

Bike share capstone project1

Scenario

You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. 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.

Skills Demonstrated

Some of the skills demonstrated for a pizza sales report are:

- Data acquisition: Obtained the data made available by Motivate International Inc.
- Data transformation: I have cleaned, filtered, and transformed data using SQL, a structured query language for manipulating and analyzing data.
- Data modeling: I designed and implemented a data model that represents the relationships and attributes of the data entities, such as biketypes, bike stations, member_casual types, and ride times and date.
- Data analysis: Performed in-depth data analysis using SQL queries, such as aggregating, grouping, sorting, filtering, and joining data.
- Data visualization: Created interactive and informative visualizations using Tableau, a business intelligence tool for creating reports and dashboards.
- Data interpretation: Derived insights and recommendations from the data analysis and visualizations, such as identifying sales trends, customer preferences, product performance.
- Communication: Able to present and communicate the data analysis and visualizations in a clear and concise manner, using appropriate language, format, and style.

Tools Used

MS OFFICE/ EXCEL: VERSION 2016

SQL BIGQUERY

© 2024 TABLEAU SOFTWARE, LLC, A SALESFORCE COMPANY. ALL RIGHTS RESERVED

SUMMARY

“Cyclistic” is a hypothetical bike-share company in Chicago aiming to convert casual riders into annual members. The marketing team, led by Lily Moreno, plans to achieve this by analyzing usage patterns and creating engaging data visualizations. The analysis revealed differences in usage between casual riders and annual members, providing insights to help shape their marketing strategy.

Solution

Findings:

I used tableau for visualization. Here I am mentioning the link:

[Click here to view the Dashboard](#)

- There were six hundred thousand more recorded rides by member riders over casual riders. But the casual riders spent more than two times as many minutes in their rides than the member riders. Casual riders also experienced a maximum average ride length much higher than that of the member riders per month.
- Casual riders experienced a peak in July while member riders experienced theirs in August. Both riders also experienced a year low in January.
- Casual riders went on the most rides on Saturdays whilst member riders went on the the most rides on Tuesday.
- The most common rideable type in the recorded period was the classic bike. Both rider types recorded the most rides with the classic bike. However, the member riders did not make use of the docked bike at all.
- Casuals have the greater max time spent riding, this could be as a result of enjoying the ride(pleasure/leisure/exploring) or as a means of exercising.

Some recommendations based on the analysis:

- Implement a limited time promotion for annual membership that loosens limits on Friday, Saturday, Sunday rides, given those are the most popular days casual riders use Cyclistic bikes
- Add more electric bikes to inventory given casual riders prefer them over classic bikes.
- Another way to convert casual riders into full time members could be targeted physical ads and campaigns.
- Targeted premium features could be offered to persuade casual users to join as members to meet their goals for riding, mostly on weekends

Approach:

The first step in the analysis is to define the objective and scope of the Project. This means coming up with a clear and specific question or problem that you want to solve with data, such as

Ask phase: How do annual members and casual riders use Cyclistic bikes differently?

Prepare phase:

Data is located in the form of CSV files in the system. And it is As a disclaimer the data has been made available by Motivate International Inc. under this [license](#). It is organized in the table form.

A good data source is ROCCC which stands for **R**eliable, **O**riginal, **C**omprehensive, **C**urrent, and **C**ited.

Reliable — high — it has 476888 rows

Original — high — provided by the company directly

Comprehensive — high — Parameters match parameters

Current — med — Data is 3 years old and somehow relevant

Cited — high — Data collected from company, hence useful

Process phase:

- I imported the tables into SQL bigquery for analysis for cleaning and analyzing.
- I used a union all query to combine the tables. After that, I saved the results into new table for analysis.
I used the new table for cleaning the data.
- At first I check for the data and removed all the duplicate rows from the data.
- Checked for all the missing and NULL values and remove the data with missing values.
- I used the date column to extract days, months, time into separate columns.
- I used time in seconds for my analysis of bike share data.
- I removed the column where started_at was bigger than ended_at.
- I checked for any typos and then corrected those typos.

NOTE: I saved all the queries I used to clean the data in the project file.

