**Google Data Analytics Capstone Project**

**Table of Contents:**

* Introduction
* Ask Phase
* Prepare Phase
* Process Phase
* Analyze Phase
* Share Phase
* Act Phase

**Introduction**

**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.

**Ask Phase**

**Guiding Questions:**

* What is the problem you are trying to solve?

The ultimate goal is to find a proper marketing strategy to turn casual riders into annual riders

* How can your insights drive business decisions?

The insights will be helpful in turning more casual riders into annual riders.

**Key Tasks**

1. Identify business task.
2. Consider key stakeholders.

**Deliverable**

* A clear statement of business task.

**Prepare Phase**

**Guiding Questions:**

* Where is your data located?

I have downloaded data for the year 2022 from the link <https://divvy-tripdata.s3.amazonaws.com/index.html> and store it locally in my laptop.

* How is the data organised?

Data is organised by month in different csv files

* Are there issue with bias or credibility in the data? Does your data ROCCC?

There is no issue with bias and credibility as it is downloaded from reliable source.

As well as data is ROCCC as it is reliable, original, comprehensive, current and cited.

* How are you addressing licensing, security, privacy and accessibility?

Data is not having any personal information like name, email, phone number so it is private and secure so there is no issue with who access the data and as it is from reliable source so it is licensed.

* How did you verify the data integrity?

All the csv file have consistent columns with same type of data.

* How does it help you to answer your question?

The data in all files will help to get some insights regarding casual riders and annual members.

* Are there any problems with data?

Data should be stored in more correct format for station Id so it would be more helpful.

**Key Tasks**

1. Download data and store appropriately.
2. Identify how it’s organised.
3. Sort and filter the data
4. Determine the credibility of the data.

**Deliverable**

* A description of all the data sources used.

**Process Phase**

**Guiding Questions:**

* What tools you are using and why?

I am using R and Excel as to import data in SQL will be more difficult task and in Excel and R both I can directly do some visualizations.

* Have you ensured your data’s integrity?

Yes I have ensured the data’s integrity.

* What steps you have taken to make sure that your data is clean?

To make sure the data is clean I removed all the duplicated rows.

* How can you verify your data is clean and ready to analyse?

By running the code of removing duplicates in R again and if it removes more row it means data was not clean but if it removes no rows then it is clean as well as I removed all the rows which had null values or no values in it’s columns.

* Have you documented your cleaning process so you can review and share those results?

Yes, this is the document.

**Key Tasks**

1. Check the data for errors.
2. Choose your tools.
3. Transform the data so you can work with it effectively.
4. Document the cleaning process.

**Deliverable:**

* Documentation of cleaning and manipulation of data.

**Analyze Phase**

**Guiding Questions:**

* How should you organize your data to perform analysis on it?

I have combined all the CSV files in one dataframe using R to perform analysis on it.

* Has your data been properly formatted?

Yes, I cleaned the data by removing the duplicates and null values rows furthermore did some cleaning in excel to make sure there is no inconsistent data.

* What surprises did you discover in the data?

Surprises I discovered in the data is that it looks like there are more members than casual but it is used more by casual

* What trends and relationships did you find in the data?

Trends and relationship between data I find is on weekdays members use this program more and casual uses more on weekends

* How will these insights help you answer business questions

These insights give clear visualization so it would be easy to plan marketing strategy

**Key Tasks**

1. Aggregate your data so its useful and accessible.
2. Organize and format your data.
3. Perform Calculations.
4. Identify trends and relationships.

**Deliverable**

* A summary of your analysis.

**Share Phase**

**Guiding Questions:**

* Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?

Yes.

* What story does your data tell?

Data says that there should be a strategy that benefits members and casual riders in its own way and due to which we can increase the number of members as well as casual riders.

* How do your findings relate to your original question?

It relates as I have compared members and casual to get more insights for the questions.

* Who is your audience? What is the best way to communicate with them?

Audience for me is the marketing team as I am a data analyst helping the marketing team to strategize the way to get more people using this program. The best way to communicate is using presentations.

* Can data visualization help you share your findings?

yes

* Is your presentation accessible to your audience?

Yes, I will make it accessible to the audience once I build it.

**Key Tasks**

1. Determine the best way to share your findings.
2. Create effective data visualizations.
3. Present your findings.
4. Ensure your work is accessible.

**Deliverable**

* Supporting visualizations and key findings

**Act Phase**

**Guiding Questions:**

* What is your final conclusion based on your analysis?

Final Conclusion is both members and casual riders are playing an important role but we need strategies which are beneficial for both.

* How could your team and business apply your insights?

It would be useful to develop a good marketing strategy to attract more members and causal riders.

* What next steps would you or your stakeholders take based on your findings?

To find good marketing strategy and how to implement it as soon as possible and check its impact.

* Is there additional data you could use to expand on your findings?

It would be more beneficial if data was included with distance covered and weather which could led to more better insights

**Key Tasks:**

1. Create your portfolio.
2. Add your case study.
3. Practice presenting your case study to a friend or family member.

**Deliverable:**

* Your top three recommendations based on your analysis

Recommendations :

1. From the data I believe that members are using this program for riding it to their work or may be from one station to another to catch bus or train as their average ride length is less compare to casual and members use it more on weekdays so for that I believe there should be offer that one ride will be free for them one day every week.
2. Another insight is casual riders uses more on weekend mostly for the purpose of enjoying their free time and they travel more than members so for them we should offer if that on weekends if they travel certain miles they will be getting discount in membership cost as well as one free ride after certain miles target is fulfilled.
3. One last recommendation is may be causal riders are those who are using it for exercise purpose then by announcing cycling race to promote the healthy lifestyle as well as our program and first 5 will get special membership discount.