**Background**

This case study is based on the Google's Data Analyst Capstone Project. This project focuses on a company called Cyclistic. Under this [licensing](https://divvybikes.com/data-license-agreement), Motivate International Inc. made the [public bike-sharing dataset](https://divvy-tripdata.s3.amazonaws.com/index.html) available for this project for the purpose of research and study. To dive deeper in the details check [Google Data Analytics Certification Capstone](http://https/www.coursera.org/learn/google-data-analytics-capstone/supplement/7PGIT/case-study-1-how-does-a-bike-share-navigate-speedy-success) page.

About Cyclistic

In 2016, Cyclistic launched a successful bike-sharing 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 form 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 approval that helped make these decisions possible was the flexibility of its pricing plans: single ride passes, Full and annual memberships. Customers who purchase single-ride or full-day pass are referred to "Casual" riders. Customer who purchase annual memberships are Cyclistic "Members".

Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno Believes that maximizing the number of annual members will be key to future growth. rather Creating a marketing campaign that targets all-new customers, Moreno believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic for their mobility needs.

Goal of the organization

The Director of the marketing has set a goal for the future: Design marketing strategies aimed at converting casual riders into annual members. To do that, however, the marketing analyst team needs to better understand:

1. How annual members and casual riders differ?
2. Why casual riders would buy a membership?
3. How digital media could affect their marketing tactics?

Monero and her team are interested in analyzing the Cyclistic historical bike trip data to identify trend.

Here is a [link](https://docs.google.com/presentation/d/1AgOLsHnojHSiwKQJHFQyQGVwUlWsBnj5HjK5AW3MVIk/edit?usp=sharing) for the PowerPoint that is based on the data and its analysis. The presentation show the story of the project along with the ease of access to the visualizations to understand the insights to be able to make data-driven decisions for the organisation.

**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 earning 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.

**Project procedure**

To meet the goals of the organization, the pathway used in this case study is as follows:

1. Act
2. Prepare
3. Process
4. Analyze
5. Share
6. Act

**Ask**

In this section, it is important to understand the business objective and ask all the important questions to clarify and understand all the vital information linked to the project. Before proceeding to the next step. From the provided documentation it is clear that:

**What is the Problem?**

* How do annual members and casual riders used Cyclistic bikes differently?

**How to solve the problem?**

* By analysing the provided data and understand the trends.

**Who are the stakeholders?**

1. Monero (Director)
2. Cyclistic executive team
3. Cyclistic marketing analytics team

**Where will be data collected?** Internal/Marketing Team

*To perform an analysis of the most recent year for this study is 2022 to study and understand the trends and patterns. The data is provided by Motivation Internation Inc. as a part of the Google Data Analyst Certification.*

**Prepare**

In this stage, All the data is collect from the provided channels, identified, renamed, protected and organised in an manner to make sure the data is unbiased.

**Step 1**

**Data source:**

In the first step, the data will be collected form the public bike-sharing [datasets](http://divvy-tripdata.s3.amazonaws.com/index.html). This data set was provided by Motivate Internation Inc under the [license](http://https/divvybikes.com/data-license-agreement). The data is available on the website name "Index of bucket" and the files on the website are named "divvy-tripdata" ending with year and either numbered as months or quarters. Quarters are named as "Q1Q2"and "Q3Q4".

**Data Collection:**

The second step in this phase is to collect the most recent data so that after the analysis is complete the data can help to set new procedures and marketing campaigns. The most recent data to work with is 2022. The files for the year 2022 are created according to their respective months and these files are available in .zip format. Download the ".zip" file for each month and save it in ".csv" format. Store all the files locally in one place and rename the dataset to organize it appropriately.

**Uploading datasets to Kaggle**

After organizing the datasets. Its time to upload the dataset to Kaggle. Click on "File" select "Upload data". Select all datasets to upload to Kaggle and name to the dataset. All the files will show up in the tab on the right side of the screen.

Check the working directory with "getwd()" to check the existing path.

