How Does a Bike-Share Navigate Speedy Success?

Google Data Analytics Capstone

S.M.M.Shamry

25th October 2022

Table of Contents

1 Introduction 3

1.1 Analytics tools 3

2 Ask 3

3 Prepare and process 4

3.1 R studio 4

4 Analyze 7

4.1 Microsoft Excel 7

4.1.1 Min & Max Ride length and Total Customer 7

4.1.2 Rides in weekdays 8

4.1.3 Average Ride Length 9

4.1.4 Average Ride Length by Days 10

4.1.5 Month ride Counts 11

4.1.6 Seasons Summary 12

5 Share 13

6 Conclusions 14

7 Recommendations 14

1 Introduction

Started Google data analysis project to improve skills and I felt more interest on that and finished the all the seven course and final capstone project. Want inform about capstone project I used R programing, Microsoft Excel and Tableau tools for analysis used,

1.1 Analytics tools

Used R studio for data cleaning and data manipulations, used Microsoft Excel for data analysis and used Tableau for data visualization. Used GitHub and medium for publish the project.

2 Ask

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: 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.

Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could

affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

Moreno has assigned answer the question about;

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

3 Prepare and process

For conducted an analysis on Cyclistic's historical trip data. The data has been made available by Motivate International Inc. under this license. Initially, View the data and inspect the variables and observation, there were some null values and negative ride length. And want to add some variables they were "ride_length" and "day_of_week".

As mentioned on introduction I were used R studio, Microsoft Excel and Tableau

3.1 R studio

Here is the R code Click

Used R programing for data cleaning and data manipulation. Initially, I install the packages and library

```
## installing the necessary packages
install.packages("tidyverse")
install.packages("lubridate")

## loading the packages
library(tidyverse)
library(lubridate)
library(ggplot2)
```

Importing the data sets and viewing the each data sets

```
## importing the data sets
january <- read.csv("C:/Users/Desktop/Excerciess/Data_analiysis_Project/2021/202101-divvy-tripdata.csv")
february <- read.csv("C:/Users/Desktop/Excerciess/Data_analiysis_Project/2021/202102-divvy-tripdata.csv")
march <- read.csv("C:/Users/Desktop/Excerciess/Data_analiysis_Project/2021/202103-divvy-tripdata.csv")
april <- read.csv("C:/Users/Desktop/Excerciess/Data_analiysis_Project/2021/202104-divvy-tripdata.csv")
may <- read.csv("C:/Users/Desktop/Excerciess/Data_analiysis_Project/2021/202105-divvy-tripdata.csv")
june <- read.csv("C:/Users/Desktop/Excerciess/Data_analiysis_Project/2021/202106-divvy-tripdata.csv")
july <- read.csv("C:/Users/Desktop/Excerciess/Data_analiysis_Project/2021/202107-divvy-tripdata.csv")
august <- read.csv("C:/Usersy/Desktop/Excerciess/Data_analiysis_Project/2021/202108-divvy-tripdata.csv")
september <- read.csv("C:/UsersDesktop/Excerciess/Data_analiysis_Project/2021/202109-divvy-tripdata.csv")
october <- read.csv("C:/UsersDesktop/Excerciess/Data_analiysis_Project/2021/202110-divvy-tripdata.csv")
november <- read.csv("C:/UsersDesktop/Excerciess/Data_analiysis_Project/2021/202111-divvy-tripdata.csv")
december <- read.csv("C:/Users/Desktop/Excerciess/Data_analiysis_Project/2021/202112-divvy-tripdata.csv")

## Discovering the columns in each data sets
colnames(january)
colnames(february)
colnames(march)
colnames(april)
colnames(may)
colnames(june)
colnames(july)
colnames(august)
colnames(september)
colnames(october)
colnames(november)
colnames(december)
```

