

**Case Study: How Does a Bike Share Navigate Speedy Success?**

**Scenario:**

I am a junior data analyst working in the **marketing analyst team** at **Cyclistic**, a bike-share company in **Chicago**. The director of marketing believes the company’s future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand **how casual riders and annual members use Cyclistic bikes differently**. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

**About Cyclistic:**

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geo-tracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with   
disabilities and riders who can’t use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to   
commute to work each day.

Until now, Cyclistic’s marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

**Stakeholders:**

* **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.
* **Cyclistic marketing analytics team:** A team of data analysts who are responsible for collecting, analysing, and reporting data that helps guide Cyclistic marketing strategy.
* **Cyclistic executive team:** A notoriously detail-oriented team which will decide whether to approve the recommended marketing program.

**Business Task:**

Our company has aimed to maximise its number of annual memberships in order achieve the overall goal of increasing annual revenue. Our Marketing Team needs to design a new marketing strategy to convert casual riders into annual members. I shall be helping my stakeholders with the following finding-

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

**Source of Data:**

This case study uses Cyclistic’s historical trip data (previous 12 months) to analyse and identify trends. The data has been made available by Motivate International Inc. under an open [license](https://ride.divvybikes.com/data-license-agreement). The data can be downloaded from [here](https://divvy-tripdata.s3.amazonaws.com/index.html).

This data is **reliable, original, comprehensive, current and cited** as it is internally collected and stored safely by Cyclistic from **Aug 2021 to July 2022**. Personally identifiable information such as credit card numbers has been removed because of data-privacy issues.

The data selected for use covers the **last 12 months from Aug 2021 to July 2022**. Each month has a separate dataset. The datasets are organized in tabular format and have **13** columns. Combined, the datasets have **5901463** rows. The **member\_casual** column will allow me to group, aggregate and compare trends between casual riders and member riders.

**Cleaning and Manipulation of Raw Data:**

All the cleaning and manipulation of data has been done in the ‘Data Cleaning and Transformations.ipynb’ notebook.

* **Tools used:**

To process the data from dirty to clean, I chose to use python. This is because python is relatively fast and thus useful in dealing with huge dataset. Python is also heavily supported by handy open-source libraries such as pandas and matplotlib.

* **Cleaning and transforming the data:**

1. After reading in and combining the 12 datasets into a single data frame, the first step in data cleaning was to identify which columns and rows have missing data. I discovered that 6 out of 13 columns had missing data. Additionally, 1272233 rows had missing values. I have also calculated the percentage and of missing values per column which turned out to be of maximum 15%. Considering this is very small portion of the whole dataset and there is no proper way of data imputation within the given deadline, I have dropped all the records having single or multiple null fields.
2. After dropping null values, I have also checked for any duplicate entries in dataset which, after turned out to be none, I am confident that I am ready to move towards column-by-column analysis and do any cleaning and transformation if required.
3. I have changed the datatype of ‘started\_at’ and ‘ended\_at’ to datetime and sorted the dataset by start time to verify if there is any records outside of our required timeline.
4. ‘ride\_year’, ‘ride\_month’, ‘ride\_day\_of\_week’, ‘ride\_week\_no’, ‘ride\_hour’, ‘weekday\_weekend’, ‘ride\_season’, ’ride\_len\_mins’ are the new columns extracted and verified from start and end time for ease of analysis.
5. I have discovered that there are some negative values as well as outliers in ‘ride\_len\_mins’ column which can hamper correct analysis. Remove all such records from the data frame.
6. After confirming no further cleaning and transformation is required, I moved towards dropping unnecessary columns from the data frame like start and end station details.
7. Finally, our cleaned dataset became ready for analysis and have 4621853 records and 13 columns. I have saved the final dataset in the same path as this file named ‘aug2021\_to\_jul2022\_ready\_for\_analysis.csv’.

**Analysis of Cleaned Data:**

In this step, I have analysed the cleaned data to answer our business task: **How do annual members and casual riders use Cyclistic bikes differently?**

All the analysis has been done in the notebook ‘analysis.ipynb’.

After giving a high level look on our dataset, I can think of mainly two categories of analysis to answer our business task -

* 1. Analysis of No. Of Bike Hires and
  2. Average Ride Duration in Minutes

with respect to different variables like Rider Category, Day of Week, Month, Season, Hour of Day, Rideable/Bike Type etc.

All the findings from our dataset along with visualizations and my recommendations are shared in the form of presentation file [here](https://1drv.ms/p/s!AkcG9O6ijlVFgmIFcCpqWmlg4ALw?e=edpXUX).

**My Top 3 Recommendations:**

Here are my top 3 recommendation based on findings from analysis done on our dataset to convert casual riders into annual members -

* Show casual riders how they can save money on long rides by opting for annual membership.
* Run special ad campaigns on weekends, summer season and afternoon/ evening hours targeting casual riders.
* Show environmental and monetary benefits of using Cyclistic Bike Share annual membership service for work commute to casual riders.

**Thank you!**