

DATA ANALYSIS ON CONSUMER BEHAVIOUR FOR A BIKE RENTAL BUSINESS USING R

1. Introduction

The following project analyses consumer behaviour for a bicycle rental business. The analysis focuses on identifying differentiated consumer trends, preferred product categories and general consumer behaviour, by using a large scale dataset of almost 6 million real observations. This type of analysis can be particularly useful for the Marketing Analytics team, to inform of beneficial approaches in marketing strategy, and to executives, to ultimately improve the service to clients and increase revenue for the company.

2. Data Sources

Source of data

The data used is public data from “divvy-tripdata”, where service use is updated monthly. It has been made available by *Motivate International Inc.* under the following license: <https://ride.divvybikes.com/data-license-agreement>

Quality and credibility of data

This data is reliable and original because it is first party in origin. It is comprehensive, including detailed information for each ride taken. For the extracted data we analyse 12 months, from June 2021 to June 2022. It is also cited.

There are no problems with bias in the observations, as we are taking the entire population of customers instead of a sample.

Because of these reasons, our data is credible and can be safely used for robust analysis.

Structure and format of the data

Data is structural and obtained as comma-separated files. Organized by observations, it includes both quantitative and qualitative information.

3. Data Processing

Data integrity

We first ensure data safety through replication, by first storing the original data set separately.

Data integrity is confirmed. Coming from an internal source we know it is accurate, consistent and can be trusted throughout the lifecycle. And data manipulation has been carried with high precaution to prevent errors and thus not compromise efficiency.

Tools used

Our main tool for analysis will be the programming language R, which will be used for loading, transforming, analysis and visualizations. RStudio or RGui are easily obtainable open-source options to run the code scripts.

This entire project has been designed to only require R from start to finish. However, we could also use Excel or a text formatting software for an early, preliminary observation of the .csv data just to know what we're working with.

Data Cleaning

First, to avoid having to check the datasets using Excel by converting the .csv files into structured rows and columns, we can directly use commands like 'readr' in our R software, loading every file one by one.

We check our data using functions like *glimpse*, *head* or *summary*. We see the variables we have, their types and start thinking of possible errors or difficulties we might have to fix with data cleaning.

To proceed we combine the 12 datasets into a single dataset, named 'trips'. In it we create some new variables that will be useful to gain business insights. We create the variable "weekday" which tells the day of the week for every ride. Similarly, we create a variable "month" for the month in which each trip was performed. We also create the variable 'duration' of every trip (which is the end time minus the start time).

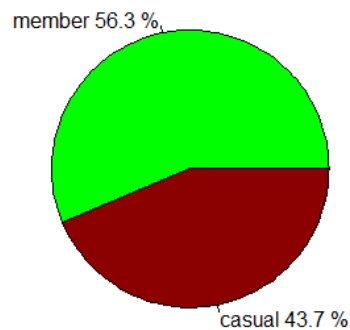
After creating the 'duration' variable, an error in the data became more evident: some trips ended at earlier time than they started, giving us a negative trip duration. A total of 139 observations had a negative time. Furthermore, some other 507 observations had a duration of 0 seconds. We calculate that these impossible observations amount to approximately 0.011% of the total dataset, which is a reasonable margin of error in business coming from technical issues at measuring the input, false positives, etc. So, we conclude these are in fact just errors. Finally, we take these observations out of our dataset to prevent them from contaminating the final results of our analysis, by creating a new dataset of only positive duration trips, 'ptrips'.

4. Data Analysis and Visualizations

Total number of trips by client type

Valid observations sum to a total of 5860776 bike trips. Most trips are done by clients with memberships, which amount to 3300919 trips (56,32% of the total) while casual clients amounted to 2559857 trips (43,68% of total).

Trips by client type



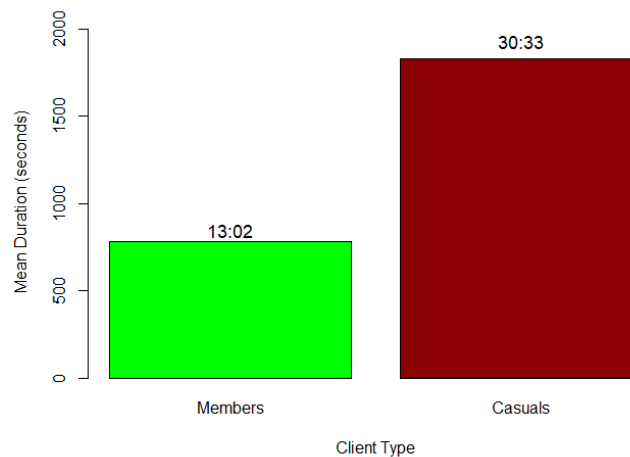
Average duration of bike trips

The average duration of the trips is 20 minutes and 42 seconds.