Regarding on my PC performance combined the data sets into 3 sets that are q_1, q_2, q_3 on each sets did rbind the data set, identified the column type, removed the (start_station_name, start_station_id, start_station_id,e nd_station_name, end_station_id, start_lat, start_lng,end_lat, end_lng) coloumn and created the new coulumn as (ride_length and “day_of_week).

```
## combining the data sets into one data set
q_1 <- rbind(january,february,march,april)
## to identify the column type
str(q_1)
## Removing the following column
q_1 <- q_1 %>%
  select(-c(start_station_name,start_station_id,start_station_id,end_station_name, end_station_id,start_lat,start_lng,end_lat,end_lng))
## calculating the ride_length
q_1 <- q_1 %>%
  mutate("ride_length"=(difftime(ended_at,started_at, units = "mins")))
## finding the day
q_1 <- q_1 %>%
  mutate("day_of_week"= format(as.Date(started_at), "%A"))

##q_2
## combining the data sets into one data set
q_2 <- rbind(may,june,july,august)
## to identify the column type
str(q_2)
## Removing the following column
q_2 <- q_2 %>%
  select(-c(start_station_name,start_station_id,start_station_id,end_station_name, end_station_id,start_lat,start_lng,end_lat,end_lng))
## calculating the ride_length
q_2 <- q_2 %>%
  mutate("ride_length"=(difftime(ended_at,started_at, units = "mins")))
## finding the day
q_2 <- q_2 %>%
  mutate("day_of_week"= format(as.Date(started_at), "%A"))

##q_3
## combining the data sets into one data set
q_3 <- rbind(september,october,november,december)
## to identify the column type
str(q_3)
## Removing the following column
q_3 <- q_3 %>%
  select(-c(start_station_name,start_station_id,start_station_id,end_station_name, end_station_id,start_lat,start_lng,end_lat,end_lng))
## calculating the ride_length
q_3 <- q_3 %>%
  mutate("ride_length"=(difftime(ended_at,started_at, units = "mins")))
## finding the day
q_3 <- q_3 %>%
  mutate("day_of_week"= format(as.Date(started_at), "%A"))
```

Combined the q_1, q_2, q_3 data sets removed annual Null values, duplicates and ride length less than and equal to zero. Finally export the data to analyses on Excel and Tableau

```
##binding the All data
annual_data <- rbind(q_1,q_2,q_3)

## removing NA
annual_data <- annual_data %>%
  na.omit()
## Removing Duplicates
annual_data <- annual_data %>%
  distinct()

## removing the ride_length less than or equal to zero
annual_data <- annual_data [!(annual_data$ride_length <= 0),]

## Exporting the data set analysis in Excel & Tableau
write.csv(annual_data, file = "annual_data", row.names = FALSE)
```

4 Analyze

For the analyze data, used Microsoft Excel.

4.1 Microsoft Excel

Here is the Excel Image Click

On the Excel we can't import more Than 5 million rows because of that used Excel power query to make connection on the data sets and analyze it. Excel power query is great feature.

4.1.1 Min & Max Ride length and Total Customer

Rider's minimum ride length 0.017 minutes and maximum ride length 55944.15 minutes, Total number of riders 5,594,410 on that Casual 2,528,664 and members 3,065,746.

Min ride length (mins)

	casual	member	Grand Total
Min of ride_len	0.017	0.017	0.017

Max ride length (mins)

