CYCLISTIC BIKE-SHARE ANALYSIS INSIGHT

SQL + TABLEAU

INTRODUCTION

This report presents an analysis of Cyclistic bike share data using SQL and Tableau to gain insights into the usage patterns of casual riders and annual members. As a junior data analyst at Cyclistic, the objective is to maximize the number of annual memberships to increase company profitability. By understanding how casual riders and annual members utilize Cyclistic bikes differently, the marketing team aims to develop a data-driven marketing strategy to convert casual riders into loyal annual members. The recommendations provided in this report are supported by compelling data insights and professional data visualizations.

METHODOLOGY

The analysis is conducted using SQL to query and extract relevant data from the Cyclistic bike share database. The data is then imported into Tableau, a powerful data visualization tool, to create visually appealing and informative visualizations. The SQL queries are designed to extract key metrics such as ride duration, ride frequency, customer types, and station popularity to understand the usage patterns of casual riders and annual members.

Below are the steps taken to derive insights from the data provided.

- Downloaded data, and stored it appropriately.
- Provided CSV Data structure is analyzed and sorted.
- Necessary Data cleaning was done by removing blank/error values (i.e., the ones which couldn't be correlated with the available variables).
- After primary data cleaning via EXCEL, SQL has been selected for data extraction from the Cyclistic database.
- The data is further imported to Tableau, converted as customer SQL, and connected relationships of various extracted data tables to the main dataset through the common table.

Some of the metrics derived from the dataset via SQL query are.,

- Rides per day data (RPD_Data)
- Duration per day data (DPD_Data)
- Customer Bike Preference data (CBP_Data)
- Station Ride Data (SR_Data)
- Trip Duration Data (TD_Data)

RIDES PER DAY DATA (RPD_DATA)

Step 1: Executed query for Rides per day by Casual customers

```
Create table Rides_per_day_casual (
started_at_date DATE,
Rides_per_day_C INT);
INSERT INTO Rides_per_day_casual (started_at_date, rides_per_day_C)
SELECT started_at_date, COUNT(*) AS Rides_per_day_C
FROM dataset
WHERE member_casual = 'casual'
GROUP BY started_at_date;
```

Step 2: Executed query for Rides per day by Member customers

```
Create table Rides_per_day_member (
started_at_date DATE,
Rides_per_day_M INT);
INSERT INTO Rides_per_day_member (started_at_date, rides_per_day_M)
SELECT started_at_date, COUNT (*) AS Rides_per_day_M
FROM dataset
WHERE member_casual = 'member'
GROUP BY started_at_date;
```

Step 3: Executed query for Rides per day by total customers

```
CREATE TABLE Rides_Per_Day_Total AS
SELECT Started_At_Date AS Started_At_Date, COUNT(*) AS Rides_Per_Day_Total
FROM Dataset
GROUP BY Started_At_Date;
```

Step 4: Joined all three tables as RPD_Data by the left join function

```
Create table Rpd_Data as
SELECT rpd_c.started_at_date, rpd_c.rides_per_day_c, rpd_m.rides_per_day_m, rpd_t.rides_per_day_total
FROM Rides_per_day_casual as rpd_c
left join rides_per_day_member as rpd_m on rpd_c.started_at_date=rpd_m.started_at_date
left join rides_per_day_total as rpd_t
on rpd_c.started_at_date=rpd_t.started_at_date
```

DURATION PER DAY DATA (DPD_DATA)

Step 1: Executed query for Duration per day by Casual customers

```
Create table Dur_per_day_C (
started_at_date DATE,
Dur_per_day_C INT);
INSERT INTO Dur_per_day_C ( Started_at_Date, Dur_Per_day_C)
SELECT started_at_date, SUM(TIMESTAMPDIFF(MINUTE,
CONCAT(started_at_date, ' ', started_at_time),
CONCAT(ended_at_date, ' ', ended_at_time))) AS Dur_per_day_C
FROM dataset
WHERE member_casual = 'casual'
GROUP BY started_at_date;
```

