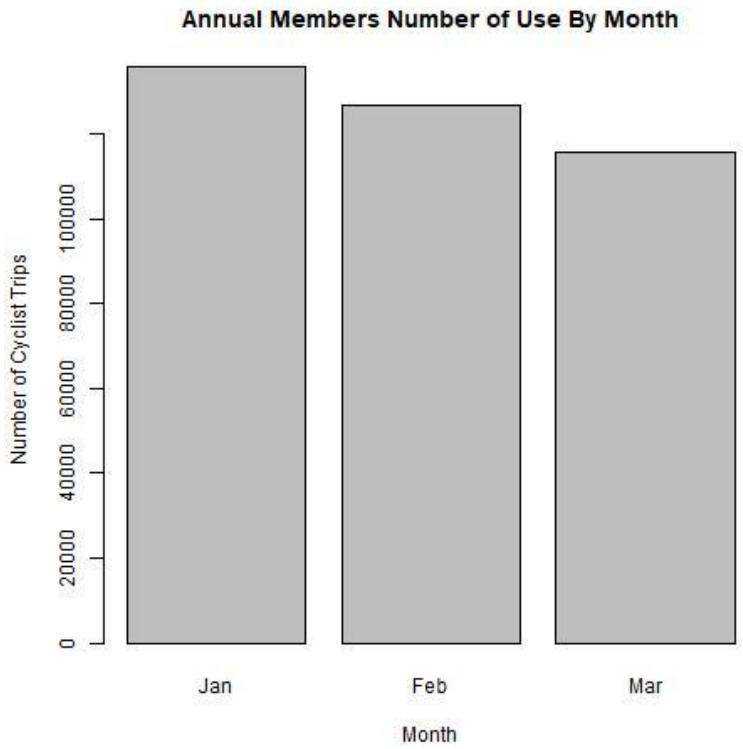
Step 5: Share

- Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?
- What story does your data tell?

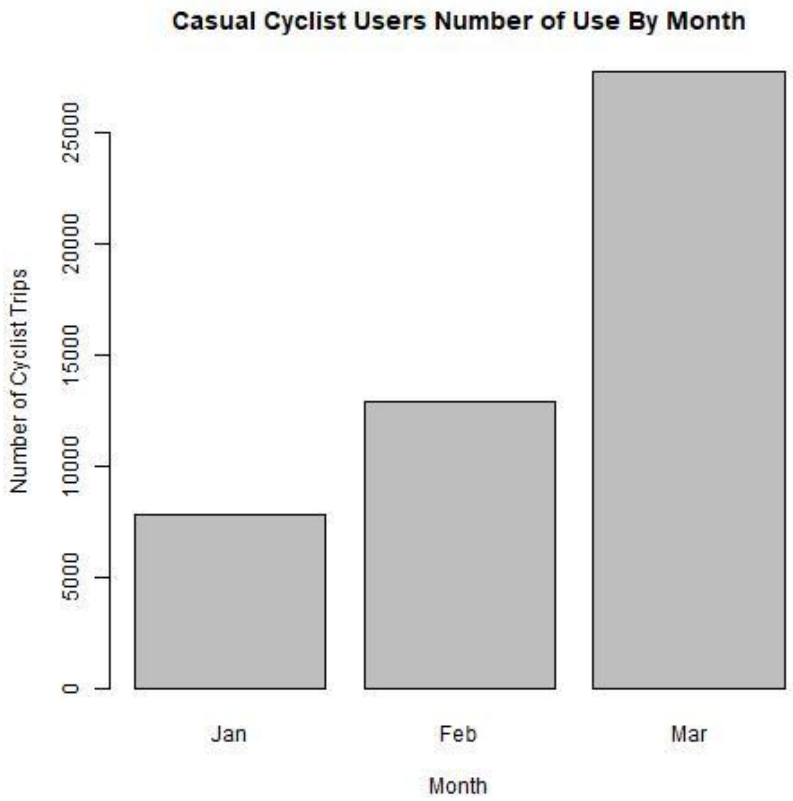
Casual Cyclist Vs. Annual Members Trips by month