	casual	member	Grand Total
Max of ride_le	55944.15	1559.93	55944.15

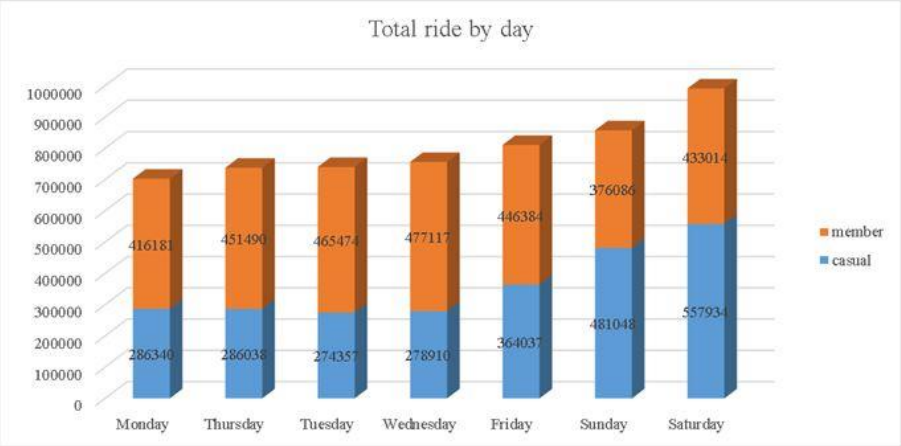
Total coustomers

	casual	member	Grand Total
Count of ride_i	2528664	3065746	5594410

4.1.2 Rides in weekdays

Most of the Casuals rides on Saturdays that 557,934 and Most of the Members rides on Wednesdays that 477,117.

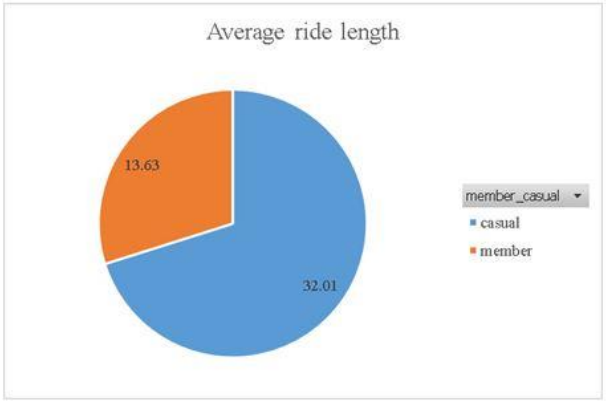
Day rides		
Count of day_of_week		
	casual	**member**
Monday	286340	416181
Thursday	286038	451490
Tuesday	274357	465474
Wednesday	278910	477117
Friday	364037	446384
Sunday	481048	376086
Saturday	557934	433014



4.1.3 Average Ride Length

The Casual Average ride length 32.01 minutes and Members Average ride length 13.63 minutes Total Average ride length 21.94 minutes.

Average ride length	
	Average of ride_length
casual	32.01
member	13.63
Grand Total	**21.94**