Step 2: Executed query for Duration per day by Member customers

```
Create table Dur_per_day_M (
started_at_date DATE,
Dur_per_day_M INT);
INSERT INTO Dur_per_day_M ( Started_at_Date, Dur_Per_day_M)
SELECT started_at_date, SUM(TIMESTAMPDIFF(MINUTE,
CONCAT(started_at_date, ' ', started_at_time),
CONCAT(ended_at_date, ' ', ended_at_time))) AS Dur_per_day_M
FROM dataset
WHERE member_casual = 'member'
GROUP BY started_at_date;
```

Step 3: Executed query for Duration per day by total customers

```
Create table Dur_per_day_total as
SELECT DATE(started_at_date) AS started_at_date,
SUM(TIMESTAMPDIFF(MINUTE,CONCAT(started_at_date, ' ', started_at_time),
CONCAT(ended_at_date, ' ', ended_at_time))) AS total_duration_per_day
FROM Dataset
GROUP BY DATE(started_at_date);
```

Step 4: Joined all three tables as DPD_Data by the left join function

```
Create table Dpd_Data as
SELECT dpd_c.started_at_date, dpd_c.dur_per_day_c, dpd_m.dur_per_day_m,
dpd_t.total_duration_per_day
FROM Dur_per_day_C as dpd_c
left join Dur_per_day_m as dpd_m on dpd_c.started_at_date=dpd_m.started_at_date
left join dur_per_day_total as dpd_t
on dpd_c.started_at_date=dpd_t.started_at_date
```

CUSTOMER BIKE PREFERENCE DATA (CBP_DATA)

Step 1: Executed query for customer bike preference by ride count

```
Create Table BRP_Data as
SELECT member_casual, rideable_type, COUNT(*) AS ride_count,
    ROUND((COUNT(*) / (SELECT COUNT(*) FROM dataset) * 100), 2) AS ride_percentage
FROM dataset
GROUP BY member_casual, rideable_type;
```

Step 2: Executed query for customer bike preference by duration

```
Create Table BRD_Data as
SELECT member_casual,rideable_type, SUM(TIMESTAMPDIFF(MINUTE,
CONCAT(started_at_date, ' ', started_at_time),
CONCAT(ended_at_date, ' ', ended_at_time))) AS total_duration_minutes
FROM dataset
GROUP BY member_casual,rideable_type;
```

Step 3: Joined two tables as CBP_Data by the left join function

```
Create table CBP_Data as
SELECT BRP_D.member_casual, BRP_D.Rideable_type, BRP_D.Ride_count, BRP_D. Ride_percentage,
BRD_D.Total_duration_minutes
FROM BRP_Data as BRP_D
left join BRD_Data as BRD_D on (BRP_D.rideable_type=BRD_D.rideable_type and
BRP_D.member_casual=BRD_D.member_casual)
```

STATION RIDE DATA (SR_DATA)

Step 1: Executed query for station-wise ride count by Casual customers

```
Create Table SR_C as
SELECT start_station_name, COUNT(*) AS total_rides_C
FROM dataset
WHERE member_casual = 'casual'
GROUP BY start_station_name;
```

Step 2: Executed query for station-wise ride count by Member customers

```
Create Table SR_M as
SELECT start_station_name, COUNT(*) AS total_rides_M
FROM dataset
WHERE member_casual = 'member'
GROUP BY start_station_name;
```

Step 3: Executed query for station-wise ride count by total customers

```
Create Table SR_T as
SELECT start_station_name, COUNT(*) AS total_rides_T
FROM dataset
GROUP BY start_station_name;
```

Step 4: Joined all three tables as SR_Data by the left join function

```
Create Table SR_Data as
Select SR_C. Start_Station_Name, SR_C. Total_Rides_c, SR_M.Total_Rides_m, SR_T.Total_Rides_t
FROM SR_C
Left Join SR_M on SR_C.Start_Station_name=SR_M.Start_Station_Name
Left Join SR_T on SR_C.Start_Station_name=SR_T.Start_Station_Name
```

