Research Report: Excel

Valentin Jimenez

02/02/2022

First Part: Analysis with Excel

Target question: 1. How do annual members and casual riders use Cyclistic bikes differently?

Cleaning data

The databases have many blank spaces, depending of what part of the analysis we are making we will decide what to do with them; I made the cells format consistent and trimmed spaces. I added some columns to make some calculations with spreadsheets. I prepared a file called **"changesLog.txt"** which contains the changes I made, and the order in which they were executed to keep historical track.

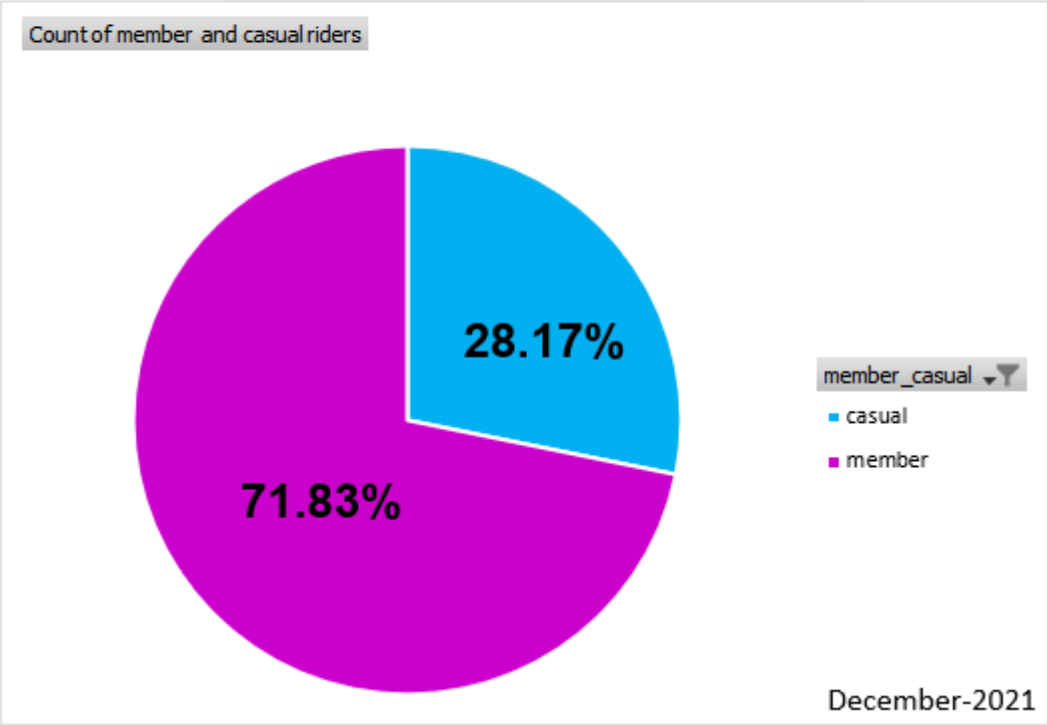
Exploring datasets

Basic analysis We begin with a very basic analysis in order to start to be familiarized with our data. This time the tool for managing spreadsheets was Excel. As an example I will show the results for the data corresponding to the month of December 2021.

For this part I built some pivot tables and pivot charts:

1. From our pie pivot chart we know that **28.17%** of riders are casual and **71.83%** of riders are member riders. We will see if the difference in percentage keeps huge for the rest of the year or if it just the case for December or some months, this will help us find which months are preferred by each type of rider.