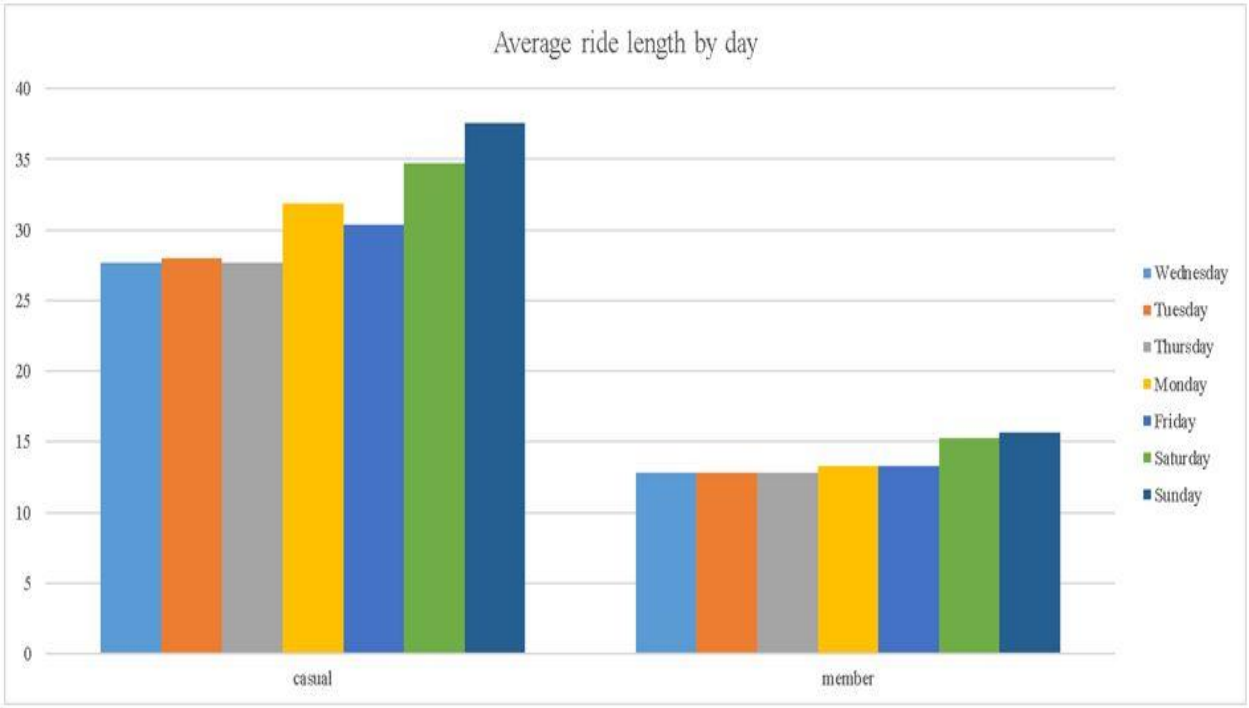


4.1.4 Average Ride Length by Days

Average ride length by days both Casual and Members were ride In Sundays 27.95 minutes in below chart we can see clear different between Casual average ride length in a day and Members average ride length in a day.

Average ride length by days

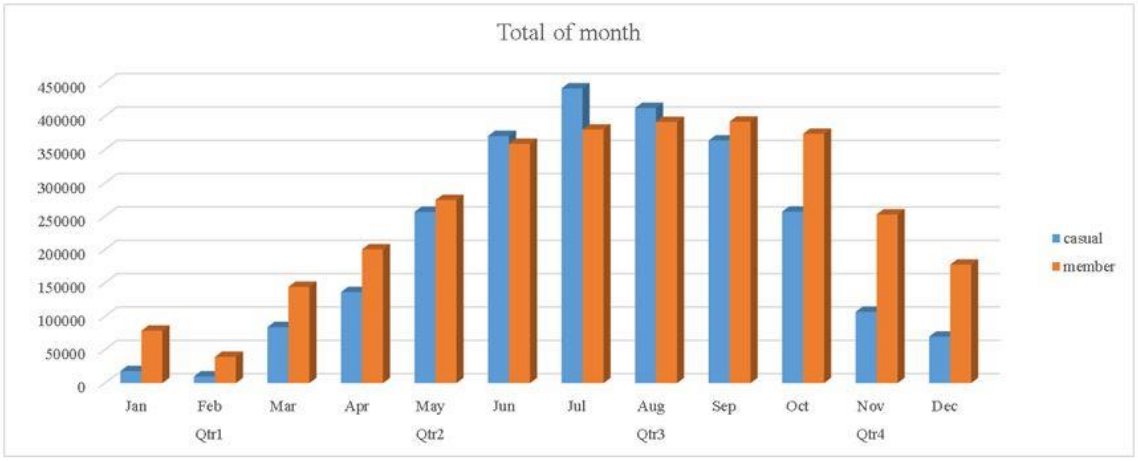
Average of ride_length								
	Wednesday	Tuesday	Thursday	Monday	Friday	Saturday	Sunday	Grand Total
casual	27.66107382	27.9754879	27.7057761	31.87912185	30.35176786	34.70996644	37.5709556	32.00577995
member	12.82020703	12.7890845	12.77703264	13.24825985	13.32608027	15.26552217	15.65922923	13.63454515
Grand Total	**18.29523128**	**18.42077076**	**18.56689814**	**20.84199872**	**20.97393273**	**26.21333837**	**27.95671381**	**21.93831267**



4.1.5 Month ride Counts

The most rides were ride in July that 2,382,613 and lest number of rides in February the 3rd quarter was the busiest quarter in 2021.

