

# How does a bike-share navigate speedy success?

## Scenario

You are a junior data analyst working on the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

## My business question

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

## The business task

The main goal of this case study is to **design marketing strategies aimed at converting casual riders to annual members of Cyclistic**. This can be done by gathering insights into the behaviour of both annual and casual riders by means of historical bike-share data.

## Data Sources

Historical datasets consisting of 2019 Q1 trips and 2020 Q1 trips will be used in this case study, this data includes information such as ride type, ride id, stations used, member type and duration of rides. **The two datasets used have a combined 791,000 rows**. This data helps solve the business task as it gives us all the data on behaviour of casual riders and annual memberships rides, from this we can gain insights into how much casual riders are using the bikes comparatively to those that hold an annual membership. The data can be located [here](#).

The data has been made available by Motivate International Inc. under this [license](#). This is public data, but data-privacy issues prohibit the use of riders' personally identifiable information

## Tools

**Google sheets** for sorting, filtering and preparing data for analysis. **SQL** was used for aggregation of data and descriptive analysis. **Tableau** was used for data visualisation.

## Cleaning and Data preparation

The datasets were sorted by date ascending and columns were filtered to eliminate null values in the cells. In the 2019 Q1 dataset 19,715 rows were deleted due to null values in the cells.

Column names were made consistent between datasets by using the same names for both 2019 Q1 and 2020 Q1 such as ride\_id, start\_time, end\_time, start\_station\_name and end\_station\_name. New columns were created to find values in the tripduration by using start and end time data and also a column for the weekday of the trip which could be used in future analysis.

120 rows in 2020 Q1 data were deleted as tripduration showed a negative value, which in an error in the data.

## Analysis

### Aggregation

Average ride length, maximum ride length, and modal weekday was calculated for both datasets. This was done by using google sheets function such as AVERAGE(), MAX() and a formula was used to find the modal weekday (=INDEX(D:D, MODE(IF(D:D<>"", MATCH(D:D, D:D, 0)))) where column D was the weekday in text.

### Descriptive analysis

Pivot tables were created for both datasets to find insights. Pivot tables were made to compare casual and members rides by average trip duration, by day of the week and number of rides for users by day of the week

## 2020 Q1

usertype	tripduration avg	hh:mm:ss
casual	5,761.26	1:36:01
member	760.63	0:12:41

Average ride length for users by day of the week								
AVERAGE of trip_start_time - Day of the week								
usertype	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	
casual	5,710.57	4,452.19	5,084.97	4,531.83	7,599.78	7,126.19	6,017.16	
member	949.34	778.63	692.03	699.54	693.23	757.32	929.99	
Grand Total	2,343.16	1,044.69	1,000.14	1,022.66	1,201.39	1,297.87	1,942.44	

Number of rides for users by day of the week								
COUNTA of ride_start_time - Day of the week								
usertype	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Grand Total
casual	14886	4835	5257	5891	4864	5147	7480	48360
member	35964	61923	69697	63978	61245	55496	30104	378407
Grand Total	50850	66758	74954	69869	66109	60643	37584	426767

## 2019 Q1

usertype	tripduration avg	hh:mm:ss
	#DIV/0!	#DIV/0!
casual	3715.88	00:37:01
member	833.47	00:13:52



### Average ride length for users by day of the week

AVERAGE of trip\_start\_time - Day of the week

usertype	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Grand Total
								#DIV/0!
casual	2,493.03	2,667.37	2,427.29	3,117.43	8,025.06	3,593.57	3,619.74	3,715.88
member	1,007.51	877.93	862.02	725.72	720.59	833.33	1,018.86	833.47

### Number of rides for users by day of the week

COUNTA of ride\_start\_time - Day of the week

usertype	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Grand Total
								0
casual	3766	1892	2728	2489	2920	3375	5992	23162
member	24233	48507	58277	57925	63983	59672	29309	341906
<b>Grand Total</b>	<b>27999</b>	<b>50399</b>	<b>61005</b>	<b>60414</b>	<b>66903</b>	<b>63047</b>	<b>35301</b>	<b>0</b>

## Exploratory Data Analysis

This section will include key insights from SQL queries and Tableau visualisations.

### SQL Analysis

SQL analysis was done on Google BigQuery, the datasets were queried to find interesting insights such as total number of rides, average trip duration, usertype ratio, peak usage hours for casual riders and the most frequent start stations for casual riders. The queries written and direct out put can be found in the appendix below.

