**CYCLISTIC PROJECT DOCUMENTATION**

My approach in this project is based on 6 phases which are Ask, Prepare, Process, Analyze, Share and Act as follows.

**ASK**

**Three questions will guide the future marketing program**:

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

2. Why would casual riders buy Cyclistic annual memberships?

3. How can Cyclistic use digital media to influence casual riders to become members?

Moreno has assigned the first question to answer: How do annual members and casual riders use Cyclistic bikes differently?

**Guiding questions**:

● What is the problem you are trying to solve?

● How can your insights drive business decisions?

**Key tasks**:

1. Identify the business task

2. Consider key stakeholders

**Business Task**: Design a new marketing strategy to convert casual riders into annual members

**Key Stakeholders**:

● **Lily Moreno**: The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels.

● **Cyclistic marketing analytics team**: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy. You joined this team six months ago and have been busy learning about Cyclistic’s mission and business goals — as well as how you, as a junior data analyst, can help Cyclistic achieve them.

● **Cyclistic executive team**: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

**PREPARE**

**Guiding questions**:

● Where is your data located?

● How is the data organized?

● Are there issues with bias or credibility in this data? Does your data ROCCC?

● How are you addressing licensing, privacy, security, and accessibility?

● How did you verify the data’s integrity?

● How does it help you answer your question?

● Are there any problems with the data?

**Key tasks**:

1. Download data and store it appropriately.

2. Identify how it’s organized.

3. Sort and filter the data

4. Determine the credibility of the data.

I used the Cyclistic’s historical trip data to analyze and identify trends. I Downloaded the previous 12 months of Cyclistic trip data available on the AMAZON AWS. This is public data that can be used to explore how different customer types are using Cyclistic bikes. But note that data privacy issues prohibit you from using riders’ personally identifiable information. This means that you won’t be able to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.

**License**: The datasets have a different name because Cyclistic is a fictional company. For the purposes of this project, the datasets are appropriate and will enable you to answer the business questions. The data has been made available by Motivate International Inc. under the following link.

Click on the license link below to see the Data License Agreement

[**License**](https://divvybikes.com/data-license-agreement)

**PROCESS**

**Guiding questions**:

● What tools are you choosing and why?

● Have you ensured your data’s integrity?

● What steps have you taken to ensure that your data is clean?

● How can you verify that your data is clean and ready to analyze?

● Have you documented your cleaning process so you can review and share those results?

**Key tasks**:

1. Check the data for errors.

2. Choose your tools.

3. Transform the data so you can work with it effectively.

4. Document the cleaning process.

Tools Used: **Excel** & **R**

**Followed these steps**:

1. Downloaded the previous 12 months of Cyclistic trip data.

2. Unziped the files.

3. Created a folder on my desktop or Drive to house the files. Used appropriate file-naming conventions.

4. Created subfolders for the .CSV file and the .XLS so that I have a copy of the original data. Moved the downloaded files to the appropriate subfolder.

5. Followed these instructions for either Excel: Launch Excel, open each file, and choose to Save As an Excel Workbook file. Put it in the subfolder I created for .XLS files.

6. Opened spreadsheet and created a column called “ride\_length.” Calculated the length of each ride by subtracting the column “started\_at” from the column “ended\_at” (for example, =D2-C2) and format as HH:MM:SS using Format > Cells > Time > 37:30:55.

7. Created a column called “day\_of\_week,” and calculated the day of the week that each ride started using the “WEEKDAY” command (for example, =WEEKDAY(C2,1)) in each file. Format as General or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday.

8. Sorted according to start date from oldest data to newest data & added filter.

9. Proceed to the analyze step.

**ANALYZE**

**Guiding questions**:

● How should you organize your data to perform analysis on it?

● Has your data been properly formatted?

● What surprises did you discover in the data?

● What trends or relationships did you find in the data?

● How will these insights help answer your business questions?

**Key tasks**:

1. Aggregate your data so it’s useful and accessible.

2. Organize and format your data.

3. Perform calculations.

4. Identify trends and relationships.

**Followed these steps for using spreadsheets**:

1. Where relevant, made the columns consistent and combined them into a single worksheet.

2. Cleaned and transformed the data to prepare for analysis.

3. Conducted descriptive analysis.

4. Run a few calculations in one file to get a better sense of the data layout.

● Calculated the mean of ride\_length

● Calculated the max ride\_length

● Calculated the mode of day\_of\_week

5. Created a pivot table to quickly calculate and visualize the data.

● Calculated the average ride\_length for members and casual riders. rows = member\_casual; Values = Average of ride\_length.

● Calculated the average ride\_length for users by day\_of\_week. columns = day\_of\_week; Rows = member\_casual; Values = Average of ride\_length.

● Calculated the number of rides for users by day\_of\_week by adding Count of trip\_id to Values.

6. Open another file and performed the same descriptive analysis steps. Explored different seasons to make some initial observations.

7. Merged all files into a full year file and did few analysis.

8. Export a summary file for further analysis.

**Followed these steps for using R Open R Studio**:

1. Imported data.

2. Made the columns consistent and merge them into a single dataframe.

3. Cleaned up and added data to prepare for analysis.

4. Conducted descriptive analysis.

5. Export a summary file for further analysis.

**For R Studio analysis refer to the R Console docx file**.

**SHARE**

**Guiding questions**:

● Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?

