Cyclistic Customer Conversion Analysis.

A Report By Daniel Gaza;

1. Introduction

1.1. Background

This report covers the certificate program's capstone project in Google Data Analytic. In March 2021, Google introduced the Data Analytics Professional Certificate. This was timed perfectly as there was a shortage of data analysts in the industry due to the supply and demand for analytics roles. The Ask, Prepare, Process, Analyze, Share, and Act phases of the data analytic process were heavily highlighted in the training. A capstone project is necessary to be taken once the Google Data Analytic certificate program has been successfully completed in order to create a portfolio, demonstrate the skills and knowledge that have been acquired. This capstone is based on a fictitious bike-share firm named Cyclistic, which was established in 2016, With the expansion of the program, the fleet now numbers 5,824 bicycles that are geo-tracked and locked into a system of 692 stations spread throughout Chicago.

1.2. Purpose of Analysis

Converting our customers (one-time hirers) to subscribers (annual member). Coming up with steps how our company can turn our customers to subscribers.

1.3. Statement of problem

Cyclistic is trying to figure out a way to convince casual users (customers) to sign up for yearly memberships (Subscribers). I have been given the responsibility by the manager to conduct analysis as a junior data analyst at Cyclistic in order to suggest actions to achieve this objective. A list of questions to ask in order to uncover potential answers. Using the five whys method, you ask "why" five times to identify the underlying cause. The fifth and final response ought to provide you with some insightful and occasionally unexpected information.

1.What are the differences in how annual members and casual riders use bicycles?
a. how clients ride their bikes on different days of the week

b. how long they spend on trips

c. How customers hire based on their age (particularly casual riders).

d. Does a customer's age influence their decision?

2. Why would a non-cyclist purchase a cyclistic annual membership

2. Data

As was already mentioned, the data for this assignment came from a fictitious business that was developed just for the capstone project. This data set includes 2019 hires. It is additionally available through Kaggle.com.

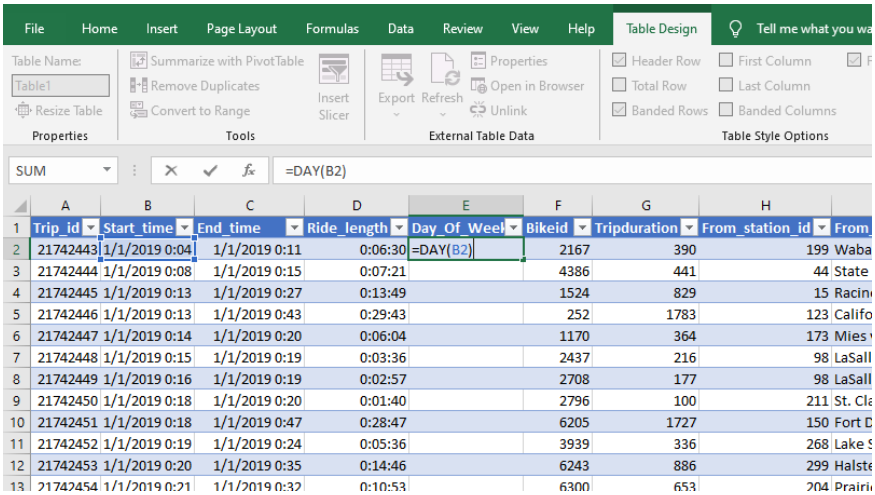
2.1. Data Cleaning

The data was already well-organized thanks to the data validation process. To further tidy up the data, consistency and spelling checks were made. Before conducting analysis, confirm consistency using Excel and SQL Server.

2.2. Data Wrangling

Originally the date had twelve(12) columns, three columns is added to make the analysis process a little better,

i. Ride_length: To determine the time each hire lasted, using the Start_time and End_time to result the results.
ii. Day_Of_Week: To determine the day each hire was made, using Start_time and the Weekday function in Excel.
iii. Age: To determine the age of customers using their birthyear