But this result changes depending on the type of client. Members have shorter trips on average, around 13 minutes 2 seconds, while casual riders have trips around 30 minutes and 33 seconds.

```
[1] "Table 1: Client types comparison"
      Total trips Mean duration (m:s)
member 3300629    13:02
casual 2559501    30:33
```

Mean trip duration by client type

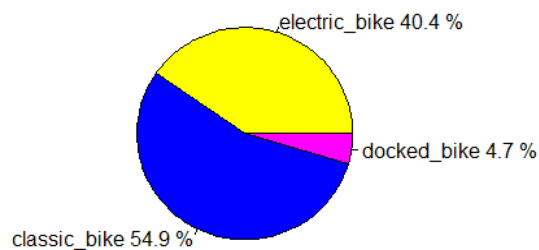


Total number of trips by bike type

We observe that the most used type of bike in the company is the classic bike, with 54,9% of all total trips corresponding to it. The electric bike also represents a sizeable share of rentals, with 40,4%. Docked bikes only represent 4,7% of the services provided by the business.

```
[1] "Table 2: Bike service comparison"
      Total trips Mean duration (m:s)
electric_bike 2368211      15:05
classic_bike  3217479      19:29
docked_bike   274440       23:07
```

Trips by bike type



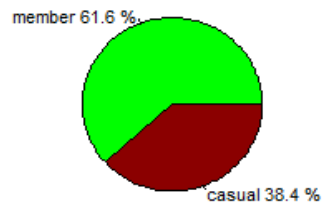
Bike type usage

We can observe a clear difference among the client type usage depending on the type of bicycle rented.

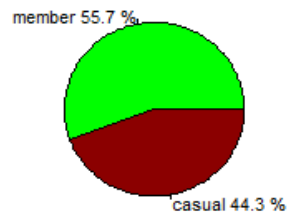
Member clients dominate in classic bike trips (61,6%) and in electric bike trips (55,7%). They are not present however in Docked bike usage, as no membership service is offered for this category.

```
[1] "Table 3: Number of trips by client type and bike type"
      electric_bike classic_bike docked_bike
member      1981071      1319558           0
casual       1236408      1048653      274440
```