TRIP DURATION DATA (TD_DATA)

Step 1: Executed query for the trip duration from start time to end time by start station to end station.

```
Create Table TD_Data as
SELECT start_station_name, end_station_name,
SUM(TIMESTAMPDIFF(MINUTE, CONCAT(started_at_date, ' ', started_at_time),
CONCAT(ended_at_date, ' ', ended_at_time))) AS total_duration_minutes
FROM dataset
GROUP BY start_station_name, end_station_name;
```

KEY FINDINGS

USAGE PATTERNS - CASUAL RIDERS

- Casual riders show higher bike usage on weekends, with an average of 51,765 rides and a total ride duration of 23,16,901.25 minutes on Sundays.
- Monday sees a slight dip in casual rider activity, with 32,235 rides and a total ride duration of 12,64,878.75 minutes.
- Usage further decreases on Tuesday, with 20,837 rides and a total ride duration of 6,19,084 minutes.
- Wednesday sees a moderate increase in casual rider activity, with 27,543 rides and a total ride duration of 9,90,475 minutes.
- On Fridays, casual riders exhibit higher usage again, with 38,327.5 rides and a total ride duration of 13,92,706.25 minutes.

USAGE PATTERNS - ANNUAL MEMBERS

- Annual members consistently utilize Cyclistic bikes throughout the week. On Sundays, there are 45,526.25 rides with a total ride duration of 8,55,903.75 minutes.
- Mondays see similar activity levels, with 45,171.25 rides and a total ride duration of 6,75,332.5 minutes.
- Annual members maintain their usage on Tuesdays, with 42,883 rides and a total ride duration of 5,92,846 minutes.
- Wednesdays show slightly higher activity, with 51,368 rides and a total ride duration of 7,16,313 minutes.
- Thursdays and Saturdays experience lower usage compared to other weekdays, with ride counts ranging from 27,741.25 to 27,543 and total ride durations ranging from 8,72,565 to 7,93,206.25 minutes.

USAGE PATTERNS BASED ON HOURLY DATA

- **Peak Hours:** The highest number of rides for both casual and member riders occur between 11 AM and 12 PM, indicating a popular time for bike usage.
- **Casual Riders' Peak:** Casual riders have a significant spike in ride counts during the evening hours (between 5 PM and 7 PM), potentially indicating leisure or recreational usage.
- **Member Riders' Consistency:** Member riders exhibit a more consistent pattern throughout the day, with a gradual increase in ride counts from morning until noon.
- **Morning Commute:** There is a noticeable increase in ride counts for both casual and member riders during the morning hours (between 6 AM and 9 AM), suggesting bike usage for daily commuting.
- **Evening Activity:** Casual riders show higher ride counts during the evening hours (between 6 PM and 9 PM), indicating bike usage for recreational purposes or after-work activities.

USAGE PATTERNS BASED ON POPULAR STATION DATA

- **Station Preference:** Wabash Ave & Grand Ave and Streeter Dr & Grand Ave are popular stations for both casual and member riders, indicating their preference for these locations.
- **Diverse Usage:** Millennium Park and Lake Shore Dr & Monroe St are frequented by both casual and member riders, suggesting that these stations attract a diverse range of users.
- **Consistent Demand:** The Theatre on the Lake station shows consistent average ride counts for both casual and member riders, indicating a stable demand throughout the day.
- **Balanced Utilization:** The average ride counts for both casual and member riders at the popular stations is relatively balanced, indicating a similar level of usage by both types of riders.
- **High Average Ride Counts:** The average ride counts for popular stations range from approximately 35,883 to 48,354 per day, indicating a significant level of bike usage and popularity among riders.

These usage patterns highlight the importance of popular stations in the Cyclistic bike-share system, suggesting that targeted marketing efforts and station enhancements could further maximize ridership and encourage membership conversions.

