Cyclistic Case Study

Case Study Roadmap - Ask

Guiding questions

- **What is the problem you are trying to solve?**
  The problems to be solved are:
  1) How casual members and annual members use cyclistic differently.
  2) Why would casual members buy annual memberships?

- **How can your insights drive business decisions?**
  Insights discovered can help in designing a new marketing strategy to convert casual members to annual members.

Key tasks

1. **Identify the business task**
   - Analyze cyclistic historical bike trip data to understand how casual riders differ from annual members.
   - How to influence casual members to become annual riders.

2. **Consider key stakeholders**
   - Primary stakeholder- Cyclistic Executive Team
   - Secondary stakeholders- Lily Moreno (director of marketing & manager), Marketing Analytics Team

Deliverable

A clear statement of the business task
Analyze cyclistic historical bike trip data to identify trends, understand the bike usage among annual members and casual riders and how to influence casual members to become annual riders.

Case Study Roadmap - Prepare

Guiding questions

- **Where is your data located?**
  Data is made available by motivate international inc and is located on www.divvybikes.com/data.

- **How is the data organized?**
  The dataset contains last 12 months trip data. The data is organized as a different sheet for data for diff months and it also shows quarterly data (in csv format).

- **Are there issues with bias or credibility in this data? Does your data ROCCC?**
  Data is Reliable, Original, Comprehensive, Current and Citied.

- **How are you addressing licensing, privacy, security, and accessibility?**
  - ➔ The data is licensed (license agreement).
  - ➔ Accessibility- Data available to the public, subject to the terms and conditions of this License Agreement
  - ➔ Privacy, Security- License contains all privacy security measures.

- **How did you verify the data's integrity?**
  Data ROCCC's. It is accurate, complete, cited and secure.

- **How does it help you answer your question?**
  Since the data is accurate and unbiased, it'll help in getting the right insights.

- **Are there any problems with the data?**
  No

Key tasks

1. **Download data and store it appropriately.**
   Data has been downloaded and stored in a proper format.

2. **Identify how it's organized.**
   The dataset contains last 12 months trip data. The data is organized as a different sheet for data for diff months and it also shows quarterly data (in csv format).

3. **Sort and filter the data.**
   Data is sorted and filtered properly in RStudio.

4. **Determine the credibility of the data.**
   Data is credible as confirmed from the license agreement and this is a case study using public data, we are going to assume the data is credible.

Deliverable

A description of all data sources used
Data is made available by motivate international inc and is located on www.divvybikes.com/data.

Case Study Roadmap - Process

Guiding questions

- **What tools are you choosing and why?**
  Since the sizes of the datasets are very large, we're using R via RStudio to prepare, process, analyze data and create visualizations.

- **Have you ensured your data's integrity?**
  Data is accurate, complete, secure as confirmed from the license.

- **What steps have you taken to ensure that your data is clean?**
  Checked the data properly, ensured there's no duplicate or irrelevant or incorrect data.

- **How can you verify that your data is clean and ready to analyze?**
  Data is complete and accurate and properly formatted.

- **Have you documented your cleaning process so you can review and share those results?**
  Yes, the cleaning process is documented.

Key tasks

- **Check the data for errors.**
  Data is error free.

- **Choose your tools.**
  RStudio is chosen for this task.

- **Transform the data so you can work with it effectively.**
  Data is transformed to fix data type inconsistencies

- **Document the cleaning process.**
  1. Before importing datasets into R, 1 column (Day of the week) was added to each of the 12 monthly .CSV files. The .CSV files were then saved as Excel (.xlsx) files.
  2. The data type of columns start_station_id and end_station_id in 4 of the .xlsx files ((y2020_12, y2021_01, y2021_02, y2021_03)) were changed to numbers using the mutate() function above.

Deliverable

Documentation of any cleaning or manipulation of data

1. Before importing datasets into R, 1 column (Day of the week) was added to each of the 12 monthly .CSV files. The .CSV files were then saved as Excel (.xlsx) files.
2. The data type of columns start_station_id and end_station_id in 4 of the .xlsx files ((y2020_12, y2021_01, y2021_02, y2021_03)) were changed to numbers using the mutate() function above.
3. Data of all years is combined into a single dataframe.
4. Columns are added for date, month, day and year.Unnecessary columns(start_lat, start_lng, end_lat, end_lng) are removed.
5. NA rows are dropped.
6. Ride length is calculated and negative ride length is removed.

