Case Study 1 - How Does a Bike-Share Navigate Speedy Success?

Serge Geukjian

9/10/2021

ASK

- BUSINESS TASK: How do annual members and casual riders use Cyclistic bikes differently?
- Key Stakeholders: Lilly Moreno, Marketing & analytics team, and Executive Team

The main goal to reach here is not to add new customers to Cyclistic, but to maximize the annual members since the annual membership is much more profitable to the company and those members will be key for the future growth of the company. So if the main uses for both non-annual and annual members are recognized and studied in dept then the company can focus on how to give a good deal to the non-annual ones and attract them to switch to annual memberships.

PREPARE

Notes:

– The data Selected in this study is for the the years 2018 and 2019

– The data was uploaded to SQLite, and a database was created

– Then the data was imported from SQLite to PowerBi for data cleaning and Analysis

- The data is located in spreadsheets, on this website either each month, quarter, or year alone depending on the size of the data and the year, which ranges from 2013 till 2021 up to date.
- The data is organized in a tabular way containing all the necessary information for the trips of the customers, mentioning the duration of each trip, the start and end locations (stations) and the type of the customer as whether he/she is a one time user or an annual subscriber. Hence the data is organized in a long structure.
- The data is collected directly by the company itself who is issuing the service and the bikes, so the data insures that it's reliable, and from the data provided it seems that it is ROCCC.
- The provider company has clear and detailed licence and terms to use for the data, so it's the right thing to acknowledge them and give them full credit with references for the data used.
- Even though there are some changes in the columns of the data over the years the main columns remain consistent and those are the columns that provide the most crucial information like the stops, the type of the user and the duration driven.

- The whole data validation process helps us verify that the data is accurate, safe and consistent, so the analysis can have a proper foundation to be built on, so this will lead the analyst to accurately track the trend and the movement of the rides.
- After some macro-level study of the data we see that it contains some empty (null) values in the "gender" and the "date of birth" columns, also there are some date of births which are very old (in the 1700s-1800s) which are supposed to be removed, so we filtered them out and put the oldest date of birth to be 1919 which is 100 years back from the data being used.

PROCESS

SQL

- SQLite was used here in order to import all the data into a database and then connect that database to PowerBi
- The data was modified in SQL so they all have the same column names, and connected to PowerBi

PowerBi

- Changed all data types of the columns so they are formatted properly. This change is done in the back end M language code of PowerBi and the code looks like this.

```
Table.TransformColumnTypes(#"Removed Duplicates",{{"BIKE_RIDE_ID",
Int64.Type}, {"START_DATE", type datetime}, {"END_DATE", type datetime},
{"BIKE_TYPE_ID", Int64.Type}, {"DURATION_sec", Int64.Type},
{"START_STATION_ID", Int64.Type}, {"START_STATION_NAME", type text},
{"END_STATION_ID", Int64.Type}, {"END_STATION_NAME", type text},
{"USER_TYPE", type text}, {"GENDER", type text}, {"MEMBER_BIRTHDAY",
Int64.Type}})
```

- After all the columns in all 2018-2019 data have the same column names and data types we appended the tables into one master data table.

```
Table.Combine({#"2018_Q1_Table", #"2018_Q2", #"2018_Q3", #"2018_Q4",
#"2019_Q1", #"2019_Q2", #"2019_Q3", #"2019_Q4"})
```

- Checked and removed duplicate data.
- Then created a custom column in our table that shows us the duration in minutes instead of seconds

```
Table.AddColumn(#"Changed Type", "DURATION_min", each [DURATION_sec]/60)
```

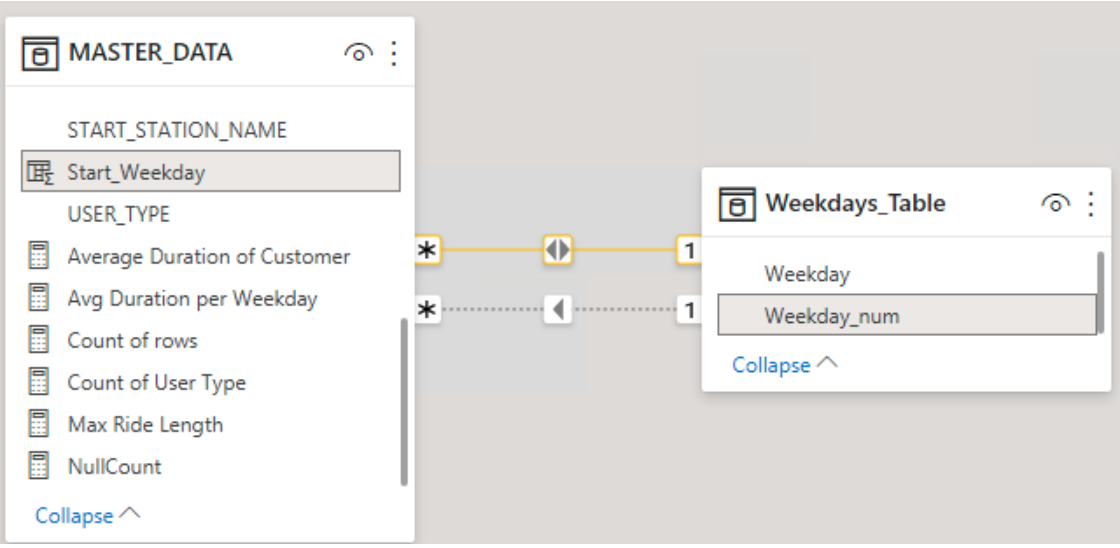
This way we can have smaller numbers on our graphs and it will look much more clean and neat. * Filtered out the member birthdays that are before 1919

```
Table.SelectRows(#"Replaced Value1", each [MEMBER_BIRTHDAY] > 1919)
```

