Cyclistic bike share Google Capstone

Kunal Raj Singh

2022-11-15

Report Deliverables :

- A clear statement of the business task
- A description of all data sources used
- Documentation of any cleaning or manipulation of data
- A summary of analysis
- Supporting visualizations and key findings
- Top three recommendations

Route for analysis :

The analysis will be based on six step process which are as follows :
* Ask
* Prepare
* Process
* Analyze
* Share
* Act
The tools that I will be using for analysis are MS Excel,Tableau,JupyterNotebook and RStudio.

ASK

Business Task - The aim is to identify how the annual members and casual riders use Cyclistic bikes differently. The main reason for this analysis is to identify any trends to help in designing marketing strategies aimed at converting casual riders into annual members. Key stakeholder for this project is Cyclistic Executive team,the Director of Marketing.

PREPARE

For the purposes of this case study, the dataset was made available by Motivate International Inc. under this license This is public data that can used to explore how different customer types are using Cyclistic bikes.The data has been anonymized and no personally identifiable information has been included in the data. While this will prevent analysis which would examine specific personal traits of individual riders, such as

individual history or areas of residence, there is still enough data to identify certain behaviours.
Note - The data-privacy issues prohibit users from using riders' personally identifiable information. This means that users won't be able to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.
Description of Data Sources used for analysis
* Data for the past 12 months from November of 2021 till October of 2022 is used to gain insights for the project.
* The data is in Structured form in .csv format with following fields.
* ride_id (String)
* rideable_type (String)
* started_at (Datetime)
* ended_at (Datetime)
* start_station_name (String)
* start_station_id (String)
* end_station_name (String)
* end_station_id (String)
* start_lat (Int)
* start_lng (Int)
* end_lat (Int)
* end_lng (Int)
* member_casual (String)
The data source for the analysis process is reliable,comprehensive,current which makes it credible for usage.

PROCESS

Tools used - MS Excel

Cleaning Process

- Removed duplicate records
- Removed records with Null values as data is fictional so reason behind Null values is uncertain. Though the sample size of data is quiet large and removed records for Null values were negligible.
- Excluded geographical coordinates
- Ensured consistency in datatypes across variables
- Alteration provided to data -
  Columns added : start_date (date dd-mm-yyyy) start_time (time hh:mm:ss) end_date (date dd-mm-yyyy) end_time (time hh:mm:ss) ride_time (time hh:mm:ss) day_of_week(number 1-Sunday,7-Saturday)
  Sorting and Filtering :

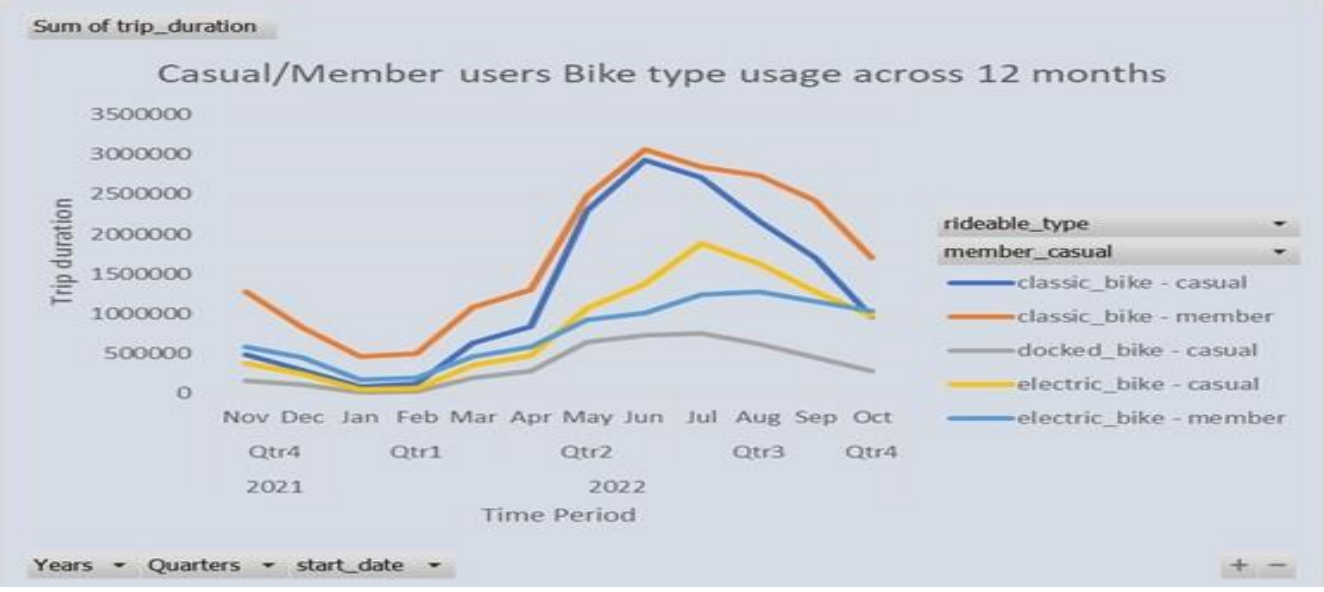
- Custom sorted dataset on the basis of start_date column.
- Filtered data for any null values in columns start_station_name,end_station_name,started_at,ended_at and member_casual.
  Formatting :
- Formatted the worksheets as a Table.
- Adjusted Column width AutoFit .
- Headers bold Data Manipulation :
- Used TRUNC() function to separate datetime to date and time.
- Subtracted end_time from start_time to generate a ride_time column
- Used WEEKDAY() function to get day of the week.