Case Study Roadmap - **Analyze**

Guiding questions

- **How should you organize your data to perform analysis on it?**
  Data is aggregated, 12 individual data frames(y2020_04, y2020_05, ..) are combined into one single data frame(all_trip_data) for analysis.

- **Has your data been properly formatted?**
  Data cleaning has been done

- **What trends or relationships did you find in the data?**
  1. Average ride time for casual riders is more than member riders. Members are using bikes for commuting and casual riders are using bikes for travel, sightseeing etc.
  2. Ride length is almost consistent for member riders on all days of week with a little increase on weekends whereas for casual riders it's low on weekdays and very high on weekends.
  3. Number of rides starts off low on mondays, increases to a peak on saturday with a small drop off on sundays.

- **How will these insights help answer your business questions?**
  These insights will help us analyze cyclistic historical bike trip data and we understand how casual riders differ from annual members. It'll also help in designing a new marketing strategy to convert casual members to annual members.

Key tasks

- **Aggregate your data so it's useful and accessible.**
  Combined data from 12 months into a single dataset

- **Organize and format your data.**
  Data is organized, negative ride length rows deleted, na fields deleted

- **Perform calculations.**
  Calculated mean, median, mode of ride length and also calculated mean ride times according to days.

- **Identify trends and relationships.**
  1. Average ride time for casual riders is more than member riders. Members are using bikes for commuting and casual riders are using bikes for travel, sightseeing etc.

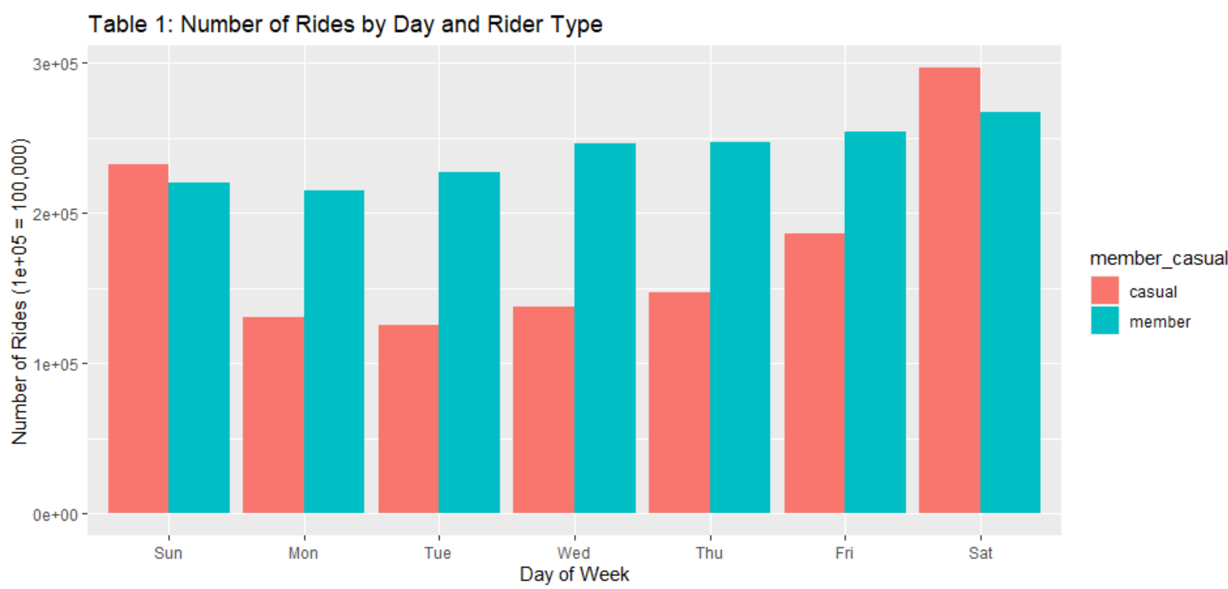
2. Ride length is almost consistent for member riders on all days of week with a little increase on weekends whereas for casual riders it's low on weekdays and very high on weekends.
3. Number of rides starts off low on mondays, increases to a peak on saturday with a small drop off on sundays.

