**Introduction:**

This relates to the first capstone project of the Google Data Analytics Certificate, in which I performed a variety of real-world data analyst functions. The case is about converting casual cyclistic riders to annual members.

**Scenario**:   
You are a junior data analyst on the marketing analyst team of Cyclistic, a bike-sharing firm in Chicago. The head of marketing feels that the business's future prosperity is dependent on increasing the number of yearly subscriptions. Your team wants to know how casual riders and yearly members utilize Cyclistic bikes differently. Using these findings, you and your colleagues will develop an innovative advertising strategy to turn casual riders into yearly members. But first, Cyclistic executives must approve your proposals, which must be supported by compelling data insights and sophisticated data visualizations.

**About Cyclistic**:  
Cyclistic established its successful bike-sharing program in 2016. Since then, the initiative has expanded to a collection of 5,824 bicycles that have been geotracked and linked to a network of 692 sites across Chicago. The bikes may be unlocked from a single location and restored to any other station in the network at any time.   
Until recently, Cyclistic's marketing approach focused on raising public awareness and appealing to a wide range of customer demographics. One technique that helped make these things feasible was the flexibility of its price structures, which included single-ride permits, full-day passes, and yearly memberships. Customers who buy single-ride or full-day passes are known as casual riders. Customers that purchase annual subscriptions become Cyclistic members.

**The Stakeholders**

1. **Lily Moreno**: The director of marketing and my manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program.
2. **Cyclistic marketing analytics team:** A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy.
3. **Cyclistic executive team**: A notoriously detail-oriented team which will decide whether to approve the recommended marketing program.

**The Business Task**

The purpose of this particular case paper is to give straightforward guidance into creating marketing tactics that convert casual riders into yearly members. To achieve this purpose, I asked the following questions:

1. In what ways do yearly and casual riders utilize Cyclistic bikes distinctly?
2. Why do casual cyclists purchase Cyclistic yearly memberships?
3. In what manner can Cyclistic leverage digital media to encourage casual cyclists to register as members?

**Preparing the Data**   
The current case study analyzes and identifies trends in Cyclistic's historically trip data (the past 12 months). Motivate International Inc. has released the data to the public under an open license. The data can be dowloaded [here.](https://divvy-tripdata.s3.amazonaws.com/index.html)

\*  
This data is trustworthy, original, thorough, and current because it was gathered and carefully preserved by Cyclistic from June 2020 to May 2021. Personal identifying information, such as credit card numbers, has been deleted due to data privacy concerns. The data chosen for usage spans the past twelve months, which runs from June 2020 to May 2021. Each month has a unique dataset. The datasets are structured in a tabular manner and contain 13 identical columns. Together, the datasets include 4073561 rows. The member\_casual column will allow me to categorize, collect, and analyze patterns among casual and member riders.

**Processing the Data from Dirty to Clean**:

**Tools:**

To go through the data from filthy to clean, I picked Python. Python is quite quick, making it suitable for dealing with large dasets. Python is also well supported by useful open-source tools like pandas and matplotlib and R.

**Cleaning the data:**

After scanning in and integrating the 12 datasets into the same dataframe, the first stage in the cleanup process was determining which columns and rows included missing data. I discovered that 6 out of 13 columns contained missing data. Furthermore, 314299 rows included missing data.  
Next, I indexed into the first and last rows with missing data and discovered that they included null values in numerous columns. I also calculated the percentage of rows that contained missing values. This was at 7.71%. After examining the columns with missing data, I realized that imputation would be an incorrect technique due to the nature of the missing data. I opted to eliminate the rows with missing values because they only make up a small fraction of our data and appear to be missing values in numerous fields.   
After removing rows with missing data, I checked to see whether any of the remaining data points were duplicates. No rows were duplicated. Following this phase, I was sure that the data was ready for additional processing and analysis.

**Transform the data**:

After that, I investigated the data overview and saw that the **started\_at** and **ended\_at** columns were texts rather than datetime values. As a result, I changed the columns to datetime using the pandas **to\_datetime()** method.  
Next, I calculated the **ride\_length** column by subtracting the **ended\_at** and **started\_at** values. This produced a time delta object, which I translated to seconds and minutes.

Furthermore, I added two columns: **day\_of\_week** and **month\_name**. These indicate the day of the week and month in which a bike was rented.   
After receiving the summary to check that the data was ready for analysis, I observed that the ride length column had negative numbers. I selected the rows with negative ride length values to investigate further. A deeper study revealed that the negative numbers were caused by the **ended\_at** day or time being shorter than the **started\_at**. Only 10,237 rows displayed this phenomenon, and they were quickly sorted away.  
Finally, I received a summary of the data and determined that it was suitable for analysis.

**Analyze data to answer questions:**

In this phase, I examined the cleansed data to see how yearly and casual riders utilize Cyclistic bikes differently.  
First, I determined the total number of bikes hired and how they were distributed amongst casual and member riders. Following that, I looked at how total bike rentals were allocated each month and then per day. This showed several intriguing tendencies, which I will address during the sharing stage.   
Afterwards, I looked at how bike rentals between the two types of rider groups contrasted in a certain month and day of the week. The purpose at this point was to determine whether casual riders preferred specific days or months over member riders.

Subsequently I wanted to examine the average ride lengths of casual and member cyclists. I noticed that casual motorcyclists ride for longer amounts of time than regular riders. I was fascinated and wanted to investigate how the average ride length differs between the two rider categories on a daily and monthly basis.   
Lastly, I studied how the type of bike leased differed across the two rider groups.