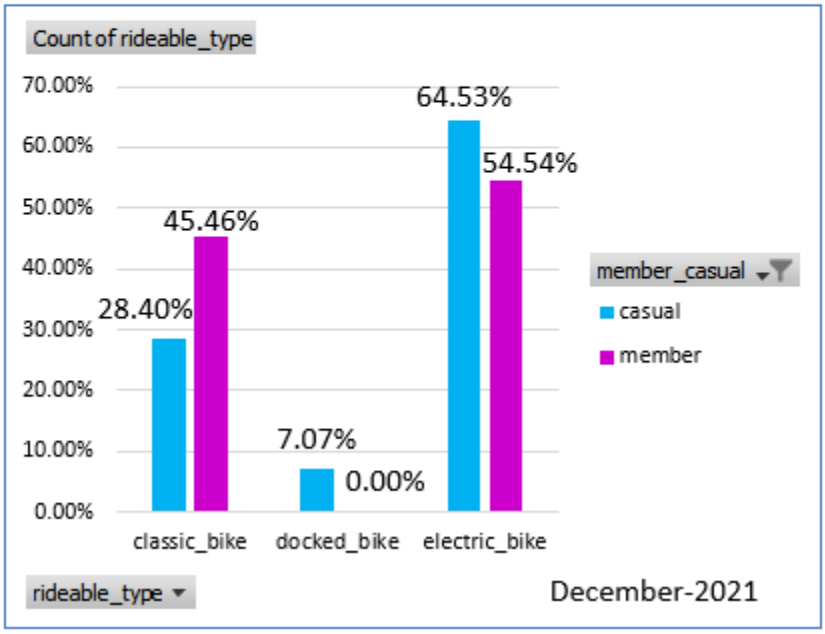
Row Labels	Count of member and casual riders
casual	28.17%
member	71.83%
Grand Total	**100.00%**



2. Analysis by Rideable Type

From the following column pivot chart, where we consider the percentage of users by each type of bike; we can see the following: docked bikes, at least for December, are by far the least used type of bikes, and electric bikes are highly preferred by casual riders, even more than member riders, so this could be a potential target.

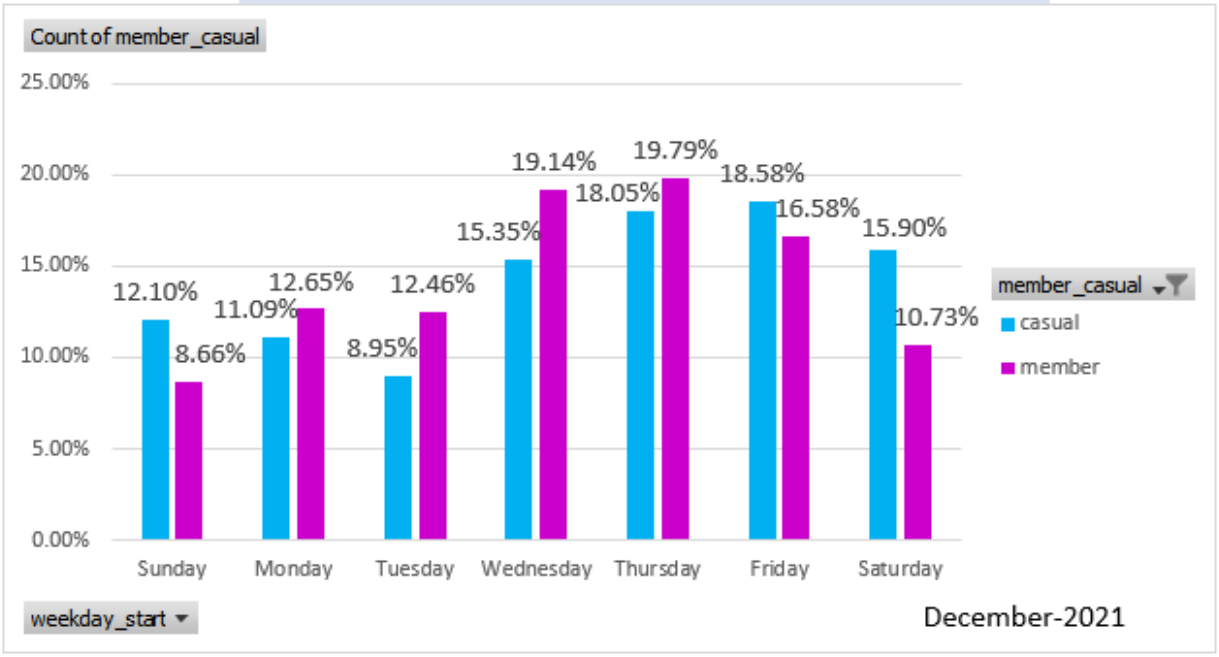
Count of rideable_type	Column Labels		
Row Labels	casual	member	Grand Total
classic_bike	28.40%	45.46%	40.65%
docked_bike	7.07%	0.00%	1.99%
electric_bike	64.53%	54.54%	57.36%
Grand Total	**100.00%**	**100.00%**	**100.00%**



3. Analysis by day of the week

From the column pivot chart about the days of the week used by each type of rider, we note that there is a significant difference in the quantity of casual riders on Saturdays and Sundays.