Deliverable

A summary of your analysis

The visualizations are as follows-

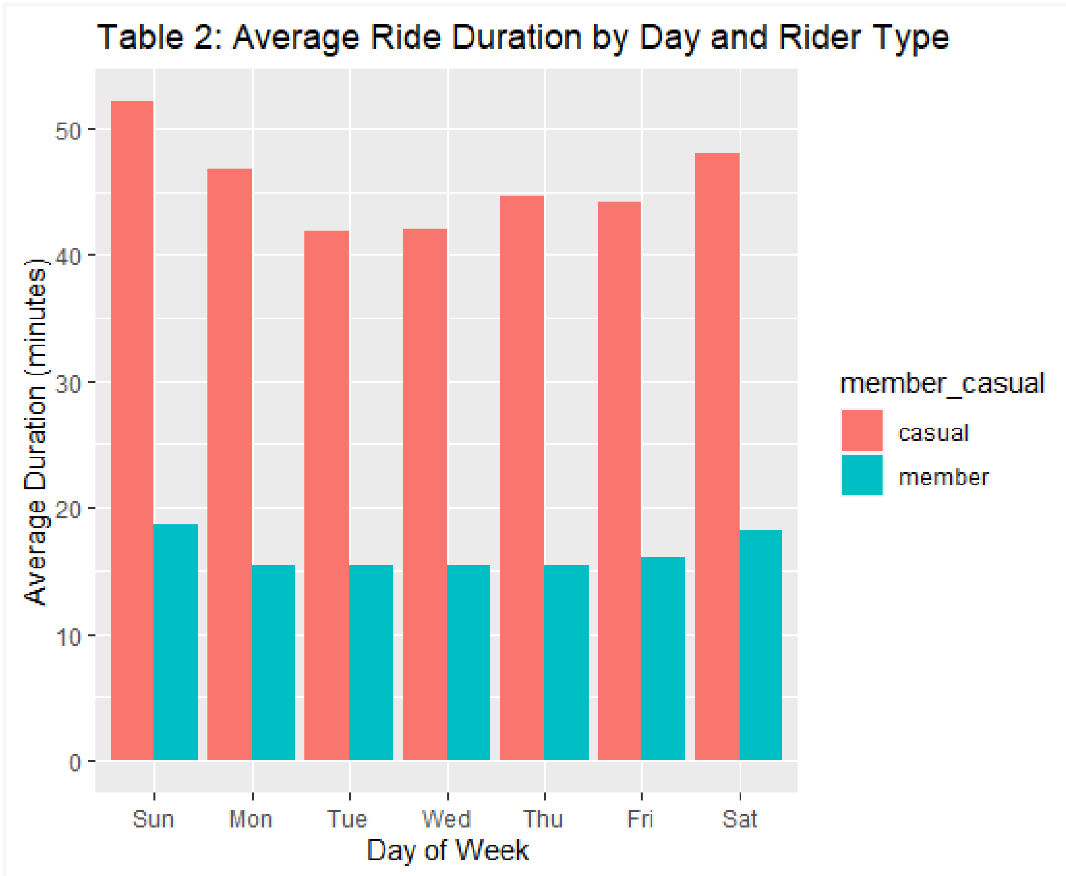
1. Number of rides by rider type



Code:

```
all_trips %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(title = "Table 1: Number of Rides by Day and Rider Type") +
  ylab("Number of Rides (1e+05 = 100,000)") +
  xlab("Day of Week")
```

