

The effect of lockdown on cycle hire in Edinburgh

Marina Kazamia

19th December 2021

1 Overview

In this report, I explore the effect of lockdown on the number of trips made using the 'Just Eat' cycle hire scheme in Edinburgh. Here, the lockdown period is considered to be the time period from March 24th to September 30th 2020. To account for the fact that the scheme is growing by regularly installing new stations across Edinburgh, three regression models have been fitted to determine the degree of the effect that the number of different stations used and the lockdown have on the number of trips. Comparing the coefficients of determination (R^2) for each model, the one with the highest R^2 value was the multiple regression model, which was then used to test the hypotheses for the regression coefficients of each parameter. For the coefficient of the date period there was sufficient evidence to reject the null hypothesis at a 0.05 significance level, however for the coefficient of the number of different stations used there was not. Thus, there are not enough evidence to suggest that the growth of the scheme contributes to the increasing popularity. To test further the effect of lockdown on bicycle usage, A/B testing was used to compare the number of trips between the samples of pre-lockdown and lockdown periods. Using a bootstrap simulation of 10000 iterations, it was shown that at a 0.05 significance level there are enough evidence to suggest that there is significant increase between the mean number of trips in pre-lockdown and lockdown periods.

2 Introduction

Context and motivation The Edinburgh 'Just Eat' cycle hire scheme has launched in September 2018 and its popularity has been rapidly increasing since then. In March 2020, however, the UK lockdown was put in place due to the COVID-19 pandemic.[1] During the lockdown, the hospitality sector, non-essential shops and leisure centres remained closed. Everyone except key-workers has been advised to stay home and work remotely. As a result, more people seek a way to exercise outdoors and occupy themselves during their free time. It would be interesting to explore whether the popularity of the 'Just Eat' cycle hire scheme has increased because of the lockdown.

Previous work Several media reports have been released not very long after the lockdown was put in place, to broadcast the abrupt increase in usage of the cycle hire scheme. Edinburgh News reported that by the 25th of June 2020 - 2 months after the lockdown had began - there have been 80% as many hires in 2020 as in the whole of 2019, while each month of 2020 counts more than double the number of hires in 2019.[2] In addition, on the 16th of October, The Edinburgh Reporter announced that the total number of trips made in 2020 using the 'Just Eat' scheme is 119% greater to the total of 2019, while the average daily number of trips have increased by 46% compared to 2019.[3] In his blog, Dr David McArthur explores the effect of the lockdown on the number of bike rentals using the Edinburgh cycle hiring scheme, as well as the daily median duration of trips.[4] Dr McArthur's investigation, however, was published just a week after the lockdown began, thus there was no room to apply any inferential techniques to the data to verify his results. Hence, he only offers descriptive visualizations on his observations.

Objectives The objective of this report is to study the effect that the lockdown has had on the number of trips made using the ‘Just Eat’ cycle hiring scheme in Edinburgh. Previous work on this topic implies that the overall usage of the bikes has increased during the lockdown. My aim is to answer whether this increase is due to the lockdown, or due to other factors such as the better weather or the fact that the scheme has been constantly expanding by installing new stations each month to accommodate more citizens.

3 Data

Data provenance The datasets used for this investigation are publicly available by the ‘Just Eat’ cycle hiring scheme under the Open Government License v3.0. There are datasets available from September 2018 until now. In this project, I explore the data available from September 2018 to February 2021. I downloaded these in the form of CSV files from the ‘Just Eat’ Cycles website.[5]

Data description The datasets available online are separated by month. So, each dataset contains a row for each trip made in that particular month. For each trip there are 13 columns. There are 2 columns to describe the start time and the end time of the trip. Both of these are in a Timestamp format. Furthermore, there is a column for the duration of the trip in seconds. Finally, the data offers lots of information on the start station and end station for the trip in the format of station ID, name, description, longitude and latitude.

Data processing Firstly, I set out to combine the monthly datasets into yearly ones by copying all of their entries into a single CSV file using the ‘copy’ command on the command prompt. Then I imported these files onto the Jupyter notebook, and decided that it was best to merge those datasets into one, too. The data was then cleaned by removing records with missing start or end time, duration and start or end station ID. A significant amount of duplicated data was also identified. Given the nature of this data, one could argue that duplicates could arise naturally since we are observing cycling trips without any key identification for each record. However, given that the start and end timestamps of each trip are precised to the millisecond, it is too unlikely that there are 1182 valid duplicates in the data. Thus, all duplicate records have been dropped except the first. I also noticed that there were a lot of records which were missing start and end station description. I decided to keep those since that column would most likely not be needed for my analysis. It is also worth noting here, that I have kept all the columns in my initial table, and then copied the ones that I thought would be needed for a specific task into a separate table.

