Case Study

Case Study: How Does a Bike-Share Navigate Speedy Success?

Introduction

This case study is my capstone project. This is my first ever case study and I will be assuming the role of a junior data analyst working in the marketing analyst team at a fictional company, Cyclistic; a bike-share company in Chicago.

About the company

Cyclistic is a fictional bike sharing company that operates a fleet of over 5,800 bicycles that are geotracked and locked into a network of over 600 stations across Chicago. The bikes can be accessed from one station and returned to any other station in the system anytime.

Over the years, Cyclistic's marketing strategy depends on building general awareness and appeal to broad consumers. Cyclistic'sss flexible pricing plans are: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as **casual riders**. Customers who purchase annual memberships are Cyclistic **members.**

Scenario

The marketing team wants to understand how casual riders and annual members use Cyclistic bikes differently, and they are also interested in creating a campaign to convert casual riders into annual members. The team is interested in analyzing the historical bike trip data to identify trends in the usage of bikes by casual and member riders.

Ask

Stakeholders

The stakeholders of this project are:

Lily Moreno, Director of Marketing at Cyclistic

The Cyclistic Marketing Analytics Team

The Cyclistic Exceutive team

Business Task:

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

At the end of this analysis, these questions will be answered with recommendations and compelling visuals of the data provided.

Prepare

The dataset used for this project is Cyclistic's historical trip from the previous year (July 2021 – June 2022), the data was made available for public use by Motivate International Inc.

Credibility of the data was verified using the ROCCC approach

Reliable: The data is complete and precise, It is relevant for the purpose of the analysis.

Original: The dataset was made available for public use by Motivate International Inc. which operates the city of Chicago's Divvy bicycle sharing service which is powered by Lyft.

Comprehensive: The data includes all information for rides taken on the system such as rider ID, start time, end time, start and end station name and ID, start and end longitude/latitude, membership type.

Current: The data is current as it is released monthly, as at July 2022, data was current to June 2022.

Cited: The data is cited and made available for public use by Motivate International Inc.

Data Limitations

A quick check and filtering of the data showed that data for some rides (mostly electric bikes) are missing in some fields, which will be taken into consideration during the data cleaning and analysis process.

Process

The tools I will be using for this project is spreadsheet and R for cleaning, analysis and visualization. Tableau will be used for some of the visualization.

Data Review

The dataset is made available for each month, thus making 12 CSV files which contains over 5 million observations and 13 fields.

The individual dataset will be imported into R and merged into a data frame to make the cleaning process easier.

The dataset was reviewed and cleaned using R due to its size and also to have an overall understanding of the fields, records and data formats. Three new column was created (ride length, month and day of the week), these new fields will provide further insight in the analysis phase.

```r
#load all packages needed for the cleaning and analysis process
library(tidyverse)
library(janitor)
library(here)
library(skimr)
library(lubridate)

#import the files singly via read_csv
jul <- read_csv("C://202107-divvy-tripdata.csv")
aug <- read_csv("C://202108-divvy-tripdata.csv")
sep <- read_csv("C://202109-divvy-tripdata.csv")
oct <- read_csv("C://202110-divvy-tripdata.csv")
nov <- read_csv("C://202111-divvy-tripdata.csv")
dec <- read_csv("C://202112-divvy-tripdata.csv")
jan <- read_csv("C://202201-divvy-tripdata.csv")
feb <- read_csv("C://202202-divvy-tripdata.csv")
mar <- read_csv("C://202203-divvy-tripdata.csv")
apr <- read_csv("C://202204-divvy-tripdata.csv")
may <- read_csv("C://202205-divvy-tripdata.csv")
jun <- read_csv("C://202206-divvy-tripdata.csv")
```

```r
#merging the indiviual datasets into one using rbind for easy view and analysis

all_trips <- rbind(jul, aug, sep, oct, nov, dec, jan, feb, mar, apr, may, jun)

#preview the merged data
ncol(all_trips) #to view number of columns
nrow(all_trips) #to view number of rows
str(all_trips) # to view columns and data types
colnames(all_trips) #list of column names
head(all_trips) #to preview the first 6 rows
tail(all_trips) #to preview the last 6 rows

#Having reviewed the data set and fields, there is a need to create two new columns
#named "ride_length" and "day_of_week", this is to give more insight into the
#dataset.

#create new column named "ride_length" in mins
all_trips$ride_length <- difftime(all_trips$ended_at, all_trips$started_at, units = "mins")

#round up ride_length to 1 decimal place
all_trips$ride_length <- round(all_trips$ride_length, digits = 1)
```

```r
#create a new column - "day_of_week"
all_trips$day_of_week <- wday(all_trips$started_at)

#assiging the dataframe to a new variable
all_trips1 <- all_trips

#For easy analysis, let's create a column for month
all_trips1$month <- format(as.Date(all_trips1$started_at), "%m")
```

```r
#The month column has been created but each month is represented by its number
#Let's change it from number to character i.e from 06 to "June"

all_trips1 <- all_trips1 %>%
  mutate(month = case_when(month == "01" ~ "January",
                           month == "02" ~ "February",
                           month == "03" ~ "March",
                           month == "04" ~ "April",
                           month == "05" ~ "May",
                           month == "06" ~ "June",
                           month == "07" ~ "July",
                           month == "08" ~ "August",
                           month == "09" ~ "September",
                           month == "10" ~ "October",
                           month == "11" ~ "November",
                           month == "12" ~ "December"))
```

```r
#Let's do the same for day_of_week column
all_trips1 %>%
  mutate(day_of_week = case_when(day_of_week == "1" ~ "Sunday",
                                 day_of_week == "2" ~ "Monday",
                                 day_of_week == "3" ~ "Tuesday",
                                 day_of_week == "4" ~ "Wednesday",
                                 day_of_week == "5" ~ "Thursday",
                                 day_of_week == "6" ~ "Friday",
                                 day_of_week == "7" ~ "Saturday"))
```

```r
#previewing the data to confirm all columns have the correct data types
str(all_trips1)

#Remove rows with NA values
all_trips1 <- na.omit(all_trips1)

#Remove duplicate rows
all_trips1 <- distinct(all_trips1)

#Preview cleaned data
View(all_trips1)
```