USAGE PATTERNS BASED ON BIKE TYPE

- **Ride Percentage:** Among casual riders, docked bikes account for 34.26% of the rides, while electric bikes contribute to 8.76% of the rides. For member riders, docked bikes have a higher ride percentage of 46.44%, compared to 10.54% for electric bikes. This indicates that both casual and member riders primarily prefer docked bikes over electric bikes.
- **Total Duration:** Docked bikes accumulate a significantly higher total duration of usage, with a combined total of 5,43,66,940 minutes. In contrast, electric bikes have a total duration of 85,30,435 minutes, indicating that riders spend more time on docked bikes overall.
- **Ride Count:** Docked bikes also have a higher ride count, with 20,19,085 rides, compared to 4,82,865 rides for electric bikes. This further highlights the popularity and usage of docked bikes among Cyclistic riders.
- **Member Preference:** Members show a higher ride percentage and total duration for both docked bikes and electric bikes compared to casual riders. This suggests that members are more likely to utilize both types of bikes more frequently and for longer durations than casual riders.
- **Potential for Electric Bike Promotion:** While docked bikes remain the preferred choice for riders, the data indicates a growing interest in electric bikes among both casual and member riders. To capitalize on this trend, Cyclistic could consider promoting electric bike usage through targeted marketing campaigns or incentives to encourage more riders to choose electric bikes.

RECOMMENDATIONS

Enhanced Marketing for Annual Memberships: Develop targeted marketing campaigns that highlight the cost-saving benefits and convenience of annual memberships, emphasizing features such as unlimited access and priority bike availability during peak hours. Leverage data insights to tailor messaging for specific customer segments.

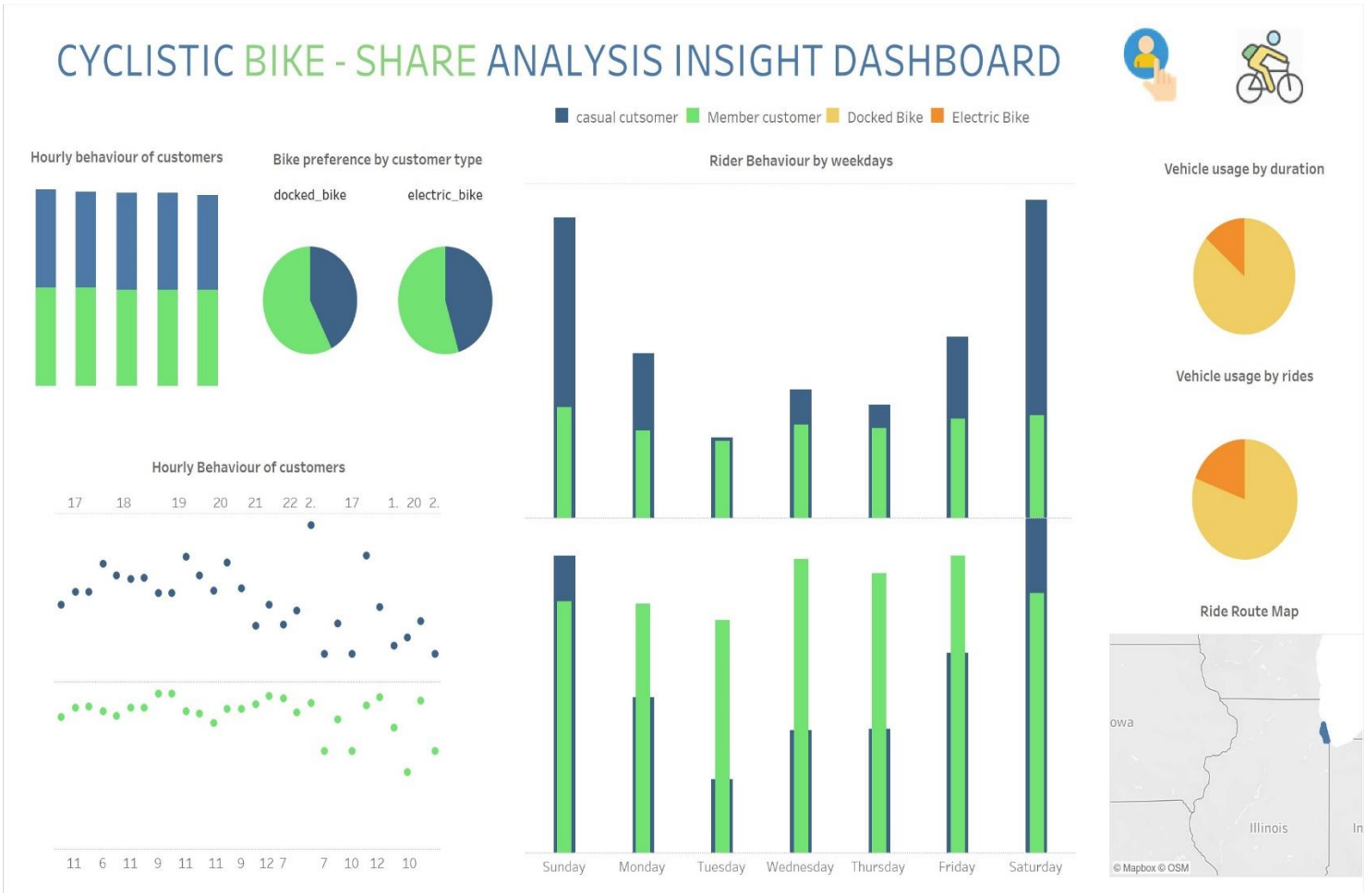
Promote Commuter Benefits: Emphasize the advantages of using Cyclistic bikes for daily commuting, such as reducing traffic congestion, saving on parking fees, and promoting a healthier lifestyle. Collaborate with local businesses to offer discounts or incentives to annual members for using the bike-share service.