● What story does your data tell?

● How do your findings relate to your original question?

● Who is your audience? What is the best way to communicate with them?

● Can data visualization help you share your findings?

● Is your presentation accessible to your audience?

**Key tasks**:

1. Determine the best way to share your findings.

2. Create effective data visualizations.

3. Present your findings.

4. Ensure your work is accessible.

**Followed these steps**:

1. Take out a piece of paper and a pen and sketch some ideas for how you will visualize the data.

2. Choose a visual form, and open tool of choice to create the visualization. Use presentation software, such as PowerPoint or Google Slides; your spreadsheet program; Tableau; or R.

3. Create data visualization, remembering that contrast should be used to draw your audience’s attention to the most important insights. Use artistic principles including size, color, and shape.

4. Ensure clear meaning through the proper use of common elements, such as headlines, subtitles, and labels.

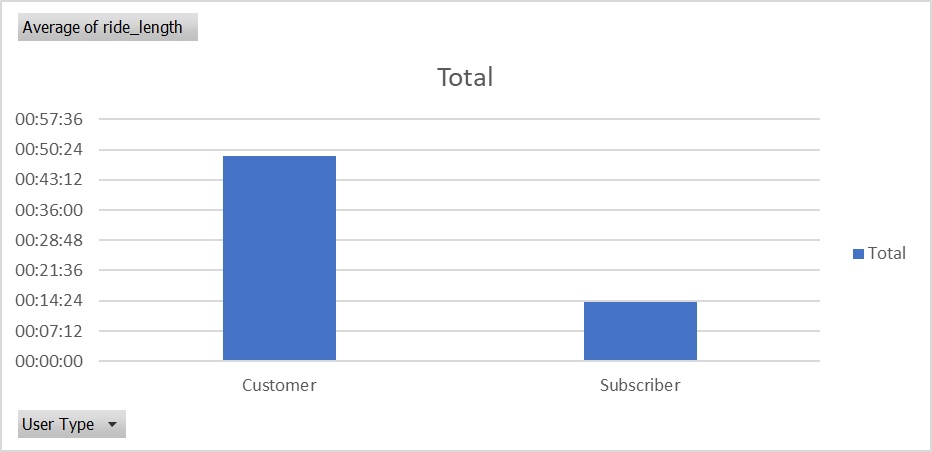
5. Refine your data visualization by applying deep attention to detail.

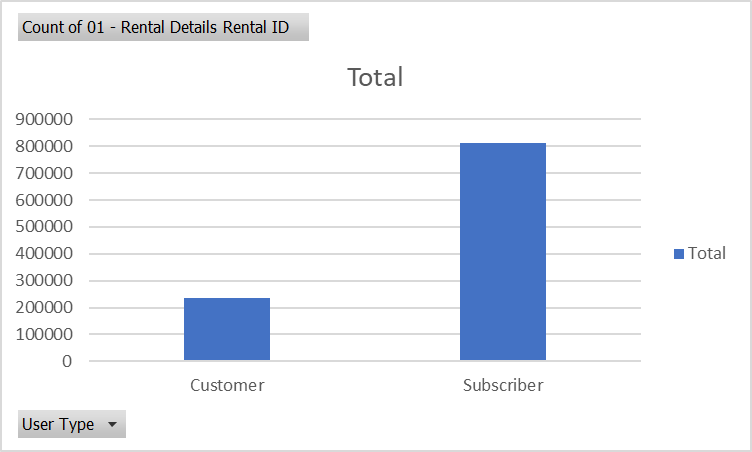
Seasons in New York City are as follows:

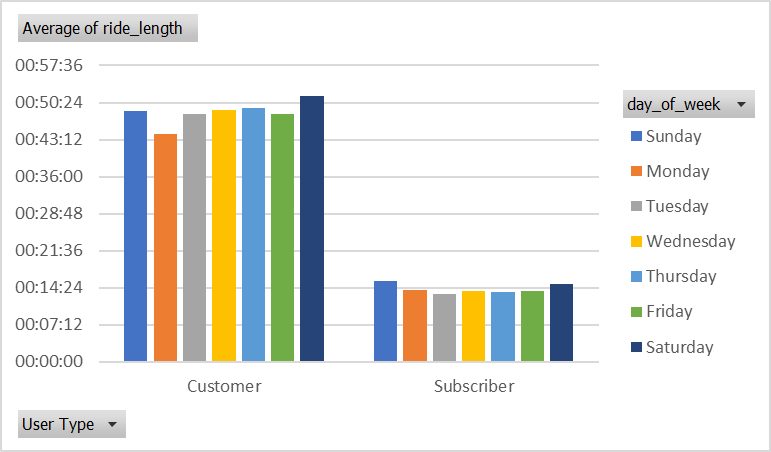
* Spring Months: March, April, May
* Summer Months: June, July, August
* Autumn Months: September, October, November
* Winter months: December, January, February