ANALYZE

Tool 1- MS Excel
Importing data(12 .csv files for each month) into PowerQuery Editor.
Transforming data. *Loaded to a PivotTable.*
Tool 2- Tableau
Created Union of all the tables *Created Calulated Fields and few parameters for analysis.*
Tool 3- Jupyter Notebook
Created a DataFrame from cleaned data. *Calculated mean,std,25%,50%,75%,max of ride_time w.r.t. day_of_week

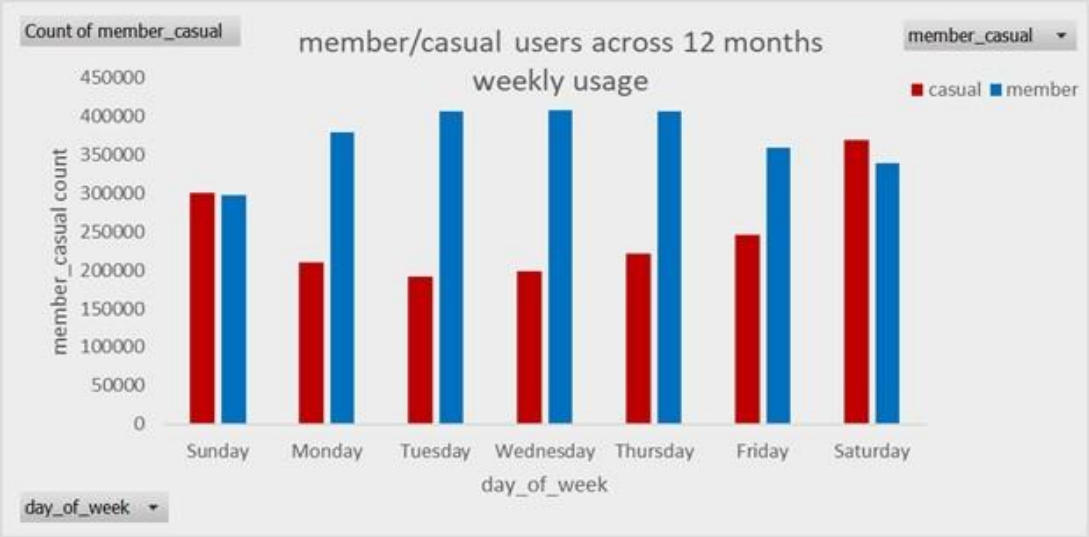
SHARE

Difference between the usage of bike type from casual riders and annual members?