ANALYZE

The data has been cleaned and descriptive analysis will be done using R.

```
#Descriptive analysis of cleaned data
#mean, median, max, min of the ride_length
mean(all_trips1$ride_length)
max(all_trips1$ride_length)
median(all_trips1$ride_length)
min(all_trips1$ride_length)

#Average ride_length for members and casuals

all_trips1 %>%
  group_by(member_casual) %>%
  summarize(avg_ride_length = mean(ride_length))

#Average ride_length for users by day_of_week

all_trips1 %>%
  group_by(day_of_week) %>%
  summarise(avg_ride_length = mean(ride_length))
```

```
#Average ride_length for members and casuals by day_of_week

all_trips1 %>%
  group_by(day_of_week, member_casual) %>%
  summarise(avg_ride_length = mean(ride_length))

#The day_of_week needs to be ordered, it should be arranged
#since it is a categorical data, else our results will look distorted.

all_trips1$day_of_week <- ordered(all_trips1$day_of_week, levels =
                                    c("Sunday", "Monday",
                                      "Tuesday", "Wednesday",
                                     "Thursday", "Friday",
                                     "Saturday"))

#Let's Check if the day_of_week is now ordered
is.ordered(all_trips1$day_of_week)
```

```
#count of number of rides per user by day_of_week
all_trips1 %>%
  group_by(day_of_week, member_casual) %>%
  summarise(no_of_rides = n())

#membership size
all_trips1 %>%
  group_by(member_casual) %>%
  summarise(no_of_memberships = n())

#Let"s export the summary file in form of csv for visualiaztion in Tableau

write.csv(cyclistic_data, "cyclistic_trips.csv")
```

From the above analysis, the average ride length is 19 mins with casual riders having the longest average ride length of 28 mins. Casuals had the most rides in total for the period under review.

Classic bikes are the most popular bike type for both user types, while casuals use all bike type, members prefer classic and electric bikes.

Casual rider has the longest rides everyday compared to members, with Sunday being the peak day for casuals while members ride equally through the week with a slight increase on Sunday.

The busiest month for both user type is July – September, while the busiest month for members is August, Casual riders' peak period is July.

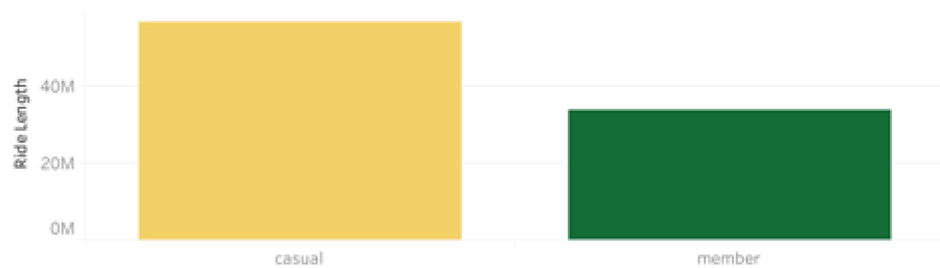
SHARE

Visualization of the data using Tableau

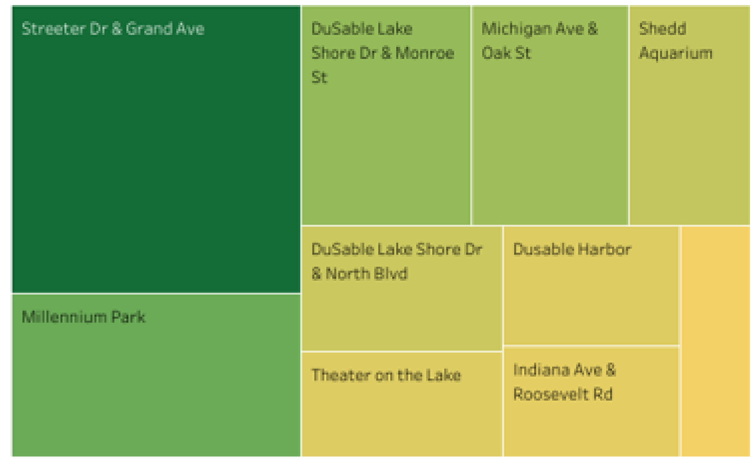
Google Data Analytics Capstone Project

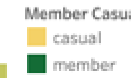
Analysis of Cyclistic, a bike sharing company