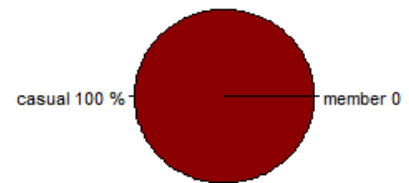
Composition of Classic bike trips



Composition of Electric bike trips



Composition of Docked bike trips

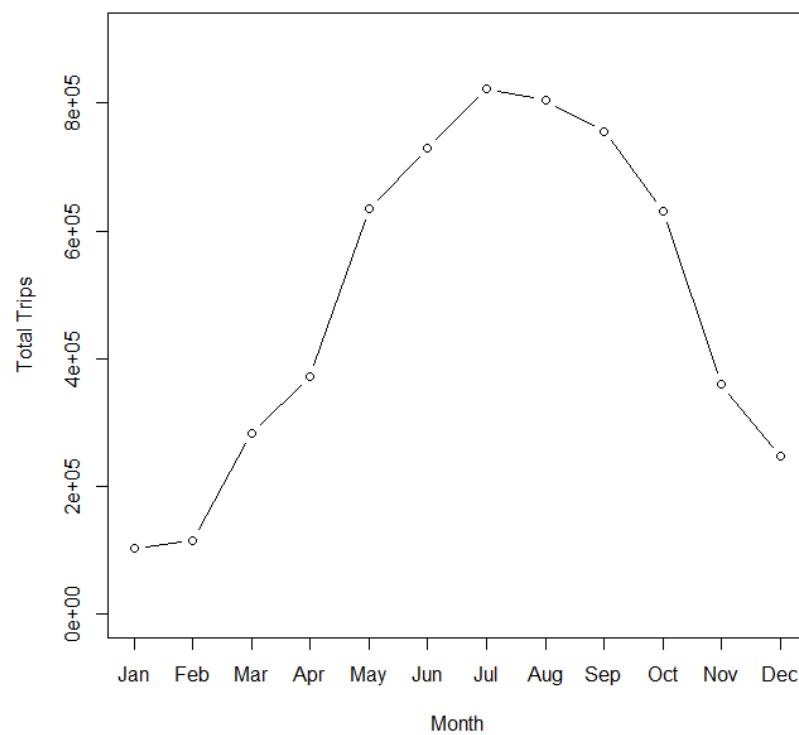


Rental trips across the year

By looking at time data, we can notice a strong degree of stationarity.

Bike rentals spike at and around summer, while being at their lowest during January and February.

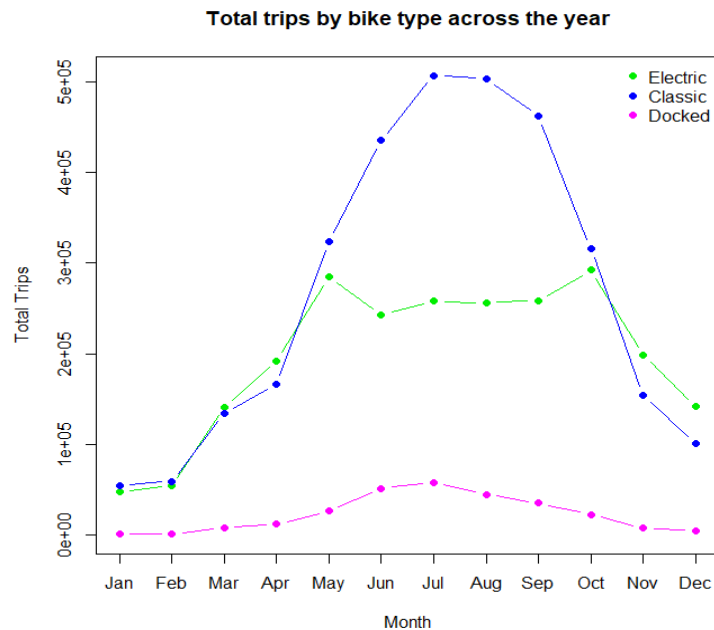
Total trips across the year



Rental trips across the year

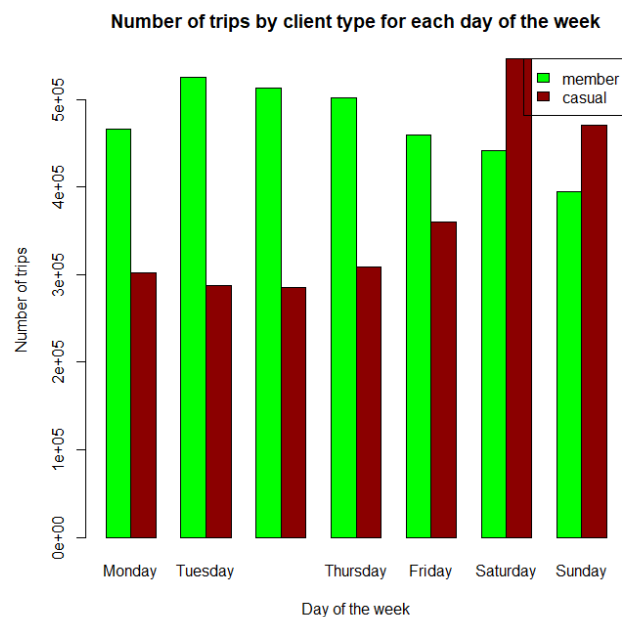
We can separate this trend by bike category to learn about the differences in stationarity for the different company's services.

By doing this we'll notice a curious and sudden roof in the usage of electric bikes around summer.