4 Exploration and analysis

I began the data exploration by looking at the overall trend of the number of trips per month from the launch of the scheme in September 2018 to February 2021, as seen in Figure 1 by the blue line plot. It was interesting to see the steep increase in number of trips from April to May 2020. The months from May 2020 to August 2020 have had quite a constant number of trips per month but these have been the higher number of trips per month ever recorded. June 2020 counts around 29500 number of trips which is almost twice the number of trips made in August 2019 - the recorded peak before Summer 2020.

Realizing that this sudden increase in the number of trips took place during the lockdown, I thought it would be more illuminating if we could see the weekly change of the number of trips from the start of 2020 to the end of September 2020. In Figure 2, we can easily see how the key-dates of the lockdown[1] correspond to the changes of the weekly number of trips. Since the start of the lockdown on March 24th until May 29th when the lockdown measures started easing, the weekly number of trips have gone up by 543% reaching a total of 9749 trips made just in the last week of May. Moreover, the most sudden increase in the number of trips per week seems to be occurring by the time that the Scottish Government has lifted the once-a-day exercise limit on May 10th. In just those 20 days from May 10th to May 29th the

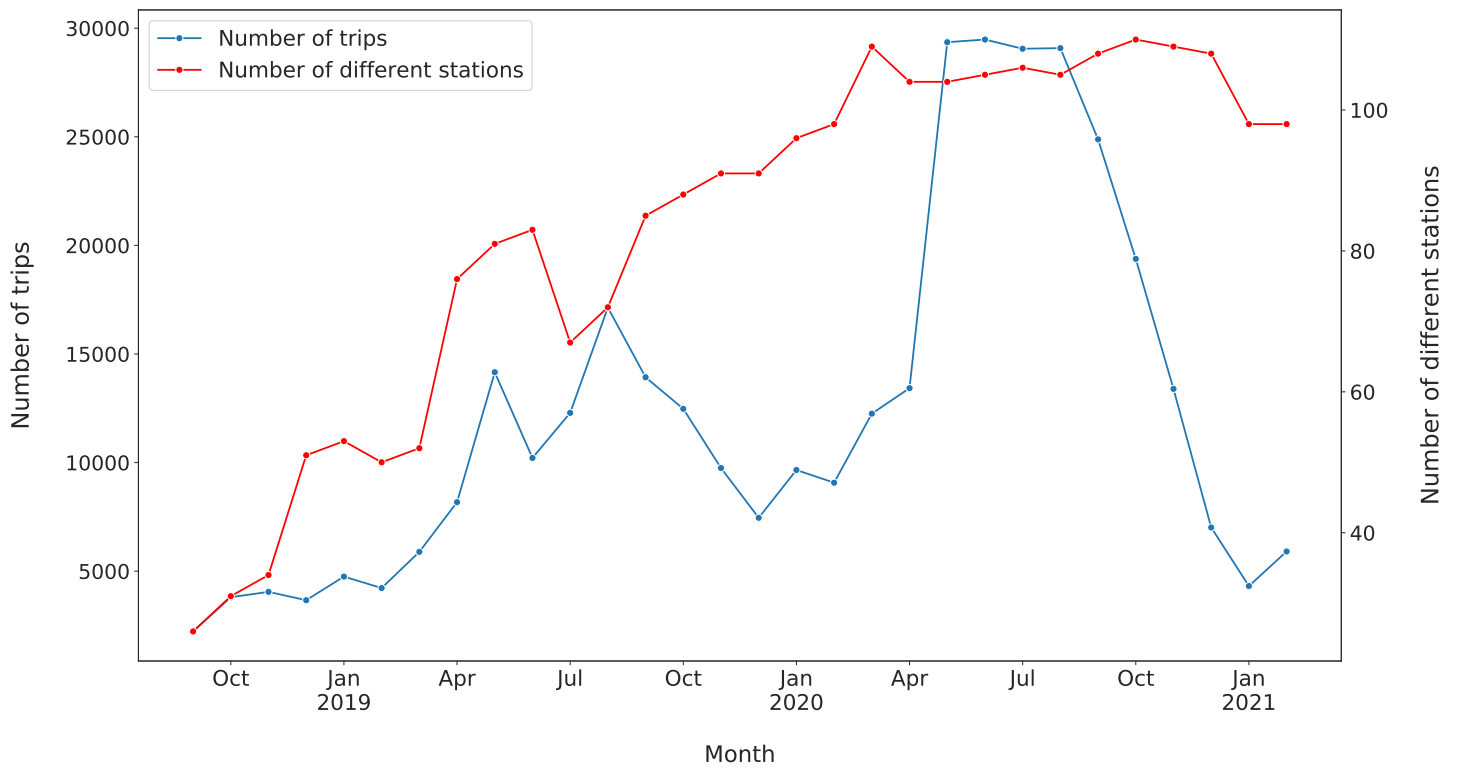


Figure 1: A timeline of how the number of trips taken (blue) and the number of different stations used (red) change by month from September 2018 to February 2021.

number of trips have increased by 55%. The data together with the timeline of the events in Edinburgh during that time suggest that the driver of this increase of popularity for the ‘Just Eat’ scheme is the lockdown. How can we be certain that this is the case?