Geographical Expansion: Identify areas with a high concentration of casual riders and consider expanding bike station networks to target these locations. By providing convenient access and promoting nearby attractions or events, Cyclistic can attract casual riders and convert them into annual members.

Improved User Experience: Enhance the user experience by integrating digital tools and features. Develop a user-friendly mobile application that provides real-time bike availability, suggested routes, and personalized recommendations based on user preferences. This will appeal to both casual riders and annual members, making the service more attractive and convenient.

CONCLUSION

"By understanding the differences in behavior between casual riders and annual members, Cyclistic can design an effective marketing strategy to convert casual riders into loyal annual members. Leveraging data insights and implementing the recommended strategies, Cyclistic has the potential to maximize annual memberships, increase company profit, and establish itself as a leading bike-share service provider in Chicago."


CYCLISTIC BIKE - SHARE ANALYSIS INSIGHT DASHBOARD
casual cutsomer
Member customer
Docked Bike
Electric Bike
Hourly behaviour of customers
Bike preference by customer type
docked_bike
electric_bike
Rider Behaviour by weekdays
Vehicle usage by duration
Vehicle usage by rides
Hourly Behaviour of customers
17 18 19 20 21 22 2. 17 1. 20 2.
11 6 11 9 11 11 9 12 7 7 10 12 10
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Ride Route Map
owa
Illinois
© Mapbox © OSM

