DIVVY BIKE SHARE

Google Capstone Project

December 31, 2022

Prepared By: K. Jones
Prepared For: Cylistic Bike Share

TABLE OF CONTENTS

INTRODUCTION

1. Statement of the Problem
2. Significance of the study
3. Deliverables
4. Scope of the Study

METHODS OF ANALYSIS

1. Sources of Data and Safety
2. Licenses
3. Cleaning and Manipulation

DATA FINDINGS AND CONCLUSIONS

1. Summary of Analysis
2. Supporting Visualizations and Key Findings
3. Recommendations

INTRODUCTION

In effort to secure longevity in the fast-paced app market of today, it is crucial to understand the habits and needs of potential members. In the past, local transportation relied solely on cab, bus and train services that relied on space capacity and operation schedules to steer annual projections, the current market can use real-time data to personalize services with the privilege of classic and electric bikes offering a more intimate relationship in a service-based market allowing for more accurate projections in shorter timeframes.

Ensuring sustainability in the future in a ever-growing market, Cyclistic , wants to understand the "why" and "how" of it's casual riders and their differences to membership holders.

1. **Statement of the Problem**
   How do annual members and casual riders use Cyclistic bikes differently?

   Objective:
   Maximize the number of annual memberships.

2. **Significance of the Study**
   Understanding the difference between members and casual riders will lead marketing strategies to increases annual membership by converting casual riders to members.

3. **Deliverables**
   Design marketing strategies aimed at converting casual riders into annual members.
   Understand how annual members and casual riders differ.
   Why would casual riders buy a membership?
   How digital media could affect their marketing tactics.
   Analyze Cyclistic historical bike trip data to identify trends.

4. **Scope of Study**
   The study is limited to users of the Divvy Bike share only. For the purposes of this study, casual rider shall be defined as riders who do not participate in the annual membership program. Also, total ride times will be determined by the column of the rider's status.

METHODS OF ANALYSIS

1. Source of Date and Storage

The date used in this analysis is collected in real time from Cylistic's records of Divvy trips or rentals. The records are both quarterly and monthly. Each record should contain geographical information, rider status, dates, bike type, ride times and bike Id. It should be noted that all records are not complete, and some locations are either no longer in existence or new in development. Personal identifying information about each rider is not recorded in the datasets.

Cyclistic historical trip data:
https://divvy-tripdata.s3.amazonaws.com/index.html

- October 2021
- December 2021
- April 2022
- July 2022

Datasets were:

- downloaded locally
- converted as Excel files
- uploaded into SQL and saved into a single data frame
- Collection of data frame saved as XML file and stored locally.

2. Licenses

Open license:

Motivational International Inc:

https://ride.divvybikes.com/data-license-agreement

3. Cleaning and Manipulation

Cleaning:

- Check for null values and exclude them from the table. Null values will cause errors in queries, null values shown for popular ride day excluded from results.

```
--Check each column for nulls

select *
from DivvyBikeRiders
where Start_Date is null
-- where Ride_type is null
-- where status is null
-- where start_time is null
-- where end_time is null
-- where Start_Date is null
-- where End_Date  is null
order by Status
```

- Rentals less than 1 minute in length were removed as test rides or rider error and rides greater than 300 minutes due to possible docking errors.

```
--Delete rides less than 1 minute from temp table but more than 300 ride minutes

delete
from DivvyBikeRiders
where TotalRideMins <= 1 or TotalRideMins > 300
```

- Ride type queries do not include rides less than 5 mins to focus on riders who ride for a longer time based on the bike chosen. Possibly identifying classic riders who ride for exercise in comparison to electric riders who ride for social scenery or physical assistance.

```
--Total number of rentals more than 5 mins per ride type

select  Ride_Type, count(Ride_type) as MemberRentals
from DivvyBikeRiders
where Status = 'member' and totalridemins >= '5'
group by Ride Type

select Ride_Type, count(Ride_type) as CasualRentals
from DivvyBikeRiders
where Status = 'casual' and totalridemins >= '5'
group by Ride_Type
```

Manipulation:

- Prepare date and time to create accurate start and end time columns for each month representing an assigned season- October = Fall, December = Winter, April = Spring, July = Summer. *October is the only month needing an update.

```
--Seperate date and time and add Start_Time and End_Time columns for October
select CONVERT(nvarchar(20), Start_Date, 108) AS Start_Time,
CONVERT(nvarchar(20), End_Date, 108) AS End_Time
from Q4Oct2021

alter table Q4Oct2021
add Start_Time datetime, End_Time datetime;

update Q4Oct2021
set Start_Time = CONVERT(nvarchar(20), Start_Date, 108),
    End_Time = CONVERT(nvarchar(20), End_Date, 108);
```