**Step 2**[**¶**](https://kkb-production.jupyter-proxy.kaggle.net/static/assets/jupyterlab-v4/jupyterlab-index-c05075b3172af98ebfae.html?session=eyJhbGciOiJub25lIiwidHlwIjoiSldUIn0.#Step-2)

**Checking the metrics and measure of the datasets**

This study is identifying the differences between the users’ groups of the cyclistic bike-sharing company. With the provided data the users of the bicycles are categorized in two groups, Annual members and Casual riders. To study the behaviour and trends of both user types there are some of the main categories to help understand the data.

* Bike usage
  1. Average frequency
  2. Average duration
  3. User bike type preference
* Time-varying preference
  1. Busiest day of the week
  2. Busiest month of the year by user type and trip duration
* Location based preference
  1. Most popular/frequent start and end station
  2. Least popular/frequent start and end station

**Process**

In this phase the data gets cleaned, any errors or inaccuracies that could get in the way of the result gets eliminated. One of the important steps in the "process" phase is that when the data is in the cleaning phase, make sure the data is transformed into a more useful format, create more complete information and remove outliers.

**Step 3**

Setup the analysis environment by loading the important packages and libraries to clean the data and check for errors.

**Loading files on Kaggle:**

Once the working environment is set, read files and rename them appropriately. The data below is save as first three letters of the month follow by "\_" connect the month to its numeric value.

**Checking the consistency the data:**

After loading the file, the first operation is to check the column names follow by checking the structure of the dataset. to make sure there are not any errors or mistakes.

All months have same colnames() and str() and the data is ready to be merged together for further anaylsis.

**Merging all the data in one dataset:**

Binding all months in one dataframe as "all\_trips" to work collectively on the whole data group. NOTE: make sure to call all the month "str()" before joining the dataset. If required change the names of the column "Header" to make sure all the elements match with their columns.

In the code lines below, running similar function that was used to check the colnames(), str(), nrow(),dim(), head() and View(). These functions would help to show the dimension, number of rows, top 5 rows of the data.

A easier way to check all these commands is to run "summary()" to check the details of the data. Summary function will show the num count of the each column, that way it would be easier to define all the columns have equal number of elements in them.

**Process**

**Step 4**

**Cleaning the data:**

As you may have notice that "started\_at" column have a a lot of data in it. It would be convenient to separate the date, month, year and day of the week in their independent columns. Separating the data now would come handy while analysing monthly, date, day of the week data.

The dataset have columns "started\_at" and "ended\_at". This data will help to find out the total duration/length of each trip in the dataset.

**Calculating length of rides**:

to calculate the length of the rides. Use data from "started\_at" and "ended\_at" and calculate the time difference with the "difftime" function.

**Removing bad time data:**

There are time when the cyclistic company to run analysis and test the bike for that reason there are data entries that are either "0" or have a negative value. This step will eliminate that data to avoid problems and confusion.

**Analyze**

In this part of the project. calculation are performed to analyze the differences in the trends for the each day of the week & monthly along with the ride length of its respective users.

**Step 6**

**Descriptive analysis:**

In this step, we will dive into details and check the outliers from the data.

The data above is the for the length duration that shows us :

*the output time is processed by dividing the original number with 60*

* Mean - (1160/60) = 19.33 mins
* Median - (611/60) = 10.3 mins
* Min- 0 mins
* Max - (2483235/60) = 28.7 days
* IQR duration - 5.8 mins (1st Qu.) - 18.5 mins(3rd Qu.)

In the Tibble above.

* It is shown that **41%** of the rides were made by casual rider and **59%** by annual members showing that the annual members are the more frequent group.
* As per the data, the total duration of the casual rider is peaked at **61%** whereas, annual members numbers are set at **39%**.
* The data shows that the average duration of casual riders is **2.3** (29 mins) times more than the annual members (13 mins).
* The standard deviation for the causal riders are **9.6** time more than the annual members.

From this data it is clear that the casual riders are leading most of the stats. To take decision for the organisation, it is required to take in depth look at the data.

**Analyzing the monthly data for members and casual:**

In the tibbles above, it shows that the casual riders ride for longer durations. There is a gradual growth in the number of riders from the beginning of the year and it peaks in July. Same trend can be seen for the duration of rides.

The members are generally using bicycles to commute from home to work, school and universities that justifies the duration of their trips.

