**A logo for a bicycle sharing company

Description automatically generatedCase study – Cyclistic bike-share analysis ( Google Data Analytics Capstone project )**

**Cyclistic – SQL server/Excel ( Data cleaning, Data wrangling ) Tableau (EDA, Visualization)**

**Dataset Source** - <https://divvy-tripdata.s3.amazonaws.com/index.html>

**License** - The data has been made available by Motivate International Inc. under [this](https://www.divvybikes.com/data-license-agreement)

license.

**About the company:**

**A map of a city

Description automatically generated**

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 2 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.

**Business Task:**

1. How do annual members and casual riders use Cyclistic bikes differently?

2. Why would casual riders buy Cyclistic annual memberships?

3. How can Cyclistic use digital media to influence casual riders to become members?

**SQL SERVER/ Microsoft Excel:**

* Import data ( importing 12 csv. files ) each representing 1 month of data gathered throughout 2022
* Combining data from all of the months into 1 dataset

USE citi

GO

INSERT INTO T1

SELECT \*

FROM T2;

#Continue merging data for each month until all 12 are combined;

#Confirming merging process was successful:

SELECT COUNT (\*) AS ALL\_RECORDS

FROM T1;

# Result: 5,667,717 results

* Cleaning/ Wrangling Data

-- Create a view named 'distance'

CREATE VIEW distance AS

-- Start with a Common Table Expression (CTE) named 'distance'

WITH clean\_data AS

(

-- Select relevant columns and perform data cleaning

SELECT

start\_lat,

start\_lng,

end\_lat,

end\_lng,

rideable\_type,

member\_casual,

-- Clean leading and trailing spaces from start\_station\_name

LTRIM(RTRIM(start\_station\_name)) AS start\_station\_name,

-- Clean leading and trailing spaces from end\_station\_name

LTRIM(RTRIM(end\_station\_name)) AS end\_station\_name,

-- Convert started\_at to DATE and name it start\_date

CAST(started\_at AS DATE) AS start\_date,

-- Convert ended\_at to DATE and name it end\_date

CAST(ended\_at AS DATE) AS end\_date,

-- Extract hour from started\_at and name it start\_hour

DATEPART(HOUR, started\_at) AS start\_hour,

-- Extract hour from ended\_at and name it end\_hour

DATEPART(HOUR, ended\_at) AS end\_hour,

-- Extract weekday name from started\_at and name it weekday

DATENAME(WEEKDAY, started\_at) AS weekday,

-- Extract month name from started\_at and name it month

DATENAME(MONTH, started\_at) AS month,

-- Calculate duration in minutes

DATEDIFF(MINUTE, started\_at, ended\_at) AS duration,

-- Calculate distance using geography type

geography::Point(start\_lat, start\_lng, 4326).STDistance(geography::Point(end\_lat, end\_lng, 4326)) / 1000.0 AS distance

-- Specify the source table (T1) and apply filtering conditions

FROM T1

WHERE

-- Exclude rows where start or end coordinates are null

start\_lat IS NOT NULL

AND start\_lng IS NOT NULL

AND end\_lat IS NOT NULL

AND end\_lng IS NOT NULL

)

-- Select all columns from the clean\_data CTE and filter by duration

SELECT \* FROM clean\_data WHERE duration > 0;

* Data Analysis Queries
  + - Summary Statistics:

-- Basic statistics for numerical columns

SELECT

COUNT(\*) AS total\_records,

AVG(duration) AS avg\_duration,

MIN(duration) AS min\_duration,

MAX(duration) AS max\_duration,

AVG(distance) AS avg\_distance,

MIN(distance) AS min\_distance,

MAX(distance) AS max\_distance

FROM distance;

# Result:



# max\_duration and max\_distance seem to be rather high and could be treated as outliers.

* + - Summary Statistics by Member Type:

-- Summary statistics by member type

SELECT

member\_casual,

COUNT(\*) AS total\_records,

AVG(duration) AS avg\_duration,

MIN(duration) AS min\_duration,

MAX(duration) AS max\_duration,

AVG(distance) AS avg\_distance,

MIN(distance) AS min\_distance,

MAX(distance) AS max\_distance

FROM distance

GROUP BY member\_casual;

# Result:



* + - Monthly Trends:

-- Monthly ride counts and average distance

SELECT

month,

COUNT(\*) AS ride\_count,

AVG(distance) AS avg\_distance

FROM distance

GROUP BY month

ORDER BY month;

# Result:



#

* + - Monthly Trends by Member Type:

-- Monthly ride counts and average distance by member type

SELECT

member\_casual,

month,

COUNT(\*) AS ride\_count,

AVG(distance) AS avg\_distance

FROM distance

GROUP BY member\_casual, month

ORDER BY member\_casual, month;

# Result:



* + - Hourly Patterns:

-- Hourly ride counts and average distance

SELECT

start\_hour,

COUNT(\*) AS ride\_count,

AVG(distance) AS avg\_distance

FROM distance

GROUP BY start\_hour

ORDER BY start\_hour;

