

Case Study

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



Introduction

This is a project for Google Data Analytics Capstone – Case Study 1 (Cyclistic bike-share analysis)

Scenario

You are a junior data analyst working in 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.

About the company

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.

Ask

Guiding questions

- What is the problem you are trying to solve?
To convert casual bike rider in to Cyclistic member by offering them annual membership
- How can your insights drive business decisions?
By providing detail analysis of bike ride data and all the potential riders to Moreno to design the strategies

Key tasks

- Identify the business task
- Consider key stakeholders

Deliverable

- A clear statement of the business task
To identify the objectives of the analysis to define the marketing strategies

Prepare

Guiding questions

- Where is your data located?
The data is located in the google cloud
- How is the data organized?
The data are in CSV. Files
- Are there issues with bias or credibility in this data? [Does your data ROCCC?](#)
This data is from Cyclistic company and there is no bias or credibility issue
- How are you addressing licensing, privacy, security, and accessibility?
The data is provided for the case study hence there is no private or personal information are shared. It is protected under data privacy.
- How did you verify the data's integrity?
By checking the format and data types
- How does it help you answer your question?
The data has all the key information such year, location and station which is essential to analyze the data.
- Are there any problems with the data?
The data looks good no problem identified.

Key tasks

- Download data and store it appropriately.
- Identify how it's organized.
- Sort and filter the data.
- . Determine the credibility of the data.

Deliverable

- A description of all data sources used
The data source provides 12 months of Cyclistic trip data.

Process

Guiding questions

- What tools are you choosing and why?
The excel is used to standardized the data and load in to SSMS for data analysis
- Have you ensured your data's integrity?
The integrity is being taken care. The data is accurate, complete, consistence and trustworthy
- What steps have you taken to ensure that your data is clean?
The data was checked thoroughly, duplicates are removed, spelling, space between the character, data type, formats are checked and necessary changes are made
- How can you verify that your data is clean and ready to analyze?
The data is verified and looks complete. The data is ready to answer the business goal
- Have you documented your cleaning process so you can review and share those results?
At every steps of the cleaning process the duplicate has been stored to review in future

Key tasks

- Check the data for errors.
- Choose your tools.
- Transform the data so you can work with it effectively.
- Document the cleaning process.

Deliverable

- Documentation of any cleaning or manipulation of data

Analyze

Guiding questions

- How should you organize your data to perform analysis on it?
The data should be sorted and filtered when performing the analysis
- Has your data been properly formatted?
The data is formatted
- What surprises did you discover in the data?
The timestamp has some values interchanged and some of the column has not appropriate data type
- What trends or relationships did you find in the data?
The data is analyzed by rideable_type and in yearly, monthly and quarterly
- How will these insights help answer your business questions?
The data shows that the casual riders drive more than members. Their usage is high on particular days of the week and particular months. Creating special offers on those days and months will help increase more members.

Key tasks

- Aggregate your data so it's useful and accessible.
- Organize and format your data.
- Perform calculations.
- Identify trends and relationships.

Deliverable

- A summary of your analysis

Steps followed using Excel

1. Combine all the excel sheets in to one was not possible because of the size of the file.
2. Cleaned and transformed the data to prepare for analysis.
3. Conducted descriptive analysis.
4. Created two columns 'ride_length' and 'day_of_week'

'ride_length' (= ended_at,D2-Started_at,C2) day_of_week(=weekday,C2,1)

Since all the files cannot be combined in one hence perform the analysis in SSMS using SQL

Steps followed using ETL

Since there is lot of data and excel files, performed Extract, Transform and Load in to data warehouse. Tool used Microsoft Integration Service

Steps followed using SQL

1. Explore your data, perhaps looking at the total number of rows, distinct values, maximum, minimum, or mean values.

2. Created column 'ride_length', and 'day_of_week'

```
alter table cyclistic_bikeshare add ridelength time (0) null
```

```
update cyclistic_bikeshare set ridelength = cast (ended_at - started_at as time(0))
```