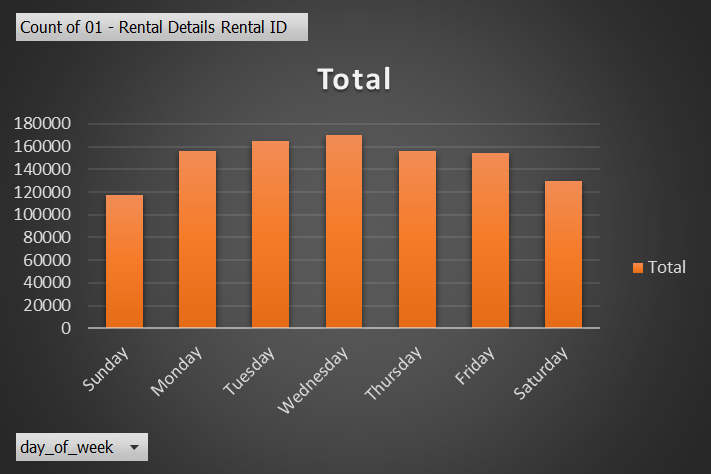
**Q2 2019**

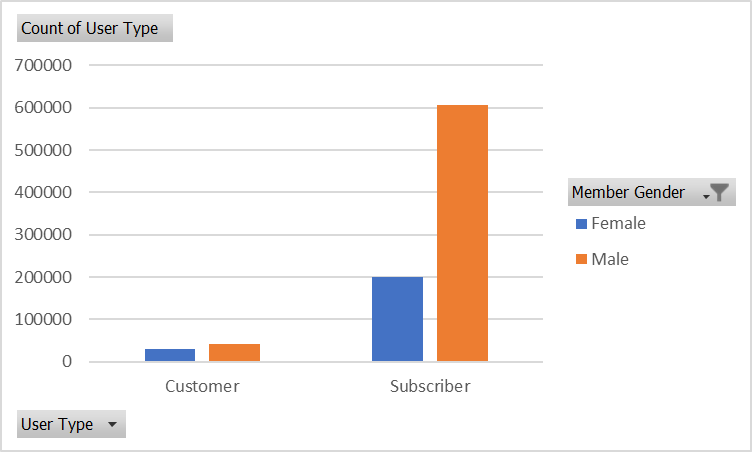
|  |  |
| --- | --- |
| **average\_ride\_length** | 00:21:50 |
| **max\_ride\_length** | 01:33:58 |
| **mode\_day\_of\_week** | 4 |