# Result:



#

* + - Hourly Patterns by Member Type:

-- Hourly ride counts and average distance by member type

SELECT

member\_casual,

start\_hour,

COUNT(\*) AS ride\_count,

AVG(distance) AS avg\_distance

FROM distance

GROUP BY member\_casual, start\_hour

ORDER BY member\_casual, start\_hour;

# Result:



* + - Weekday Patterns:

-- Weekday ride counts and average distance

SELECT

weekday,

COUNT(\*) AS ride\_count,

AVG(distance) AS avg\_distance

FROM distance

GROUP BY weekday

ORDER BY CASE WHEN weekday = 'Sunday' THEN 1

WHEN weekday = 'Monday' THEN 2

WHEN weekday = 'Tuesday' THEN 3

WHEN weekday = 'Wednesday' THEN 4

WHEN weekday = 'Thursday' THEN 5

WHEN weekday = 'Friday' THEN 6

WHEN weekday = 'Saturday' THEN 7 END;

# Result:



* + - Weekday Patterns by Member Type:

-- Weekday ride counts and average distance by member type

SELECT

member\_casual,

weekday,

COUNT(\*) AS ride\_count,

AVG(distance) AS avg\_distance

FROM distance

GROUP BY member\_casual, weekday

ORDER BY member\_casual, CASE WHEN weekday = 'Sunday' THEN 1

WHEN weekday = 'Monday' THEN 2

WHEN weekday = 'Tuesday' THEN 3

WHEN weekday = 'Wednesday' THEN 4

WHEN weekday = 'Thursday' THEN 5

WHEN weekday = 'Friday' THEN 6

WHEN weekday = 'Saturday' THEN 7 END;

# Result:



* + - Rideable Type Distribution:

-- Count of rides by rideable type

SELECT

rideable\_type,

COUNT(\*) AS ride\_count

FROM distance

GROUP BY rideable\_type;

# Result:



* + - Rideable Type Distribution by Member Type:

-- Count of rides by rideable type and member type

SELECT

member\_casual,

rideable\_type,

COUNT(\*) AS ride\_count

FROM distance

GROUP BY member\_casual, rideable\_type;

# Result:



* + - Member Type Distribution:

-- Count of rides by member type

SELECT

member\_casual,

COUNT(\*) AS ride\_count

FROM distance

GROUP BY member\_casual;

# Result:



**Analysis: How Annual Members and Casual Riders Use Cyclistic Bikes Differently**

**Monthly Usage Trends:**

A line drawing of a triangle

Description automatically generated

Annual Members:

Consistently high ride counts throughout the year.

Possibly commuting regularly, as patterns are relatively stable month-to-month.

Casual Riders:

More variability in ride counts, peaking during warmer months or special events.

May use bikes sporadically for leisure or specific occasions.

**Hourly Patterns:**

A graph with a line and pointy line

Description automatically generated with medium confidence

Annual Members:

Peak ride counts during typical commuting hours (early morning and late afternoon).

Indicates likely use for daily work commuting.

Casual Riders:

More evenly distributed ride counts throughout the day.

Suggests usage for non-commuting purposes, such as leisure or occasional transportation.

**Weekday Usage Patterns:**

A graph of a bar chart

Description automatically generated

Annual Members:

Consistent high ride counts during weekdays, peaking on weekdays.

Strong indication of weekday commuting behavior.

Casual Riders:

Relatively consistent ride counts across weekdays.

Usage not heavily concentrated on weekdays, indicating varied patterns.

**Rideable Type Preference:**

A group of orange and brown circles

Description automatically generated

Annual Members:

Likely to prefer standard bikes, possibly for daily commuting needs.

Casual Riders:

More diverse usage, showing relatively even distribution between standard and electric bikes.

Indicates a variety of purposes, including leisure rides and exploration.

**Ride Duration and Distance:**

**A red and white dotted line

Description automatically generated with medium confidence**

Annual Members:

Longer average ride durations and distances.

Suggests more purposeful, utilitarian use, possibly for commuting or longer trips.

Casual Riders:

Shorter average ride durations and distances.

Implies more spontaneous and leisure-oriented usage.

**Seasonal Variations:**

A screenshot of a color chart

Description automatically generated

Annual Members:

Usage remains relatively stable across seasons.

Indicates consistent reliance on bike commuting regardless of weather conditions.

Casual Riders:

Seasonal fluctuations with higher usage in warmer months.

Suggests a preference for fair weather and outdoor activities.

**Overall Insights:**

A close-up of a graph

Description automatically generated

Annual members exhibit more consistent and predictable usage patterns, indicating a primary use case for commuting.

Casual riders demonstrate diverse and sporadic usage patterns, suggesting a mix of leisure, recreational, and occasional transportation needs.

Understanding these differences can inform targeted marketing strategies aimed at converting casual riders into annual members.

Tableau Project page [HERE](https://public.tableau.com/views/citi_dashboard/Dashboard1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)