### Tableau Analysis

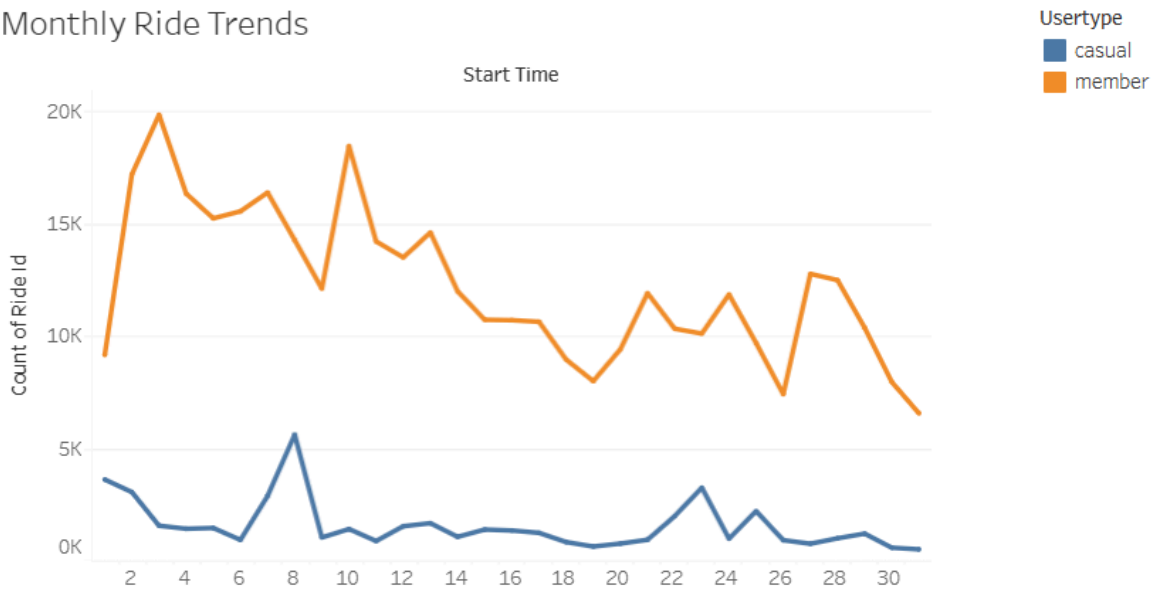
Tableau was used to create insight driven visualisations that compared the behaviour of casual riders and members. Dashboards including these visualisations were made for both the 2019 and 2020 datasets. A link to these dashboards can be found here, [2019](#) and [2020](#).

### Ride Trends Over Time

A line chart on **monthly ride trends** was created to compare the number of riders per day for each month of the quarter for casual versus member riders, as well as a **peak usage hours heat map** that showed what times casual and member riders use the service.

*Insights:* These visualisations helped identify seasonal trends and gave insights to the consistency in member rides while casual riders tend to have peaks and troughs.

## Monthly Ride Trends



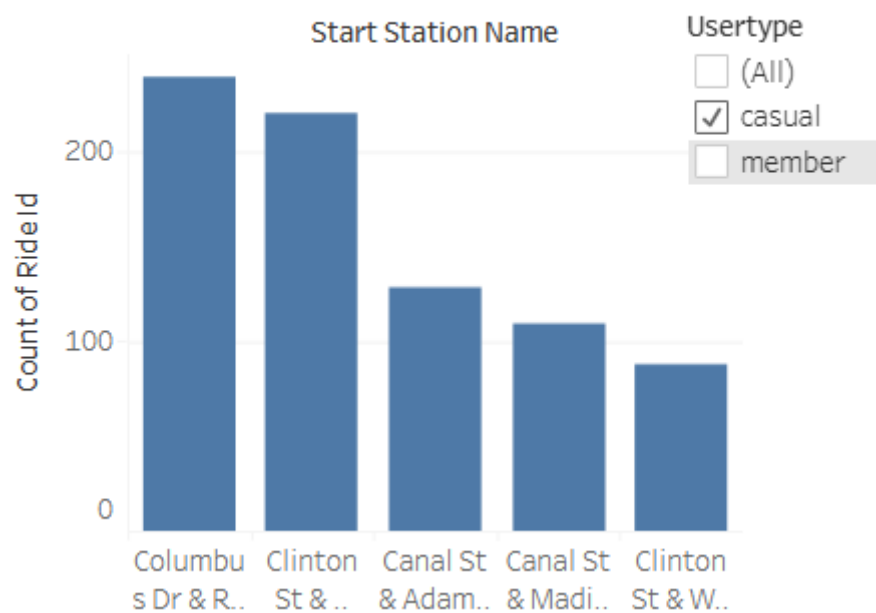
*Line chart of Monthly Ride Trends for January 2020*

## Popular Stations

A box chart was created to show the top 5 most popular stations for casual versus member riders, the tableau dashboard gave option to filter between casual and member rides, and also to view both as a stacked bar chart.

