Case Study FB

Federico Bereziuk

2023-12-20

Cyclistic Case Study

Ask

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

Business task:

- Find ways to convince casual riders to buy annual memberships

Data Source:

- Provided by Company

Stakeholder:

- Lili Moreno

Prepare

Data Location:

- Google Cloud

SQL Engine:

- BigQuery

Cheking integrity on Start Station

```
SELECT (totalrecords - startnotnull) AS startidnull,
     ROUND(((checkstart.totalrecords - checkstart.startnotnull)/checkstart.totalrecords), 2) AS startn
FROM (
      SELECT COUNT(start_station_id) AS startnotnull,
            COUNT(ride_id) AS totalrecords
      FROM `cyclistic-405915.Stations_Trips.2308_trips`
      ) AS checkstart
```

Results:
* July   16,0%
* August   15,4%
* September 15,2%
* October   15,7%

Checking integriy on Start Hour and Weekday

```
SELECT (totalrecords - startnotnull) AS startidnull,
       ROUND(((checkstart.totalrecords - checkstart.startnotnull)/checkstart.totalrecords), 4) AS startr

FROM (
SELECT COUNT(started_at) AS startnotnull,
        COUNT(ride_id) AS totalrecords
FROM `cyclistic-405915.Stations_Trips.2310_trips`
  ) AS checkstart
```