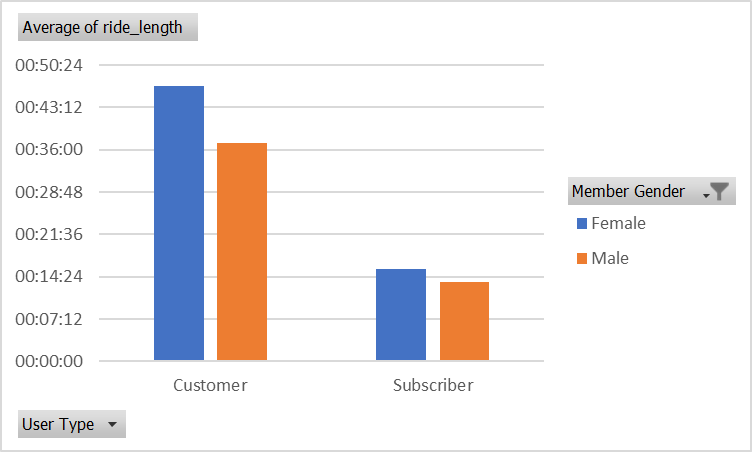


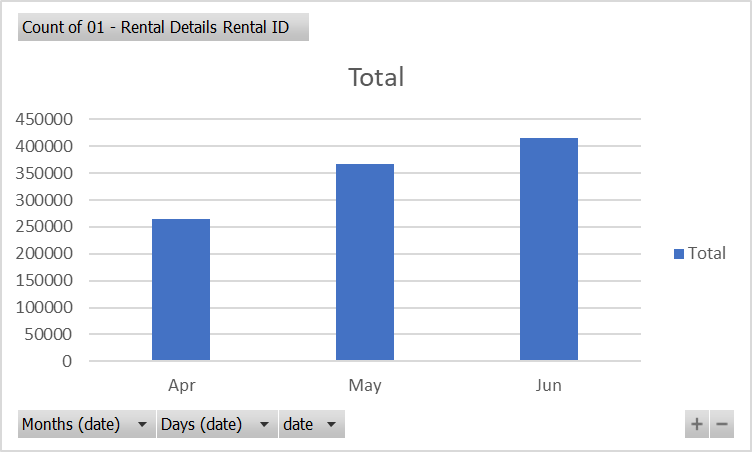


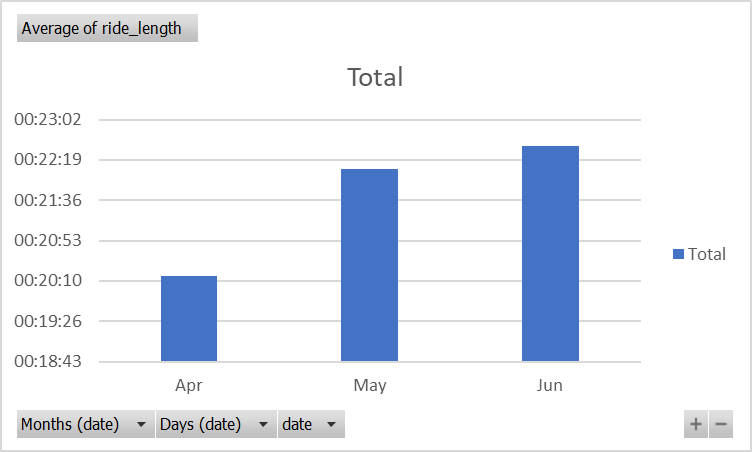






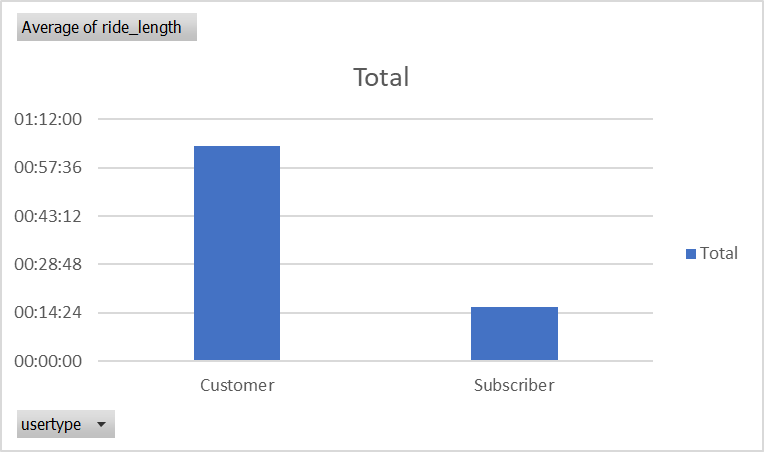




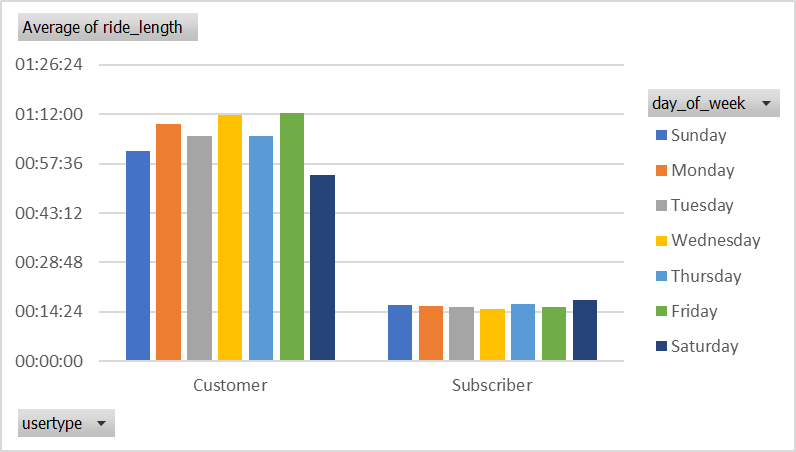


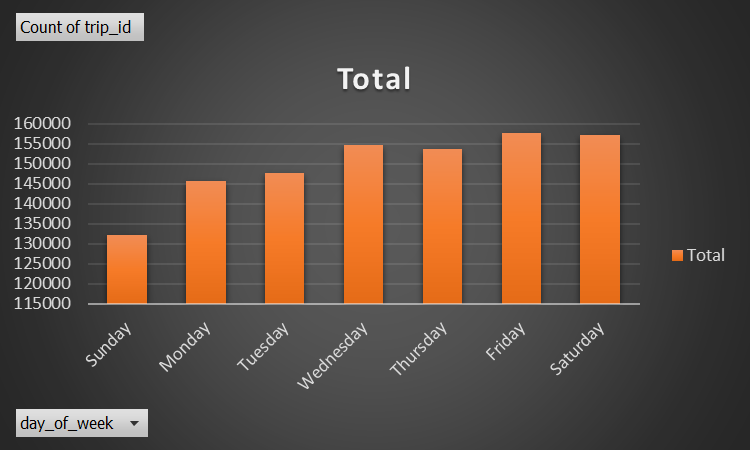
**Q3 2019**

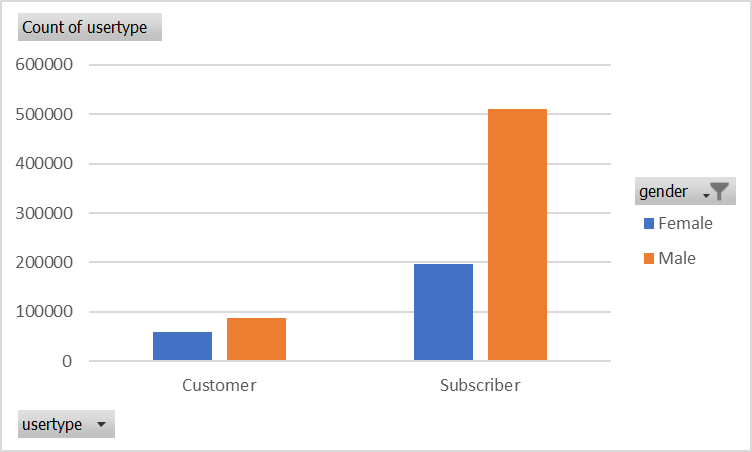
|  |  |
| --- | --- |
| **average\_ride\_length** | 00:31:29 |
| **max\_ride\_length** | 19:43:54 |
| **mode\_day\_of\_week** | 6 |