	A	B	C	D	E	F	G	H	
1	Trip_id	Start_time	End_time	Ride_length	Day_Of_Week	Bikeid	Tripduration	From_station_id	From
2	21742443	1/1/2019 0:04	1/1/2019 0:11	0:06:30	=DAY(B2)	2167	390	199	Waba
3	21742444	1/1/2019 0:08	1/1/2019 0:15	0:07:21		4386	441	44	State
4	21742445	1/1/2019 0:13	1/1/2019 0:27	0:13:49		1524	829	15	Racin
5	21742446	1/1/2019 0:13	1/1/2019 0:43	0:29:43		252	1783	123	Califo
6	21742447	1/1/2019 0:14	1/1/2019 0:20	0:06:04		1170	364	173	Mies
7	21742448	1/1/2019 0:15	1/1/2019 0:19	0:03:36		2437	216	98	LaSall
8	21742449	1/1/2019 0:16	1/1/2019 0:19	0:02:57		2708	177	98	LaSall
9	21742450	1/1/2019 0:18	1/1/2019 0:20	0:01:40		2796	100	211	St. Cla
10	21742451	1/1/2019 0:18	1/1/2019 0:47	0:28:47		6205	1727	150	Fort D
11	21742452	1/1/2019 0:19	1/1/2019 0:24	0:05:36		3939	336	268	Lake S
12	21742453	1/1/2019 0:20	1/1/2019 0:35	0:14:46		6243	886	299	Halste
13	21742454	1/1/2019 0:21	1/1/2019 0:32	0:10:53		6300	653	204	Prairi

Fig.2.1. Using Excel to create Day_Of_Week column, similar steps were used in creating the listed columns above.

3.Exploratory Data Analysis

3.1. Total Hires in 2019.

Assessing the number of hires in 2019 shows us we had a total number of 2.99 million hires,

Fig. 3.1

525,830 are customers and 2.46 million are subscribers, remembering the goal is to convert the one time hirers(customers) to subscribers.

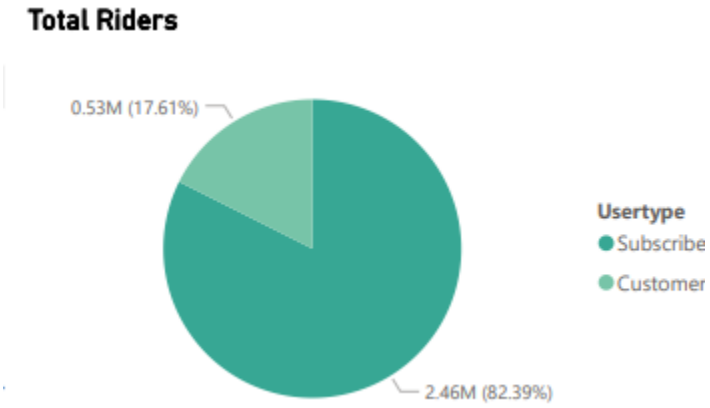


Fig.3.2

3.2. How Subscribers and Customers use our Bikes.

Analyzing the data in-depth to identify potential solutions to the issue and provide an answer to the question, What are the differences between annual members (subscribers) and casual riders (customers) in their use of bicycles?

A number of columns, including Day of Week, Age, and Months, are used to generate the results for this question in order to determine whether there is a substantial difference between the two groups.

3.2.1. Day_of_Week according to hires for customers

Our company operates seven days a week, looking into the days of the week to see if there are any notable distinctions that may be investigated or exploited in achieving our aim. recalling customer hires and arranging them by weekdays using a SQL query (Fig. 3.3)