- Create a temp table with columns needed for queries based on ride types, status, start and end date, start and end time.

```
--Create Table

create table DivvyBikeRiders(
Ride_Type nvarchar(255),
Status nvarchar(255),
Start_Date datetime,
End_Date datetime,
Start_Time datetime,
End_Time datetime)

select *
from DivvyBikeRiders

--Add data to table
insert into DivvyBikeRiders

select Ride_Type, Status, Start_Date, End_Date, Start_Time, End_Time
from Q3July2022
union

select Ride_Type, Status, Start_Date, End_Date, Start_Time, End_Time
from Q2Apr2022
union

select Ride_Type, Status, Start_Date, End_Date, Start_Time, End_Time
from Q4Oct2021
union

select Ride_Type, Status, Start_Date, End_Date, Start_Time, End_Time
from Q1Dec2021
;
```

- Add a column that calculates the total ride time in minutes.

```
--Use Start_Time and End_Time to calculate Total Ride Minutes

select Start_Time, End_Time, DATEDIFF(MINUTE,Start_Time, End_Time) as TotalRideMins
From DivvyBikeRiders

--Add RideMins column to table

alter table DivvyBikeRiders
add TotalRideMins int;

-- Add data to new column

update DivvyBikeRiders
set TotalRideMins =  DATEDIFF(minute,Start_Time, End_Time)

select*
from DivvyBikeRiders

--Delete rides less than 1 minute from temp table but more than 300 ride minutes

delete
from DivvyBikeRiders
where TotalRideMins <= 1 or TotalRideMins > 300
```

- Separate date into 3 separate columns.

```
--Seperate Date for future use and update table

select datepart(YEAR, Start_date) as Year,
       datepart(MONTH, Start_date) as Month,
       datepart(DAY, Start_date) as Day
from DivvyBikeRiders

alter table DivvyBikeRiders
add Year int, Month int, Day int;


Update DivvyBikeRiders
set Year = datepart(YEAR, Start_Date),
    Month = datepart(month, Start_date),
    Day = datepart(day, Start_date);

select*
from DivvyBikeRiders
```

- Update day and month from int to char.

```
-- Change Day for int to char and update table

select Day,
case
when Day = '1' then 'Sunday'
when Day = '2' then 'Monday'
when Day = '3' then 'Tuesday'
when Day = '4' then 'Wednesday'
when Day = '5' then 'Thursday'
when Day = '6' then 'Friday'
when Day = '7' then 'Saturday'
end as RideDay
from DivvyBikeRiders

alter table DivvyBikeRiders
add RideDay nvarchar(255);

update DivvyBikeRiders
set RideDay = (
case when Day = '1' then 'Sunday'
when Day = '2' then 'Monday'
when Day = '3' then 'Tuesday'
when Day = '4' then 'Wednesday'
when Day = '5' then 'Thursday'
when Day = '6' then 'Friday'
when Day = '7' then 'Saturday'
end)

select *
from DivvyBikeRiders
```

- Add a column to reflect seasonal use.

```
--Add column to reflect season
select MONTH,
case
when MONTH = '12' then 'Winter'
when MONTH = '4' then 'Spring'
when MONTH = '7' then 'Summer'
when MONTH = '10' then 'Autumn'
end as RideSeason
From DivvyBikeRiders

 alter table DivvyBikeRiders
add RideSeason nvarchar(255);


Update DivvyBikeRiders
set  RideSeason = (
case
when MONTH = '12' then 'Winter'
when MONTH = '4' then 'Spring'
when MONTH = '7' then 'Summer'
when MONTH = '10' then 'Autumn'
end)

select *
from DivvyBikeRiders
```

DATA FINDINGS AND CONCLUSIONS

1. Summary of analysis:

```
-- Total number of rides by status

select Status, count(status) as RideTotals
from DivvyBikeRiders
where ride_type != 'docked_bike'
group by status
```

