Hi Data Folks,

I am currently learning data analytics through the [Google Data Analytics Professional Certificate](https://www.coursera.org/professional-certificates/google-data-analytics) and have reached the final course: [Google Data Analytics Capstone: Complete a Case Study.](https://www.coursera.org/learn/google-data-analytics-capstone/home/welcome" \t "_blank)

In this capstone project, I will be working on a case study for **Cyclistic**, a fictional bike-share company. To answer key business questions, I will follow the data analysis process: **Ask, Prepare, Process, Analyze, Share, and Act**.

Excited to share my first case study project with you all!

**Scope of work (SOW)**

* Data analyst team wants to understand how **casual riders** and **annual members** use Cyclistic bikes differently.
* From these insights, team will design **a new marketing strategy** to convert **casual riders** into **annual members.**
* **Executive** approval is necessary for recommending **new** **strategies**, so **compelling data** insights and **professional visualizations** are **essential**.

**My JD (Job Description) as Data Analyst on this project**

Answer the question, how do annual members and casual riders use Cyclistic bikes differently?

My report will include the following key deliverables:

1. **Business Task:** A clear statement defining the business problem and objectives.
2. **Data Sources:** A detailed description of all data sources used in the analysis.
3. **Data Cleaning & Preparation:** Documentation of any data cleaning, transformation, or manipulation performed.
4. **Analysis Summary:** A concise summary of key insights derived from the data.
5. **Visualizations & Findings:** Supporting charts, graphs, and key takeaways from the analysis.
6. **Recommendations:** My top three data-driven recommendations based on the findings.

**Ask**

**Business Goal**

Marketing Director wants to design marketing strategies aimed at converting **casual riders** into

**annual members**. In order to do that, the marketing data analyst team needs to better understand.

* How do annual members and casual riders use Cyclistic bikes differently?
* Why would casual riders buy Cyclistic annual memberships?
* How can Cyclistic use digital media to influence casual riders to become members?

**Stakeholders**

* Cyclistic Executive Team
* Lily Moreno (Director of Marketing)
* Cyclistic Marketing Analytics Team

**Prepare**

**Data sources:**

For my analysis, I’ll be working with Cyclistic bike trip data from the past 12 months. The data has been made available This data will help me understand how different types of customers—**annual members** and **casual riders**—use the bike-sharing service.

Here’s what you need to know about the data:

* It’s **publicly available** and provided by Motivate International Inc.( [Download the previous 12](https://divvy-tripdata.s3.amazonaws.com/index.html) [months of Cyclistic trip data here.](https://divvy-tripdata.s3.amazonaws.com/index.html)) under ([license](https://www.divvybikes.com/data-license-agreement)).
* Even though Cyclistic is a fictional company, this dataset is real and perfect for this case study.
* **Privacy is a priority!** The data does not include any personal details like names or payment information. So, I can’t see where riders live or if they’ve bought multiple passes—but that’s okay! There’s still plenty to **explore**.

**Data Organization:**

The dataset is organized in CSV format, containing the following key attributes:

* Trip start and end times
* Trip duration
* Start and end station details
* Bike types (traditional, reclining, hand tricycles, cargo bikes)
* User categories (**annual members** and **casual riders**)

With this data, I’ll dive into trends and patterns to uncover insights about Cyclistic’s riders. Let’s see what the numbers reveal! 🚴‍♂️📊

**Process**

I started by unzipping the downloaded files. Then, I created a folder named "Cyclistic Data" on my desktop/Drive to organize the files properly.

To maintain a structured file system, I followed appropriate file-naming conventions and created two subfolders:

* CSV Files (to store the .csv file)
* XLS Files (to store the .xls file)

After organizing the folders, I moved the downloaded files into their respective subfolders to ensure I have a copy of the original data for future reference.

This setup helps keep the data well-organized and easily accessible.

**Data Cleaning & Manipulation** **(Dataset:** *Divvy\_Trips\_2020\_Q1.xls*)

**Creating the ride\_length Column**

* Added a new column ride\_length in column N.
* Calculated the ride duration using the formula:

*=D2 - C2* (where **D** is ended\_at and **C** is started\_at).

* Formatted ride\_length as **HH: MM: SS** using:
* Format > Cells > Time > 37:30:55

**Creating the day\_of\_week Column**

* Added a new column **day\_of\_week** in column **N**.
* Extracted the weekday number of each ride using the formula:  
  *=WEEKDAY(C2,1*) (where **C** is started\_at).
* Formatted day\_of\_week as **General** (1 = Sunday, 7 = Saturday).