*Insights:* This visualisation helped to identify where casual riders typically start/end trips and helped identify if popular casual rider stations were near transit hubs, parks, or tourist spots.

## Most Used Stations

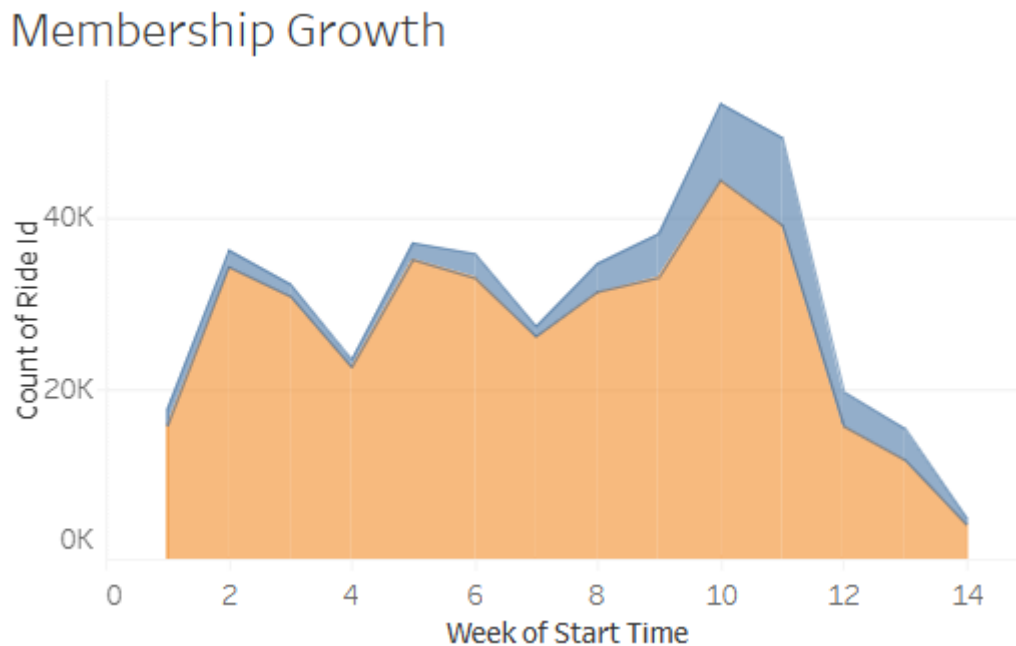


*Screenshot of bar chart of Most Used Stations from 2019 Q1*

## Membership Growth Analysis

A stacked area chart was created to show trend of casual versus member rides over weeks of the quarter. The stacked charts could be used to compare growth in 2019 versus 2020 and identify any significant changes, such as a rise in people becoming members or a lack in casual rides.

Insights: This chart gave insight into the specific times casual riders converted to members. The charts may have showed when external factors impacted memberships (End of 2020 Q1 showed a massive drop in rides likely to COVID-19)



*Screenshot of Membership Growth during 2020 Q1*

## Key Findings & Recommendations

### Findings

- When and where casual riders are most active.
- Differences in ride duration & station usage.
- Trends that could help convert casual riders into members.

### Recommendations

- **Targeted Promotions:** Offer membership discounts at peak casual rider stations.
- **Pricing Adjustments:** Show casual riders potential cost savings if they join as members.
- **Marketing Strategy:** Send personalized email/app notifications to frequent casual riders.
- **Partnerships:** Work with local businesses near popular casual stations for member perks.

## Conclusion

In conclusion, key insights into the behaviour of casual riders and members were found from data provided in 2019 and 2020. Findings from the insights gained can be used to help improve member conversion for the company by implementing such recommendations stated above. Based on these findings, some future analysis that could be implemented may involve analysing the growth in members after these strategies have been put in place,

## Appendix

### SQL Queries

#### Total number of rides

```
SELECT COUNT(ride_id) AS total_rides
FROM `cyclistic-453614.Cyclistic2019.2019Q1`
```

#### Average Trip Duration

```
SELECT AVG(tripduration_s) AS avg_tripduration
FROM `cyclistic-453614.Cyclistic2019.2019Q1`
```

#### Ratio of casual to members

```
SELECT usertype, COUNT(*) AS count
FROM `cyclistic-453614.Cyclistic2019.2019Q1`
GROUP BY usertype;
```