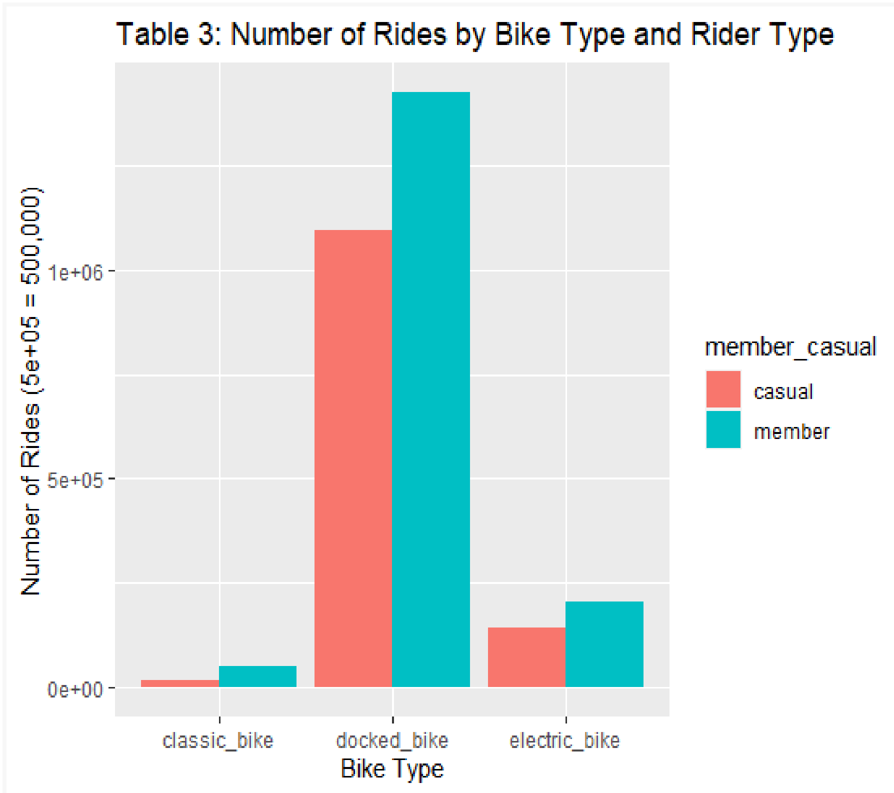
2. Average ride duration



Code:

```
all_trips %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length/60)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(title = "Table 2: Average Ride Duration by Day and Rider Type") +
  ylab("Average Duration (minutes)") +
  xlab("Day of Week")
```

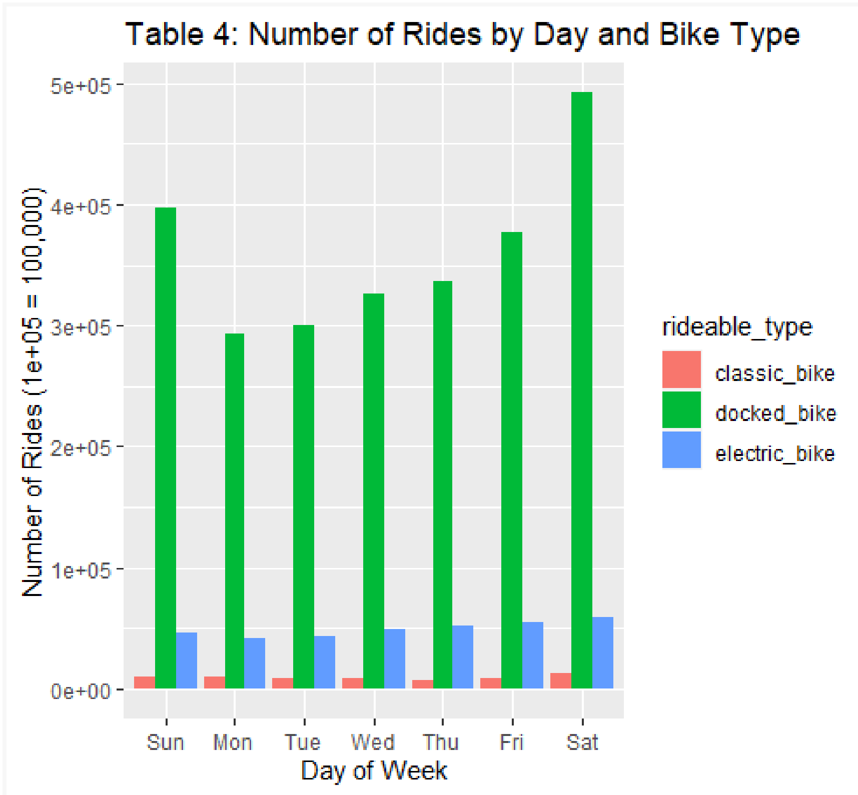
3. Number of rides by bike type and rider type



Code:

```
all_trips %>%
  group_by(member_casual, rideable_type) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(member_casual, rideable_type)  %>%
  ggplot(aes(x = rideable_type, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(title = "Table 3: Number of Rides by Bike Type and Rider Type") +
  ylab("Number of Rides (5e+05 = 500,000)") +
  xlab("Bike Type")
```

