

Cyclistic bike-share analysis

Alan F. de Souza Almeida

Google data analytics - Capstone project

Structure of presentation

- 1 Scenario
- 2 About the data
- 3 Analysis
- 4 Conclusion

Scenario

- 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
- There are a good flexibility of its pricing plans, but they are divided in two classes: casual and member riders
- Cyclistic's finance analysts Moreno have concluded that members are much more profitable than casual riders. So the main goal is to maximize the number of annual members

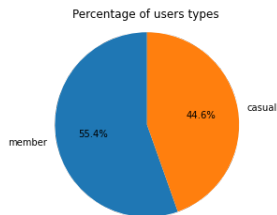
About the data

- Data was obtained from Cyclistic's historical trip and we used only the 12 months of 2021 ¹ data.
- Initially 5,595,093 rows of data was available containing attributes such as: ride-id, bike-type, datetime-ride, stations, gps coordinates of trip and type of user.
- I cleaned the data that had some type of inconsistency like missing information (in general that problem was in stations data) and duplicates in ride-id
- I created new columns such as the ride duration and about the day that the ride was made.
- Converted the data into the proper type.

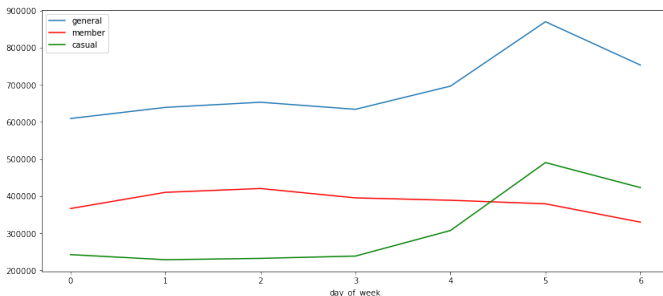
¹<https://divvy-tripdata.s3.amazonaws.com/index.html>

Analysis

- After the data clean stage we see some interesting statistics:
 - the most common day to ride was Saturday (as expected)
 - mean duration of a ride was 19 minutes
- The percentage distribution of users are:

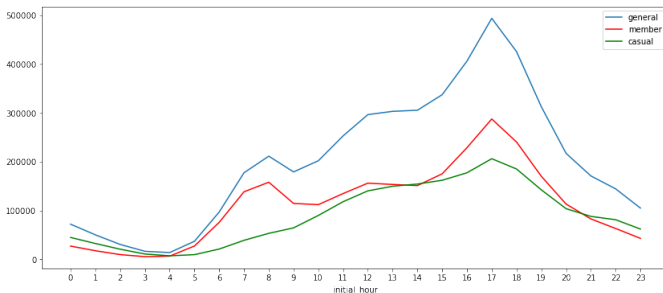


- In which we can see that the percentage of casual users is almost the same as members
- By exploring the statistics other interesting point is that the mean ride duration of casual users are 28 minutes and for member the number is 13 minutes. How can we explain this?
 - We can plot the number of rides per day (0 = monday and 6 = sunday) per category (general, casual and member)



- We can observe that the trips for casual rides grow continually from friday to sunday, which can be explained by the main use of bikes for them, that is for entertainment purposes. So, naturally, the mean value of ride duration is explained because people tends to ride longer in this kind of activity.
- As for member's users the use of bikes consists in day-to-day activities such as go to work and market which explains the constant use of bikes over the days.

- Other aspect that we can visualize is the quantity of bike rides per hour of day



- we can observe in the member's curve that there are two main points of maximum rides that is around 7 am and 17 pm that is usually when people go to work and then come home.
- as from casual members there is a crescent tendency from 5 am to 17 pm probably because that's the time people choose to ride on weekends and then come home.

Conclusion

- From this analysis we can describe both types of users as:
 - Casual: People that contract the services for entertainment purposes and usually rides on weekends.
 - Member: People who sign the annual services because the use of bicycles is constant on day-to-day activities such as go to work, household chores (go to market for example) or for entertainment
- How to maximize the number of members?
 - offering discount to casual riders whom have duration rides greater than a certain value (suggestion: mean value of rides duration of members)
 - propaganda about how use bicycles is healthier for our body and the environment (and maybe offer a voucher for whom watches the propaganda)