

## Scenario:

I am a data analyst working at Cyclistic, a bike-share company in Chicago. Cyclistic allows people to pick up bikes and dock them at various stations. Riders who have an annual subscription are called members while riders who are single-ride or full-day pass users are considered casual riders.

The director of marketing is looking to maximise the number of annual memberships as they are more profitable than single-ride or full-day passes. This strategy is believed to be the key to future growth.

Management believes that the best approach to do this is through converting casual riders into members as casual riders are already customers and aware of Cyclistic's services.

## Ask/Business Task:

To understand the differences in bike usage among casual and member users in order to recommend how to encourage casual riders to become members. Prepare/Understanding the Dataset

I will be using the public dataset located here. The data has been made available by Motivate International Inc. under this license. This is a public dataset therefore, there is no personal data and no way to know how often the same rider uses the biking service.

Across the various tables, the column names varied slightly however the key columns within the 2020-2021 dataset which are worth noting are the start and end stations, start and end ride times as well as the member and casual rider columns. The station latitude and longitude columns were also useful for data visualisation.

## Process:

I started by creating a data cleaning rundown script. Among other things I:

- Removed duplicate records with the use of unique ride ids
- Deleted rows with incomplete or unavailable data
- Deleted rows with inconsistent times where the start time was later than the end time
- Used the clean name function to make sure all columns were unique and had no title spacing
- Added a column for the duration of the trip in minutes and a column for the starting hour e.g. If the ride start time was '2020-07-22 15:38:23' the starting hour would be 15
- Added 2 weekday columns for the respective weekdays the ride started and ended on.

My main challenges were working with the date format. They had various date formats and I found working with dates to be the most challenging part of this case study as certain functions turned dates into string or number formats which were then not readable as dates.

I read up on different ways to analyse outliers from more familiar methods like standard deviation to other methods like DB Scan Clustering and Isolation Forest. I then explored standard deviation and box plot in Python.

Using the box plot, I found that the mean was 14 while the upper range was 53. To get rid of extreme outliers I used the Empirical rule and deleted data above or below three standard deviations of the mean. This meant that I deleted data that was more than 966 minutes in trip duration which is approximately 16 hours or 2 working days.

I saved the extreme outliers in another table and I was surprised to find that a large majority of the longer trips were done by casual riders. This means that members are more frequently taking shorter rides while casual riders take fewer rides which are longer in duration. After creating box plots and removing the extreme outliers I then had a better understanding of the data.

## What the Data Tells us so Far:

As expected, this data tells us that most people use the service for short trips that last around 14 mins however the data also suggests that there are people who use the bikes for long periods at a time. It may also be true that bikes are not being properly docked once the ride is over meaning that the actual end time isn't able to be recorded.

## Daily Use

The key difference in usage, shown below, is that casual riders use bikes most on weekends and least at the beginning of the week, whereas members use the bikes more evenly across the week. Members also use classic and electric bikes more often throughout the week whereas casual riders mostly use docked bikes.

## Hourly Use

Throughout the week, members and casual riders use the bikes most often around 5/6 pm on weekdays and 1/2 pm during the weekends. Casual riders begin trips fairly frequently in the early afternoons on weekdays while members start journeys quite often around 7/8 am on weekdays.

## Monthly Use

Overall bikes are used most often and for longer periods from July to September, casual riders use the service particularly more during Saturdays in August. As expected, there are fewer users in the winter months. It is also interesting to note that casual riders use the service a fair amount between March and June with a slight dip in May. Electric bikes are fairly popular during autumn for casual riders while classic bikes are almost only used by casual riders during March and February.

## Business Task Questions:

- How do annual members and casual riders use Cyclistic bikes differently?
- Why would casual riders buy Cyclistic annual memberships?
- How can Cyclistic use digital media to influence casual riders to become members?

## To Directly Encourage Casual Riders to Become Members:

- Introduce late fees for full-day passes and cap the length of time a person can rent a bike with a single-ride pass. A lot of riders are using the service less often but for longer amounts

of time. In the most extreme cases, they are using it for months or years with one pass. Keep flexible access exclusively for members.

- Introduce priority access for members during peak times.
- Increase the price of single-ride and full-day passes:
  - competitively and across the board
  - specifically, during peak hours e.g., during the summer and weekends.
  - specifically at peak stations e.g., central and north Chicago.
- Consider introducing a monthly or weekly pass as casual riders might not want to be tied down to an annual subscription. The monthly pass could encourage more casual users to become members because the more they use the monthly pass, the more they may feel more comfortable getting an annual subscription. Price the monthly pass so that 12 individual months is more than an annual pass.
- Introduce an off-peak annual membership which is cheaper than the regular annual membership. This would allow a person to only use the pass during off-peak times, this could be the middle ground for people who don't want to commit to the full membership.
- Paid advertisements should be concentrated in the most popular start stations which are in Central and North Chicago.
- Market the membership in summer and on weekends when casual riders use the service most and introduce a promotional annual membership deal.
- Explain in the marketing how the annual membership makes financial sense e.g. if membership is \$50 and each full-day ride is \$5, a person will only get 10 rides throughout the year before they are paying more than the membership cost.

## Other Ideas and Considerations:

To encourage casual riders all year round, introduce a cheaper off-peak single-ride or full-day pass during winter and advertise a promotional pass for casual riders in May as there is a dip in usage during this month and on weekdays before 12 pm or after 7 pm.

With further information we could determine if the same person or a group of people are riding together, if the latter is true then offering a discounted annual membership for friends, students and families could encourage customers to get memberships in groups.