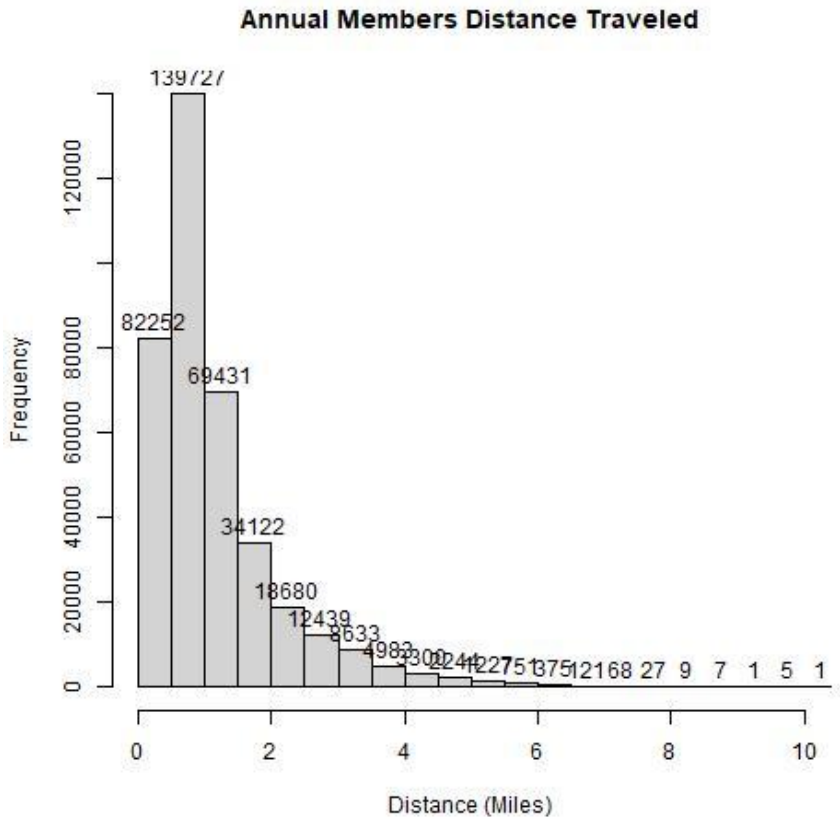
Closer look: Annual Member trips by Month



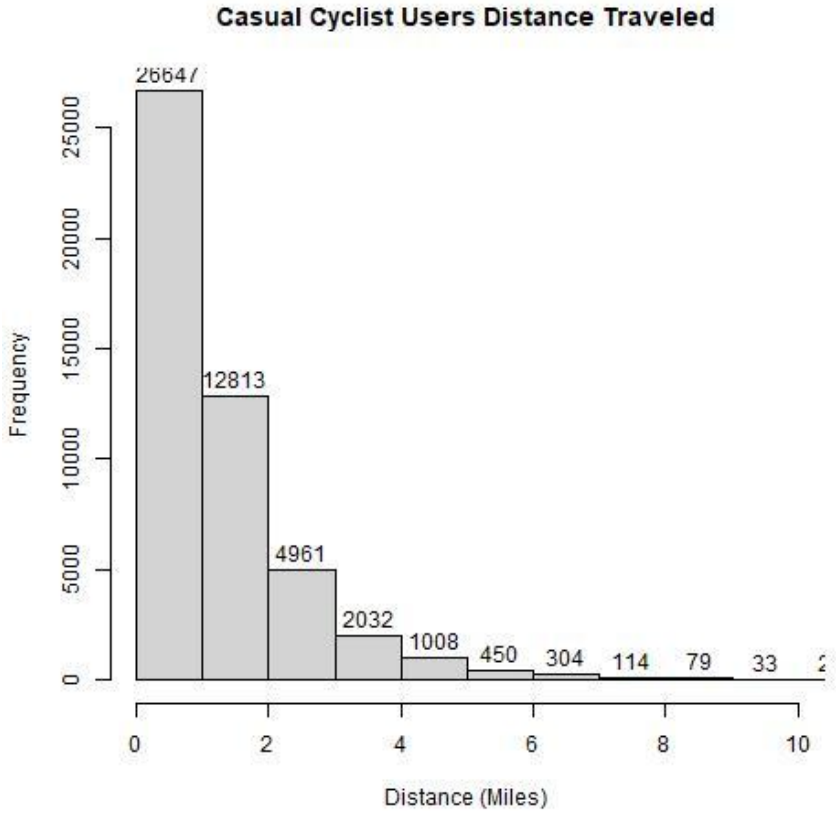
Closer look pt2: Casual Cyclist by Month



The Frequency of Distances Traveled by Annual Members



The Frequency of Distances Traveled by Casual Cyclist



Step 6: Act

- Concluding Final Analysis
- How could your team and business apply your insights?
- Is there additional data you could use to expand your findings?