After cleaning the data, I saved the cleaned data into separate table for analysis where I analyzed the following measures:

- **No of rides:** I selected rideable type and used count function for no of rides from the table and grouped by rideable type, order by no of rides descending order so that the most Rides comes first

```
SELECT
    rideable_type,
    COUNT
    (ride_id) AS no_of_user_types
FROM `premium-axis-403220.Cyclistic_bike_share_2022.2022_bike_share`
GROUP BY rideable_type
ORDER BY no_of_user_types DESC;
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW
Row	rideable_type	no_of_user_types		
1	electric_bike	2498292		
2	classic_bike	2129542		
3	docked_bike	142529		

- **Most patronized bike types:** Calculated the number of rides by Bike type and user type. Grouped the data by bike type and user type and ordered by descending.

```
SELECT
    rideable_type AS bike_type,
    member_casual AS user_type,
    COUNT(ride_id) AS no_of_rides
FROM
    `premium-axis-403220.Cyclistic_bike_share_2022.2022_bike_share`
GROUP BY rideable_type, member_casual
ORDER BY no_of_rides DESC;
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON	EXECUTION
Row	bike_type	user_type				
1	electric_bike	member				1430914
2	classic_bike	member				1420966
3	electric_bike	casual				1067378
4	classic_bike	casual				708576
5	docked_bike	casual				142529

- **Average length of ride in seconds:** Calculated the average ride length by using Avg function by its user types and used the round function for 2 decimal places.

```
SELECT
    member_casual AS user_type,
ROUND
    (AVG(ride_length), 2) AS avg_ride_length
FROM
    `premium-axis-403220.Cyclistic_bike_share_2022.2022_bike_share`
GROUP BY member_casual;
```

Query results			
JOB INFORMATION		RESULTS	CHART PREVIEW
Row	user_type	avg_ride_length	
1	casual	1246.75	
2	member	723.86	

- **Most active days:** Calculated the no of rides in a day by its user and bike type.

```
SELECT
    day_started as day_of_week,
COUNT(day_started) AS no_of_rides_per_day,
    member_casual as user_type,
    rideable_type as bike_type
FROM
    `premium-axis-403220.Cyclistic_bike_share_2022.2022_bike_share`
GROUP BY member_casual, day_started,rideable_type;
```

Query results					
JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON
Row	day_of_week	no_of_rides_per_day	user_type	bike_type	
21	Tuesday	76926	casual	classic_bike	
22	Wednesday	224036	member	electric_bike	
23	Wednesday	133371	casual	electric_bike	
24	Thursday	227345	member	electric_bike	
25	Thursday	145309	casual	electric_bike	
26	Friday	206062	member	electric_bike	
27	Friday	159554	casual	electric_bike	
28	Saturday	189114	member	electric_bike	
29	Saturday	203176	casual	electric_bike	
30	Sunday	164625	casual	electric_bike	
31	Sunday	161816	member	electric_bike	
32	Monday	200647	member	electric_bike	
33	Monday	131575	casual	electric_bike	
34	Tuesday	129768	casual	electric_bike	
35	Tuesday	221894	member	electric_bike	

- **Max ride length:** calculated the maximum ride length by its user and bike type using MAX function.

```
SELECT
    max(ride_length) as max_ride_length,
    member_casual as user_type,
    rideable_type as bike_type

FROM
    `premium-axis-403220.Cyclistic_bike_share_2022.2022_bike_share`

GROUP BY member_casual,rideable_type;
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON	EXECUTION
Row	max_ride_length	user_type		bike_type		
1	82127	casual		docked_bike		
2	84509	casual		classic_bike		
3	69947	member		classic_bike		
4	28831	member		electric_bike		
5	28823	casual		electric_bike		

- **Most active month by usertypes:** calculated the most active days by its user and bike type:

```
SELECT
    count(month_started) as no_of_rides_per_month,
    member_casual as user_type,
    rideable_type as bike_type

FROM
    `premium-axis-403220.Cyclistic_bike_share_2022.2022_bike_share`

GROUP BY member_casual,rideable_type;
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON	EXECUTION
Row	no_of_rides_per_mor	user_type		bike_type		
1	142529	casual		docked_bike		
2	1420966	member		classic_bike		
3	708576	casual		classic_bike		
4	1430914	member		electric_bike		
5	1067378	casual		electric_bike		