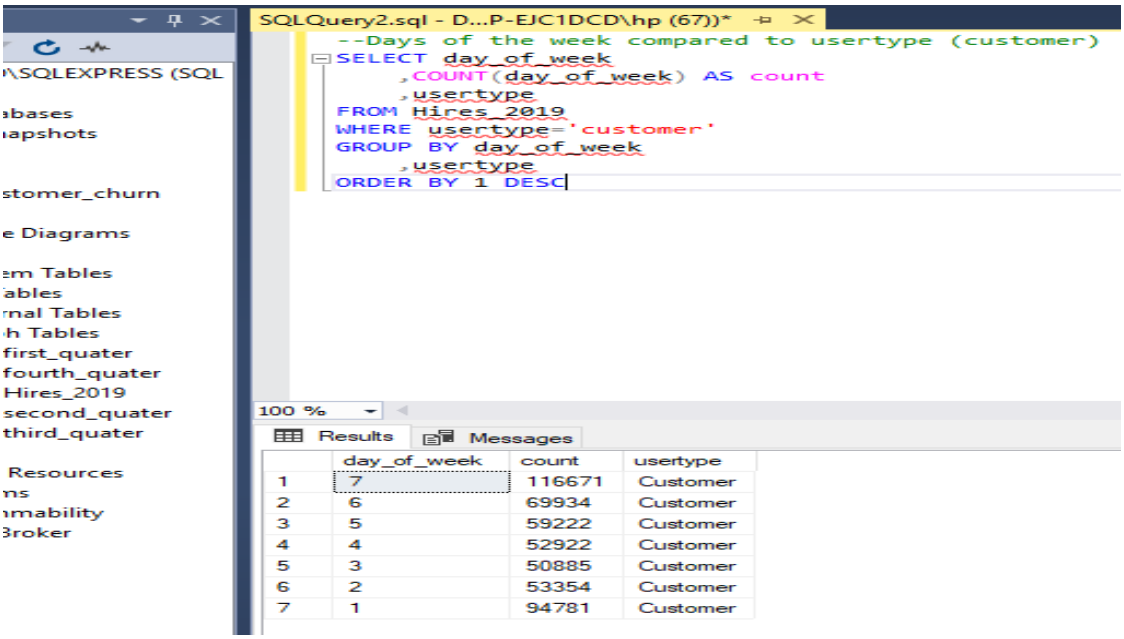


Fig. 3.3 using SQL Server to result hires by customers.

The remaining digits correspond to the remaining days of the week, with 1 denoting Sunday and 7 denoting Saturday in the days of the week column. According to our findings, clients hire more frequently on the weekends, with Sunday, Friday, and Saturday recording the highest numbers.

3.2.2. Day_of_Week according to hires for subscribers

We know from the study above that our customers hire more frequently on the weekends, so we want to find out if this is also true with our subscribers. The results below were created using a SQL query. (fig. 3.4.

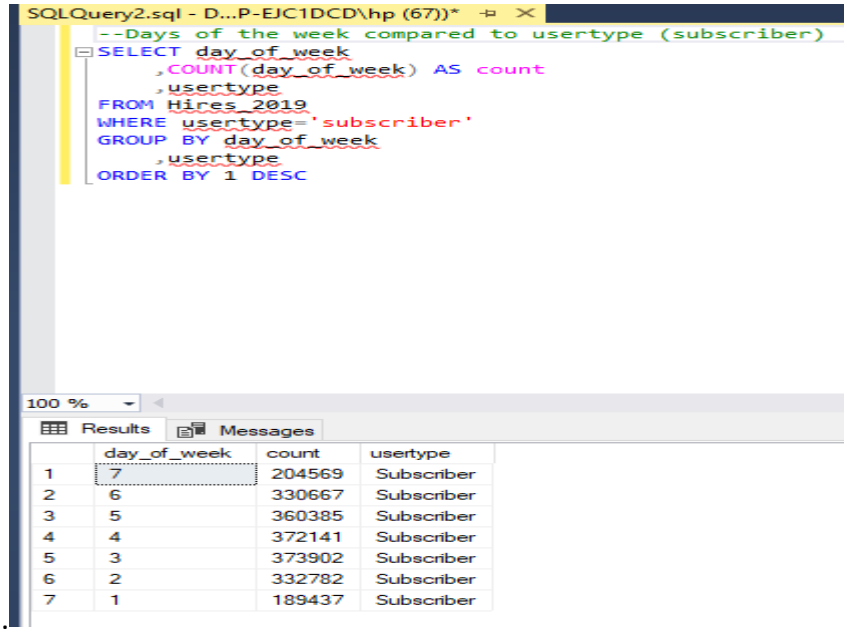


Fig. 3.4.

According to the inquiry and results above, subscribers hire more workers during the week than do our clients.

I imported the data into Power Bi where I will be constructing a dashboard to show the entire analysis results for easy interpretation. The statistics appear a bit complicated and pretty hard to understand, so to be able to detect the discrepancies without too much work. Here, I compared the days of the week to the hires made, categorized by usertype, using a line chart (subscribers & customers) Fig.3.5.