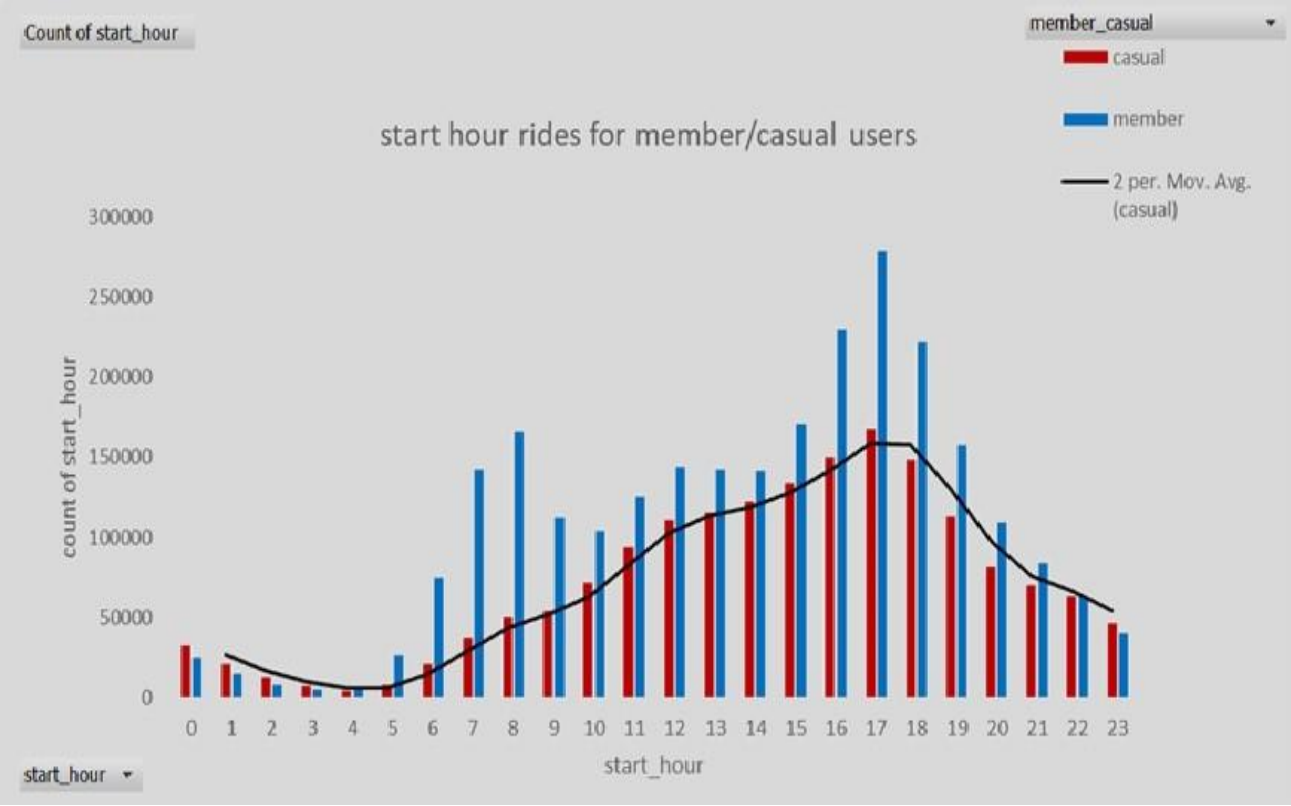
The casual users preferred classic bike over electric bikes.