Type of rider and day of the week

A revealing result is that members represent the majority of clients the majority of the week, particularly on weekdays. But on weekends, the dominance is reversed as casual riders tend to be majority.



```
[1] "Table 4: Bike trips by day of the week"
      Sunday Monday Tuesday Wednesday Thursday Friday Saturday
Total trips 864748 768128 811755 798297 810362 819735 987105
```

The mean duration of trips is also longer on weekends. This is consistent with the dominance of casual clients as was reported previously.

```
[1] "Table 5: Bike trips - weekdays vs weekends"
      mean_duration
workdays 1108.483 secs
weekends 1529.464 secs
[1] "Bike trips during weekdays are on average 37.98 % longer than on weekends."
```

```
[1] "Table 6: Client type by day of the week"
      Sunday Monday Tuesday Wednesday Thursday Friday Saturday
member 394650 466086 524769 512565 501778 459766 441015
casual 470098 302042 286986 285732 308584 359969 546090
```

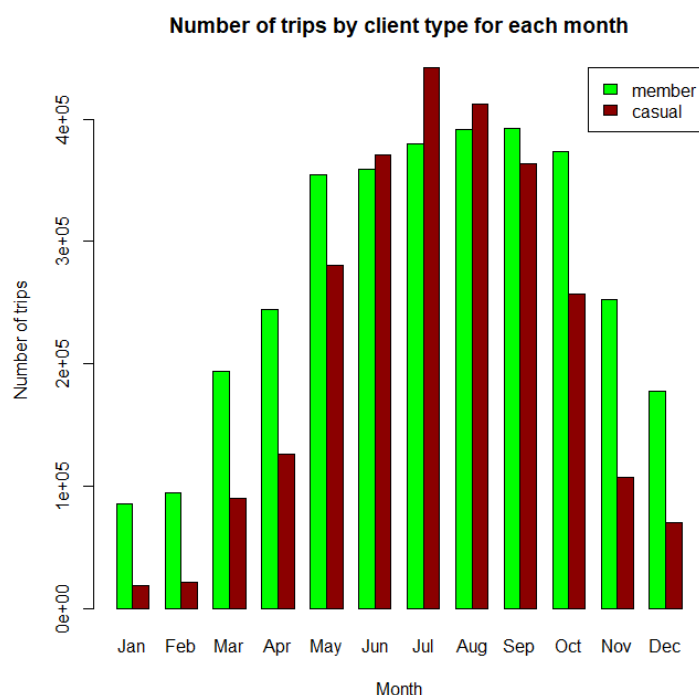
Type of rider and month of the year

Trips increase significantly in summer and drop notably on winters for both types of clients. However, members show a greater stability across the year.

It is only in the months of June, July and August that casual trips surpass in number the membership trips. This trend is consistent with holidays.

```
[1] "Table 7: Bike trips by month"
      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Total trips 103765 115604 284024 371218 634810 729529 822328 804245 756040 631156 359892 247519
```

```
[1] "Table 8: Client type by month"
      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
member 85248 94190 194150 244820 354423 358893 380317 391637 392200 373953 253008 177790
casual 18517 21414 89874 126398 280387 370636 442011 412608 363840 257203 106884 69729
```



5. Sharing findings

In conclusion, we have looked at a year's worth of data, and with very little useful variables but a large number of observations, we identified crucial trends on how clients make use of the rental services.

We now know how clients with a membership differ from casual clients. They are more stable across the year and are prominent on most months except around holidays. Memberships are also prominent on weekdays but this dominance is reversed on weekends, where casual become a majority.

This has consequences given the average duration of trips depending on the type of client. Casual clients have trips which are one average more than twice as long as those from members.

From this we can see that memberships are correlated with work and study time both across the year and during the week, and are more likely used, proportionally, as a form of commuting. Casual riders are the majority on weekends and holidays, while also taking longer trips, as these are more likely related to leisure activities. Memberships also make use of a slightly larger share of classic bikes than electric bikes, which hints at a preference for light cardio exercise (13 minutes on average) during their commuting, while for casual riders the preference of classic over electric exists but it's not as strong.

We can not ignore the importance seasonal tourism has on the casual service. A potential line of action would be to consider a partnership with tourism-based businesses, such as local hotels or key touristic destinations, by offering special discounts on the casual service to their visitors during holidays. Thus, taking advantage of the consumer preferences we detected to attract more clients to the bike renting service.

The marketing team is also advised to start a campaign specifically promoting its classic bikes as a form of healthy lifestyle, locating the advertisement around the main office areas across the city, promoting light cardio exercise during commuting would attract this client segment we have detected.

In the same line to promote a healthy lifestyle, partnerships with gyms or other fitness institutions can be done to increase the share of clients using docked bikes. These represent only 4,7% of the service provided and there is currently no membership option for them. Since members do have a stronger tendency for classic bikes over electric bikes, the preference for exercising can open a new business opportunity for docked bike memberships.