**Google Data Analytics Professional Certificate Capstone Project**

**How does a bike-share navigate speedy success?**

In this case study, you will perform data analysis for a fictional bike-share company in order to help them attract more riders. Along the way, you will perform numerous real-world tasks of a junior data analyst by following the steps of the data analysis process: Ask, Prepare, Process, Analyze, Share, and Act. By the time you are done, you will have a portfolio-ready case study to help you demonstrate your knowledge and skills to potential employers!

**Goal:** Stakeholder (director of marketing Lily Moreno) wants to maximize the number of annual memberships for the success of the company’s future. So she wants to design marketing strategies that can convert casual riders into annual members.

**Task:** Understand how casual riders and annual members use Cyclistic bikes differently. From these insights, design a new marketing strategy to convert casual riders into annual members with compelling data insights and professional data visualizations.

Cyclistic is the bike-share program with more than 5,800 bicycles and 600 docking stations. They have reclining bikes, hand tricycles, and cargo bikes. This makes bike-share more inclusive for people with disabilities and riders who can’t use a standard two-wheeled bike.

Most riders opt for traditional bikes (~8% use the assistive options). Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.

**Time Period Location:** Cyclistic launched in 2016, and has been successful, growing to a fleet of 5,824 bicycles geotracked and locked into a network of 692 stations across Chicago.

Bikes can be unlocked from one station and returned to any other station in the system any time.

They have flexible pricing plans: Single-ride passes, full-day passes, and annual memberships.

**Casual riders:** Customers who purchase single-ride or full-day passes

**Cyclistic members:** Customers who purchase annual memberships.

Cyclistic’s finance analysts concluded that annual members are much more profitable than casual riders. The stakeholder believes maximizing the number of annual members is the key to future growth. So there is a chance to convert casual riders into members.

**To meet the goal, I need to understand:**

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

------------------------------------------------------------------------------------------------------------------

**Ask:**

Business Task and Key questions to ask:

The problem I am being asked to solve is to discover what factors, patterns, trends, and insights contribute to the differences between casual riders and Cyclistic annual members, and how these findings can be used to influence marketing strategies that could lead to more casual riders obtaining memberships. The data that will be provided for me comes from a company that started in 2016 and has seen significant success since its opening in Chicago. The analysis will need to consider the flexible pricing plans (from single-ride to full-day to annual memberships), the types of bikes people ride (traditional, assistive, accessible bikes for people with disabilities, etc.), and how casual riders differ from cyclistic members in a variety of different scenarios.

From all of these findings, I will be able to deduce why casual riders would buy a Cyclistic membership based on various patterns and trends within the data, and be able to determine how digital media could influence casual riders to become members. This will allow the stakeholder (the director of marketing Lily Moreno) to come up with marketing strategies (i.e. TV commercials, radio advertisements, billboards, signs, newspaper advertisements, etc.) that can lead to an increase in the number of Cyclistic memberships. Knowing what most annual members go for and what most casual riders go for would be crucial for determining what factors would influence a decision to get an annual membership.

Key questions using the SMART methodology include:

1. What times of year do most people tend to get their Cyclistic memberships?
2. Are there certain types of bikes that casual riders and Cyclistic members prefer and why?
3. What kind of bikes and pricing plans would lead to an increase or decrease in casual riders and Cyclistic memberships?
4. How do accessible bikes impact the number of annual memberships from 2016 to the present? How do they affect the number of casual riders?

The deliverables will be a report and data visualizations that highlight key insights, trends, and patterns between casual riders and Cyclistic members that can help reveal what might contribute to a casual rider getting a membership, and what key differences are there between casual riders and annual members. **One important thing to note is that only the data from the past 12 months will be analyzed (December 2020 to November 2021)**

------------------------------------------------------------------------------------------------------------------

**Prepare:**

Description of data sources used:

The data is contained in 12 .csv files each representing bike ride data for the Cyclistic company from December 2020 to November 2021. Thus it represents a year’s worth of data. Each row in each of the CSV files represents one ride for a given rider. It tells what type of bike it was (electric, classic or docked bike), the start date/time and end date/time of the ride, the start and end stations (if one was identified), if the rider was a casual rider or a Cyclistic annual member, and the latitude and longitude of the start and end locations of the ride. All 12 CSV files are structured this way. There are quite a large number of rides in each dataset with blank start and end stations identified. There are also some missing end latitude and longitudes within the data. This may cause problems with determining the popular locations where riders start and end their ride. Luckily every ride appears to have a start and end date/time, and identifies what type of bike they rode and if they were a casual rider or a Cyclistic annual member.