**Manipulation in column of user\_type**

* First Replace column header from user\_member to user\_type
* Then **378407** cells member has been replaced with Annual Member and **48480** cells of Customer replace with Casual Rider through Find and Replace tool (Home > Find & Select > Replace).

**Dataset:** *Divvy\_Trips\_2019\_Q1.csv*

Performed the same descriptive data formatting and manipulation on this data sheet.

1. Created aride\_length **column in Column M** using the formula =C2 - B2, where:

* **Column C** is ended\_at.
* **Column B** is started\_at.

1. Created aday\_of\_week **column in Column N** using the formula =WEEKDAY(B2,1), referencing:

* **Column B** as started\_at.

1. Renamed the heading of Column **J** from usertype to user\_type.
2. Renamed the heading of Column **D** from usertype to user\_type.
3. Updated entries in Column **J (**user\_type**)** using the **Find and Replace** tool (**Home > Find & Select > Replace**):

* Replaced **341,906** instances of Subscriber with Annual Member.
* Replaced **23,163** instances of Customer with Casual Rider.

1. Formatted Columns **B (**start\_time**) and C (**end\_time**)** to meet SQL requirements:

* Selected both columns, right-clicked, and chose **"Format Cells..."**.
* Navigated to the **"Number"** tab and selected **"Custom"**.
* Entered yyyy:mm:dd hh:mm:ss in the **"Type"** field.
* Clicked **"OK"** to apply the new format, aligning dates to yyyy:mm:dd hh:mm:ss for SQL data upload.

**Dataset:** *Divvy\_Trips\_2019\_Q2.csv*

Data Formatting for Data Cleaning and Manipulation:

1. Aligning Column Names in Divvy\_Trips\_2019\_Q2.csv with Divvy\_Trips\_2019\_Q1.csv

* Column-A: 01 - Rental Details Rental ID > trip\_id
* Column-B: 01 - Rental Details Local Start Time > start\_time
* Column-C: 01 - Rental Details Local End Time > end\_time
* Column-D: 01 - Rental Details Bike ID > bikeid
* Column-E: 01 - Rental Details Duration In Seconds Uncapped > tripduration
* Column-F: 03 - Rental Start Station ID > from\_station\_id
* Column-G: 03 - Rental Start Station Name > from\_station\_name
* Column-H: 02 - Rental End Station ID > to\_station\_id
* Column- I: 02 - Rental End Station Name > to\_station\_name
* Column-J: User Type > user\_type
* Column-K: Member Gender > gender
* Column-L: 05 - Member Details Member Birthday Year > birthday

1. Applied data cleaning and manipulation to Divvy\_Trips\_2019\_Q2.csv as did for Divvy\_Trips\_2019\_Q1.csv. Additionally, updated entries in Column J (user\_type) using the **Find and Replace** tool (**Home > Find & Select > Replace**)."

* Replaced **812642** instances of **Subscriber** with **Annual Member**.
* Replaced **235935** instances of **Customer** with **Casual Rider**.

**Dataset:** *Divvy\_Trips\_2019\_Q3.csv*

Data Formatting for Data Cleaning and Manipulation:

1. Aligning Column Names in Divvy\_Trips\_2019\_Q3.csv with Divvy\_Trips\_2019\_Q1.csv

* Column-D: bikeid > bike\_id
* Column-J: usertype > user\_type

1. Applied data cleaning and manipulation to Divvy\_Trips\_2019\_Q2.csv as did for Divvy\_Trips\_2019\_Q1.csv. Additionally, updated entries in Column J (user\_type) using the **Find and Replace** tool (**Home > Find & Select > Replace**)."

* Replaced **713548** instances of **Subscriber** with **Annual Member**.
* Replaced **335047** instances of **Customer** with **Casual Rider**.

**Dataset:** *Divvy\_Trips\_2019\_Q4.csv*

Data Formatting for Data Cleaning and Manipulation:

1. Aligning Column Names in Divvy\_Trips\_2019\_Q3.csv with Divvy\_Trips\_2019\_Q1.csv

* Column-D: bikeid > bike\_id
* Column-J: usertype > user\_type

1. Applied data cleaning and manipulation to Divvy\_Trips\_2019\_Q2.csv as did for Divvy\_Trips\_2019\_Q1.csv. Additionally, updated entries in Column J (user\_type) using the **Find and Replace** tool (**Home > Find & Select > Replace**)."

* Replaced **597860**instances of **Subscriber** with **Annual Member**.
* Replaced **106194** instances of **Customer** with **Casual Rider**.

