Coke-Zero Bike Usage in Galway (2019-2020)

An Analysis by Brian Carey

Introduction

Bike-Sharing schemes are becoming increasingly popular around the world as a cost-effective way to reduce traffic congestion, improve the cardiovascular health of its users and reduce carbon emissions. Bike-sharing is a short-term rental service where users can pick up a bike at an unattended station and drop it off at another. By sharing with others through a publicly available scheme, individuals can use bicycles on an "as-needed" basis, without the costs and responsibilities associated with ownership. The Coke-Zero scheme was launched in Galway in November 2014, starting out with 19 stations and 195 bikes. The scheme is operated by An Rothar Núa, as part of the National Transport Authority (NTA). Today it has 22 stations. In Ireland, there are similar schemes in Dublin, Cork and Limerick.

Galway has the worst traffic congestion in the country. A report by Intirx[1] found that it ranked 61 out of more than 1,060 in worst performing cities worldwide, the only Irish town or city in the top 100. A bike sharing scheme that integrates with public transport would go some way to improving this ranking.

The NTA kindly provided me access to this dataset. It was already in good shape and did not require any cleansing. I used Excel and PowerBI for this analysis.

Who Uses It?
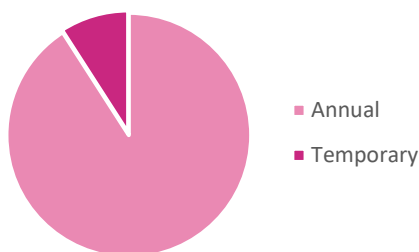
The Coke-Zero scheme has two available passes: The Annual Subscription and the Temporary pass.
The Annual Subscription costs €10 for the year, while the Temporary pass costs €3 for three days of usage. It would be fair to assume that users of the Annual pass are local commuters and users of the 3-day passes are tourists or short-term visitors. 93 per cent of users are annual subscribers, while 7 per cent use the 3-day pass. There were some instances where the distance covered amounted to 0, which suggests that the trip was initiated by accident or that, the user started and finished at the same station. Interestingly, almost 33 per cent of customers who travelled 0 were Temporary customers. This makes sense as a tourist may prefer to explore Galway on a leisurely cycle, without going anywhere in particular. I decided to filter out these instances as I feel they did not add much to analysing commuting patterns of cyclist.
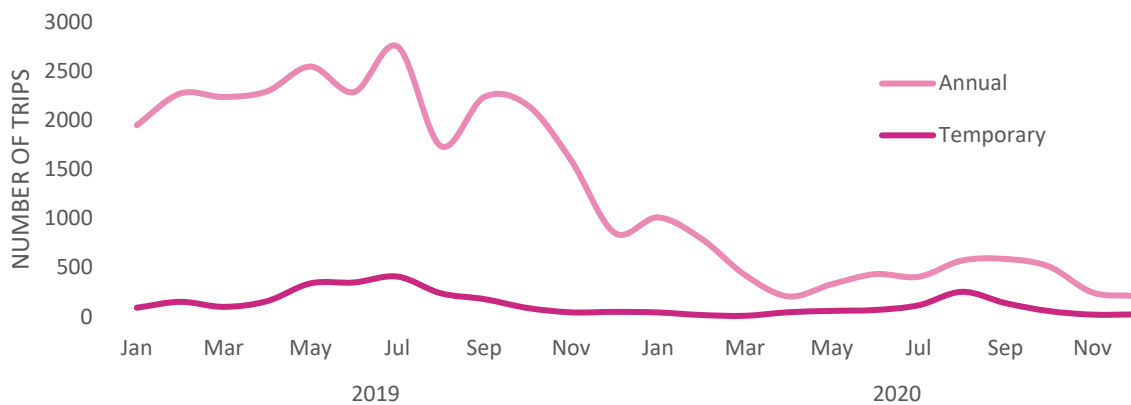
Looking at the graph above, there are slight peaks in Temporary usage in the summer months, which reaffirms the assumption that these are tourists. Generally there is more activity in the Summer than the Winter. In 2020, we can see there is a dramatic reduction of usage due to the pandemic.