Results:
* July   0%
* August   0%
* September 0%
* October   0%

Process

Join July, August, September and October tables identify weekday

```
WITH w2307 AS (
      SELECT
      CASE WHEN tmc.daystart = 1 THEN 'Sunday'
           WHEN tmc.daystart = 2 THEN 'Monday'
           WHEN tmc.daystart = 3 THEN 'Tuesday'
           WHEN tmc.daystart = 4 THEN 'Wednesday'
           WHEN tmc.daystart = 5 THEN 'Thrusday'
           WHEN tmc.daystart = 6 THEN 'Friday'
           WHEN tmc.daystart = 7 THEN 'Saturday' END AS weekdaytrip,
           'July' AS tripmonth,
      tmc.members,
      tmc.casuals,
      ROUND((tmc.casuals/(tmc.members+tmc.casuals)), 2) AS caspercent
      FROM (
    SELECT EXTRACT(DAYOFWEEK FROM t2307.started_at) AS daystart,
           COUNT(CASE WHEN member_casual = 'member' THEN 1 END) AS members,
           COUNT(CASE WHEN member_casual = 'casual' THEN 1 END) AS casuals,
    FROM `cyclistic-405915.Stations_Trips.2307_trips` AS t2307
    GROUP BY daystart
    ORDER BY daystart) AS tmc),

    w2308 AS (
      SELECT
      CASE WHEN tmc.daystart = 1 THEN 'Sunday'
           WHEN tmc.daystart = 2 THEN 'Monday'
```

```
        WHEN tmc.daystart = 3 THEN 'Tuesday'
        WHEN tmc.daystart = 4 THEN 'Wednesday'
        WHEN tmc.daystart = 5 THEN 'Thrusday'
        WHEN tmc.daystart = 6 THEN 'Friday'
        WHEN tmc.daystart = 7 THEN 'Saturday' END AS weekdaytrip,
        'August' AS tripmonth,
  tmc.members,
  tmc.casuals,
  ROUND((tmc.casuals/(tmc.members+tmc.casuals)), 2) AS caspercent
  FROM (
SELECT EXTRACT(DAYOFWEEK FROM t2308.started_at) AS daystart,
        COUNT(CASE WHEN member_casual = 'member' THEN 1 END) AS members,
        COUNT(CASE WHEN member_casual = 'casual' THEN 1 END) AS casuals,
FROM `cyclistic-405915.Stations_Trips.2308_trips` AS t2308
GROUP BY daystart
ORDER BY daystart) AS tmc),

w2309 AS (
  SELECT
  CASE WHEN tmc.daystart = 1 THEN 'Sunday'
        WHEN tmc.daystart = 2 THEN 'Monday'
        WHEN tmc.daystart = 3 THEN 'Tuesday'
        WHEN tmc.daystart = 4 THEN 'Wednesday'
        WHEN tmc.daystart = 5 THEN 'Thrusday'
        WHEN tmc.daystart = 6 THEN 'Friday'
        WHEN tmc.daystart = 7 THEN 'Saturday' END AS weekdaytrip,
        'September' AS tripmonth,
  tmc.members,
  tmc.casuals,
  ROUND((tmc.casuals/(tmc.members+tmc.casuals)), 2) AS caspercent
  FROM (
SELECT EXTRACT(DAYOFWEEK FROM t2309.started_at) AS daystart,
        COUNT(CASE WHEN member_casual = 'member' THEN 1 END) AS members,
        COUNT(CASE WHEN member_casual = 'casual' THEN 1 END) AS casuals,
FROM `cyclistic-405915.Stations_Trips.2309_trips` AS t2309
GROUP BY daystart
ORDER BY daystart) AS tmc),

 w2310 AS (
  SELECT
  CASE WHEN tmc.daystart = 1 THEN 'Sunday'
        WHEN tmc.daystart = 2 THEN 'Monday'
        WHEN tmc.daystart = 3 THEN 'Tuesday'
        WHEN tmc.daystart = 4 THEN 'Wednesday'
        WHEN tmc.daystart = 5 THEN 'Thrusday'
        WHEN tmc.daystart = 6 THEN 'Friday'
        WHEN tmc.daystart = 7 THEN 'Saturday' END AS weekdaytrip,
        'October' AS tripmonth,
  tmc.members,
  tmc.casuals,
  ROUND((tmc.casuals/(tmc.members+tmc.casuals)), 2) AS caspercent
  FROM (
SELECT EXTRACT(DAYOFWEEK FROM t2310.started_at) AS daystart,
```

```
            COUNT(CASE WHEN member_casual = 'member' THEN 1 END) AS members,
            COUNT(CASE WHEN member_casual = 'casual' THEN 1 END) AS casuals,
    FROM `cyclistic-405915.Stations_Trips.2310_trips` AS t2310
    GROUP BY daystart
    ORDER BY daystart) AS tmc)


SELECT *
FROM w2307
        UNION ALL SELECT * FROM w2308
        UNION ALL SELECT * FROM w2309
        UNION ALL SELECT * FROM w2310
ORDER BY CASE WHEN weekdaytrip = 'Monday' THEN 1
                WHEN weekdaytrip = 'Tuesday' THEN 2
                WHEN weekdaytrip = 'Wednesday' THEN 3
                WHEN weekdaytrip = 'Thrusday' THEN 4
                WHEN weekdaytrip = 'Friday' THEN 5
                WHEN weekdaytrip = 'Saturday' THEN 6
                WHEN weekdaytrip = 'Sunday' THEN 7 END
```

Join tables by Start Station

```
# Define views for each month
WITH s07 AS (
  SELECT startstation, "July" AS tripmonth, casuals, casual_ratio
FROM `cyclistic-405915.Stations_Trips.2307_stations`
),

s08 AS (
  SELECT startstation, "August" AS tripmonth, casuals, casual_ratio
FROM `cyclistic-405915.Stations_Trips.2308_stations`
),

s09 AS (
  SELECT startstation, "September" AS tripmonth, casuals, casual_ratio
FROM `cyclistic-405915.Stations_Trips.2309_stations`
),

s10 AS (
  SELECT startstation, "October" AS tripmonth, casuals, casual_ratio
FROM `cyclistic-405915.Stations_Trips.2310_stations`
)

SELECT * FROM s07
UNION ALL SELECT * FROM s08
UNION ALL SELECT * FROM s09
UNION ALL SELECT * FROM s10
ORDER BY casual_ratio DESC
```

