

Data Analytics Case Study: How Does a Bike-Share Navigate Speedy Success?



Title	Maximizing the number of annual memberships by converting the casual riders to annual members.
Client / Sponsor	<u>Cyclistic</u> – Bike share company in Chicago
Key Stakeholders	<ul style="list-style-type: none">▪ Lily Moreno – Director of Marketing▪ Cyclistic Marketing Analytics Team – Responsible for creating marketing strategies▪ Cyclistic Executive Team – Responsible for approving the recommended marketing program.
Industry Focus	Marketing
Problem Statement	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.
Business Use Case	1. How do annual members and casual riders use Cyclistic bikes differently?
Goals / Metrics	Top three key differences in the usage patterns of casual and member riders.
Deliverables	<ol style="list-style-type: none">1. A clear statement of the business task2. A description of all data sources used3. Documentation of any cleaning or manipulation of data4. A summary of your analysis5. Supporting visualizations and key findings6. Your top three recommendations based on your analysis
Are Dataset Available?	Yes
Dataset List	The previous 12 months of Cyclistic trip data

Prepare

Key Points:

- Data was shared through Amazon cloud storage link.
- Data was organized in 32 zipped files, containing:
 - Trip Data from 2013 to May, 2021 in .csv file format.
 - Station Data from 2013 to 2017 in .csv file format.
- Data is split differently for different years; quarterly/bi-yearly or monthly.
- Data is made available by Motivate International Inc. under license: <https://www.divvybikes.com/data-license-agreement>
- Data should be able to provide different types of riders' behaviors and usage patterns. And this in turn will help us to solve the business use case in discussion.
- Data file problems:
 - File – type for “Divvy_Trips_2019_Q1” and “Divvy_Trips_2019_Q2” missing.
Solved - Data type is .csv
 - File “Divvy_Stations_2014-Q1Q2” was in .xlsx file format.
Solved – File was opened in spreadsheets and exported to .csv format to keep file type consistency.

Conclusion:

After the prepare stage, we have 2 master spreadsheets contained compiled information:

- combined_station_data.csv
- combined_trips_data.csv

Process

Key Steps:

- In this step, we will be processing our data before analysis and visualization.
- We will be completing the ETL process in Python using Pandas library.
- Cleaning and Manipulation steps have been documented in the Python notebook.

Analyze

We did preliminary descriptive analysis in SQL, and got a few initial insights in the difference in behavior for casual and member riders.

Key Finding:

```
SELECT usertype,
       ROUND(AVG(distance/tripduration)*3600,2) AS `Average_Speed_kmph`,
       ROUND(AVG(tripduration)/60,2) AS `Average_Speed_of_Travel_mins`
FROM `bigquery-sandbox-311309.casestudy1.tripdata`
GROUP BY usertype
LIMIT 10
```

usertype	Average_Speed_kmph	Average_Speed_of_Travel_mins
casual	9.93	24.73
member	12.18	15.99

```

SELECT * FROM
(
  SELECT
    usertype,
    weekday,
    int64_field_0
  FROM `bigquery-sandbox-311309.casestudy1.tripdata`
)
PIVOT
(
  COUNT(int64_field_0) AS No_Of_Trips
  FOR weekday in (
    0 AS Monday,
    1 AS Tuesday,
    2 AS Wednesday,
    3 AS Thursday,
    4 AS Friday,
    5 AS Saturday,
    6 AS Sunday)
)

```

usert ype	No_Of_Trips_ Monday	No_Of_Trips_ Tuesday	No_Of_Trips_ Wednesday	No_Of_Trips_ Thursday	No_Of_Trips_ _Friday	No_Of_Trips_ Saturday	No_Of_Trips_ _Sunday
casual	10238	9468	10034	10457	13207	17360	14010
memb er	14595	14727	15182	15250	16488	16883	13895

Though the SQL queries provided us initial trends and relationships, it is difficult to understand the same in the current table format. Thus we will represent the same in visualization in the next step of data analytics.

Share

The results of above analysis findings and other insights were share with the marketing team using Tableau – public.

<https://public.tableau.com/app/profile/utsav.shrivastava/viz/GoogleAnalytics-Cyclistic/Dashboard>

Key Features:

- The Viz compares the riding patterns between casual and member riders.
- The casual riders are identified in yellow and member riders in green.
- The Dashboard is divided in 3 parts with gives us insights in 3 main categories:
 1. Map of Chicago city showing all the trips that the riders undertook in 2020-2021. The Map can be filtered through the day of the week, to compare the weekly trends.
 2. The two dial charts are for trip start and end time in a day. It provide us the insights on daily trend in the trips for the two categories of the riders. Again, the charts can be filtered by the day of the week.
 3. Finally, the third part include a series of bubble and bar charts to provide descriptive analysis results, like no. of total trips, speeds, etc for the two types of riders.

Act

Analyzing and visualizing the riders' data, gives us three main insights:

1. Casual Members use the bikes mostly near the Chicago City Centre and may indicate being used by mostly tourists in the city.
Annual members, however, use the bikes both in and around the City Centre.
2. Casual rides prefer starting their ride around late mornings and ending them around evening.
Annual members generally start their ride early morning and end at similar time to casual riders. Casual riders also tend to ride longer distances and at slower speeds, than the annual members.
3. The number of casual riders increases towards the end of the week, from Friday to Sunday.
Number of rides by Annual members seems to be quite uniform, with a slight dip in Sunday.