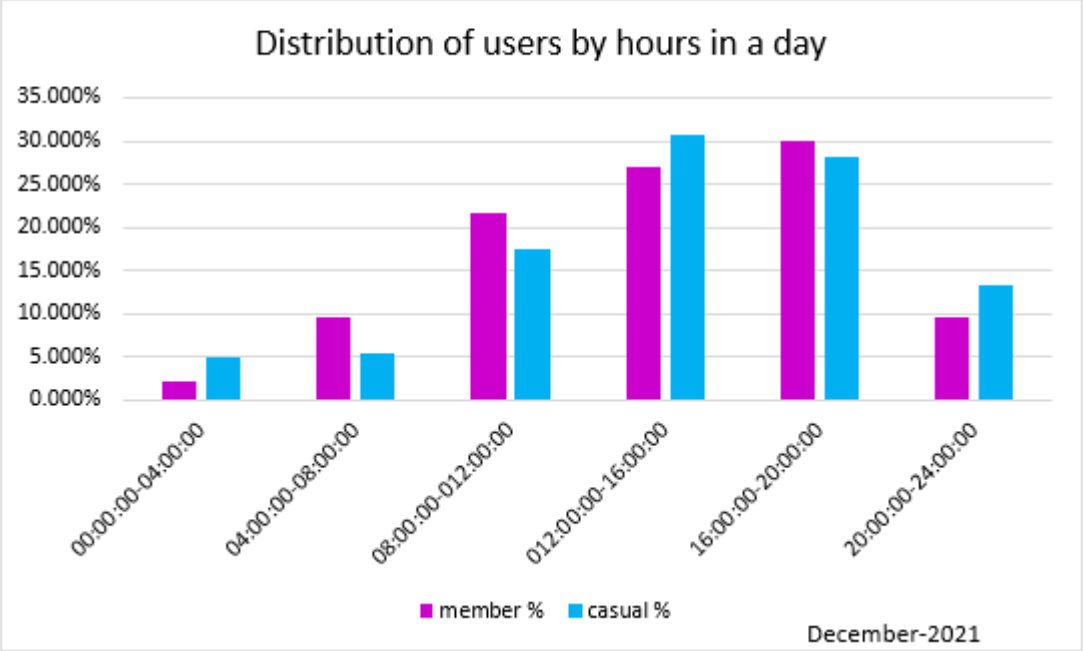
Count of member_casual	Column Labels		
Row Labels	casual	member	Grand Total
Sunday	12.10%	8.66%	9.63%
Monday	11.09%	12.65%	12.21%
Tuesday	8.95%	12.46%	11.47%
Wednesday	15.35%	19.14%	18.07%
Thursday	18.05%	19.79%	19.30%
Friday	18.58%	16.58%	17.14%
Saturday	15.90%	10.73%	12.18%
Grand Total	**100.00%**	**100.00%**	**100.00%**



4. Analysis by hours in a day

In this analysis the pattern for the two types of riders is pretty similar with no big differences, just note that **78.4%** of members and **76.2%** of casual riders use the service between 8 a.m. and 8 p.m. We will see if the trend is kept for the rest of the year.

hours in a day	member	member %	casual	casual %
00:00:00-04:00:00	3949	2.221%	3393	4.865%
04:00:00-08:00:00	17088	9.611%	3819	5.476%
08:00:00-012:00:00	38468	21.635%	12169	17.450%
012:00:00-16:00:00	47878	26.928%	21443	30.748%
16:00:00-20:00:00	53190	29.915%	19579	28.075%
20:00:00-24:00:00	17229	9.690%	9335	13.386%
Total	177802	100.00%	69738	100.00%