As it is natural for any young initiative as the ‘Just Eat’ cycle hire scheme, the scheme has been growing and installing new stations around Edinburgh regularly. While exploring the data, I have found that the number of different stations used per month have increased from 26 in September 2018 to 110 in October 2020. While, there is no data available about the actual number of stations accessible to the public at any given time, Edinburgh Reporter states that the number of stations across Edinburgh in October 2020 were 75 more than in the year before.[3] In Figure 1, we can see that even though there has been a constant, gradual increase in the number of stations across Edinburgh over the years, one of the lowest number of trips per month was recorded in January 2021 when there have been available even more stations than in Summer 2020.

Of course, other reasons could also be behind the variability of the number of trips, such as the weather. To make my analysis as accurate as possible, I have decided to define my ‘lockdown period’ as the period from March 24th to September 30th. In that way, we constraint the variability of weather and of number of daily positive cases for COVID-19 (from mid-October, stricter measures were enforced to combat the second wave of the pandemic[1]). Moreover, I want to investigate the effect that the number of different stations used and the date period have on the number of trips. By coding the date period as 0 for ‘pre-lockdown’ and 1 for ‘lockdown’, I have fitted three different regression models to the data. In Table 1, you can find the coefficient of determination, R^2 , for each. The multiple regression model and the linear regression model for the number of trips on the date period code have an almost identical coefficient of determination of 0.481 and 0.479 respectively. These rank higher than the one of the model for the number of trips on the number of stations where $R^2 = 0.442$. This suggests that the data is best modelled using one of the former models, with the date period being a better predictor for the number of trips than the number of stations used. But, how confident can we be that this is the case?

Using the stasmodels library, I have obtained the regression coefficients and their corresponding 95% confidence intervals and p-values for each parameter of the multiple regression model. We get that the point estimator for the coefficient of the number of stations used parameter is 13.58 with a 95% confidence interval (CI) of (-42.24, 69.40) and a p-value of 0.627. Clearly, there are not enough