Total Ride Length by User Type

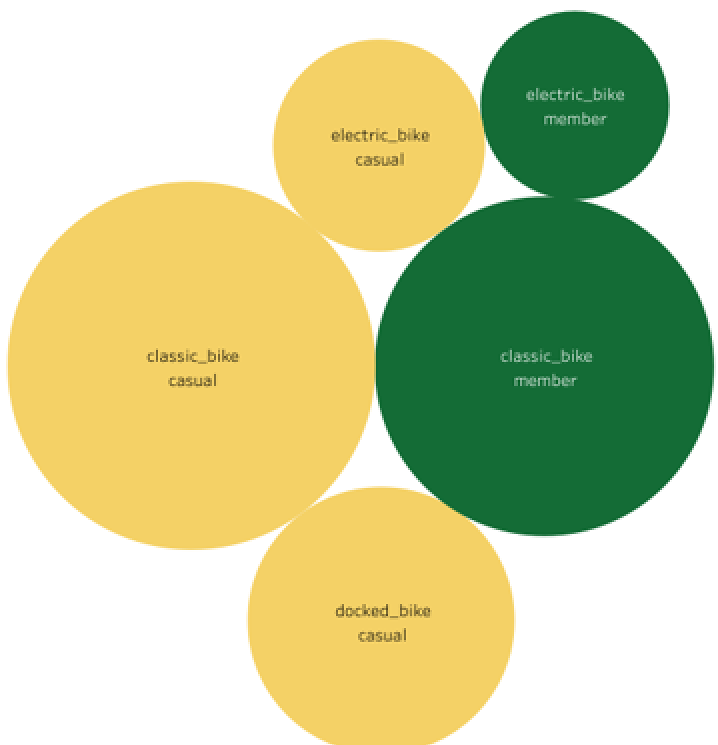


Popular Start Station for Casuals

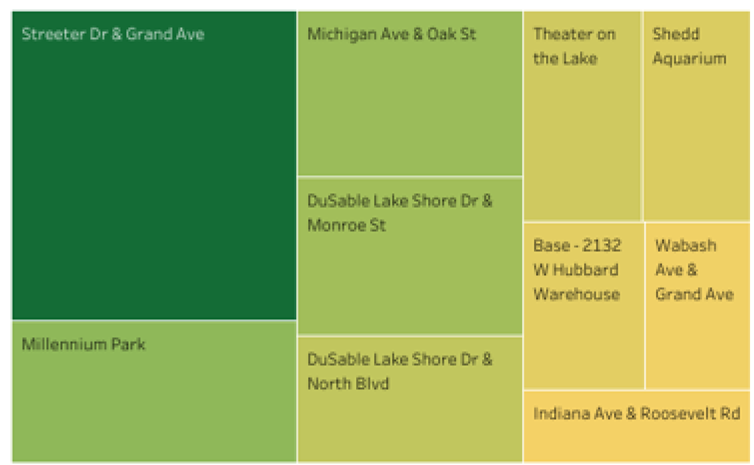




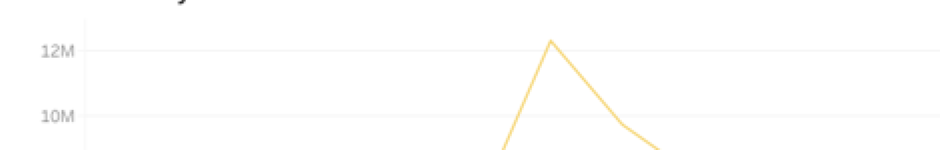
Bike Usage by User Type



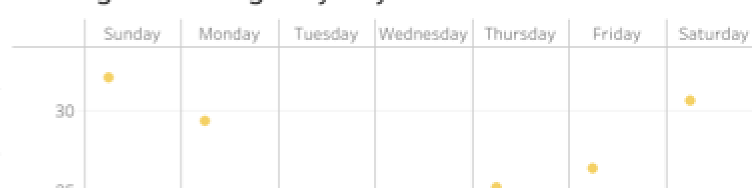
Popular End Station for Casuals



Total Rides by Month



Average Ride Length by Days



ACT

Recommendations based on the analysis and visualizations.

- Offer membership program discounts on peak days for casual riders.
- Enhance membership program benefits, details of enhancement can be printed on the back of receipts or tickets issued to customers at the point of purchase/payment at the popular stations known for casual rider influx to create awareness.
- Increase the availability of classic bikes - preferred bike type for casuals; Introduce different membership scheme to cater for low-income earners, and also introduce loyalty points which can be tied to the different schemes.

Thank you.