Weekly usage analysis throughout 12 months time



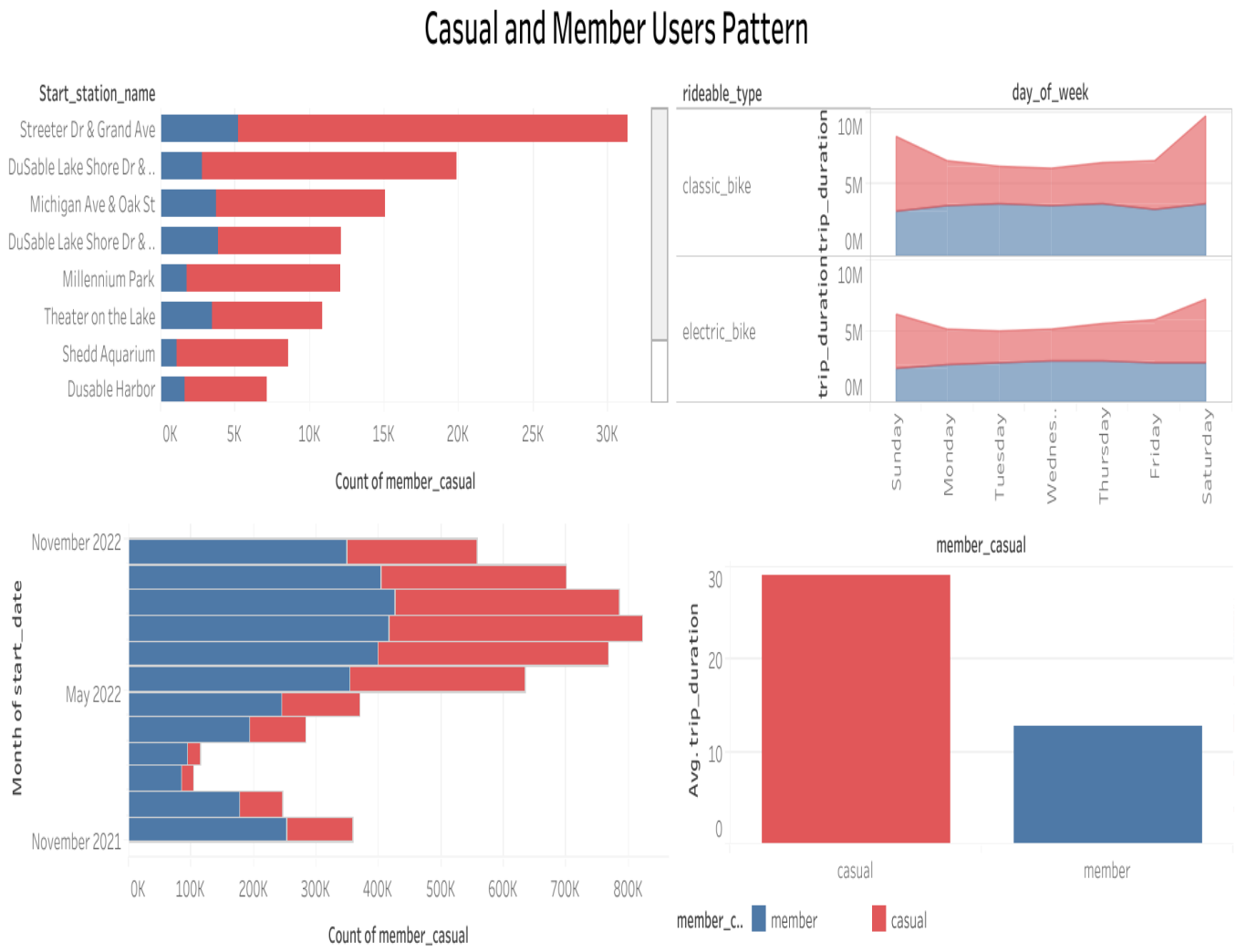
On Saturday and Sunday the usage of casual riders is observed to be more than other weekdays.

Start hour analysis



3P.M. to 6P.M. seemed to be a quiet spike in casual users.

Dashboard for Casual and Members Patterns



Interactive dashboard

ACT

Top three recommendations :
1- Generate some sort of marketing strategy which offers some perks or discount to users from the month of June till November as a spike in casual users is observed.
2-Provide promotional schemes at the top 30 stations where causal riders are using the bike most like Streeter Dr & Grand Ave,DuSable Lake Shore Dr & Monroe St,DuSable Lake Shore Dr & North Blvd,Michigan Ave & Oak St,Wells St & Concord Ln,etc
3-Offer a discount on membership on rides taking above 24 mins on Saturdays and Sundays as the casual rides average ride_time during weekends is 24 mins.
4-Membership flash sale at 3P.M. to 6P.M. since the demand peaks during those hours.