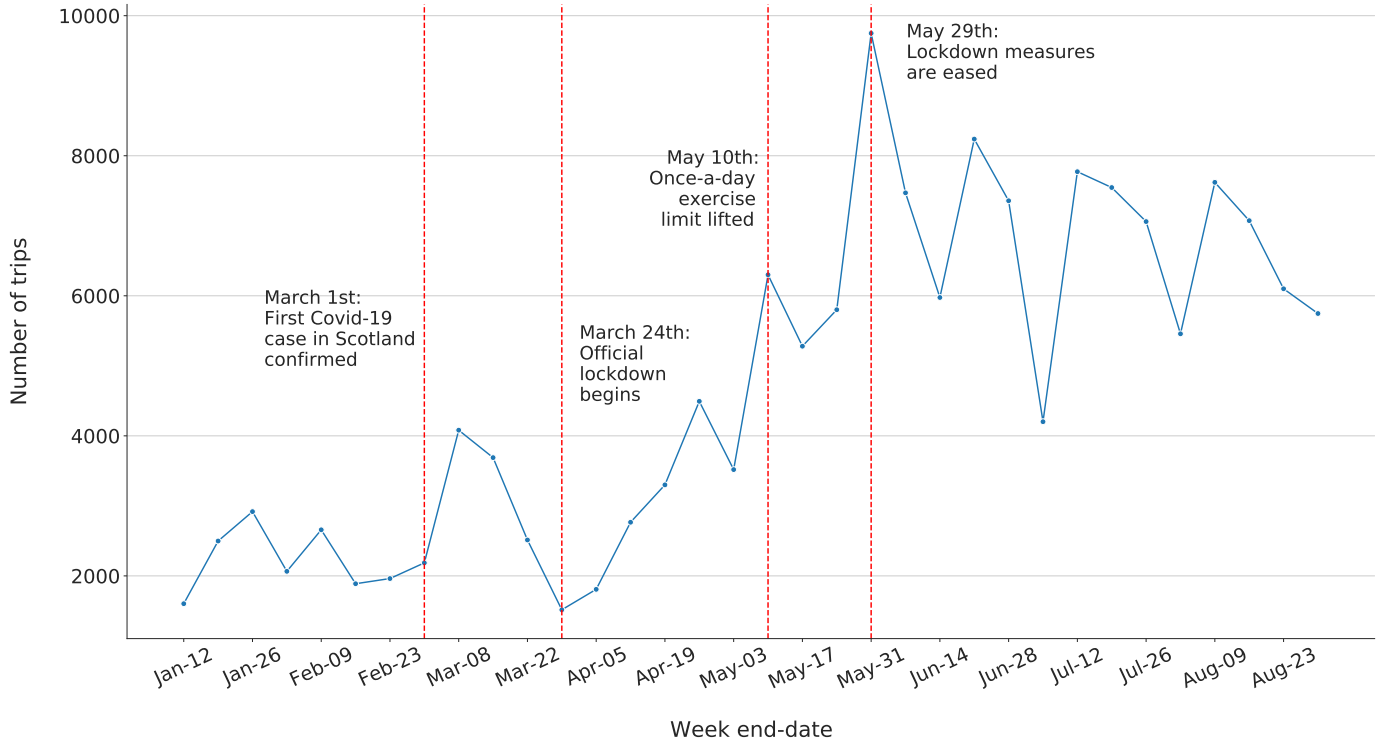


Figure 2: Total number of trips per week from January 2020 to August 2020. The red dotted lines show the key-dates of the lockdown in 2020 for Edinburgh.

evidence to suggest that the number of stations used have an effect on the number of trips at the 0.05 significance level, since $p > 0.05$ and the 95% CI contains 0. Thus we cannot reject the null hypothesis for this parameter. Moreover, the point estimator for the coefficient of the date period code parameter is 2340.15 (95% CI: -40.90 - 4721.19; $p = 0.054$). Here, we can see that the 95% CI is wide again, and 0 is contained in that range, but the p-value of 0.054 indicates that we can reject the null hypothesis at the 0.05 significance level. Thus, it is safe to say now that the parameter with the most effect on the number of trips is the date period - in other words, if there has been a lockdown at that period or not - and not the number of stations publicly available for use.

In Figure 3, we compare the number of trips made per week between my lockdown period (in orange) and the same dates just a year before (referred to as the pre-lockdown period shown in blue). These are illustrated by the number of weeks passed from March 24th. To examine how significant the change on the number of trips per week from the pre-lockdown to the lockdown period is, an A/B test is conducted to compare the two samples. The hypotheses are defined as follows:

H₀: The mean weekly number of trips in lockdown is the same as the mean weekly number of trips in the pre-lockdown period.

H_a: The mean weekly number of trips in lockdown is different from the mean weekly number of trips in the pre-lockdown period.

To test the null hypothesis, the bootstrap technique was used. For 10000 iterations, a sample of size 27 (equal to the size of each population) was taken with replacement from each of the pre-lockdown and lockdown samples. Then, the difference of the mean number of trips per week of was calculated and stored at each iteration. These differences were used to plot the histogram of Figure 4. The differences calculated at each iteration form a normal distribution of mean 2877.35 with a 95% CI of (2057.19, 3665.27). The p-value $P(d < 0)$ is 0, so we can reject the null hypothesis at a 0.05 significance level, and conclude that the lockdown has had a significant effect on the increase of the number of trips per week. In addition, we can verify the results of the bootstrap by observing the population means for each period through the dotted lines on Figure 3. The actual difference is thus $d = \mu_{\text{lockdown}} - \mu_{\text{pre-lockdown}} =$

Table 1: Coefficients of determination for different regression models

| Model | R^2 |
|--|-------|
| Number of trips per week on number of stations used and lockdown code (mult. regression) | 0.481 |
| Number of trips per week on number of different stations used (regression) | 0.442 |
| Number of trips per week on date period code (regression) | 0.479 |

$5757.704 - 2875.815 = 2881.889$. The actual mean is almost equal with the the mean of the bootstrap distribution, thus the bootstrap simulation has been successful at estimating the population mean and its results are to be trusted.