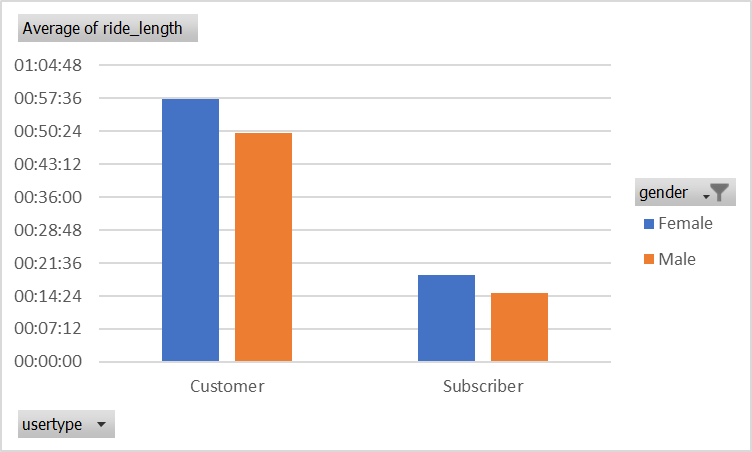


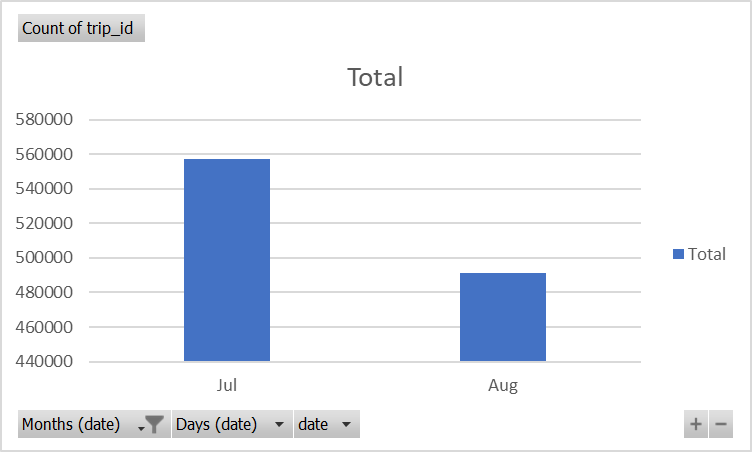


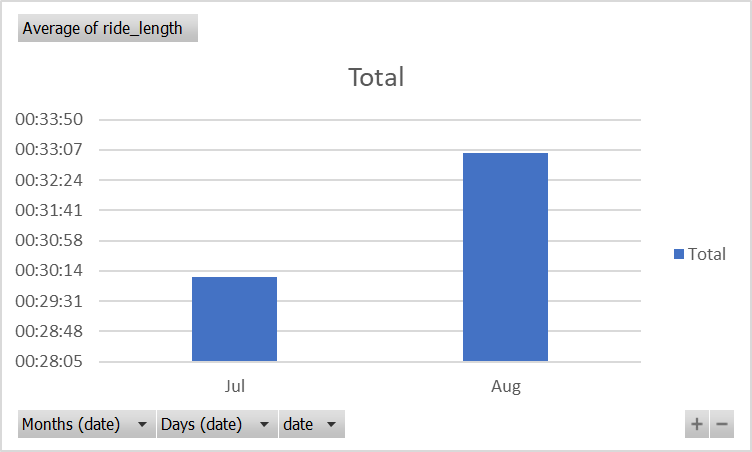






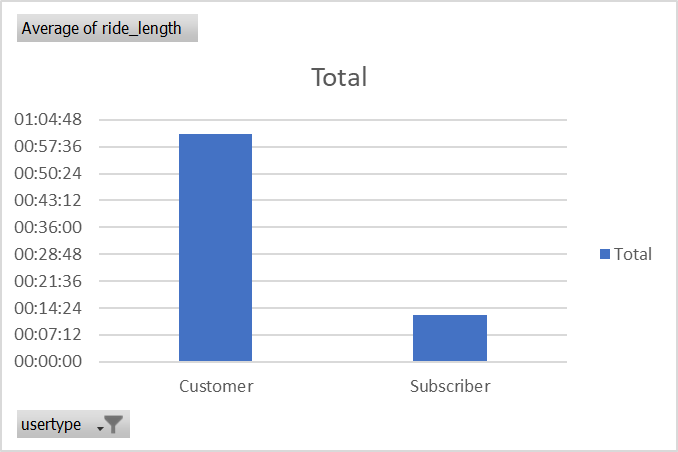


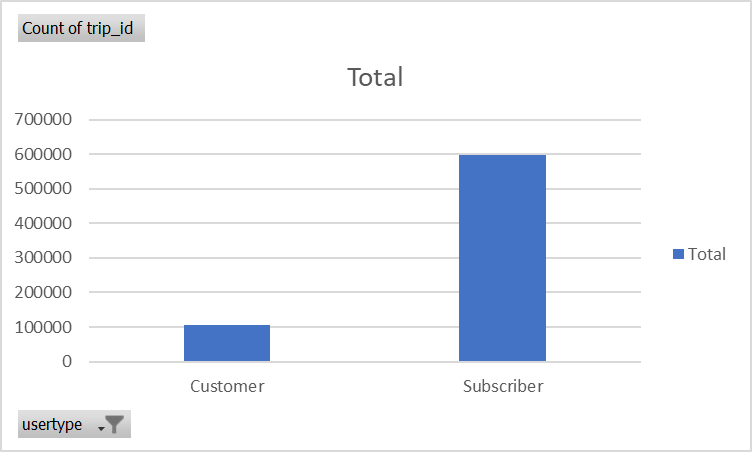


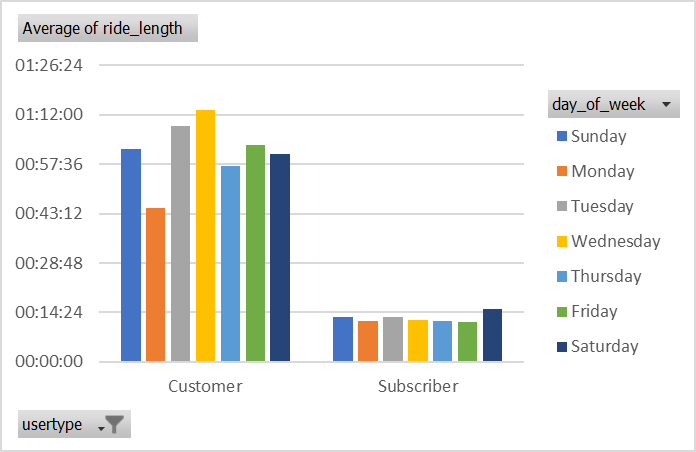