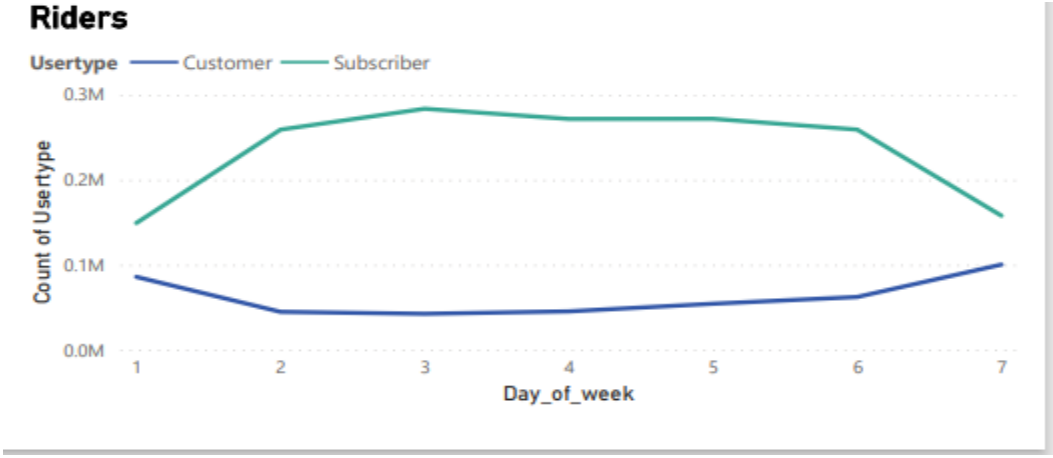


Fig.3.5. Hires according to weekdays made by customers and subscribers

Above we can easily note what our query results is showing us for customers we can see, how hires drops from Sunday to Monday and picks up around Friday although to Saturday. Subscribers, Goes up on Monday and drops on Saturday.

3.3. How customers hire based on their age

Based on analysis of the days of the week, we can see how different hirers use our bikes. However, at this point, we are unsure of what factors our hirers consider when choosing how to use our bikes. With the data at our disposal, we plan to dig even deeper to determine whether age influences the hirers' choice of use.

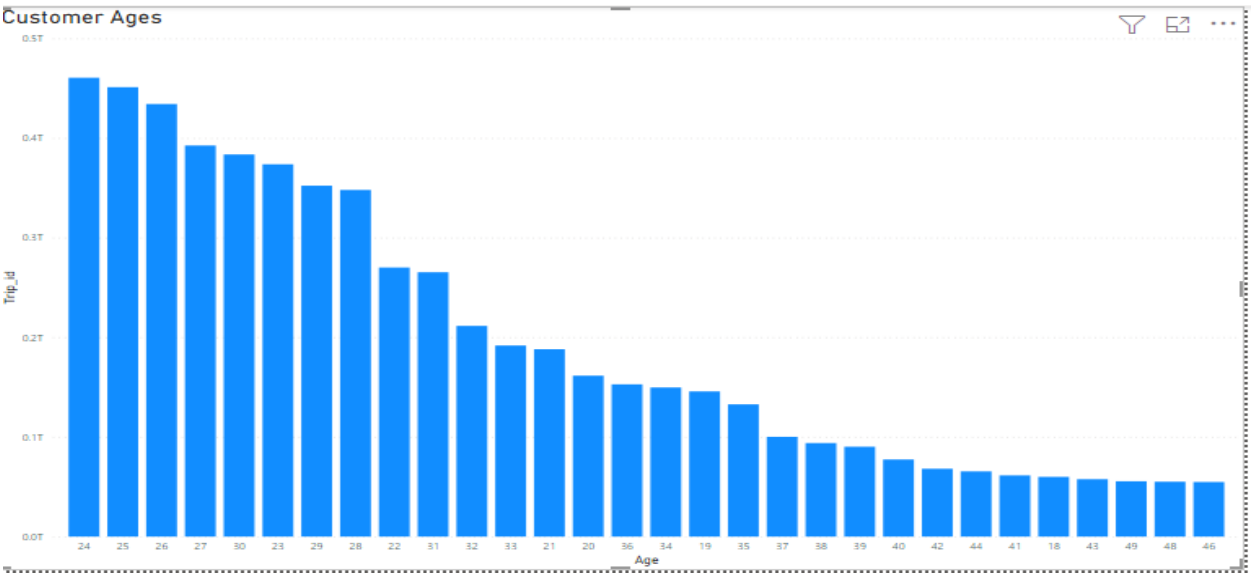


Fig.3.6. A chart showing the ages of customers.

