**Cyclistic Bike-Share Data Analysis**

**Business Task:**

How do annual members and casual riders use Cyclistic bikes differently? This was analyzed through answering the following questions:

* Is there a difference in the amount of time spent riding?
* Are there differences regarding the type of bike they choose to ride?
* Are there differences in patterns regarding which day of week the cyclists choose to ride?

**Data Sources:**

The data was obtained from the following site. <https://divvy-tripdata.s3.amazonaws.com/index.html>

The data contains information about each of the rides of the month which includes the variables ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, start\_station\_id, end\_station\_name, end\_station\_id, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual. Ride\_id lists the numerical id of the ride. Rideable type indicates whether the rider rode a classical or an electric bike. Started\_at indicates the time the ride started and ended\_at indicates the time the ride ended. Start\_station lists the station the ride started, and end\_station lists the station the ride ended. Start\_lat lists the latitude the ride started, start\_lng lists the longitude the ride started, end\_lat lists the latitude the ride ended, and end\_long lists the longitude the ride ended. Member\_casual indicates whether the rider has a casual or annual membership. Additionally, two other variables were added during the data cleaning process: ride\_length and day\_of\_week. Ride\_length indicates how long the ride lasted and day\_of\_week indicates the day of week the ride was performed.

The year 2022 was the year of focus. Only data from the months January to April and October to December were analyzed as the other months contained files which were too big for bigquery to process.

**Data Cleaning:**

The column ride\_length was added to the dataset which subtracted the values from ended\_at and started\_at and used the [ss] timing format. The column day\_of\_week was added which applied the weekday function weekday function on the started\_at column.

**Analysis:**

The months from January to April and October to December were processed by bigquery. There were four separate queries written: first save, ride\_length, rideable\_type, and day\_of\_week, each focusing on answering each of the questions.

The average ride length for casual riders was roughly 1600 seconds while the average ride length for annual members was roughly 692 seconds. At first glance, it was surprising that on average, casual riders rode longer than annual members because we would expect more commitment from annual members. However, it was found that there were many more rides by annual members than casual riders, so the difference could be attributed to annual members choosing to ride more often than casual riders. Data on the identity of the rider would need to be collected to see if this were the reason. If so, this would indicate that annual members are riding for shorter times, but more often.

In usages between classic and electric bikes, both exhibited a preference towards electric bikes. Casual members used classic bikes around 36.2% of the time while annual members used classic bikes around 47.8% of the time, and the larger share of use of classic bikes for the annual members was statistically significant.

The weekdays (Monday through Friday) featured a larger portion of annual members, while the weekends featured a larger portion of casual members. This was aligned with expectations as we would expect annual members to take riding more seriously and thus exhibit more willingness to ride even during the weekdays.

**Visualizations:**

**Recommendations:**

Unfortunately, it is difficult to ascertain if these trends are correlative or causative, so short of causative analysis, it is entirely possible that many of these recommendations could fall short. To determine the specific *causes* which lead one to become annual members, polls should be conducted on riding habits focusing on the frequency of riding, ride lengths, bike types, and the days people choose to ride. Furthermore, data should be collected on the identity of the riders, as this will enable us to see if the difference in average ride times is made up by annual members choosing to ride more often.

Assuming there was a causal effect, to gain more annual memberships, there are a couple of things which can be done. If it were found that annual members were riding more often, there should be additional resources dedicated to advertising the benefits of riding regularly. Furthermore, as there is a preference towards the classic bikes of annual members relative to casual members, more resources can be put into improving the quality of the classic bikes. Finally, although I doubt that a more even distribution of rides throughout the week for annual members causes annual memberships, rather than the other way around, it might be beneficial to get riders to participate more often during the weekdays.

**Code:**

First Save:

WITH

  Cyclists AS (

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Jan2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Feb2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.March2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.April2021`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Oct2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Nov2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Dec2022`

  )

  SELECT COUNT(member\_casual) AS numMembers

  FROM

    Cyclists

  WHERE

    member\_casual = 'casual'

  UNION ALL

  SELECT COUNT(member\_casual)

  FROM

    Cyclists

  WHERE

    member\_casual = 'member'

ride\_length:

WITH

  Cyclists AS (

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Jan2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Feb2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.March2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.April2021`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Oct2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Nov2022`

    UNION ALL

      SELECT

        \*

      FROM

        `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Dec2022`

  )

  SELECT AVG(ride\_length) AS rideLengthAvgs

  FROM

    Cyclists

  WHERE

    member\_casual = 'casual'

  UNION ALL

  SELECT AVG(ride\_length)

  FROM

    Cyclists

  WHERE

    member\_casual = 'member'

rideable\_type:

WITH

  Cyclists AS (

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Jan2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Feb2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.March2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.April2021`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Oct2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Nov2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Dec2022`

  )

  SELECT COUNT(rideable\_type) AS ridingType

  FROM

   Cyclists

  WHERE

    rideable\_type = "classic\_bike" AND member\_casual = 'casual'

  UNION ALL

  SELECT COUNT(rideable\_type)

  FROM

   Cyclists

  WHERE

    rideable\_type = "electric\_bike" AND member\_casual = 'casual'

  UNION ALL

  SELECT COUNT(rideable\_type)

  FROM

   Cyclists

  WHERE

    rideable\_type = "classic\_bike" AND member\_casual = 'member'

  UNION ALL

  SELECT COUNT(rideable\_type)

  FROM

   Cyclists

  WHERE

    rideable\_type = "electric\_bike" AND member\_casual = 'member'

day\_of\_week:

WITH

  Cyclists AS (

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Jan2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Feb2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.March2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.April2021`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Oct2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Nov2022`

    UNION ALL

    SELECT

      \*

    FROM

      `trusty-entity-390703.GoogleAnalytics\_Cyclist\_Project.Dec2022`

  )

  SELECT COUNT(day\_of\_week) as weekday

  FROM

    Cyclists

  WHERE

    member\_casual = "casual" AND day\_of\_week = 1

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "casual" AND day\_of\_week = 2

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "casual" AND day\_of\_week = 3

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "casual" AND day\_of\_week = 4

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "casual" AND day\_of\_week = 5

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "casual" AND day\_of\_week = 6

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "casual" AND day\_of\_week = 7

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "member" AND day\_of\_week = 1

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "member" AND day\_of\_week = 2

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "member" AND day\_of\_week = 3

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "member" AND day\_of\_week = 4

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "member" AND day\_of\_week = 5

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "member" AND day\_of\_week = 6

  UNION ALL

  SELECT COUNT(day\_of\_week)

  FROM

    Cyclists

  WHERE

    member\_casual = "member" AND day\_of\_week = 7