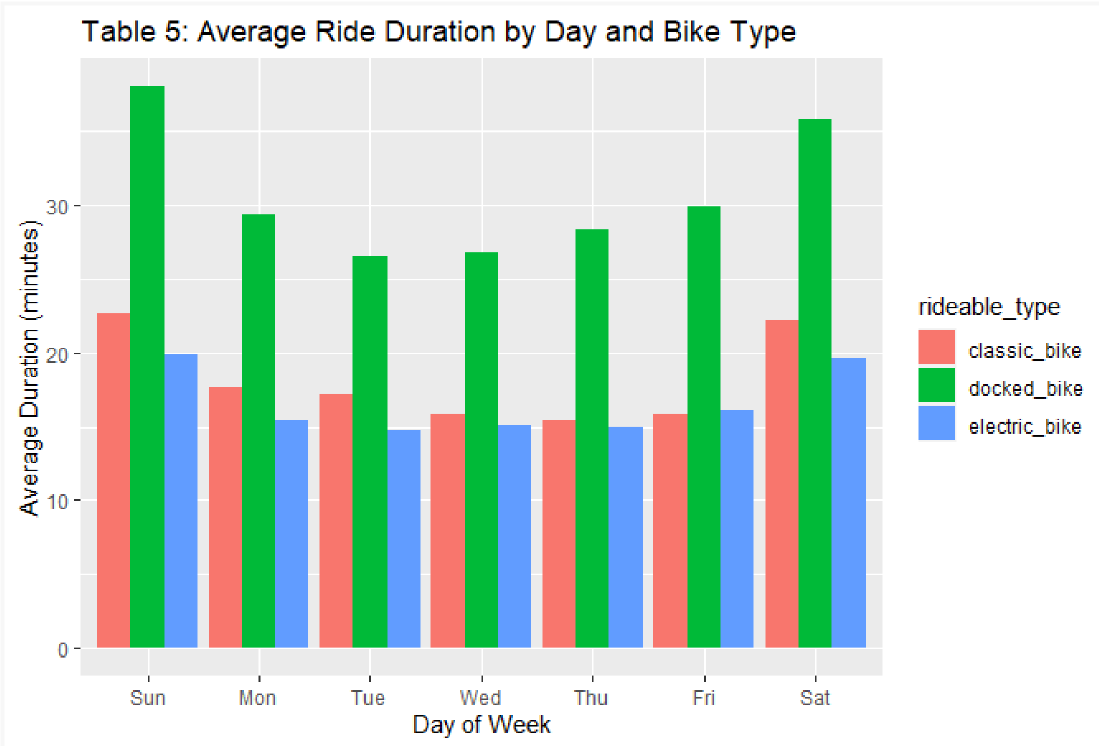
4. Number of rides by day and bike type



Code:

```
all_trips %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(rideable_type, weekday) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(rideable_type, weekday)  %>%
  ggplot(aes(x = weekday, y = number_of_rides, fill = rideable_type)) +
  geom_col(position = "dodge") +
  labs(title = "Table 4: Number of Rides by Day and Bike Type") +
  ylab("Number of Rides (1e+05 = 100,000)") +
  xlab("Day of Week")
```