Join tables to analyze hour

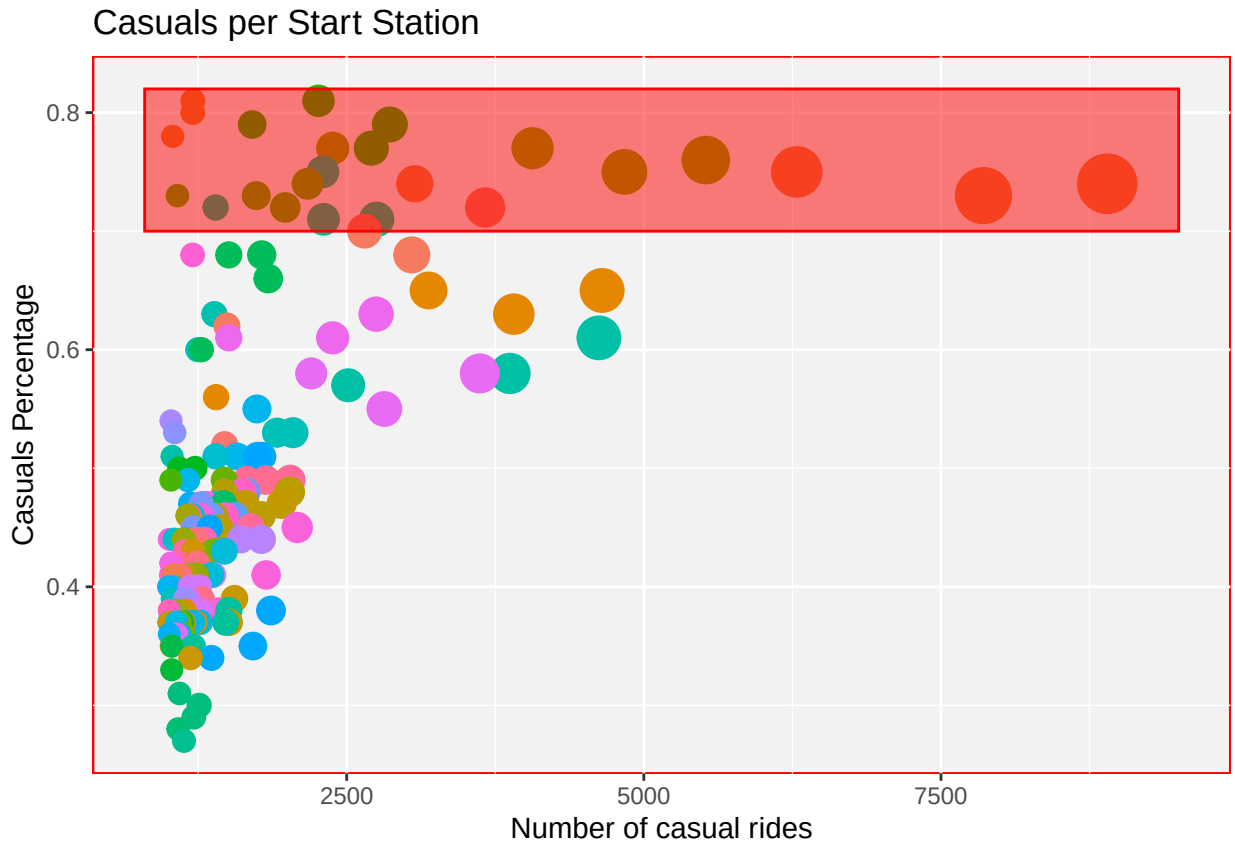
```sql
WITH h07 AS (
SELECT tmc.starthour, "July" AS tripmonth, tmc.members, tmc.casuals,
      ROUND((tmc.casuals/(tmc.members+tmc.casuals)), 2) AS caspercent
      FROM (
            SELECT EXTRACT(HOUR FROM t2307.started_at) AS starthour,
                  COUNT(CASE WHEN member_casual = 'member' THEN 1 END) AS members,
                  COUNT(CASE WHEN member_casual = 'casual' THEN 1 END) AS casuals,
             FROM `cyclistic-405915.Stations_Trips.2307_trips` AS t2307
            GROUP BY starthour
            ORDER BY starthour) AS tmc
),

h08 AS (
  SELECT tmc.starthour,"August" AS tripmonth, tmc.members, tmc.casuals,
      ROUND((tmc.casuals/(tmc.members+tmc.casuals)), 2) AS caspercent
      FROM (
            SELECT EXTRACT(HOUR FROM t2308.started_at) AS starthour,
                  COUNT(CASE WHEN member_casual = 'member' THEN 1 END) AS members,
                  COUNT(CASE WHEN member_casual = 'casual' THEN 1 END) AS casuals,
            FROM `cyclistic-405915.Stations_Trips.2308_trips` AS t2308
            GROUP BY starthour
            ORDER BY starthour) AS tmc
),

h09 AS (
  SELECT tmc.starthour,"September" AS tripmonth, tmc.members, tmc.casuals,
      ROUND((tmc.casuals/(tmc.members+tmc.casuals)), 2) AS caspercent