- Created a custom table that contains the weekday number and day

```
Table.RenameColumns(#"Changed Type",{{"Column1", "Weekday_num"}, {"Column2",
"Weekday"}})
```

- Using the PowerBi Tools we added the starting weekday number to the master data table and connected that to the “Weekday” Table so that in our graphs we can display the name of the weekday instead of only the numbers



Analyze

- The Data is stored in a long tabular form with over 7 Million rows of data, and all columns have their proper data type formats.
- Calculated the average duration per weekday, the maximum ride length, and the count of the users per weekday and put it in a matrix table that splits it into customers(one time users) and subscribers separately

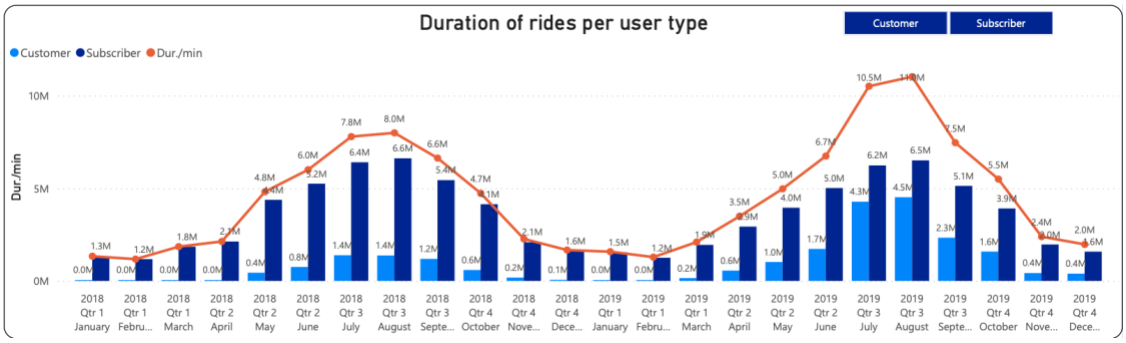
```
Avg Duration per Weekday = AVERAGEX(MASTER_DATA,MASTER_DATA[DURATION_min])
Max Ride Length = MAX(MASTER_DATA[DURATION_min])
Count of User Type = COUNTA(MASTER_DATA[USER_TYPE])
```

- The peak period where the duration of rides is the most is during Q3 mostly in July and August where both the subscribers and the one time users use the company’s services more with the subscribers being more than the one time customers. Also on a micro level the average duration of the rides are longer during the weekends but during the weekdays the total duration of the rides are more than the weekends, this means that more people tend to drive during the week for a shorter duration but on the weekends the people who do drive they drive for a longer duration.
- These separations of the graphs and the numbers will help us know when the count and the average riding duration is more for the subscribers and the one time users.

Share

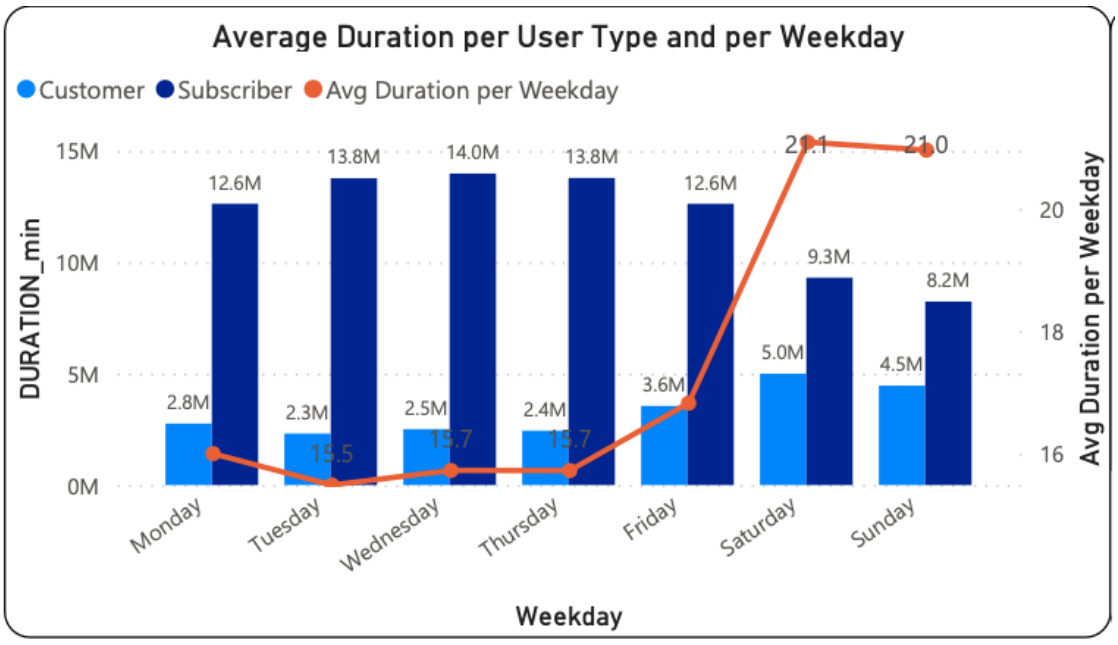
- The main story here is that subscribers mostly use the ride for their daily transportation needs like to get to their jobs, to visit relatives,to go to school/university, or to buy groceries for example, whereas on the other hand the customers may be tourists or maybe teenagers for fun activities.