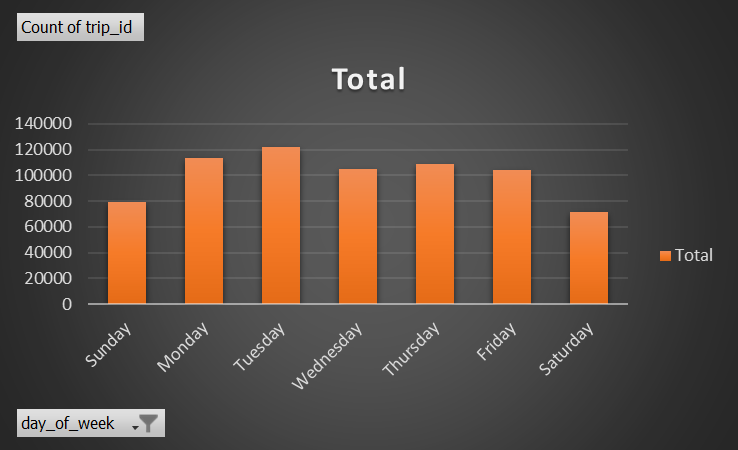
**Q4 2019**

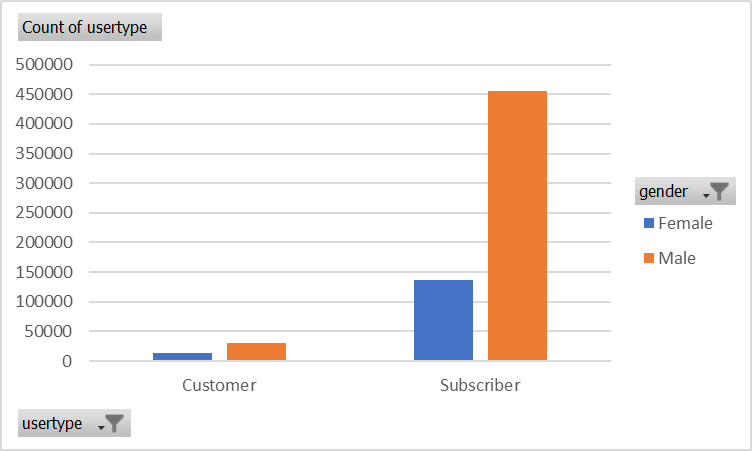
|  |  |
| --- | --- |
| **average\_ride\_length** | 00:19:49 |
| **max\_ride\_length** | 07:58:22 |
| **mode\_day\_of\_week** | 3 |