I created a Calendar Table, which acts as a lookup table for date groupings:

```
Date Table = ADDCOLUMNS(
    CALENDAR("1-1-2019","31-12-2020"),
    "Year", YEAR([Date]),
    "Monthnumber", FORMAT ( [Date], "MM" )…
```

Other columns include "MonthNameShort", "MonthNameLong" etc. These columns allow greater flexibility and increased granularity when analysing the data. A relationship could then be established between the new "Date Table" and the original "DateStart" Column. It was necessary to create "MonthNumber" columns so I could order the months in their natural order, instead of

Which Stations are the most popular?

Not surprisingly, the two most popular stations are the two NUIG stations, Kingfisher and Park and Ride. The third most popular station is Bodkin. There is good cycling infrastructure around NUIG and going from there to Bodkin would be a straightforward journey over the Quincentennial bridge, which has a dedicated cycle lane. There is also a large student population in Gort na Coribe, not far from Bodkin. The cycle lane on the Quincentennial bridge was recently revamped in 2020 by the NTA.

It would be interesting to know why the three least popular stations are Newtownsmith, Gaol Road and County Hall.

The original dataset did not provide Latitudes and Longitudes to the stations, so I manually input these using data bikeshare.ie. The coordinates where then assigned to their corresponding stations using VLOOKUP().
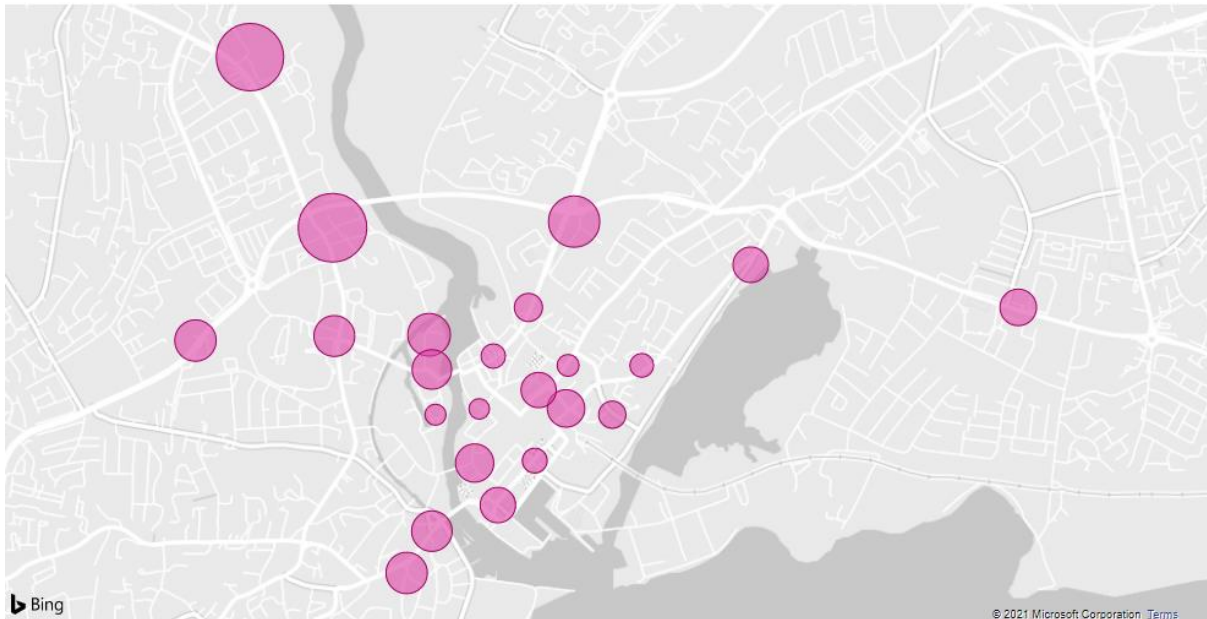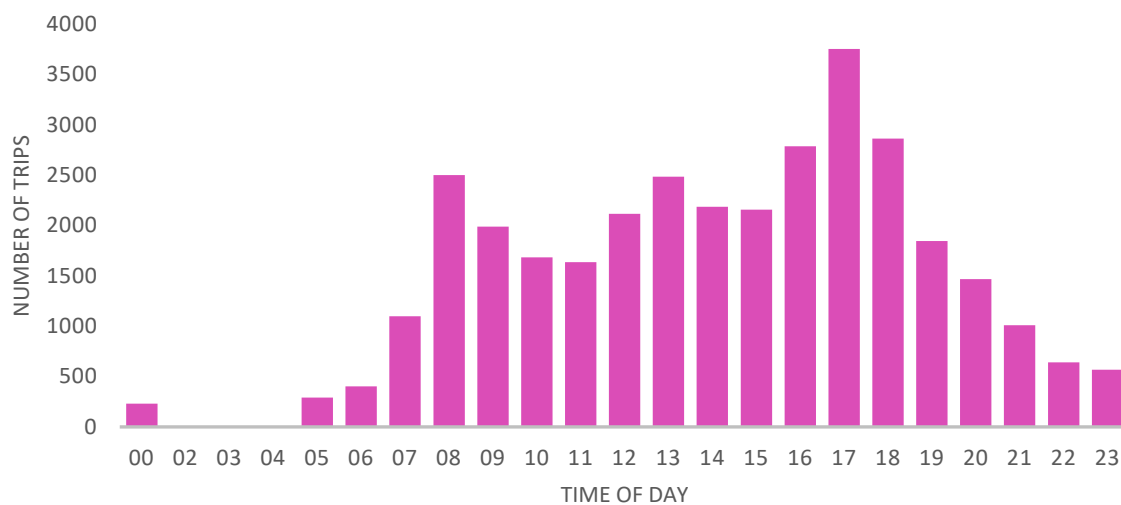
Figure 1: Number of Trips by Start Station