**Analyzing Weekday data:**

The data from the column "day\_of\_the\_week" is reordered by **Sunday-Saturday** instead of **1-7** for ease of in depth analysis for both type of users.

In the tibble above:

* casual riders ride length is **2x** than the annual members for the whole week.
* **Saturday-Sunday** has the longest ride duration for casual riders.
* **Wednesday** is the least favourite day for causal riders.

**Analyzing data by bike preference:**

There are 3 type of bike in the dataset:

* Electric bike
* Classic bike
* Docked bike

In this section the data is calculated and analyzed by each bike type used by the members and casual riders.

In the tibble above it is clear that:

* Casual riders prefer **electric bikes** more than the classic
* Members prefer **classic bike** for their use as compared to the electric bikes.
* **Docked bikes** are available for the casual rider but **NOT** available for members.

**Summary of the analysis**

1. **Summary data of ride\_length**
   * **19.4 minutes** for an **average** ride duration.
   * **10.3 minutes** for the **median** ride duration.
   * **0 seconds** for the **minimum** ride duration.
   * **28.7 days** for the **maximum** ride duration.
   * **5.8-18.5 minutes** is duration's IQR, meaning 50% of the duration fall into tis range.
2. **Analysis of difference in biking usage between members and casual riders**
   * Annual members made **59%** of the rides in the year 2022. Whereas casual members made **41%**.
   * Total duration for casual riders are **61%** and the members ride duration are **39%**.
   * The average duration of casual riders is **2.3** times more than the annual member riders.
3. **Analysis of bike usage for members and casual riders: by month**
   * **Jan - Apr** the number of rides for casual riders were less than half as compared to the annual members.
   * Towards the end of the year **Oct - dec** the number or rides for casual riders followed the same trend as the beginning of the year.
   * **May-Sept** is the popular time for both type of riders.
   * Summer months are more preferred by casual riders.
4. **Analysis of bike usage for members and casual riders: by week**
   * **Sat-Sun** are more preferred by casual riders as the ride duration is high and frequency of riders is higher as well.
   * **Tues-Wed** least preferred by casual riders. the ride duration tends to lower down as compared to the rest of the week.
   * The number of casual riders is double the count of annual member riders for the whole week.
   * The trend for the whole week is same for the annual members.
5. **Analysis of bike usage for members and casual riders by bike type preference**
   * **Electric Bikes** are preferable by casual riders.
   * **Classic Bikes** are the first choice for annual members.
   * **Docked Bikes** were only available for casual riders.

**Share**

In this phase, all the summary of the analysis is shared with the stakeholders. The data is interpreted in a form of visualizations to help and understand the data in a convenient and effective way.

Here is a [link](https://docs.google.com/presentation/d/1AgOLsHnojHSiwKQJHFQyQGVwUlWsBnj5HjK5AW3MVIk/edit?usp=sharing) for the PowerPoint that is based on the data and its analysis. The presentation show the story of the project along with the ease of access to the visualizations to understand the insights to be able to make data-driven decisions for the organisation.

**Setting the output directory:** The working directory is changed to the "working" to store the output data and visualizations.

**Monthly used of bikes by casual and members:**

* **Jan-Feb** is the has the lowest count of each user type.
* Frequency and duration for each type of user is trending up and it peaks in the month of **July** for **casual riders** and for **members** the peak time is in **August**.
* In the second half of the year. The trend is doing downward for both users but the count of **members** is almost double as compared to **casual riders**.
* From the trends in the graph, it is clear that casual riders longer in summer season and member riders longer in colder months.

**Ride by month**

In the bar graph above, It is clear that the subscription based **annual member** have higher count of rides for each month and reaching the summit in the month of August with **450,000 riders**(approx.). In the beginning and the end the year the number of rides for casual rider is drastically low than their counter parts. However, in the month of July the **casual riders** are at its highest reaching **400,000+ riders**.

**Step 3**

**Analyzing data based on the day of the week for member and casual riders**

Casual Vs. Members: Average weekday based data

As per the bar graph, the casual riders’ average ride duration is almost double than the annual members.