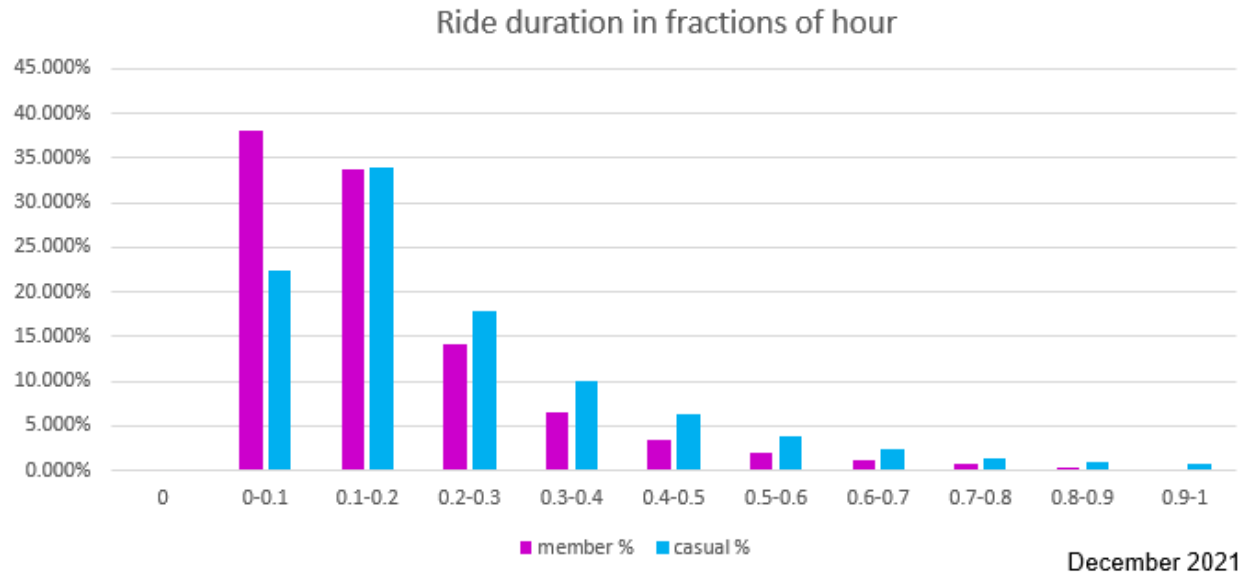


5. Analysis by ride duration in fractions of hour

The following tables consider the interval open on the left and closed on the right: for example 0-0.1 means the interval is greater than 0 and less or equal to 0.1. Also the first row corresponding to 0, means that the rider didn´t start the ride, for example, maybe the rider canceled the service, changed bicycle, or something else.

If we calculate the ride duration and sort it in a descending order it is plain to see that only **0.14%** of the rides last longer than **1** hour. So we make our basic analysis for records less or equal to 1 hour. From the following column chart, Ride duration in fractions of hour, we see that **71.8%** of member riders prefer to use the service for rides of length less or equal to **12 minutes**, which are 0.2 of hour, while the percentage of casual members in this same range is **56.4%**.

ride duration in fraction of hour	member	member %	casual	casual %
0	12	0.007%	9	0.013%
0-0.1	67475	38.118%	15062	22.452%
0.1-0.2	59710	33.732%	22777	33.952%
0.2-0.3	25010	14.129%	11959	17.826%
0.3-0.4	11532	6.515%	6784	10.112%
0.4-0.5	6120	3.457%	4300	6.410%
0.5-0.6	3340	1.887%	2504	3.733%
0.6-0.7	1987	1.123%	1542	2.299%
0.7-0.8	1124	0.635%	944	1.407%
0.8-0.9	454	0.256%	667	0.994%
0.9-1	251	0.142%	538	0.802%
Total	**177015**	**100.00%**	**67086**	**100.00%**



6. Location analysis

This part considers the coordinates given in the datasets, the distance is calculated considering radial distance, i.e. the distance traveled from each stations using the classical euclidean distance: the square root of the sum of the squares of the respective differences in the x and y coordinates.

After an initial analysis it turned out that more than **99%** of members and more than **98%** of casual riders travel distances less than or equal to **0.1** units.

So after analyzing this subset of riders I get something interesting. In every range, except for the interval 0.01-0.02, the casual riders are always a bigger number than member riders; however, for the interval 0.01-0.02 there are only **0.775%** of casual riders.

units traveled	member	member %	casual	casual %
0	9328	5.272%	5218	11.423%
0-0.01	50709	28.660%	13871	30.365%
0.01-0.02	58563	33.099%	354	0.775%
0.02-0.03	26460	14.955%	11976	26.217%
0.03-0.04	13627	7.702%	6114	13.384%
0.04-0.05	7547	4.265%	3452	7.557%
0.05-0.06	4418	2.497%	1925	4.214%
0.06-0.07	2736	1.546%	1294	2.833%
0.07-0.08	1841	1.040%	735	1.609%
0.08-0.09	1006	0.569%	462	1.011%
0.09-0.1	700	0.396%	280	0.613%
Total	**176935**	100.000%	**45681**	100.000%