      FROM (
            SELECT EXTRACT(HOUR FROM t2309.started_at) AS starthour,
                  COUNT(CASE WHEN member_casual = 'member' THEN 1 END) AS members,
                  COUNT(CASE WHEN member_casual = 'casual' THEN 1 END) AS casuals,
            FROM `cyclistic-405915.Stations_Trips.2309_trips` AS t2309
            GROUP BY starthour
            ORDER BY starthour) AS tmc
),

h10 AS (
  SELECT tmc.starthour,"October" AS tripmonth, tmc.members, tmc.casuals,
      ROUND((tmc.casuals/(tmc.members+tmc.casuals)), 2) AS caspercent
      FROM (
            SELECT EXTRACT(HOUR FROM t2310.started_at) AS starthour,
                  COUNT(CASE WHEN member_casual = 'member' THEN 1 END) AS members,
                  COUNT(CASE WHEN member_casual = 'casual' THEN 1 END) AS casuals,
            FROM `cyclistic-405915.Stations_Trips.2310_trips` AS t2310
            GROUP BY starthour
            ORDER BY starthour) AS tmc
)

SELECT * FROM h07
      UNION ALL SELECT * FROM h08
      UNION ALL SELECT * FROM h09
      UNION ALL SELECT * FROM h10
ORDER BY starthour,
```

```
      CASE WHEN tripmonth = 'July' THEN 1
       WHEN tripmonth = 'August' THEN 2
       WHEN tripmonth = 'September' THEN 3
       WHEN tripmonth = 'October' THEN 4 END
```