#### Peak Usage Hours for Casual Riders

```
SELECT EXTRACT(HOUR FROM start_time) AS hour, COUNT(*) AS ride_count
FROM `cyclistic-453614.Cyclistic2019.2019Q1`
WHERE usertype = 'casual'
GROUP BY hour
ORDER BY ride_count DESC
LIMIT 5;
```

## Most Frequent Start Stations for Casual Riders

```
SELECT start_station_name, COUNT(*) AS ride_count
FROM `cyclistic-453614.Cyclistic2019.2019Q1`
WHERE usertype = 'casual'
GROUP BY start_station_name
ORDER BY ride_count DESC
LIMIT 5;
```

## 2020 Q1 Query Results

total\_rides = 426767

avg\_tripduration = 1327.28 s

Row	usertype	count
1	casual	48360
2	member	378407

Row	hour	ride_count
1	14	5398
2	15	5267
3	16	5110
4	13	5007
5	17	4663

Row	start_station_name	ride_count
1	HQ QR	3646
2	Lake Shore Dr & Monroe St	1590
3	Streeter Dr & Grand Ave	1530
4	Shedd Aquarium	998
5	Millennium Park	779

## 2019 Q1 Query Results

**total\_rides = 365069**

**avg\_tripduration = 1016.34 s**

Row	usertype ▼	count ▼
1	casual	23163
2	member	341906

Row	hour ▼	ride_count ▼
1	15	2723
2	16	2555
3	14	2520
4	13	2484
5	17	2341

Row	start_station_name ▼	ride_count ▼
1	Streeter Dr & Grand Ave	1219
2	Lake Shore Dr & Monroe St	1142
3	Shedd Aquarium	834
4	Millennium Park	627
5	Michigan Ave & Oak St	386

Screenshots of both dashboards are shown below, but links to the dashboards can be found here, [2019](#) and [2020](#).

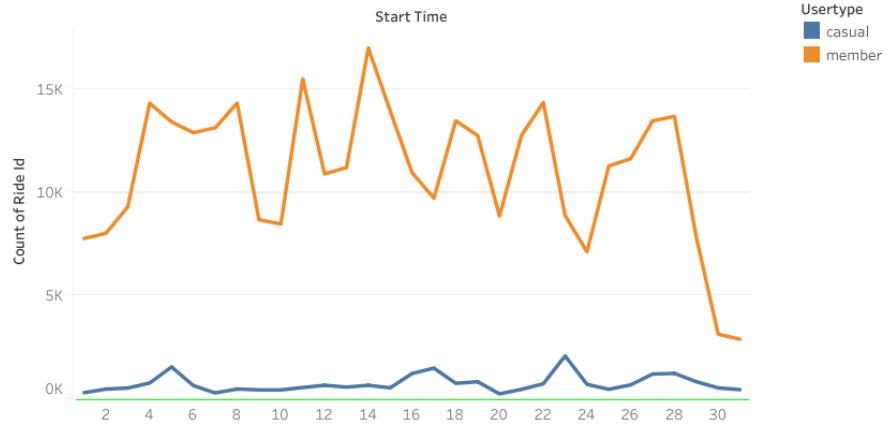
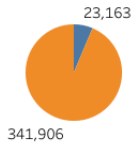