From the chart(fig.3.6.) we can see our customer fall under Minnellas and Gen Z

3.4. Station Analysis

This analysis's objective is to determine how we could persuade more of our clients to sign up for subscription services. We will also probe our stations deeper to determine which ones attracted the most people. In 2019

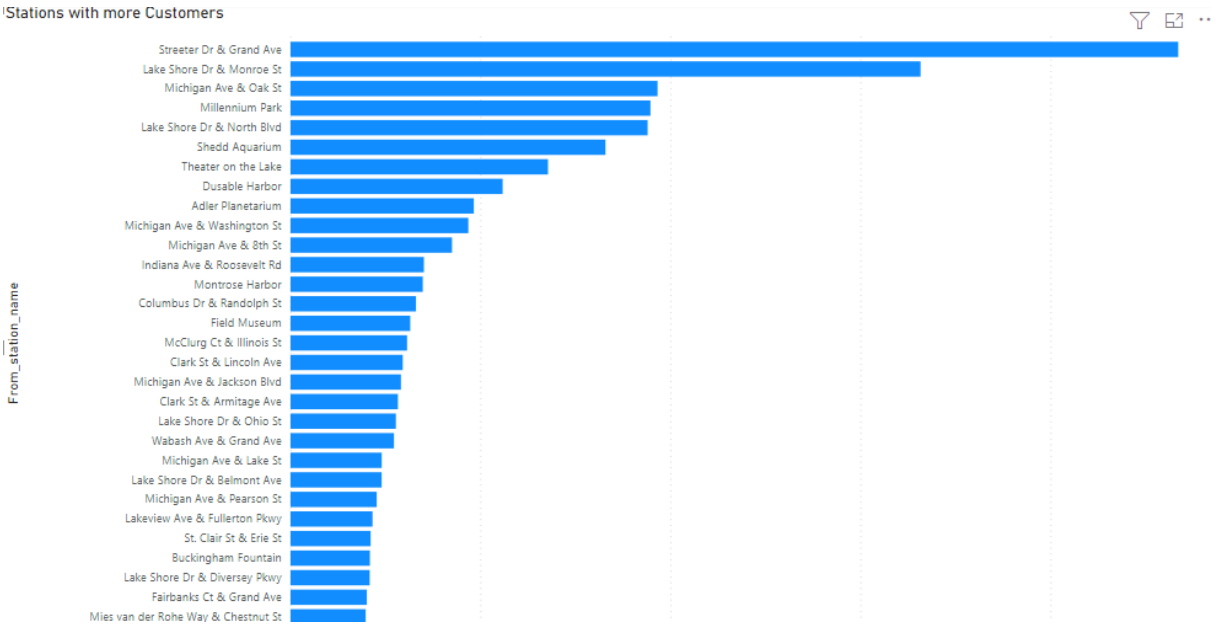


Fig.3.7. Stations with more customers hires.

The above bar chart shows us the stations with more hires in 2019, these are the stations which our target will be most focused on.

3.5. Hiring made in Months

Using the 5 Whys concept, we formulated questions that we would use to analyze our available dataset to better understand how our customers behave in order to come up with data-driven steps to achieve our goal of converting one-time hires to annual members. Analyzing the months and hiring decisions made during the months We want to see if there is any pattern in our data that can be identified.