**Purpose of These Manipulations**

* To analyze how **Annual Members** and **Casual Riders** use the service.
* Understanding ride duration and patterns based on the day of the week.
* Preparing clean data for further analysis in the next step.

This cleaning process ensures consistency and accuracy before moving forward with the **Analyze** phase.

**Analyze**

**(Dataset:** *Divvy\_Trips\_2019\_Q1, Q2, Q3 & Q4)*

Follow these steps for using SQL

Open your SQL tool of choice, then complete the following steps:

1. Import your data.

2. Explore your data, perhaps looking at the total number of rows, distinct values,

maximum, minimum, or mean values.

3. Where relevant, use JOIN statements to combine your relevant data into one table.

4. Create summary statistics.

5. Investigate interesting trends and save that information to a table.

**Organizing and Preparing Data**

* Checked for missing values and inconsistencies.
* Sorted and filtered columns to confirm data integrity.
* Verified dataset bias.
* Identified unusual patterns like canceled rides, notably at Station 675.

**Descriptive Analysis & Key Calculations**

* Calculated mean and max ride\_length.
* Found the mode of day\_of\_week.

**Creating Pivot Tables**

* Compared average ride\_length for members vs. casual riders.
* Analyzed average ride\_length by day of the week for each user type.
* Counted rides per day of the week.

**Seasonal Analysis & Observations**

* Repeated descriptive analysis on another dataset for seasonal variations.
* Considered data formatting, unexpected patterns, trends, and business insights.
* Uncovered high cancellations at Station 675, suggesting issues for casual riders.

**Share**

**Observations**

**Annual member’s usage pattern of Cyclistic services.**

Annual members booked almost 40% of the rides compared to the other days of the week on the first day of the week. On the second day, the booking reached 80% of ride bookings compared to other days of the week, then from the third day, it gradually went down until the sixth day of the week. On the seventh day of the week, it almost reached the level of the first day.

**Casual User’s usage pattern of Cyclistic services.**

Casual members' booking rate was higher on the first day of the week compared to their usage on other days. From the second day, ride bookings declined throughout the week and remained consistent at the same number of bookings until the sixth day, then increased on the seventh day of the week.

**Main findings between both User Type Uses**

1. **Annual Members**

* Higher usage during weekdays (Monday to Friday) as they rely on the service for daily commuting.
* Familiar with the service, making it a convenient and time-saving commuting option.
* Knowledge of nearby docking stations makes it easier for them to use.
* Cost-effective and budget-friendly for regular use.
* Lower usage on weekends and public holidays due to rest days and preference for other leisure activities.

1. **Casual Users**

* Higher booking rates on weekends and public holidays due to more free time.
* Likely to try the service out of curiosity or for leisure purposes.
* Increased bike availability on weekends may encourage more casual usage.
* Significant drop in bookings from Monday to Friday.
* Possible reasons for low weekday usage include difficulty in booking, lack of trust in the service, or unfamiliarity with how to use it effectively.

1. **Ride Cancellation Observations**

* Analysis of ride length data showed 481 rides lasting between 0:00:00 and 00:00:01, indicating immediate cancellations.
* All these rides were booked from station ID 675.
* Further investigation revealed that station ID 675 was primarily used by casual riders.
* Nearly 3,768 ride attempts were recorded from this station, with an average ride duration of just 00:00:03.
* This suggests that casual users at this station faced booking or usability issues, leading to immediate ride cancellations.

**Act**

**Recommendations**

**1. Improve First-Time User Experience:** Many casual riders struggle with booking and using the service, leading to cancellations. To address this:

* Simplify the booking process with a step-by-step guide in the app.
* Provide an easy-to-access tutorial at docking stations.
* Offer a **"First Ride Free"** campaign to let new users experience the service risk-free.

**2. Targeted Promotions & Discounts:** Since casual users are more active on weekends, incentivize them to become annual members by:

* Introducing **weekend-to-weekday transition discounts** (e.g., "Get 50% off your first month as an annual member").
* Providing **family or group plans** to encourage frequent use.
* Sending personalized membership offers via email or app notifications after their first few rides.

**3. Optimize Station ID 675 & Customer Engagement:** With high cancellations and ride failures at Station 675, it's essential to:

* Investigate and **fix potential issues** (e.g., bike availability, technical errors).
* Reach out to users who attempted rides from this station, offer them a **discounted trial membership**, and ask for feedback.
* Organize a **community event** at this station, allowing casual riders to learn about the benefits of annual membership firsthand.

**What result could be come out?**

***“With this all you can make your consumer your regular customers and increase the sale and revenue of your products”***