3. Calculate the day_of_week

```
alter table cyclistic_bikeshare add day_of_week int null
```

```
update cyclistic_bikeshare set day_of_week = DATEPART(weekday, started_at)
```

4. Query all the data to upload to Tableau. Select all the needed data using the query below

select

ride_id, rideable_type, started_at, ended_at,

ride_len, day_of_the_week, start_station_name, start_station_id, end_station_name, end_station_id, member_casual

from casestudycyclistic.CyclisticData.all_cyclistic_data_bkp where ride_len >= 0;

Guiding questions

- Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?
Yes. Casual riders use more than the members. Members use it to commute to work while Casual riders use it for leisure.
- What story does your data tell?
Data tells the difference in usage by riders.
- How do your findings relate to your original question?
Findings provide details about the usage by casual riders. We can create offers to make them members.
- Who is your audience? What is the best way to communicate with them?
Marketing team. Create a Presentation
- Can data visualization help you share your findings?
Yes
- Is your presentation accessible to your audience?
Yes

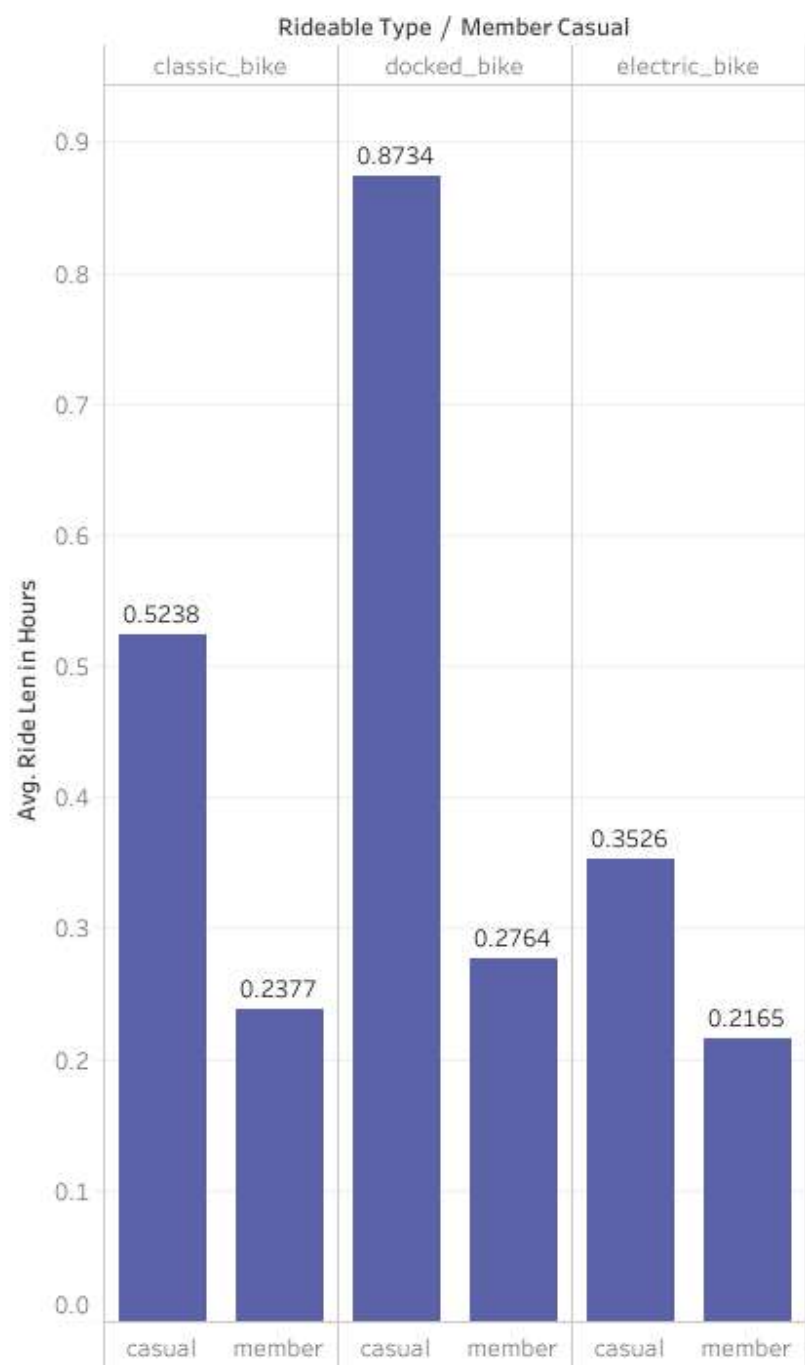
Key tasks

- Determine the best way to share your findings.
- Create effective data visualizations.
- Present your findings.
- Ensure your work is accessible.

Deliverable

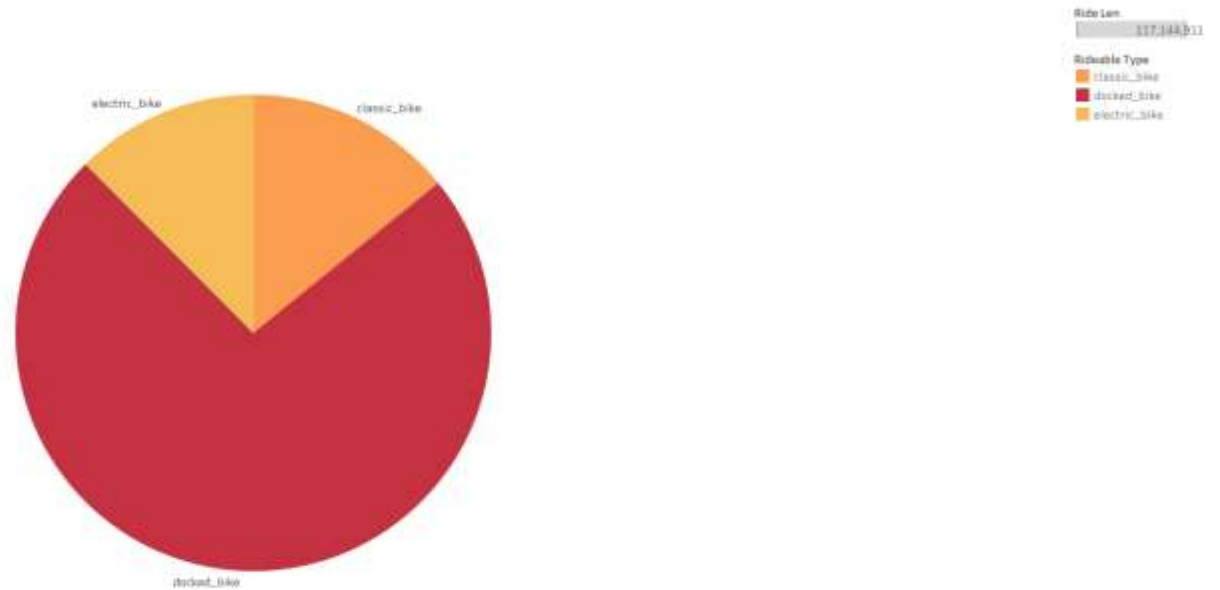
- Supporting visualizations and key findings