Furthermore, the ride ID column makes it easy to determine the number of rides there were for a given month. One thing that will need to be determined is if there are identical ride IDs among the CSV files, so joining the tables together would be a bit tricky. Since the columns in each CSV file are identical and are in the same order, I foresee SQL Unions being used to combine all 12 CSV files into one table. I can create additional columns indicating the Month and Year of a given ride. This way each ride can be categorized by the Month/Year it started on. This helps separate each ride out for each of the 12 months that need to be analyzed. Additional columns can be created during the data cleaning and data analysis process that will make it easier to identify trends and patterns within the dataset. Overall, the data is fairly credible, but the missing data on start and end stations for a given ride, as well as the missing data on some end latitudes and longitude might cause some inaccurate analysis on where the most popular start and end station locations are for particular riders.

**Due to the large number of rows in each CSV file, and the fact they will all need to be combined together, SQL will be used to analyze it. Tableau will be used to create the data visualizations.**

------------------------------------------------------------------------------------------------------------------

**Process:**

See the Google Doc Bike\_Share\_Data\_Cleaning\_Changelog

------------------------------------------------------------------------------------------------------------------

**Analyze**

See the Bike\_Share\_Portfolio\_Project\_Queries.sql file for all analysis and calculations

------------------------------------------------------------------------------------------------------------------

**Share**

Created 2 Tableau Dashboards. Here are the links:

* <https://public.tableau.com/app/profile/william.melahouris/viz/BikeSharePortfolioProjectDashboard1/Dashboard1>
* <https://public.tableau.com/app/profile/william.melahouris/viz/BikeSharePortfolioProjectDashboard2/Dashboard1?publish=yes>

------------------------------------------------------------------------------------------------------------------

**Act**

Here are my conclusions as well as my top 3 recommendations for marketing based on my analysis of the data:

Based on my analysis of the Cyclistic Bike Share ride data from December 2020 through November 2021, I have made the following conclusions. Both casual and membership riders greatly prefer classic bikes, followed by electric, and hardly anyone uses docked bikes (though casual riders preferred them more than membership riders). ELectric bikes tend to be more popular for both casual and membership riders from May through October. The most popular start stations for the casual riders include Millenium Park, Michigan Ave & Oak St, and Shedd Aquarium while the most popular end stations for casual riders include Streeter Dr & Grand Ave, Millennium Park, and Michigan Ave & Oak St. Casual riders tend to ride much more on the weekend (Friday, Saturday, and Sunday) compared to during the week (Monday, Tuesday, Wednesday, and Thursday) which is when membership riders tend to ride more. This is most likely due to work schedules, and the fact that a good amount of Cyclistic membership riders (about 30%) use bikes to commute to work each day. Both casual and membership riders tend to go on bike rides much more in the warmer months (May, June, July, August, September) compared to the colder months (December, January, February). While both casual and membership riders tend to ride for longer periods of time on the weekends compared to the weekdays, casual riders tend to go for much longer bike rides on average each day compared to membership riders.

To increase the number of casual riders obtaining memberships at Cyclistic, there are 3 key recommendations I have made for marketing to consider. A bigger push in advertisement for classic bikes throughout the entire year, as well as electric bikes in the warmer months (May through September) will help ensure that riders get the bikes they prefer. Docked bikes hardly get used, though they seem to be used more in December. Promoting docked bikes more or offering a special discount or upgrading the docked bike models would help increase interest in that bike. Special membership deals should be offered more on the weekends for casual riders as opposed to the weekdays, especially in the summertime when there is a vast increase in the number of casual riders. Casual riders mostly ride on the weekends. Promotions, deals, advertisements, discounts, and other marketing strategies to increase membership should be focused at bike stations that riders frequently start and end at. The most popular start stations include Millenium Park, Michigan Ave & Oak St, and Shedd Aquarium. The most popular end stations for casual riders include Streeter Dr & Grand Ave, Millennium Park, and Michigan Ave & Oak St.