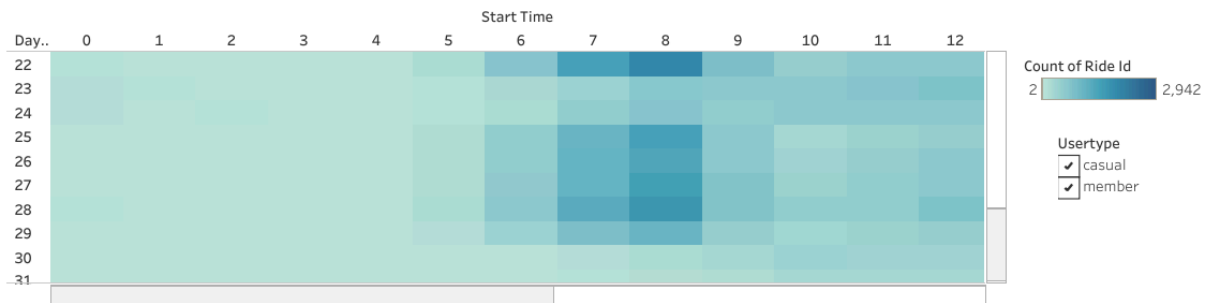
# Cyclistic Dashboard 2019 Q1

Monthly Ride Trends

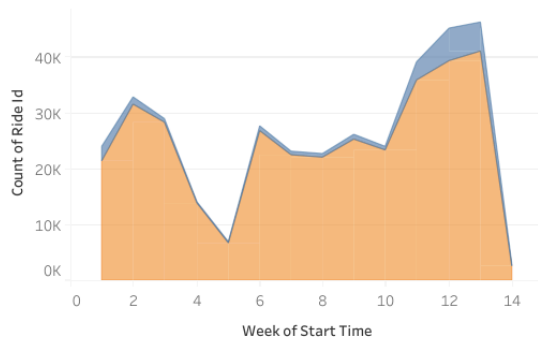
Casual vs.  
Member Ride  
Comparison



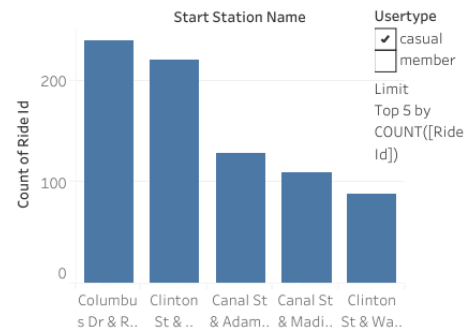
Peak Hours Heat Map



Membership Growth



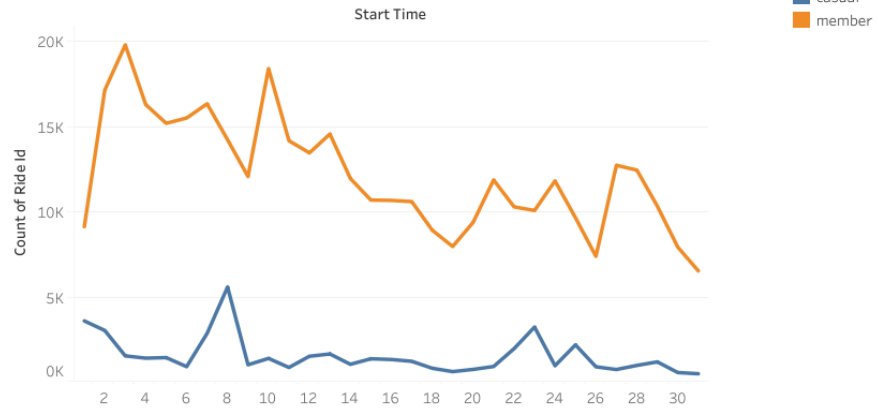
Most Used Stations



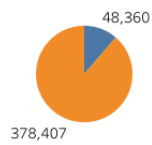


# Cyclistic Dashboard 2020 Q1

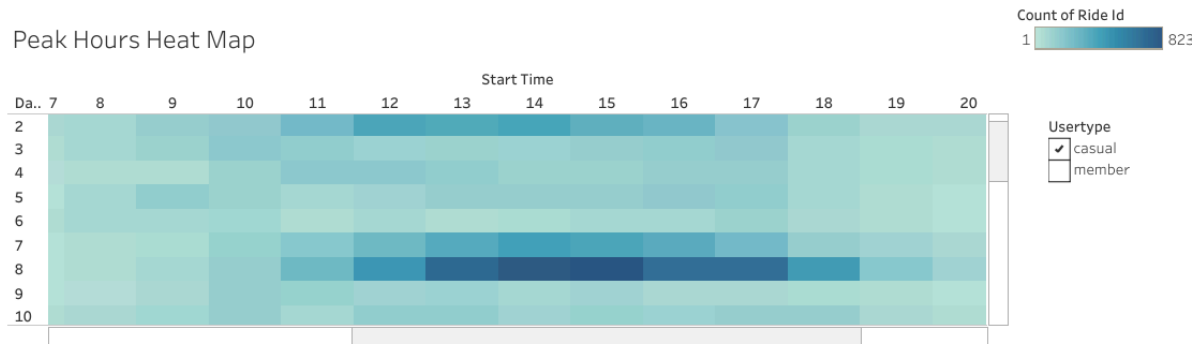
Monthly Ride Trends



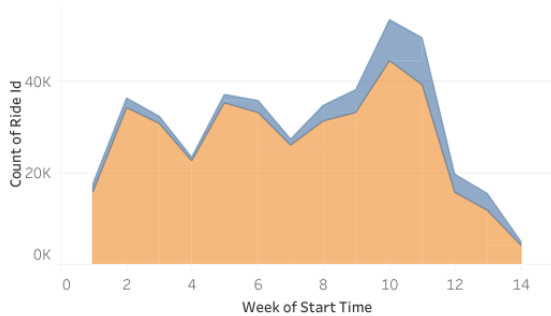
Casual vs. Member Ride Comparison



Peak Hours Heat Map



Membership Growth



Most Used Stations