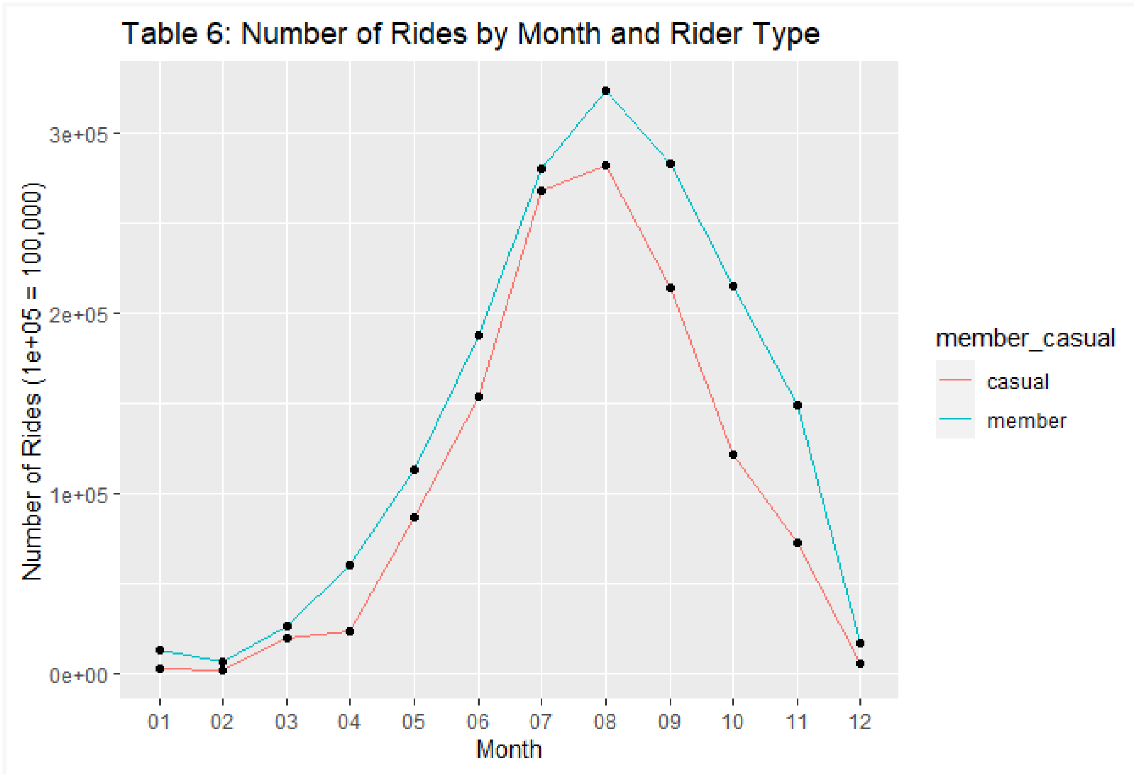
5. Average duration by bike type



Code:

```
all_trips %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(rideable_type, weekday) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length/60)) %>%
  arrange(rideable_type, weekday)  %>%
  ggplot(aes(x = weekday, y = average_duration, fill = rideable_type)) +
  geom_col(position = "dodge") +
  labs(title = "Table 5: Average Ride Duration by Day and Bike Type") +
  ylab("Average Duration (minutes)") +
  xlab("Day of Week")
```

6. Number of rides by month and rider type



Code:

```
all_trips %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(member_casual, month)  %>%
  ggplot(aes(x = month, y = number_of_rides, group = member_casual)) +
  geom_line(aes(color = member_casual)) +
  geom_point() +
  labs(title = "Table 6: Number of Rides by Month and Rider Type") +
  ylab("Number of Rides (1e+05 = 100,000)") +
  xlab("Month")
```

Case Study Roadmap - Share

Guiding questions

- **Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?**
  Yes, we were able to find out that casual riders use bikes for leisure and mostly on weekends. Whereas members are using bikes for commuting daily.

- **What story does your data tell?**
  The data shows that:
  1. Both user types scarcely ride Cyclistic bikes during the colder months.
  2. Ride length for casual riders is more than cyclistic members since casual riders use bikes for leisure whereas members use bikes for commuting.
  3. Members rent bikes on a more consistent basis throughout the entire week, whereas casual rentals are low Monday through Thursday and peak towards the weekend.

- **How do your findings relate to your original question?**
  The findings help us understand the bike usage among annual members and casual riders.

- **Who is your audience? What is the best way to communicate with them?**
  Our audience is the stakeholders i.e. the Cyclistic Executive Team and the best way to share findings with them is through the presentation.

- **Can data visualization help you share your findings?**
  Yes, visualizations will help understand the facts easily.

Key tasks

1. **Determine the best way to share your findings.**
   The best way to share findings is through an effective presentation.

2. **Create effective data visualizations.**
   Effective Data visualizations are created in R for easy understanding of data..

3. **Present your findings.**
   - Avg of ride = 30 mins
     Casual = 46 mins
     Member = 16 mins
   - Most popular day to rent a bike is Saturday.

- Bike rentals start off at a low on Mondays, peak on Saturdays with a slight drop off on Sundays (Table 1).
- Members rent bikes on a more consistent basis throughout the entire week, whereas casual rentals are low Monday through Thursday and peak towards the weekend (Table 1).
- On any day of the week, casual users ride 2.7x to 3x longer than members (Table 2).
- The docked bike option is far more popular than both classic bikes and electric bikes, both in terms of number of rentals (Table 3, Table 4) and average ride duration on each type of bike (Table 5).
- Bike rentals follow a seasonal pattern for both types of users (Table 6). Since Chicago experiences inclement weather, lowest usage is in the winter with rentals starting to ramp up in the spring. Peak usage is in the summer (August) before it starts to decline again during the Fall.

4. **Ensure your work is accessible.**
   The trends and findings will be accessible to the audience through presentation.

Deliverable

Supporting visualizations and key findings
Visualizations and key findings are given above.

Case Study Roadmap - Act

Your top three recommendations based on your analysis

My recommendations are:

- Run campaigns and promotions (e.g. referral discounts, group membership, priority access etc) for annual memberships in the summer when ridership is at its annual peak so that it reaches maximum number of people and introduce a promotional annual membership deal.
- Introduce a monthly pass for casual riders who might not want an annual subscription. The monthly pass could encourage more casual users to become members.
- Introduce late fees for full-day passes and cap the length of time a person can rent a bike with a single-ride pass. Keep flexible access exclusively for members.