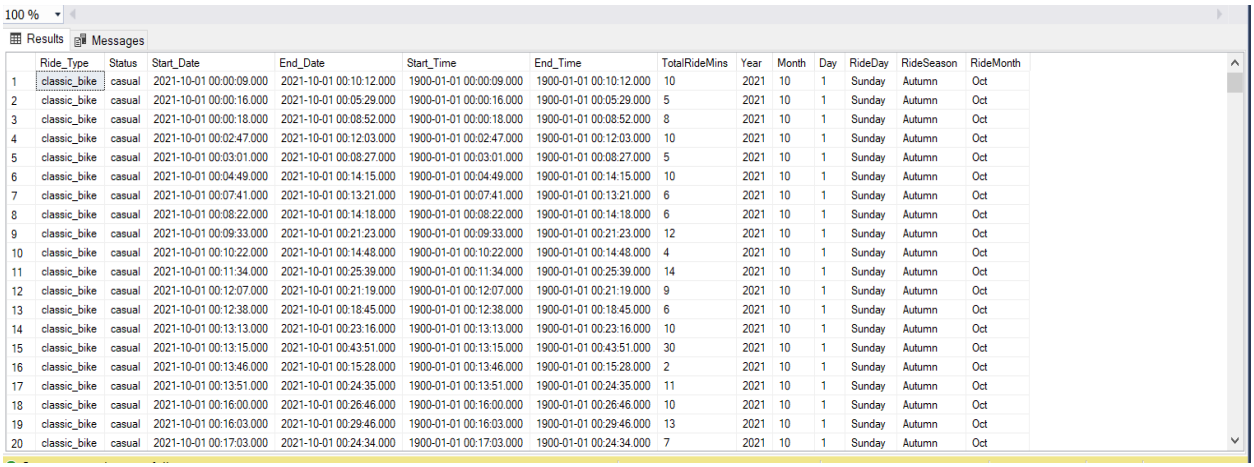
Results | Messages

	Status	RideTotals
1	member	1176285
2	casual	765240

Casual riders:

- Ride less often than members.
- Ride for a longer time than members.
- Ride more in the warmer months.
- Prefer electric bikes over classic
- Do not dock bikes properly or as consistently as members.
- Popular ride day varies by season. Most popular at the beginning of the week during summer months.

Data table cleaned with new columns:



100 %

Results | Messages

	Ride_Type	Status	Start_Date	End_Date	Start_Time	End_Time	TotalRideMins	Year	Month	Day	RideDay	RideSeason	RideMonth
1	classic_bike	casual	2021-10-01 00:00:09.000	2021-10-01 00:10:12.000	1900-01-01 00:00:09.000	1900-01-01 00:10:12.000	10	2021	10	1	Sunday	Autumn	Oct
2	classic_bike	casual	2021-10-01 00:00:16.000	2021-10-01 00:05:29.000	1900-01-01 00:00:16.000	1900-01-01 00:05:29.000	5	2021	10	1	Sunday	Autumn	Oct
3	classic_bike	casual	2021-10-01 00:00:18.000	2021-10-01 00:08:52.000	1900-01-01 00:00:18.000	1900-01-01 00:08:52.000	8	2021	10	1	Sunday	Autumn	Oct
4	classic_bike	casual	2021-10-01 00:02:47.000	2021-10-01 00:12:03.000	1900-01-01 00:02:47.000	1900-01-01 00:12:03.000	10	2021	10	1	Sunday	Autumn	Oct
5	classic_bike	casual	2021-10-01 00:03:01.000	2021-10-01 00:08:27.000	1900-01-01 00:03:01.000	1900-01-01 00:08:27.000	5	2021	10	1	Sunday	Autumn	Oct
6	classic_bike	casual	2021-10-01 00:04:49.000	2021-10-01 00:14:15.000	1900-01-01 00:04:49.000	1900-01-01 00:14:15.000	10	2021	10	1	Sunday	Autumn	Oct
7	classic_bike	casual	2021-10-01 00:07:41.000	2021-10-01 00:13:21.000	1900-01-01 00:07:41.000	1900-01-01 00:13:21.000	6	2021	10	1	Sunday	Autumn	Oct
8	classic_bike	casual	2021-10-01 00:08:22.000	2021-10-01 00:14:18.000	1900-01-01 00:08:22.000	1900-01-01 00:14:18.000	6	2021	10	1	Sunday	Autumn	Oct
9	classic_bike	casual	2021-10-01 00:09:33.000	2021-10-01 00:21:23.000	1900-01-01 00:09:33.000	1900-01-01 00:21:23.000	12	2021	10	1	Sunday	Autumn	Oct
10	classic_bike	casual	2021-10-01 00:10:22.000	2021-10-01 00:14:48.000	1900-01-01 00:10:22.000	1900-01-01 00:14:48.000	4	2021	10	1	Sunday	Autumn	Oct
11	classic_bike	casual	2021-10-01 00:11:34.000	2021-10-01 00:25:39.000	1900-01-01 00:11:34.000	1900-01-01 00:25:39.000	14	2021	10	1	Sunday	Autumn	Oct
12	classic_bike	casual	2021-10-01 00:12:07.000	2021-10-01 00:21:19.000	1900-01-01 00:12:07.000	1900-01-01 00:21:19.000	9	2021	10	1	Sunday	Autumn	Oct
13	classic_bike	casual	2021-10-01 00:12:38.000	2021-10-01 00:18:45.000	1900-01-01 00:12:38.000	1900-01-01 00:18:45.000	6	2021	10	1	Sunday	Autumn	Oct
14	classic_bike	casual	2021-10-01 00:13:13.000	2021-10-01 00:23:16.000	1900-01-01 00:13:13.000	1900-01-01 00:23:16.000	10	2021	10	1	Sunday	Autumn	Oct
15	classic_bike	casual	2021-10-01 00:13:15.000	2021-10-01 00:43:51.000	1900-01-01 00:13:15.000	1900-01-01 00:43:51.000	30	2021	10	1	Sunday	Autumn	Oct
16	classic_bike	casual	2021-10-01 00:13:46.000	2021-10-01 00:15:28.000	1900-01-01 00:13:46.000	1900-01-01 00:15:28.000	2	2021	10	1	Sunday	Autumn	Oct
17	classic_bike	casual	2021-10-01 00:13:51.000	2021-10-01 00:24:35.000	1900-01-01 00:13:51.000	1900-01-01 00:24:35.000	11	2021	10	1	Sunday	Autumn	Oct
18	classic_bike	casual	2021-10-01 00:16:00.000	2021-10-01 00:26:46.000	1900-01-01 00:16:00.000	1900-01-01 00:26:46.000	10	2021	10	1	Sunday	Autumn	Oct
19	classic_bike	casual	2021-10-01 00:16:03.000	2021-10-01 00:29:46.000	1900-01-01 00:16:03.000	1900-01-01 00:29:46.000	13	2021	10	1	Sunday	Autumn	Oct
20	classic_bike	casual	2021-10-01 00:17:03.000	2021-10-01 00:24:34.000	1900-01-01 00:17:03.000	1900-01-01 00:24:34.000	7	2021	10	1	Sunday	Autumn	Oct

2. Supporting visualizations and key findings

Casual riders ride more in during summer and autumn seasons, dropping below 100,000 riders during the month December (which is significant giving that December is not the coldest month during the Chicago winter season https://freetoursbyfoot.com/weather-in-chicago-in-january/#:~:text=HOW%20COLD%20IS%20CHICAGO%20IN,F%20(%2D9%C2%B0C).). While members maintain over 230,000 riders over the course of three seasons/months autumn, spring and summer. Close to 50% of casual riders ride during the summer months as seen in the table below.

```
--Total Rides by status and season

select  RideSeason, status, count(Status) as TotalRides
from DivvyBikePercents
--Where Status = 'casual' and Ride_Type != 'docked_bike'
group by RideSeason, status
order by RideSeason
```

Rentals By Season

	Autumn	Spring	Summer	Winter
Member	362,913	236,462	404,349	172,561
Casual	250,017	122,218	393,547	67,886

Percentage of Rentals By Season

	Autumn	Spring	Summer	Winter
Casual	29.99%	14.66%	47.21%	8.14%
Member	30.85%	20.10%	34.38%	14.67%

While members ride more often than casual riders, casual riders on average ride close to 50% longer than members. Meaning that they ride more minutes per rental. The chart below shows the average of both sets of riders. This even extends through the winter months.

```
--Average ride mins per season
select RideSeason, round(avg(totalRideMins), 0) as Casual_Avg
from DivvyBikeRiders
where status =  'casual'
Group by RideSeason;

select RideSeason, round(avg(totalRideMins), 0) as Member_Avg
from DivvyBikeRiders
where status =  'member'
Group by RideSeason;
```