Rider_data



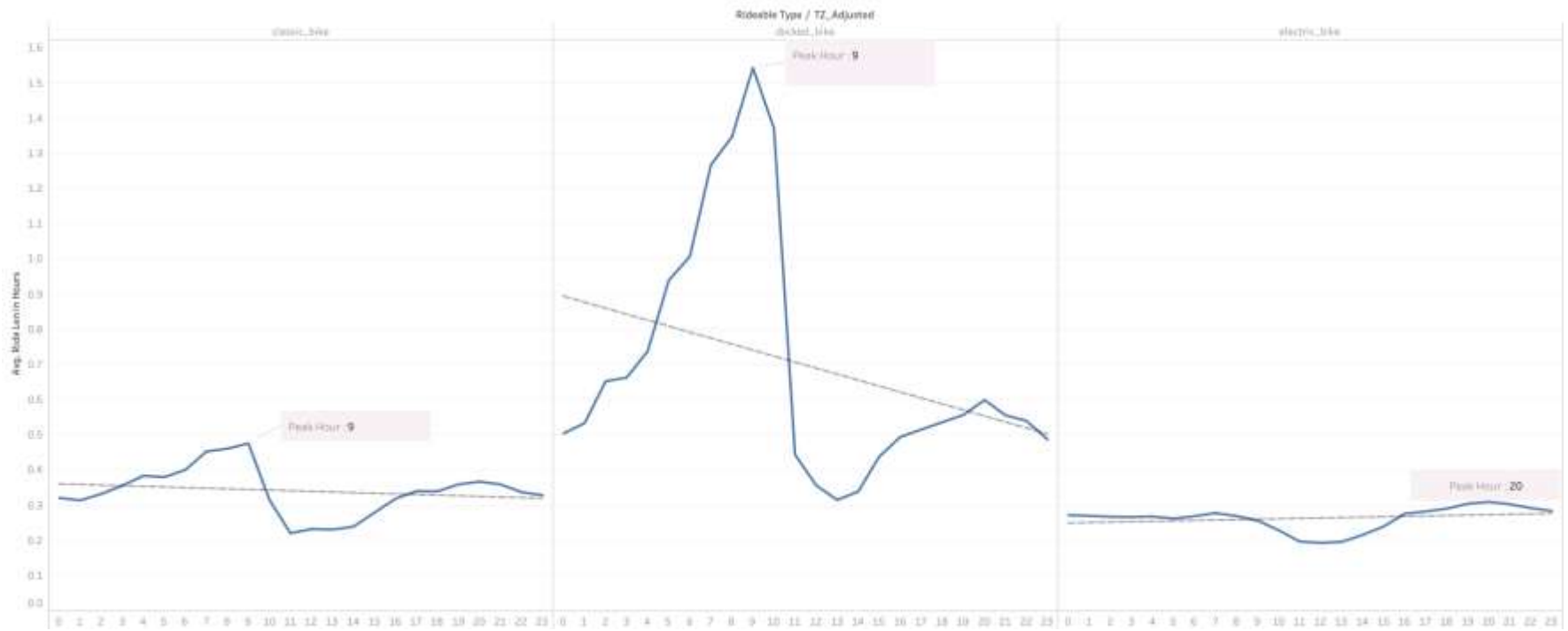
Rider data shows that average ride length of Casual riders is high in every bike type compared to Members. Casual riders ride more than the Members.