When do people cycle?

Naturally, most of the bike usage is during the rush hours. Over 17 per cent of trips are made on Thursdays, while just under 9 per cent of trips were made on Sundays. Temporary customers' usage was much more evenly spread throughout the week. During July of 2019 (the highest usage rate in this dataset) almost one fifth of trips made were on Tuesdays. There were 24,511 trips made in 2019, compared to just 5,454 in 2020, although 2020 was not a typical year.

The original Date Start column was in a Date-Time format e.g. "20/04/2019 19:55:00". I created a new column and used the INT() function to extract the date: "20/04/2019". To create the time only column, I subtracted the date-only column from the original date-time

column, leaving the time: "19:55:00". Similarly to the Date Table, I created a Time Table which broke the time into minutes and hours.

How long do people spend cycling?

When a customer rents a bike, they get the first half an hour free. Additional charges apply when the customer exceeds this time limit. Up to 1 hour costs an extra €0.50, up to 2 hours costs €2 and up to 3 hours costs €3.50. Nearly

Time (mins)	Fees	Number of Trips	Total Fees
30-60	€0.50	1895	€947.50
60-120	€2.00	849	€1,698.00
120-180	€3.50	353	€1,235.50
>180	€6.50	85	€552.50
		Total Fees Raised =	€4,433.50

98 per cent of trips made have been under half an hour but 85 trips have exceeded the 3-hour mark, which would cost the customers €6.50 each.