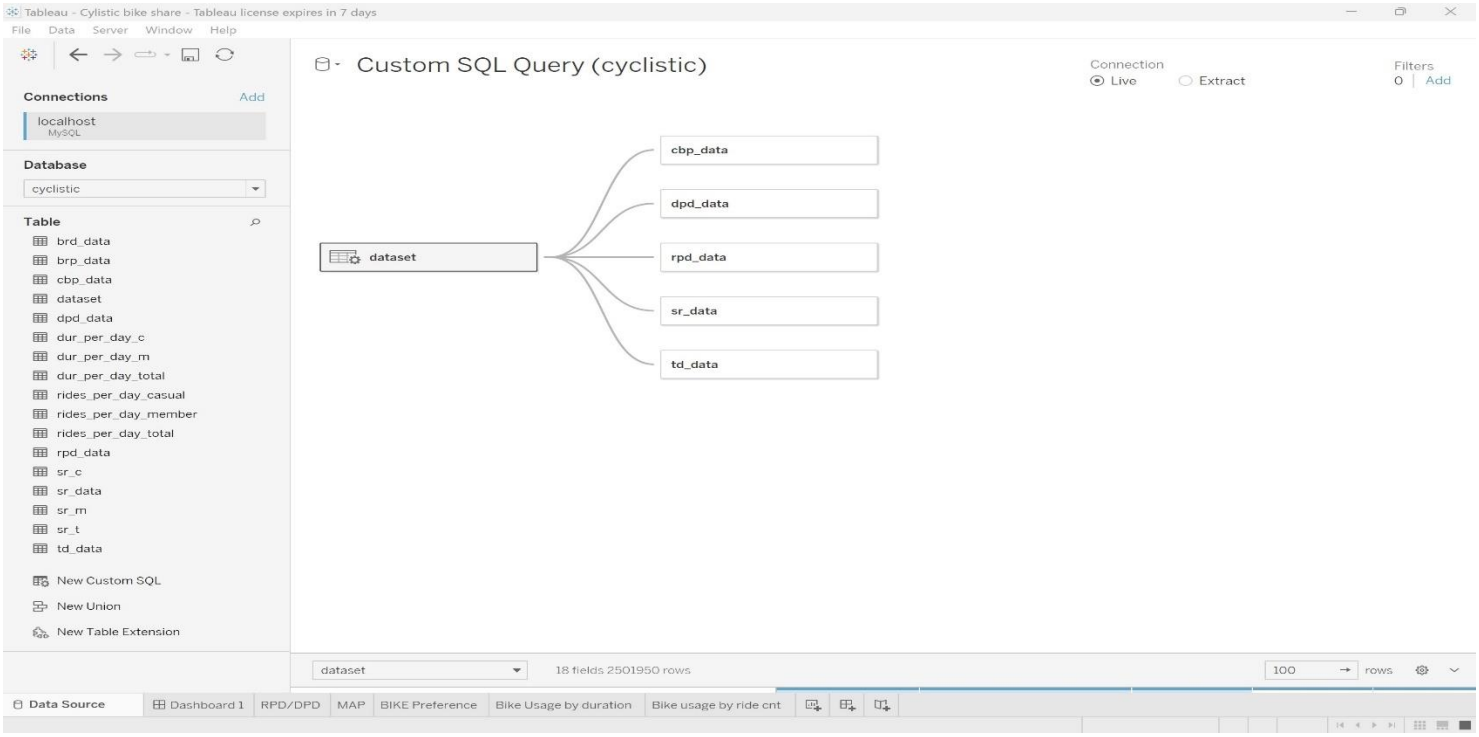


Tableau - Cylistic bike share - Tableau license expires in 7 days
File Data Server Window Help
Connections
Add
localhost
MySQL
Database
cyclistic
Table
brd_data
brp_data
cbp_data
dataset
dpd_data
dur_per_day_c
dur_per_day_m
dur_per_day_total
rides_per_day_casual
rides_per_day_member
rides_per_day_total
rpd_data
sr_c
sr_data
sr_m
sr_t
td_data
New Custom SQL
New Union
New Table Extension
Custom SQL Query (cyclistic)
Connection
Live
Extract
Filters
0 | Add
dataset
cbp_data
dpd_data
rpd_data
sr_data
td_data
dataset
18 fields 2501950 rows
100
rows
Data Source
Dashboard 1
RPD/DPD
MAP
BIKE Preference
Bike Usage by duration
Bike usage by ride cnt