Month rides			
Count of ride_id			
	casual	**member**	**Grand Total**
Qtr1	**112275**	**262655**	**374930**
Jan	18117	78711	96828
Feb	10130	39488	49618
Mar	84028	144456	228484
Qtr2	**764114**	**834186**	**1598300**
Apr	136590	200602	337192
May	256888	274691	531579
Jun	370636	358893	729529
Qtr3	**1218459**	**1164154**	**2382613**
Jul	442011	380317	822328
Aug	412608	391637	804245
Sep	363840	392200	756040
Qtr4	**433816**	**804751**	**1238567**
Oct	257203	373953	631156
Nov	106884	253008	359892
Dec	69729	177790	247519
Grand Total	**2528664**	**3065746**	**5594410**

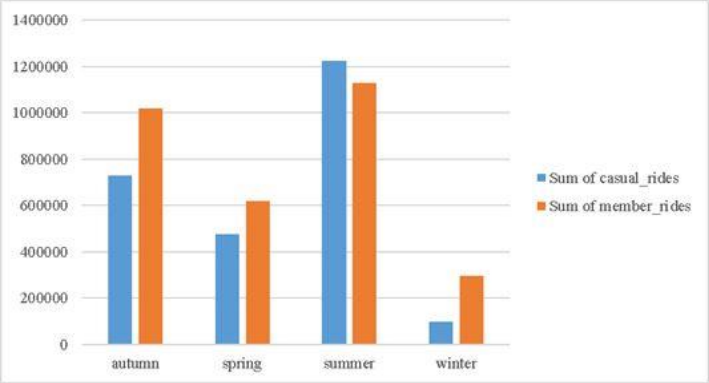


4.1.6 Seasons Summary

Most of ride were ride in summer season and least of ride were ride in winter season.

