**Google data analytics capstone project**

**Diagram, icon

Description automatically generated with medium confidence**

Andrés Peña

**Preface**

Hello everyone! The following work refers to the optional capstone project included in the Google data analytics professional certificate. It is a great opportunity for students to apply recently acquired knowledge and to showcase in a more streamlined manner the abilities and style of work a potential prospect might have. I hereby claim full rights and responsibility for the analysis and insights coming from the project.

There is a specific set of steps taught in the course that help the analyst achieve more standardized work. They are: ask, prepare, process, analyze, share and act. They will be covered in detail as the analysis progresses.

For context, the information and data gathered to perform the analysis is purely fictional. It is also legally made available by Motivate International Inc.

**Index**

[**Context** 4](#_Toc95993412)

[**Ask** 6](#_Toc95993413)

[**Prepare** 7](#_Toc95993414)

[**Process** 8](#_Toc95993415)

[**Analyze** 9](#_Toc95993416)

[**Share** 13](#_Toc95993417)

[**Act** 14](#_Toc95993418)

# **Context**

**Scenario**

You are 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.

**Characters and teams**

* Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations. 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.
* 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 learning 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.

**About the company**

In 2016, Cyclistic launched a successful bike-share 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 from 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 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.

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 than 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 program and have chosen Cyclistic for their mobility needs.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends. (this entire section was taken from the “Case\_guidelines.pdf” located in the original “CASE STUDY” folder)

# **Ask**

The first step of the Google data analytics analysis workflow is the “ask” stage. Here the analyst builds the base of the analysis by asking the right questions, avoiding bias and vague guidelines. This is accomplished by following the SMART structure. Questions need to be specific, measurable, attainable, relevant and timely.

In this case we have the following questions provided by the stakeholders, they will guide the analysis and insights.

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

**Understanding the business task**

As the information given in the context indicates, annual members are considered the most profitable, so more incentives need to be established for casual riders to migrate to an annual plan of services, but first, key differences between casual riders and annual members must be understood in order to provide the best solutions and services to the customers. Data will be cleaned to make its contents more reliable. Also, simple data visualizations and graphs will be developed so the process of insight sharing can be more streamlined.

# **Prepare**

In the prepare stage of the process the focus will mainly be on understanding the data, how reliable is it? What format is it stored in? How big is it? Are there any problems with the data?

In this scenario the data is in a .csv format for each month containing key information describing the rides made by users. To start working with the yearly data, monthly compilations must be merged into a single data frame to simplify the analysis; the consolidated yearly data is to big for an excel document (5.000.000 + observations), so other tools must be used. In this case, the method of choice will be data frames in R programming language, developed in R studio.

To begin the process of preparing the data, columns are checked for duplicates. The column named “ride\_id” is the primary key of the data frame so it is evaluated first and found it has no duplicates. After it is known that there are no repeated rides in the data frame, next step is to check for “NULL” values, or “NA”. Given the fact that a ride can start anywhere, not necessarily at a docking station, “NA” values referring to stations will still be considered in the analysis. Then, other functions are used to understand the data frame. Among them, these can be highlighted:

* Str(): using to analyze the general structure of the data frame. We can obtain information such as the data type of each column. The mean for numeric columns and much more relevant information.
* Head(): using head we can get a small preview of the contents of the columns. This for comprehension purposes.
* Glimpse(): If glimpse is used we would get a small preview very similar to head().

# **Process**

In this stage, graphs and visualizations were made to facilitate the analyze stage. For the process section refer to the Rmarkdown file located in the CASE STUDY folder, there you will be able to review the R code and the output graphs.

# **Analyze**

In the analyze section we can see the graphs and get insights from them in order to make it easier for the marketing department to take decisions. First, a bar chart was created in order to better understand the proportion of casual rides vs annual member rides. Here are the results:

Chart, treemap chart

Description automatically generated

Out of the graph we can easily tell that annual members made more rides in the year of 2021. Based on this percentage, the marketing department can set future goals for their marketing plan. Given that their objective is to convert casual riders to annual members, it would be reasonable to aim for a 60% 40% or 65% 35% goal for the next year.

After this graph, I decided to go deeper into analyzing the behaviour of casual and annual riders. With that in mind, I developed a histogram that shows where the concentration lies in regard to the time spent in each ride vs the average time for a ride of each group.

For casual rides, we have the following:

Chart, histogram

Description automatically generated

As seen in the graph, the highest concentration of time for the rides happens between 0 and 20 minutes, and the distribution starts diminishing as time increases. It is also very important to highlight the mean of the group being 25.64 minutes per ride because that is going to function as the comparing factor between casual and annual rides. For the annual member rides we see the following:

Chart, histogram

Description automatically generated

In the annual histogram there is a higher concentration in the lower time limits, and we can also see a steeper decline towards the higher time marks. The mean for the time spent in each ride is 13.19 minutes which is almost half of the mean for casual member ride time. Based on that, we can assume that casual riders tend to stay on longer rides, because tourists fall in that group. Making longer and longer trips to get to know the rest of the city.

After those comparisons are made, we proceed to compare the biketype usage, just to check if there are any wild discrepancies to guide us better through the analyze phase. This is the outcome:

Chart, bar chart

Description automatically generated

Nothing much to conclude by looking at the graph. The difference in bike usage is proportional to the difference of total rides. One very important thing is that annual members choose not to use the docked bike type, maybe that is a benefit of the annual membership.

And lastly I decided to develop a Chicago heatmap, where we could analyze what parts of the city are the most frequent for casual riders. I decided to leave out the annual members because the goal of the marketing team is going to be to target the casual members and convert them into annual members, so aiming the publicity for:

Map

Description automatically generated

In the image above, the red color indicates a high level of density for the area. In other words the point that is marked in red indicates the most concurred spot that casual riders use to start a ride. So in this example our focus as an organization should be to advertise to casual riders specifically on the navy pier station.

# **Share**

In the share stage is where a format to share the findings is established. In particular I like to work with R markdown documents, because not only you can treat it as a report with the most relevant info but you can also make it a lot easier for other team members to reproduce your analysis independently. To take a look at the R markdown format, head over to case study folder and look for the Rmarkdown file.

# **Act**

Now for the final stage, act. Here is where based on all the data analysis and the graphs that were created I provide a set of steps to take, and the important thing is that these steps need to have data to back them up. In this scenario, we have the following:

* Create an advertising plan designed to target the casual riders highlighting all the benefits that come from an annual membership. Let them know that even if they are tourists sightseeing in Chicago, they can still reap the benefits of an annual membership, specifically for those prolonged rides that we see from casual riders. Also, make them understand that the annual plan can be cancelled at any time, so there is no one year commitment that would act as an entry barrier for those that are not staying around that long.

The marketing campaign should be addressed at the station located in the navy pier as shown in the heat map, this is the one that had the most activity throughout the 2021 year. And lastly, set a goal for the entire marketing campaign for 2022 aiming towards a new proportion of casual vs. annual rides of 40% casual vs 60% annual rides.