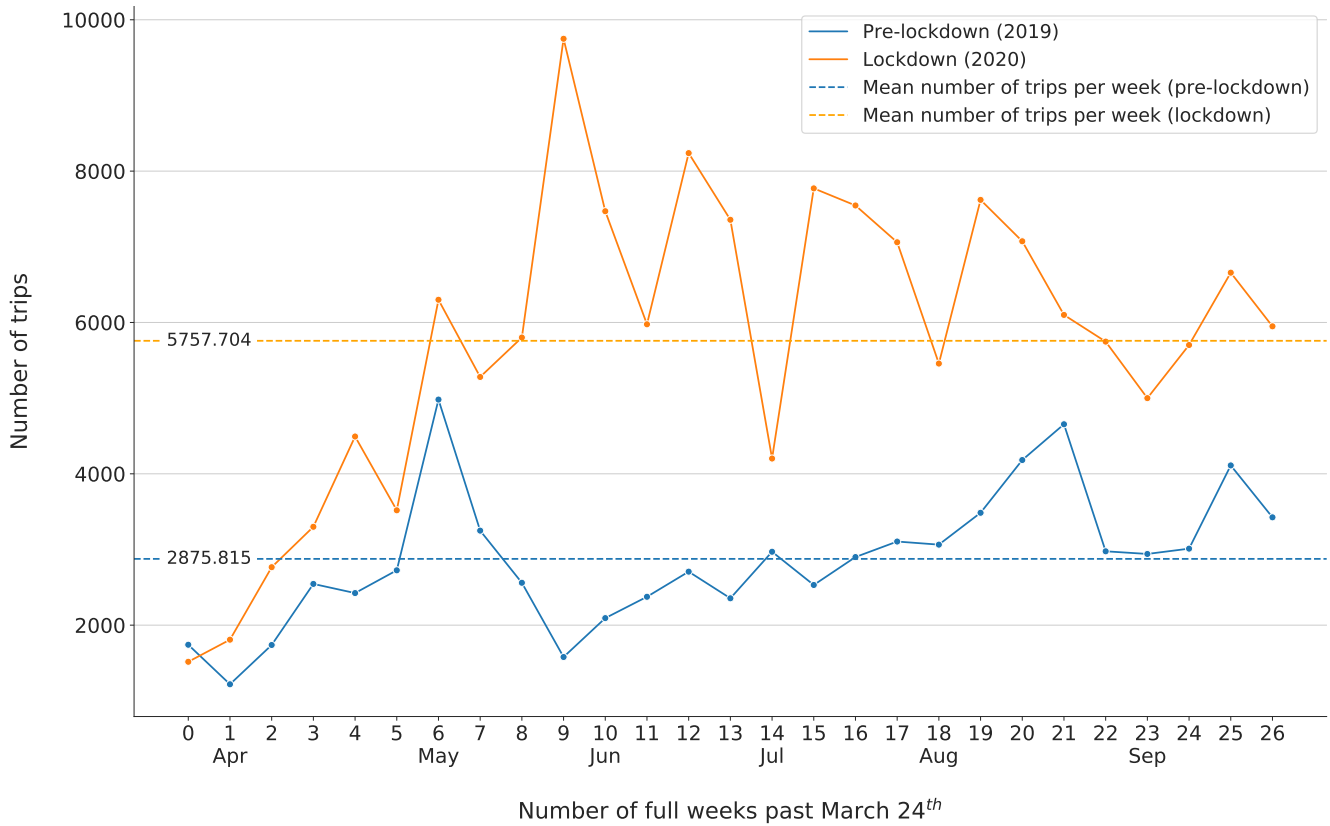


Figure 3: Comparison of the total number of trips made per week from the 24th of March between the pre-lockdown period (orange) and the lockdown period (blue). The mean number of trips are shown by the dotted lines. The x-axis of the graph indicates the number of full weeks passed from March 24th.