Average Ride Minutes for Casual Riders

Autumn	21
Spring	22
Summer	22
Winter	17

Average Ride Minutes for Members

Autumn	12
Spring	11
Summer	13
Winter	10

Looking at what days are more poplar for casual riders by season/month, the chart below displays that casual riders and members vary on not only seasonal usage, but also choice of day to ride, concluding that the purpose of the rental may vary by status. Casual riders tend to ride more on Mondays and Tuesdays possible due to tourist activity.

```sql
-- Look at popular ride days per season

select max(Rideday) as PopularDay, count(status) as Total_Rentals, Status, RideSeason
from DivvyBikeRiders
where Ride_Type != 'docked_bike'
and RideDay is not null
group by RideDay, Status, RideSeason
order by Total_Rentals desc
```

Popular Ride Day By Season

		Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Autumn	Casual	10,522	11,016	8,701	5,618	6,244	6,593	5,465
	Member	14,709	11,436	9,435	12,094	13,973	14,243	10,952
Spring	Casual	2,769	1,770	2,864	2,351	2,544	2,776	2,161
	Member	7,286	3,842	5,417	8,058	8,990	8,339	7,212
Summer	Casual	10,213	15,951	16,363	11,953	7,788	8,264	10,053
	Member	12,075	12,076	11,306	8,534	11,539	13,108	14,360
Winter	Casual	2,543	3,687	3,812	3,957	1,497	1,386	1,240
	Member	9,500	10,882	9,381	7,491	3,165	6,087	5,969

In the four charts below, it is noted that total ride minutes are consistent with the number of rides where as total ride minute account for the majority of casual riders July/Summer usage times for both classic and electric bikes, almost matching that of members and beating out member riders for electric bike usage in July. Providing more support for the argument that the boost in casual riders in the warmer month may be more in alignment with the high tourist and event season.

```sql
-- Look at bike type rentals per month

select  status, RideMonth, Ride_Type, count(Ride_type) as TotalRentals
from DivvyBikeRiders
where Ride_Type != 'docked_bike'
group by Ride_Type, status, RideMonth
order by TotalRentals
```

Total Ride Minutes by Status

		Casual	Member
classic_bike	Apr	46,350	115,995
	Jul	152,142	212,165
	Oct	103,214	205,528
	Dec	19,449	78,898
electric_bike	Apr	64,184	120,467
	Jul	211,458	192,184
	Oct	124,755	157,385
	Dec	43,688	93,663

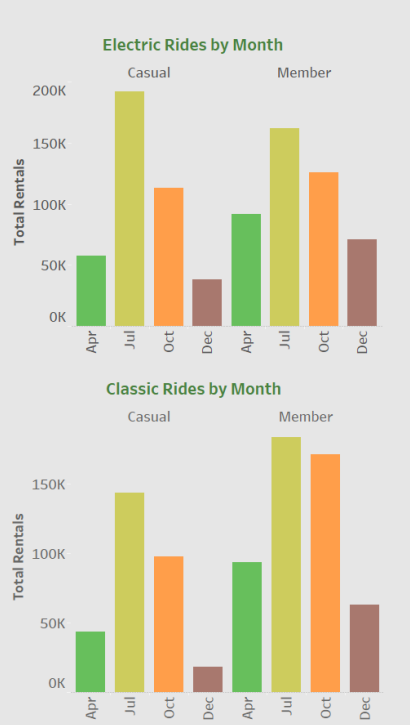
Percent of Total Ride Minutes by Status

		Casual	Member
classic_bike	Apr	28.55%	71.45%
	Jul	41.76%	58.24%
	Oct	33.43%	66.57%
	Dec	19.78%	80.22%
electric_bike	Apr	34.76%	65.24%
	Jul	52.39%	47.61%
	Oct	44.22%	55.78%
	Dec	31.81%	68.19%

```sql
--Total number of rentals more than 5 mins per ride type

select  Ride_Type, count(Ride_type) as MemberRentals
from DivvyBikeRiders
where Status = 'member' and totalridemins >= '5'
group by Ride Type

select Ride_Type, count(Ride_type) as CasualRentals
from DivvyBikeRiders
where Status = 'casual' and totalridemins >= '5'
group by Ride_Type
```



3. Recommendations

- Appeal to casual members during spring season highlighting summer membership benefits. With summer memberships beginning in mid-April.

- Consider seasonal packages. Include a seasonal analysis as part of package that shows potential savings based on current members testimonies on savings

- Include incentives when purchasing a membership during popular summer events such as Water Flicks, Taste of Chicago, Chinatown Summer Fair and Lollapalooza. For instance, casual riders can earn additional points or a discounted rate for becoming a member during the selected events dates.