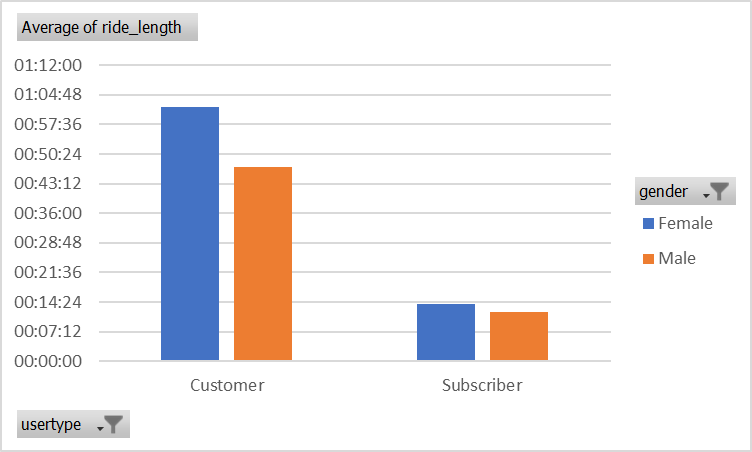


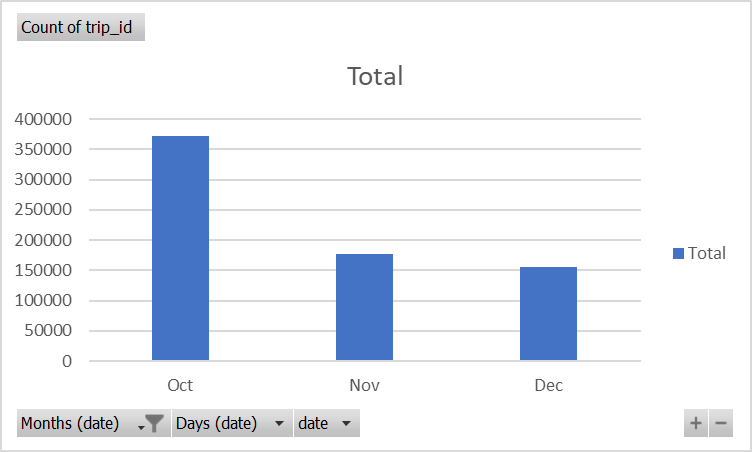


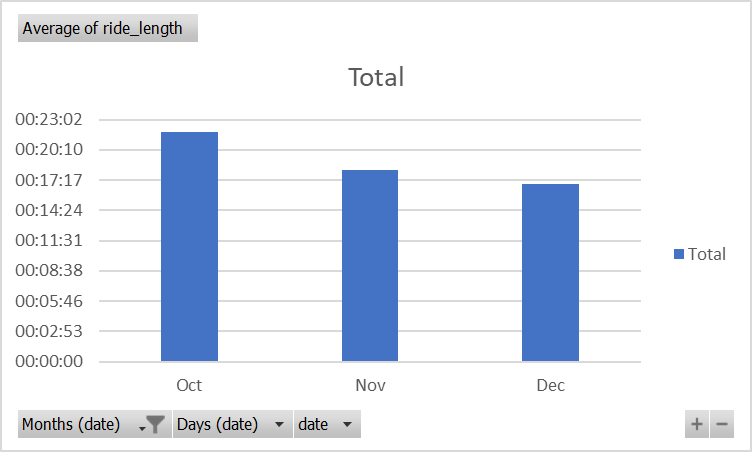






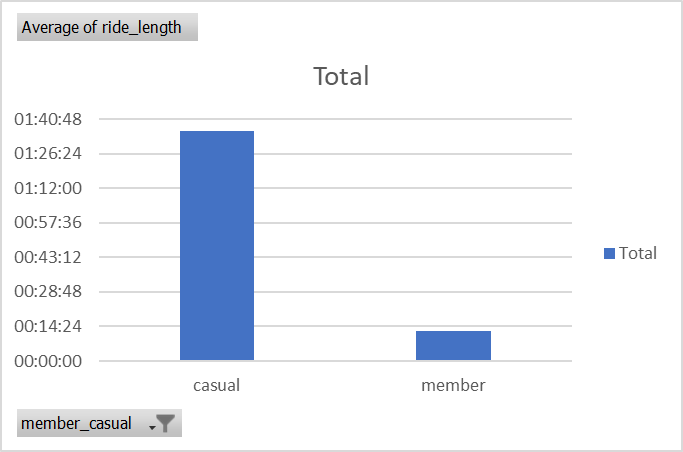


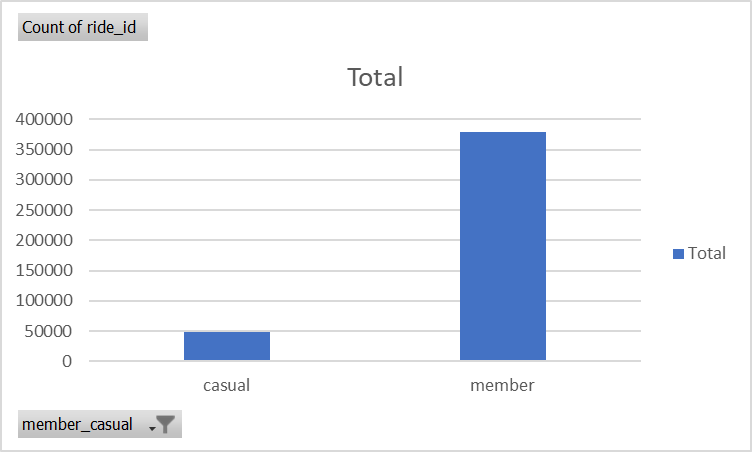


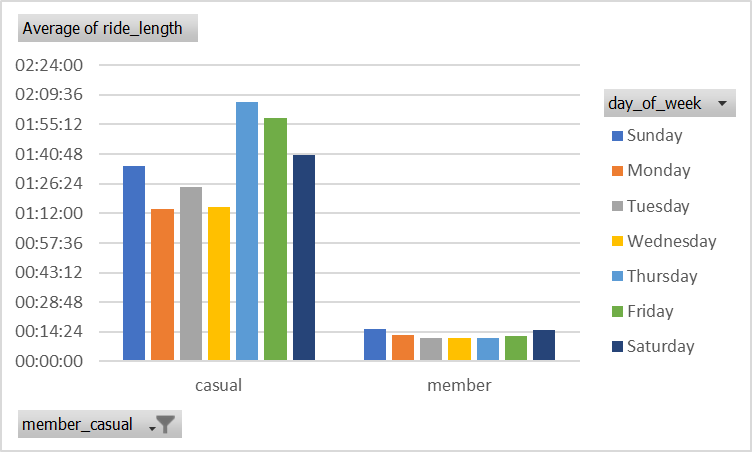


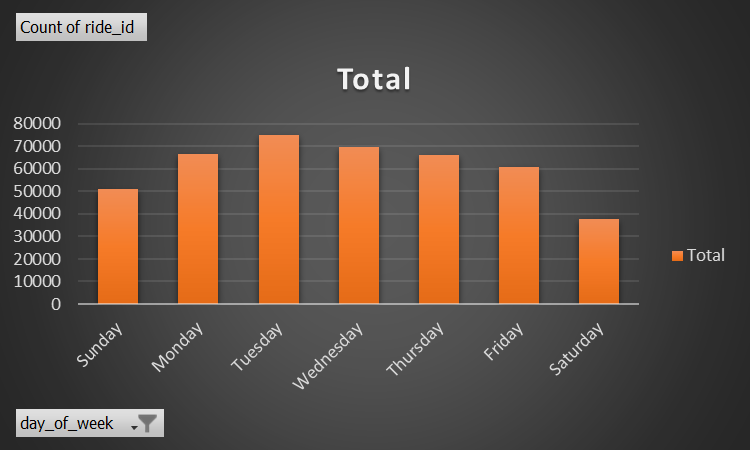
**Q1 2020**

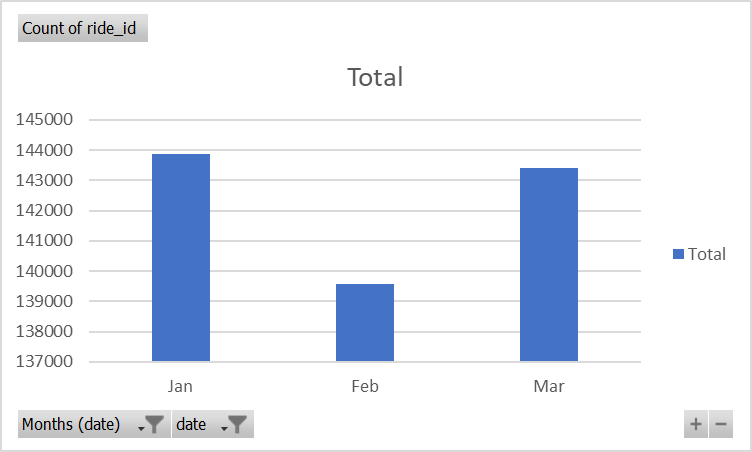
|  |  |
| --- | --- |
| **average\_ride\_length** | 00:22:07 |
| **max\_ride\_length** | 15:30:24 |
| **mode\_day\_of\_week** | 3 |

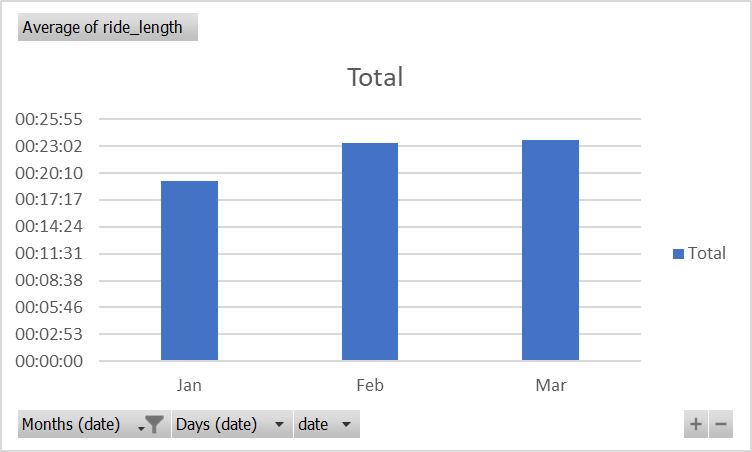












**ACT**

**Conclusion**:

* Overall, the no. of Casual users are comparatively very less than members
* Average ride length of casual users is significantly more as compared to members
* In terms of being Members, Male individuals are comparatively way more than Female individuals
* During summer season the no. of users are more than rest of the season
* During Winter season the no. of users are less than rest of the season

**Recommendations**:

* Creating a loyalty program is a great method to get users to use your bike sharing service again and again. Reward repeat customers, referrals, or even social media activity engagement. These might come in the form of discounts, cost-free services, or exclusive benefits only available to members, including early access to new events.
* Promote local routes and trails on your website, blog, or social media accounts to highlight the distinctive features of your area. Post pictures or videos of beautiful rides, point out difficult parts, or suggest places to halt along the road.
* Potential customers consider recommendations and testimonials from previous customers while making decisions. To build audience trust and credibility, gather and display good reviews on the website and social media channels. For your website to appear higher in search results when prospective clients research for bike sharing or similar services in your region, local SEO optimization is essential. Make sure your website has calls to action that are obvious, locally relevant content, current contact information, and relevant keywords.