Analyze

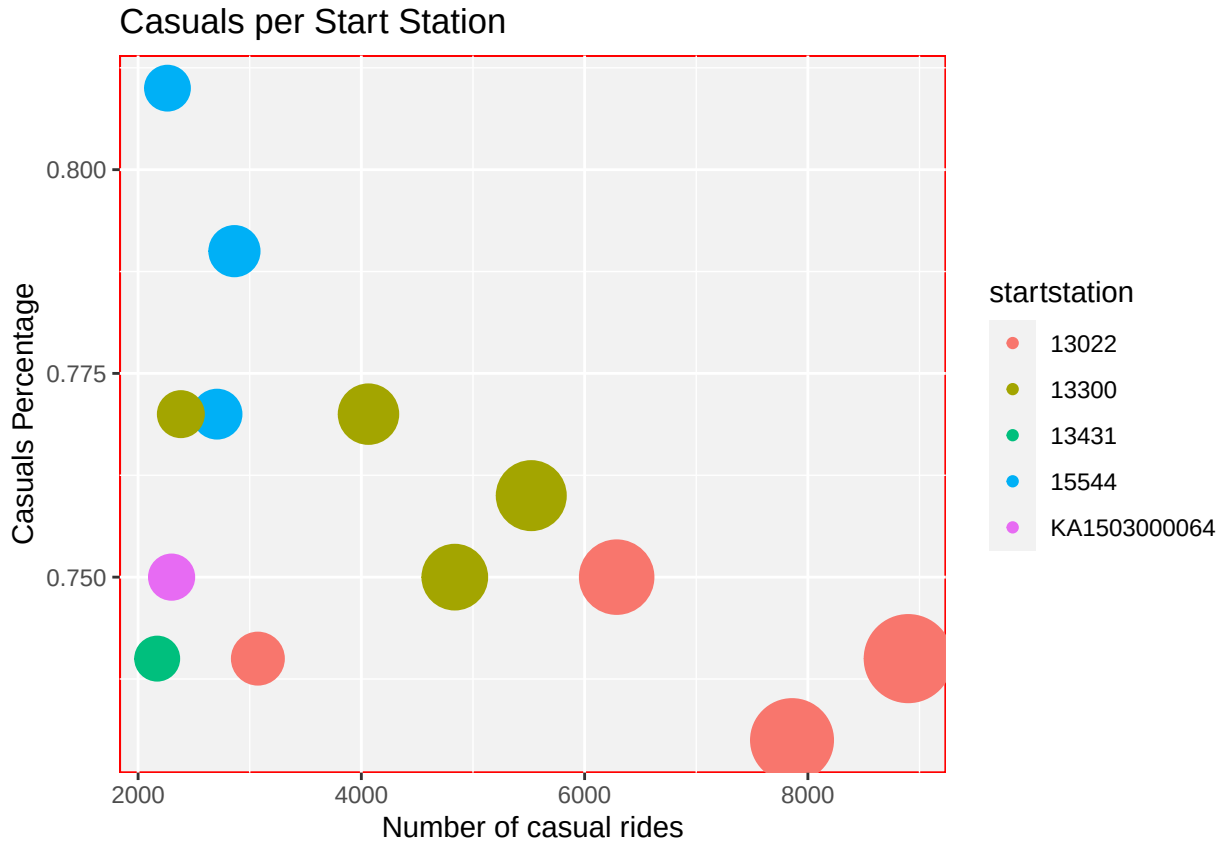
Casuals per station

```
fa <- fullstations4 %>%
  filter(casual_ratio > 0.2) %>%
  filter(casuals > 1000)
ggplot(data = fa) +
  geom_point(mapping = aes( x = casuals, y = casual_ratio, size = casuals, color = startstation))+
  scale_size_area(max_size = 10)+
  guides(size = "none", color = "none")+
  theme(panel.background = element_rect(fill = 'gray95', colour = 'red'))+
  labs(title = 'Casuals per Start Station')+
  labs(x = 'Number of casual rides')+
  labs(y = 'Casuals Percentage')+
  annotate("rect", xmin = 800, xmax = 9500, ymin = 0.70, ymax = 0.82, alpha = 0.5, color = "red", fill =
```



