DIVVY a bike-share workflow

| Title | DIVVY case study |
| --- | --- |
| Industry focus | Marketing |
| Problem statement | In what way do annual members and casual riders differ in using DIVVY bikes and what we can do to convert casual riders into annual members? |
| Business use case (what are you solving for) | * How do annual members and casual riders use bikes differently? * Why would casual riders buy divvy annual memberships? * How digital media could affect their marketing tactics |
| Deliverables | A report containing the following:   * A clear statement of the business task * A description of all data sources used * Documentation of any cleaning or manipulation of data * A summary of my analysis * Supporting visualizations and key findings * My top three recommendations. |
| Dataset list | Dataset is available for download [here](https://divvy-tripdata.s3.amazonaws.com/index.html). |
| Dataset License | Lyft Bikes and Scooters, LLC (“Bikeshare”) operates the City of Chicago’s (“City”) Divvy bicycle sharing service. Bikeshare and the City are committed to supporting bicycling as an alternative transportation option. As part of that commitment, the City permits Bikeshare to make certain Divvy system data owned by the City (“Data”) available to the public, subject to the terms and conditions of this [License Agreement](https://ride.divvybikes.com/data-license-agreement). |

# Ask

## Guiding questions

* What is the problem you are trying to solve?

How to **convert** **casual riders** to **annual members**.

* How can your insights drive business decisions?

I am hoping that my insights into the three questions discussed earlier, will aid the marketing team to understand casual riders and annual members better

## Key tasks

* Identify the business task
  1. Understand the difference between casual riders and annual members
  2. Why casual riders might buy annual memberships
  3. How to use digital marketing to nudge more casual riders to buy annual memberships
* Consider key stakeholders

Marketing team, Executive team.

## Deliverable

To understand how digital marketing can increase the conversion rate from casual riders to annual members by understanding why a casual rider might want to convert to annual membership and also to understand the difference between how casual riders and annual members use the bicycles.

# Prepare

## Guided questions

* Where is your data located?   
  [Click here.](https://divvy-tripdata.s3.amazonaws.com/index.html)
* How is the data organized?
  1. For the period of 2013 to the 1st quarter of 2020, the data is organized **quarter-wise**.
  2. For the period of the 2nd quarter of 2020 to July 2022, the data is organized **month-wise**.
* Are there issues with bias or credibility in this data? Does your data ROCCC? 
  1. The data source is **original** and **reliable** because it is made available by a **first-party source**.
  2. There is **no evidence of bias**.
  3. The data is **comprehensive** enough for our use case.
  4. At the time of this analysis, the data is **current** as the most recent data is **less than 3 months old**.
  5. The sources **are cited** in the [licensing agreement](https://www.divvybikes.com/data-license-agreement).
* How are you addressing licensing, privacy, security, and accessibility?

This [link](https://www.divvybikes.com/data-license-agreement) refers to the licensing terms we have agreed upon.

* How did you verify the data’s integrity?   
  There are **no corrupt** files.
* How does it help you answer your question?   
  Since the data source follows ROCCC and its integrity has been established, it can guarantee an accurate data analysis.
* Are there any problems with the data?

None.

## Key tasks

* Download data and store it appropriately. *(Done)*
* Identify how it’s organized. *(Done)*
* Sort and filter the data. *(Done)*
* Determine the credibility of the data. *(Done)*

## Deliverable

A description of all data sources used. *(Done)*

# Process

## Guided questions

* What tools are you choosing and why?

R Studio.

* Have you ensured your data’s integrity?

Data integrity has been maintained during the cleaning phase.

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

I have checked each column and cleaned the data. The steps are listed in the Report.rmd file

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

Yes.

## Key tasks

* Check the data for errors. *(Done)*
* Choose your tools. *(Done)*
* Transform the data so you can work with it effectively. *(Done)*
* Document the cleaning process. *(Done)*

## Deliverable

Documentation of any cleaning or manipulation of data. *(Done)*

Analyze

## Guiding questions

* How should you organize your data to perform analysis on it?
  + For Ride Duration analysis, the dataset has been organized into two datasets one for casual riders and the other for member riders.
  + A subset dataset was created only containing data pertaining to member types and ride duration.
  + For Binding membership type to location, the data has been organized into a view.
    - Start\_location\_membership
  + Splitting casual and annual members data in different data sets for comparison purposes.
* Has your data been properly formatted? *(Yes)*
* What calculations need to be performed on the data?
  + Ride duration needed to be converted from seconds to minutes for both member types.
  + Descriptive statistics were performed thereby generating the following observations
    - Minimum
    - Maximum
    - 1st Interquartile Range
    - Median
    - Mean
    - 3rd Interquartile Range
* What surprises did you discover in the data?
  + Median ride duration differs significantly between the two types of riders: 15 mins for casual riders to 9.4 mins for member riders.
  + The casual members have a more variation in ride duration where as the members have a more predictable ride duration of around 9.4 mins.
  + While the number of member rides that have a ride duration closer to the median are two times more when compared with the casual riders.
* What trends or relationships did you find in the data?
  + One notable trend that came to light was that the members prefer to use the classic bike more than the e bike. It is the same case with the casual riders too but the preference is not as pronounced as the member riders.
  + We were able to generate a view to depict the station names where the highest number of casual members start their journey. This might provide insights into where to focus the specific marketing activities such as billboards and ad campaigns.

## Key tasks

* Aggregate your data so it’s useful and accessible. *(Done)*
* Organize and format your data. *(Done)*
* Perform calculations. *(Done)*
* Identify trends and relationships. *(Done)*

## Deliverable

A summary of your analysis

A PowerPoint presentation has been created depicting the summary of the analysis to be presented to the stakeholders.

# Share

## Guiding questions

* Were you able to answer the question of how annual members and casual riders use Divvy bikes differently? *(Yes)*
* Can data visualization help you share your findings?
  + The visualizations that were produced using R will be used for the time being.
* Is your presentation accessible to your audience?
  + The link to the report will be published.

## Key tasks

* Determine the best way to share your findings.  *(Done)*
* Create effective data visualizations. *(Done)*
* Present your findings. *(Done)*
* Ensure your work is accessible. *(Done)*

## Deliverable

Supporting visualizations and key findings

# Act

## Guiding questions

* What is your conclusion based on your analysis?
  + The conclusion has been detailed in the .ppt presentation
* Is there additional data you could use to expand on your findings?
  + For obvious reasons, Payment or revenue related data have been kept out of the freely available dataset. With access to those data one can arrive at more significant findings that could have a bigger impact on the business.

## Deliverable

Your top three recommendations based on your analysis *(Done)*