Rideable_data



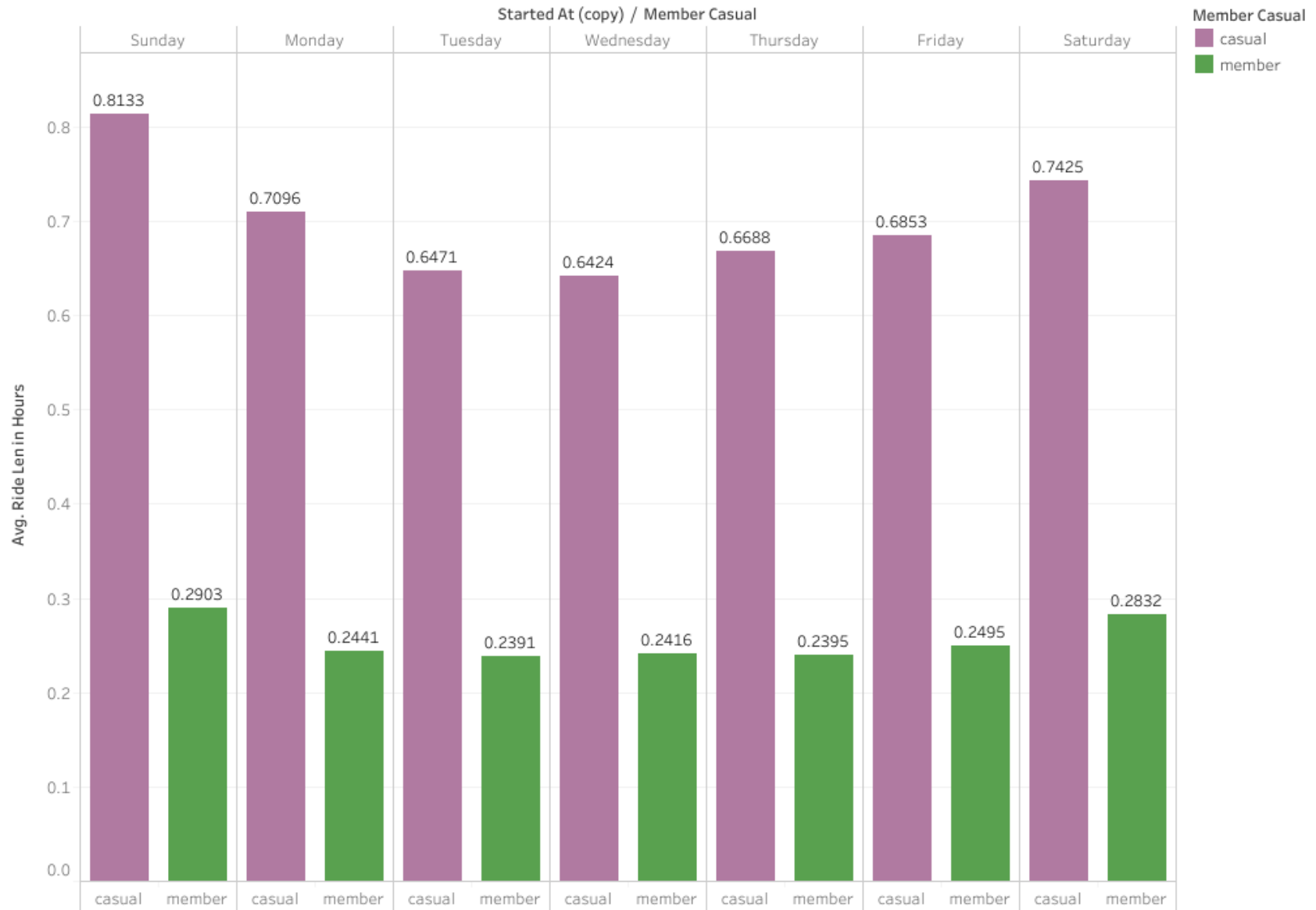
Rideable data shows that the ride length for 73% of riders prefer docked bike, 14% prefer classic bike and 12% prefer electric bike.