- Our main audience here are the subscribers since the main goal of the company is to convert one time users to annual subscribers, best way to get to them is through their contact information that they provide during registration, and to understand maybe through a survey the reasons on for what they use the rides on a daily basis.
- The company having more than 7 million rows of data, data visualization is key here to see and interpret the trends and the timings of the peaks of the rides per user type.
- The visualizations are on an open source PowerBi dashboard that can be easily accessible for anyone.
- Below is the visualization that shows us the trend over the two years of 2018-2019 of the duration(min) ridden by the customers and subscribers:



Here it clearly shows that Summertime, specially during July/August, is when the company has the most users specially subscribers, it also shows that there has been a significant increase in the duration of the customer from 2018 to 2019 with more than double the numbers.

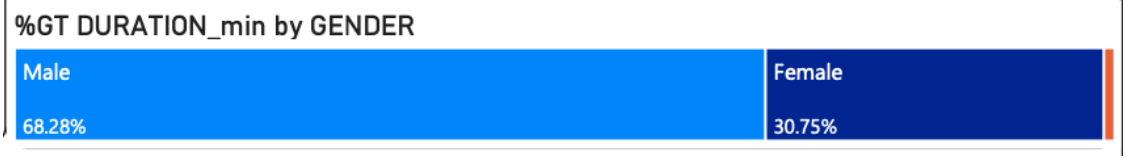
- Next i visualized the average duration ridden over the 2 years throughout the weekdays:



The average duration length ridden is most during the weekend where it starts increasing on Friday and peaks on Saturday and Sunday, we also notice that the avg duration of subscribers are increasing during the weekend unlike the customers where their highest numbers are on the weekend where they double the average duration that they usually drive during the weekdays.

- Finally i did a small matrix table where we see some key statistical measures of the customers and subscribers:

User Type	Customer				Subscriber			
Weekday	Avg Dur	Dur/Min	Count per User	Max Ride Length	Avg Dur	Dur/Min	Count per User	Max Ride Length
Monday	49.05	2,757,388.32	56214	93,462	13.95	12,628,332.08	905245	136,727
Tuesday	47.16	2,311,762.58	49023	123,677	13.92	13,773,129.50	989317	150,944
Wednesday	49.19	2,506,221.53	50949	110,735	14.02	13,983,804.70	997417	93,813
Thursday	44.22	2,439,233.28	55159	127,791	14.12	13,786,043.73	976532	225,960
Friday	56.28	3,551,583.78	63105	132,324	14.06	12,627,054.03	898297	120,361
Saturday	45.96	4,996,800.60	108730	119,555	16.34	9,309,659.68	569678	98,548
Sunday	49.48	4,460,675.25	90150	110,498	15.97	8,235,321.47	515606	116,369
Total	**48.64**	**23,023,665.35**	**473330**	**132,324**	**14.41**	**84,343,345.20**	**5852092**	**225,960**



Here we see the average duration in numbers where the average of customers are more than double than the subscribers, this means that even though the customers are much less than the subscribers both in count and in duration driven they tend to drive longer a duration per ride. We also noticed that the male users are more than two times more than

the female users. (*caution here that 1% of our data contained users that had not identified there gender*)

- *To access the full report click* here

Act

- The company clearly is on the rise throughout the years in the duration of the rides especially by the customers. The subscribers' duration shows little change. Customers are using the bikes more during the summer and specifically during the weekends with longer rides, comparing to the subscribers, so the customers are the one time users who are tending to use the rides for more adventurous and leisure purposes comparing to the subscribers who are using the rides for the everyday rides.
- The stakeholders can use this data to offer the customers some deals for the summertime and weekends in order to attract more customers to switch to annual subscriptions.

Top 3 recommendations based on analysis

- Make a new quarterly subscription deal with a very affordable price that is for the lowest duration driven months, from November till February, so this way the duration driven in these months wont decline at a very rapid pace after the summer.
- Introduce a new weekend-only annual subscription with a slightly lower price comparing to the annual subscription, this way people will be encouraged to take either one of the two which will result in more annual users.
- Make a digital wallet campaign where it aims to give credits to users when driven more than a specific duration during the weekdays, and customers can later use these credits to either renew their membership or to gift them to a new user for a one time ride