Causals per start station filtered

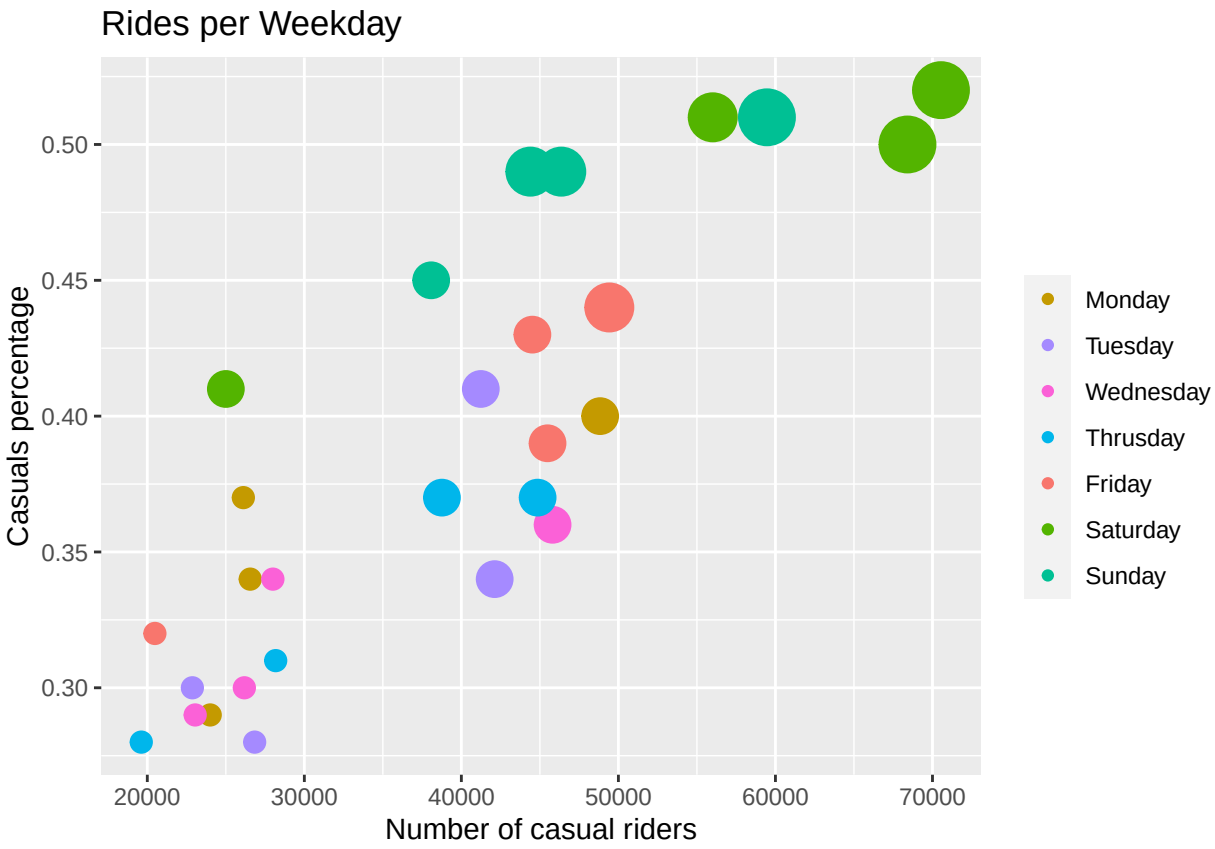
```r
# Plot Stations with more than 72% of casuals
fa <- fullstations4 %>%
  filter(casual_ratio > 0.72) %>%
  filter(casuals > 2000)
ggplot(data = fa) +
  geom_point(mapping = aes( x = casuals, y = casual_ratio, size = casuals, color = startstation))+
  scale_size_area(max_size = 15)+
  guides(size = "none")+
  theme(panel.background = element_rect(fill = 'gray95', colour = 'red'))+
  labs(title = 'Casuals per Start Station')+
  labs(x = 'Number of casual rides')+
  labs(y = 'Casuals Percentage')
```



Analyze Rides per weekday

```r
library(tidyverse)
library(ggplot2)
library(lubridate)
setwd("D:/Portofolio2")

weekday3 <- read.csv("weekday.csv")

ggplot(data=weekday3) +
  geom_point(mapping = aes(x = casuals, y = caspercent, color=weekdaytrip, size=casuals*caspercent))+
  guides(size= "none")+
  scale_color_discrete(breaks=c('Monday', 'Tuesday', 'Wednesday', 'Thrusday', 'Friday', 'Saturday', 'Su
```

```
  scale_size_binned(range=c(1,10))+
  labs(x='Number of casual riders')+
  labs(y = 'Casuals percentage')+
  labs(title = 'Rides per Weekday')+
  labs(shape="")+
  labs(color="")
```



Analyze rides per hour

```
# Set libraries
library(tidyverse)
library(ggplot2)
library(lubridate)
library(ggforce)
# Import the dataset
setwd("D:/Portofolio2")
hour4 <- read.csv("hour_analysis.csv")
# Generate Plot
ggplot(data = hour4) +
  geom_point(mapping = aes( x = starthour, y = caspercent, size = caspercent, color = tripmonth))+
  scale_size_area(max_size = 5)+
  guides(size = "none")+
  theme(panel.background = element_rect(fill = 'gray95', colour = 'red'))+
  labs(title = 'Casuals per Hour')+
  labs(x = 'Hour of the day')+
  labs(y = 'Casuals Percentage')+
```

```
  labs(color = "")+
  scale_color_discrete(breaks=c('July', 'August', 'September', 'October'))+
  annotate("rect", xmin = -0.35, xmax = 4.55, ymin = 0.44, ymax = 0.612, alpha = 0.2, color = "red", fi
```