Hourly_data



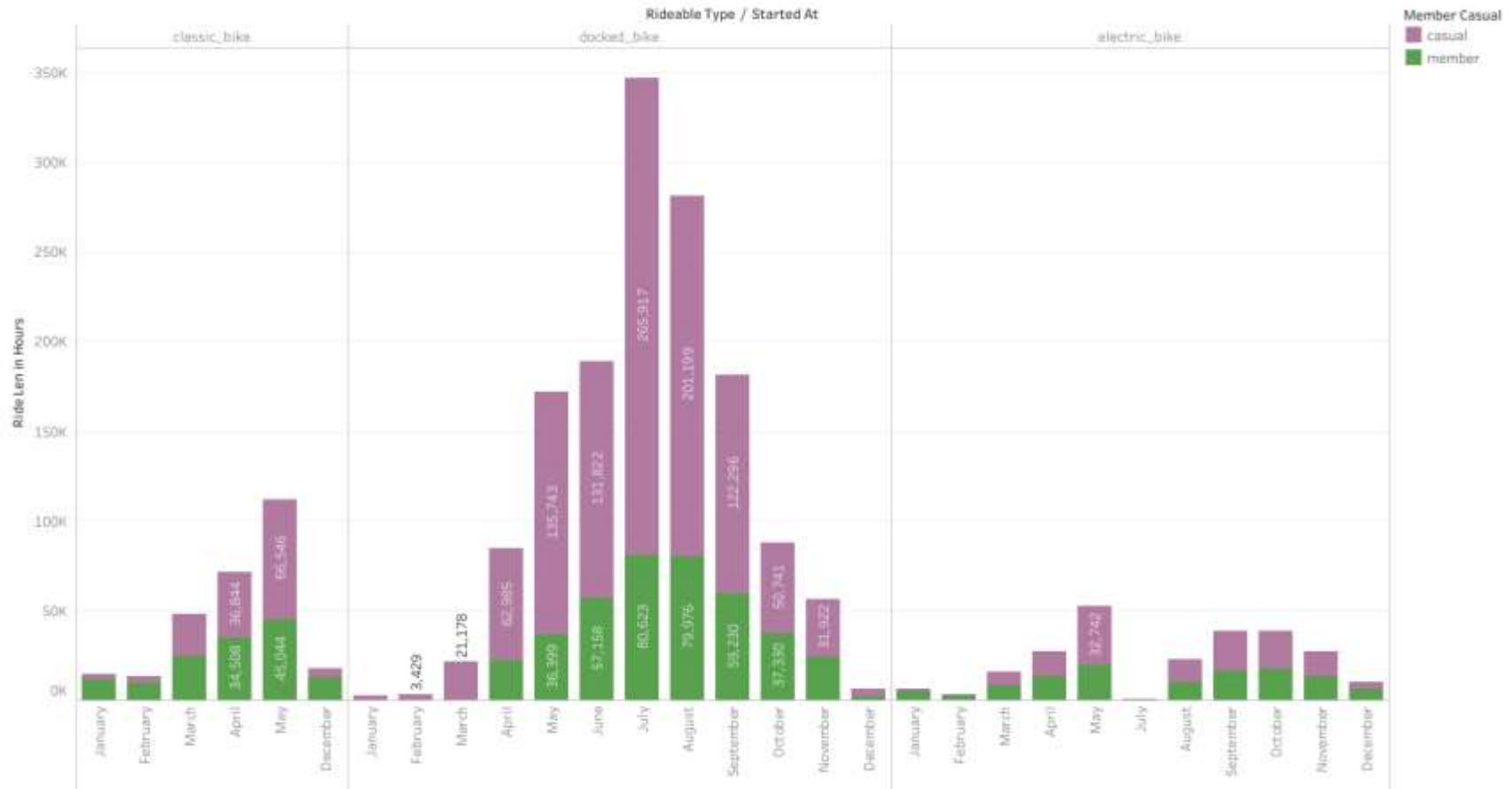
The hourly data shows that the peak usage in a day is around 9A.M to 10A.M.

Weekly_data



Casual riders use bikes more on Weekends and Member riders use consistently throughout the week.

Yearly_data



The peak usage of bikes is around summer, mostly May to July. The latest usage is around December to February. This may be because of the Chicago weather.

Act

Findings

- 80% of Casual riders use docked bikes compared to the rest.
- Usage of casual riders is twice that of the members.
- Members mostly use the bikes to commute to work since the pattern of usage is more in weekdays and same throughout the year.

Recommendations

- The usage of casual riders are more on Weekends. So provide offers/promotions for weekends, so they will become members.
- Riders prefer Docked bike more than electric and classic bike. Increase more docking stations for the ease of user's comfort.

- The usage is more in July and August months. So start creating more offers during spring and summer months.
- The peak usage in a day is at 9 A.M, so increase the availability of docked bikes during that time.
- Additional data of rider age, price of the bikes can give us more detailed view for further findings.