* For both groups weekends have the longest rides duration, **Sat-Sun**.
* **Members** have approximately same ride duration from **Mon-Fri**.
* **Casual rider's** ride duration reduce towards the middle of the week **Wed** and continue to rise towards the end of the week.

The tibble above shows:

* Based on numbers **electric bike**(1.2 million) is the favourite bike for casual riders.
* **Classic bike** is the first preference of annual members (1.7 million) users approx., followed by **electric bikes** at (1.6 million) approx.

**Analysis for members and casual based on: Percentage by ride length**

Member Vs. Casual: Percentage of Rides by User Type

* The data shows that the casual rides have lower number of rides **41** however they have higher count of total ride durations **61**.
* Annual riders have used the bike sharing service **18% more** than the casual riders but the total duration of annual rides is **39%**.

The tibble above shows:

* Based on numbers **electric bike** (1.2 million) is the favourite bike for casual riders.
* **Classic bike** is the first preference of annual members (1.7 million) users approx., followed by **electric bikes** at (1.6 million) approx.

**Analysis for members and casual based on : Percentage by ride length**

Member Vs. Casual: Percentage of Rides by User Type

* The data shows that the casual rides have lower number of rides **41** however they have higher count of total ride durations **61**.
* Annual riders have used the bike sharing service **18% more** than the casual riders but the total duration of annual rides is **39%**.

**Act**

The final stage of the analysis is to derive conclusion and take actions by recommending solution based on the problem statement and results. The goal of the organisation to maximize the annual memberships and increased ride durations to earn profits. The analysis is brought into consideration to make *Data-driven decision*. Here are the key insights from the observation to all the analysis into prospective:

* Based on the number of riders, Annual members have *18%* more users of the bike sharing service as compared to the casual riders.
* Monthly: casual riders prefer warmer weather where annual member ride in cold and warm weather.
* Weekly: Casual riders and annual members like to take longer rides on the weekends *Saturday-Sunday*.
* From the overall analysis it is clear that the number of casual riders are significantly lower than the annual members but the total ride duration of casual riders is approximately 22% higher than annual members

**Business and Analysis team Insights**

Form the summary of the analysis, the marketing team will develop a ad-campaign to convert the casual riders into annual riders and attract new riders to the cyclistic bike share service.

**Recommendation**

The marketing director, Moreno asked the marketing analysis team to analyze the data and to come with strategy to maximize the number of annual subscription members to maximize the profit for the company.

* New promotional offers for casual riders during the weekdays to utilize the service during the week so they can use the service uniformly and reflect a constant trend for all days of the week. Using bike-share service regularly can also convert the casual user to an annual member.
* Offer one week free trial to the casual users with all the perks and offers of the annual member to have a complete experience of the membership. Boosting this promotional offer during the summer can help the company to attract new users and existing casual riders to convert to subscription.
* An rewards/credit plan for the annual members sto refer friends and colleagues to try the service. With every sign up from invitee will earn points, which they can redeem to get a free ride with a time and/or distance limits.
* After the free trial period send a survey to the casual riders to collect data for their experience and what expectations do they have from Cyclistic.
* Creating a mobile application for the users with features like:
  1. Compete with friend and family and social competitions.
  2. Route tracking and previous routes.
  3. Application would allow a secure payment method and pay via mobile application.
  4. Calorie meter in the application to monitor calorie burn from each trip.
* Collaborating with the fitness and workout experts to personalize a bike trip from beginner to expert to help the users use this platform as a workout. This procedure will attract users to create a training plan that can be used to get in shape.
* During the slow hours(*non-busy*) offer discount to casual rides to encourage them to use the service.
* Design a campaign to promote health and environment friendly means of travel to reduce carbon emissions
* Use machine learning to recognise the normal destinations of the users and promote bike-sharing service with estimated travel time details rather than driving. To encourage the users to use the service.
* Creating short campaigns for platforms like Instagram, Tiktok, Facebook, Youtube as people spend immense time on these platforms.

**Conclusion**

The best way to pitch new plans to the organisation is to use the data backed up by research to convince a company and to adapt new strategies to promote a product or service can be difficult and uphill journey because of many business and market uncertainties. This project provides marketing design plans to promote the Cyclistic bike share service to its end existing and potential users.

Thank you,

Gurmeet Singh