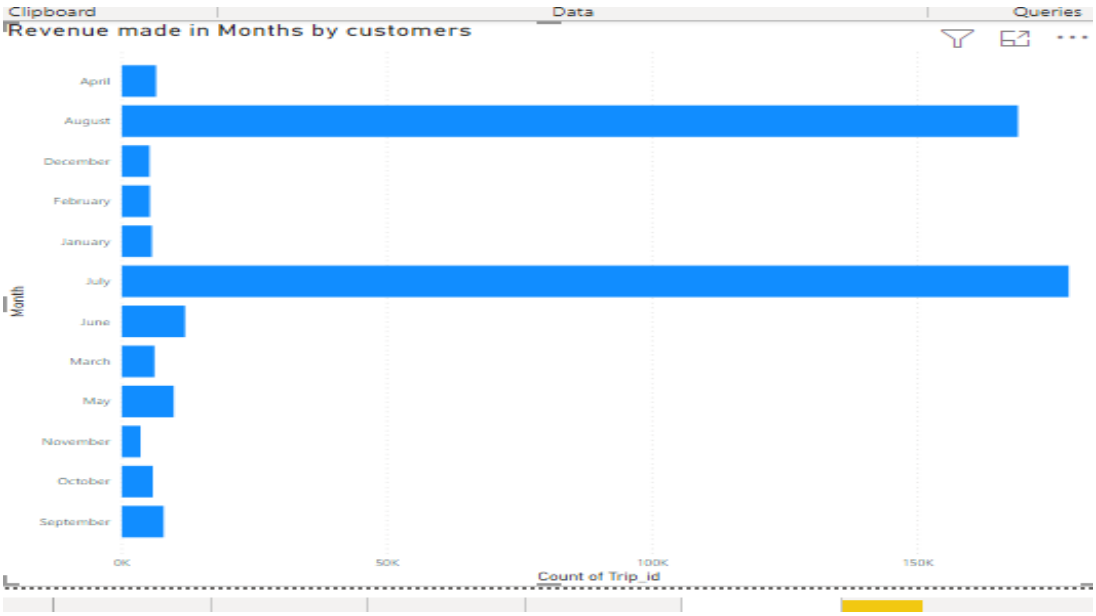


Fig.3.8. A clustered bar chart, showing hires made in each month.

From the above visual we can see our best performing month by customers are august, July, June, May, and September.

4. Results And Discussion.

Keeping in mind that the goal of this analysis is to convert customers to subscribers, the first thing we had to look at was how our hirers use our bikes in different ways, which will allow us to come up with the best approach to use in meeting this goal. To aid in this process, sub-questions were created.

First, we accessed hires according to how they fall into the days of the week, and then we grouped them by hirer type (customer & subscribers) to see which days provide the most revenue for our hirer type, as shown in fig. 3.5. The line chart, with the "green line" denoting subscribers, shows that hires begin to rise on Monday and drop on Saturday, with the lowest hires occurring on Sunday and Saturday (Note: in the graph 1 represents Sunday and 7 Saturday the rest numbers and days fall accordingly). The "blue line" indicates that customers hire more on weekends, as the line drops from Monday (1) to Friday (6) and rises on Saturday (7). Since we've seen how our various hirers (users) operate, our focus has been solely on customers.

To better understand our customers (one-time hirers), we conducted additional research using our customers' ages to see if their age influences how they use our services. Customers are more likely to be Minnellas and Generation Z. This group has the most hires throughout the week, but especially on weekends. Knowing the ages of our customers is a huge plus because it will allow us to tailor a campaign specifically to this demographic for maximum effect, and a preferred medium to deliver the message to our intended target.

Third, because we already know how our customers hire based on the days of the week and the age group they belong to, we will apply the campaign to all 638 stations, but we will prioritize stations that have brought us the most customers so far. We can see from our above bar chart fig.3.7 which stations will receive the most attention, but don't forget about the other stations.

Finally, we went a step further by drilling down to see hires made by month to see if it could help us reach our goals. We discovered that customers (one-time hirers) made more hires during the months of June, July, August, and September. In 2019, these months saw the most hiring.

5.Recommedations

We looked into different aspects of our record in order to find patterns that we can use to come up with ways to convert our customers into annal subscribers. Due diligence has been performed, and the results of our analysis have provided us with data-driven steps that I am recommending to my manager for the best informed decisions to be made in order to meet this goal.

- Since customers(casual riders) are more on the weekend. I would recommend, Cyclistic promoting weekend activities as biking racing, exercise for members.
- Promoting summer activities such as bike competition for members, during the summer period.
- Discovered Gen Z and Minnellas (those born between 1975 and 2000) have more weekend hires than any other age group under the customer category. For optimum reach, the advertisement should be run on social media.
- Stations with more customers hiring rate, I recommend should be the target.