5 Discussion and conclusions

Summary of findings Fitting a multiple regression model and testing the null hypothesis against the p-value for the t-statistic corresponding to each of the coefficients, we have found that the number of stations used each week does not have a significant effect on the number of trips made in that week. Whereas, the time period when those trips were made has an effect on the weekly number of trips as we were able to reject the null hypothesis at a 0.05 significance level. In, other words, it was shown that there is a positive correlation between the number of trips and lockdown at a 95% confidence level - even

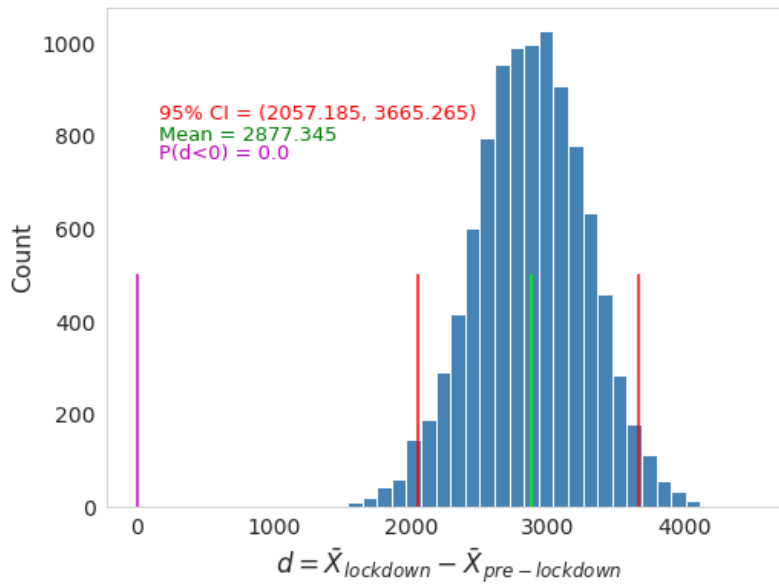


Figure 4: Bootstrap simulation of A/B test of 10000 iterations and sample size of 27. The mean difference between the sample means is 2877.35 (green), at a 95% CI of (2057.19, 3665.27) (red) and a p-value of 0 (purple).

though the 95% CI contained the value 0. To test further the degree of effect that the lockdown has on the weekly number of trips, I have used A/B testing. Here, I was able to reject the null hypothesis at the 0.05 significance level again, and show that there is significant difference between the mean weekly number of trips in lockdown and the mean weekly number of trips before the lockdown.

Evaluation of own work: strengths and limitations This project has been a challenging task, but I found it enjoyable to work on it. A strong point of the project is the quality of the visualisations presented. They are clear, detailed and color-blind friendly, however, they could be better if I managed to add minor ticks to the x-axes to show the dates missing from the labels. In addition, another strength is the effort that has been put to make the analysis as accurate as possible - first by ruling out the different waves of the pandemic which were of different intensity, second by ruling out the high variability of the weather during the seasons, and third by using regression to determine whether or not the growth and expansion of the 'Just Eat' scheme contribute to the increase of the number of trips. However, this also has been a limitation due to the fact that the coefficient of determination on any of our regression model was below 0.5, making the models not so suitable for the data.

Comparison with any other related work There has not been another published work which analyzes the effect of lockdown on the number of trips made using the 'Just Eat' bicycles. However, there have been many media reports which provide descriptive statistics on the change of the cycle hire scheme, as for example in the article by Edinburgh Reporter.[3] The article claims that there has been a 119% increase in the number of trips by October in 2020 compared to the whole of 2019, but from the data, I have calculated that the increase is only by 70%.

Improvements and extensions The project could be improved by constructing a more suitable multiple regression model for the data. This could be done by taking the whole period of the lockdown, i.e. including the second wave as well, and account also for seasonality either by coding the different seasons as a categorical variable or by including data on the weather. In this way, it is likely that we will get a coefficient of determination that is closer to 1 than the one I have used for this project. Some possible extensions on this project, could be to verify the results theoretically, and also to observe the difference in number of trips for each day of the week between the two periods.

References

- [1] Wikipedia contributors. Timeline of the COVID-19 pandemic in Scotland (2020), 2021. [Online; accessed 10-April-2021].
- [2] Alastair Dalton. New cycle hire records hit by booming Edinburgh and Glasgow bike hire schemes, 2020. [Online; accessed 10-April-2021].
- [3] Phyllis Stephen. More and more people using hire bikes in Edinburgh, 2020. [Online; accessed 10-April-2021].
- [4] David McArthur. How are people using hired bikes during lockdown?, 2020. [Online; accessed 10-April-2021].
- [5] Serco. Just Eat Cycles historical data, 2021. [Online; accessed 25-March-2021].