I initially encountered problems when trying to graph the Duration column. It was originally in a Time format, so I converted the column to a Number format. I then multiplied the result by 1,440, to give the time in minutes. It gave the time in a decimal format, which is sufficient for analysis.
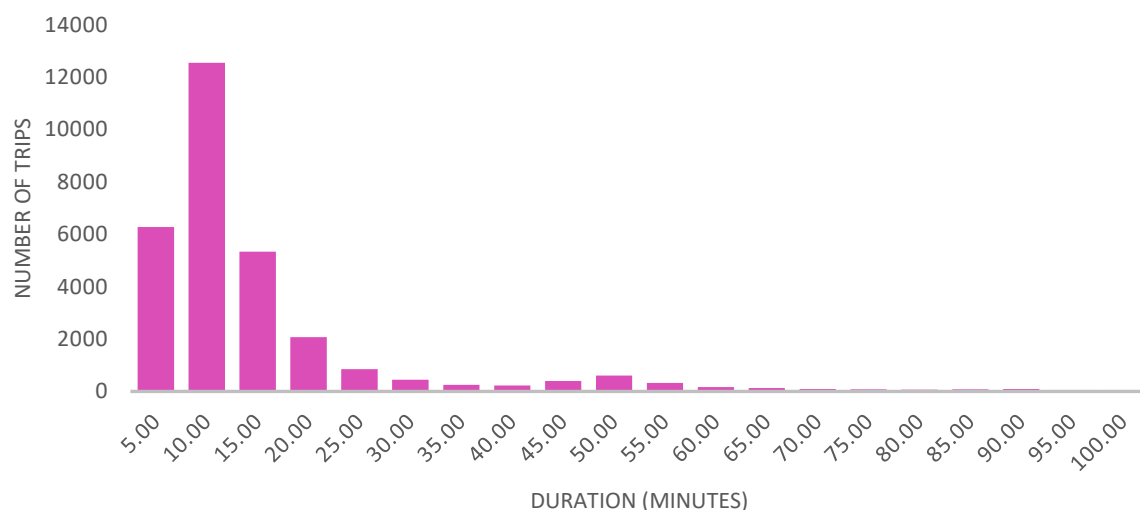
The above histogram shows the distribution of cycle durations. We can see that most journeys take under fifteen minutes.

The histogram was made using the Data Analysis Toolpak. Bins were divided into groups of five minutes. There are more entries past hundred-minute mark, but for the sake of visualising the data, these were excluded from the histogram.

How Could Usage Be Increased?

In Copenhagen, urban planners have built more cycle lanes along with safer intersections for cyclists. It is estimated that the number of Copenhageners commuting by bicycle is at 60 per cent in 2018. Cycling has always been part of Danish tradition, but can the same be

applied to Galway? Dedicated cycle lanes, with protective bollards and away from traffic would be ideal. Building cycle lanes, is a good place to start, but it can be overused as a quick fix. Dutch empirical evidence[2] has found that widening a road to place a cycle lane can encourage motorists to increase their speed. The same research suggests that on busy inner-city roads where the speed of motorists is low, mixing cyclists with traffic may in fact be safer instead putting in a cycle lane.

In Galway it seems the cycle lanes were built in places where it was easiest to do so, but can abruptly stop at difficult and dangerous intersections.

In July 2020, Galway City Councillors voted against plans for implementing a cycle lane in Salthill. Salthill would be an ideal place for a dedicated cycle lane but in its current state, it is an unpleasant place to cycle.

In November 2020 it was announced that there are plans to construct a cycle lane from Ballybane to Briarhill.

Conclusions

- It would have been more beneficial to my learning if I received a "messier" dataset. Typically, this stage of analysis takes the longest and I feel I have skipped a key part of data analysis.
- There was a lot of copying-and-pasting of data, which was time-consuming. Data for 2019 and 2020 were in two different sheets and I just copy-and-pasted them together. I did not have knowledge all the functions of Power Query to transform the data in the most efficient manner.
- I feel the decision to filter out Distance = 0 was justified. When analysing commuting patterns, it would not have told us much.
- Excel, is a powerful tool but sometimes, when transforming thousands of rows of data, the application would freeze. Pandas, a library of Python would be able to handle large datasets with less fuss but I was less comfortable using it.
- I could have looked more in depth at the less popular stations.

The Coke-Zero scheme does not represent the cycling habits of the entire population of Galway, but it does give us an insight into the routes and patterns of nearly 30,000 trips.

Looking at the bikeshare.ie website, it gives live updates on the supply of bikes at each station. It would make an interesting project to see if one could try and predict the number of bikes at a station depending on time, day of the week and the weather. When there are imbalances, the bikes are redistributed on the back of a truck. Instead of reacting to where there is an abundance of bikes and where there is a scarce supply, it would be more efficient to predict which stations need how many and when.

Galway has the potential to be great a Cycle-Friendly City. It is encouraging to hear that there are plans to build more cycle lanes in the city, but there is a lot more work yet to be done.

Sources:

[1] https://fora.ie/galway-bad-traffic-3707118-Nov2017/

[2] https://www.sciencedirect.com/science/article/abs/pii/S0925753515001472