Seasons total

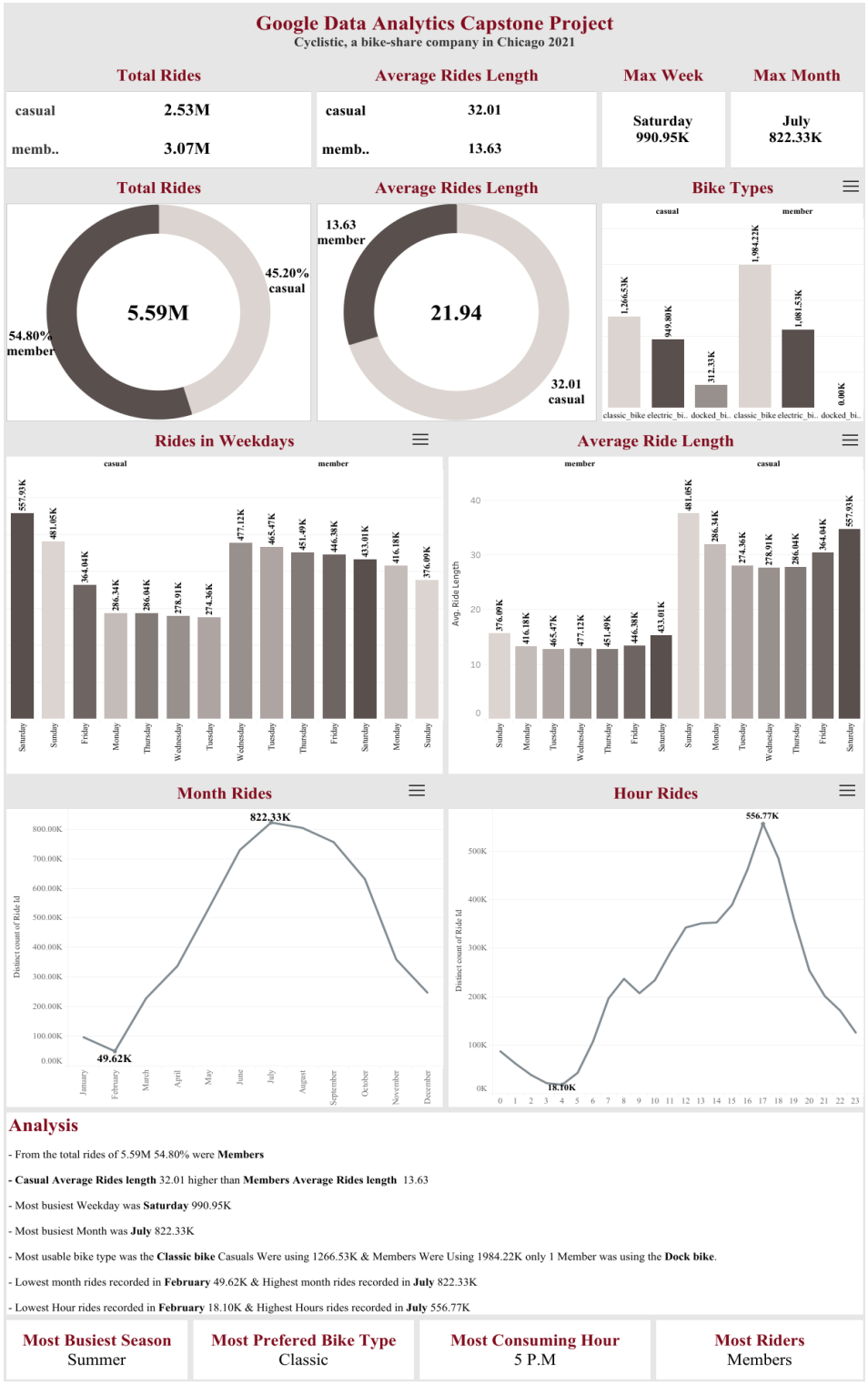
Row Labels	Sum of casual_rides	Sum of member_rides
autumn	727927	1019161
spring	477506	619749
summer	1225255	1130847
winter	97976	295989
Grand Total	**2528664**	**3065746**



5 Share

Here is final dashboard Click.

Used Tableau to visualize and present the data to research team.



6 Conclusions

Rider's minimum ride length 0.017 minutes and maximum ride length 55944.15 minutes, Total number of riders 5,594,410 on that Casual 2,528,664 and Members 3,065,746. **The most of the riders Were Members.**

Most of the Casuals rides on Saturdays that 557,934 and Most of the Members rides on Wednesdays that 477,117. **Casual riders ride more than Member riders in Weekdays**.

The Casual Average ride length 32.01 minutes and Members Average ride length 13.63 minutes Total Average ride length 21.94 minutes. **The Casual riders using services more than Member riders in 2021.**

Average ride length by days both Casual and Members were ride In Sundays 27.95 minutes in below chart we can see clear different between Casual average ride length in a day and Members average ride length in a day. **The Casual riders using services more than Member riders in day.**

The most rides were ride in July that 2,382,613 and lest number of rides in February the 3rd quarter was the busiest quarter in 2021. **Busiest month was July**.

Most of ride were ride in summer season and least of ride were ride in winter season. **More riders using the service in summer seasons.**

7 Recommendations

This study revealed the study on Members and Casual riders use Cyclistic bikes differently and presented some recommendations for the Design marketing strategies aimed at converting casual riders into annual members.

1. Notify: have to inform about annual member packages to the casual riders for that all the riders are coming to the services stations we can implement some marketing strategy in service stations.
    a. Displaying Billboards
    b. Distributing leaflet
    c. Conducting events
2. Sales promotions: Giving discounts and special offers for Annual members.
3. Social media Marketing: E mail marketing, Facebook marketing, video marketing, tele marketing can use this kind of strategies to attract the Casual riders.
4. Can Give Seasonal (summer) offer for the Annual members.