Edinburgh Cycle Hire in the Pandemic

1 Overview

This report analyzed the effect of lockdown and other factors on Just Eat bicycle usage. We cleaned Just Eat cycle data, with new columns added, particularly the expected cycling distance and the temperature for each day. We found that bicycle usage increased under lockdown, and more cycle hires were used just outside of central Edinburgh. This is likely because more people are using their bikes for leisure. There was also a more considerable number of rides in the late morning during the pandemic. Finally, we found if we could predict cycling usage through regression, and we found it could be, with an $R^2 \approx 0.50$ based on that day's temperature and if we were in lockdown or not. In particular, we found that higher temperatures caused an increase in bicycle usage and lockdown increased weather impact.

2 Introduction

Context and motivation This study will be analysing the effect of a once-in-a-lifetime pandemic, which has impacted nearly everything in our day to day life, from the economy to going to work. We will analyse this impact via Just Eat bicycle usage in the city of Edinburgh. As bikes are a common mode of transport and map out many peoples day to day travel. Through this lens, we can see how factors, such as duration and start/end stations can give us a deeper insight into the habits of people in Edinburgh and how they have changed over the course of the coronavirus pandemic.

Previous work

- The link between bike sharing and subway use during the COVID-19 pandemic: The case-study of New York's Citi Bike [1], a paper which looks through the cycling usage in NYC and how the coronavirus has impacted it, it was found that bicycle usage was signif cantly reduced during the pandemic.
- The Great Bicycle Boom of 2020 [2], an article which covers the signif cant increase in bicycle usage over the lockdown. It was found that suppliers are f nding it diff cult to keep up with demand and that March 2020 had threefold increase in the ridership of US trials than during March 2019.
- Change of Bike-share Usage in Five Cities of United States during COVID-19 [3] which found that cycle usage is positively correlated with recreational time.
- Weather factor impacts on commuting to work by bicycle [4]. This found that the likelihood of cycling increased in the absence of rain and under higher temperatures.

Objectives We are setting out to answer what the impact is of the coronavirus pandemic on bicycle usage, in particular we are trying to ask the following:

- What is the impact lockdown on the number of trips made during each month?
- How has coronavirus changed the usage of bicycle usage at certain times?
- In what ways has the pandemic impacted the stations people travel to?
- How has the duration of trips changed from before to after the pandemic?
- Can we predict cycle usage, in particular can we do this through weather data and how has lockdown impacted this?

3 Data

Data provenance The data sets used for this project were generated by Just Eat Cycles themselves and were available on their website[5], which were downloaded as a CSV file format. The datasets are available under the Open Government Licence[6] - a licence that allows the use of the data however the user sees fit, providing acknowledgement is given to the original source of the data and a link to the licence is provided.

Two APIs were also used in order to get additional data, the first of these was the World Weather Online History API, this API was suggested by "//TODO: Team Name?", according to the Terms and Conditions we have right to access, use, and view the services for personal and commercial use.[7] We also got the expected cycling distance, this was calculated with the Google Maps Distance Matrix API. This gave us the expected distance travelled between two points according to Google Maps. This API gives us the right to use it for personal use. [8]

Data description The data came in the form of 30 CSV files which had 371,611 records in total. There were 13 columns in each file: *started at* (the timestamp of the start of the trip), *ended at* (the timestamp of the end of the trip), *duration* (length of the trip in seconds), *start station id* (unique number for the station), *start station name*, *start station description* (gives a small description on the location of the station), *start station latitude*, *start station longitude*, *end station id*, *end station name*, *end station description*, *end station latitude*, *end station longitude*.

The expected cycling distance file was a matrix of the epexcted distance between each possible station (around 200 start and end stations in total).

The WorldWeatherOnline API gave us the following 25 columns: *date time*, *maxtempC*, *mintempC*, *totalSnow cm*, *sunHour*, *uvIndex*, *moon illumination*, *moonrise*, *moonset*, *sunrise*, *sunset*, *DewPointC*, *FeelsLikeC*, *HeatIndexC*, *WindChillC*, *WindGustKmp*h, *cloudcover*, *humidity*, *precipMM*, *pressure*, *tempC*, *visibility*, *winddirDegree*, *windspeedKmph*, *location*

Data processing There was little cleaning that needed to be done with the data. The missing values in the dataset were in the station description columns, which were removed as this information would not be needed to explore the data. We removed a handful of entries that visited "Smarter Travel Station" - a station that does not appear on Just Eat's station map[9] and so was considered an anomaly in the data. Duplicate entries were checked for, but there weren't any.

As the data came in the form of multiple files, they needed to be joined together into one Data Frame to allow for easy use in the exploration. They were each converted into a Data Frame and then concatenated together. That Data Frame was then split into two: one for before introducing lockdown and one for during the lockdown. As stated on the Government website[10], the lockdown began on the 23^{rd} March 2020. As this announcement was made in the evening of the 23^{rd}, we decided to use the 24^{th} as the lockdown's start point.

For the WorldWeatherOnline data frame, all columns except *date time* and *tempC* was dropped. This was then merged with our cycling data frame. The expected cycling distance was the same, where we merged based on the start and end stations. From this we added the column's *expected cycling distance* and the *speed*(expected distance / time). A second csv file was also created which grouped the data frame by each date to get the number of rides each day.

4 Exploration and analysis

From the data that was supplied on the Just Eat website, there were only a small number of properties that could be analysed; the number of trips made, the duration of the trips, and the stations that trips started or ended at.

In regards to the number of trips made, we saw a significant increase during the lockdown period. As can be seen in Figure 1, there was a particularly large spike from May 2020 to October 2020 - six consecutive months where the usage of Just Eat bikes exceeded the previous numbers consistently. In the spring and summer of 2019 we see a smaller spike, which is to be expected as the number of outdoor activities would increase as the weather becomes warmer. The difference between the spike period of 2019 and 2020 is that the initial lockdown had begun to ease in Scotland during May 2020. After having been under a strict lockdown for two months, we can see that either the number of people who decided to spend time outside increased or people were outside more frequently.

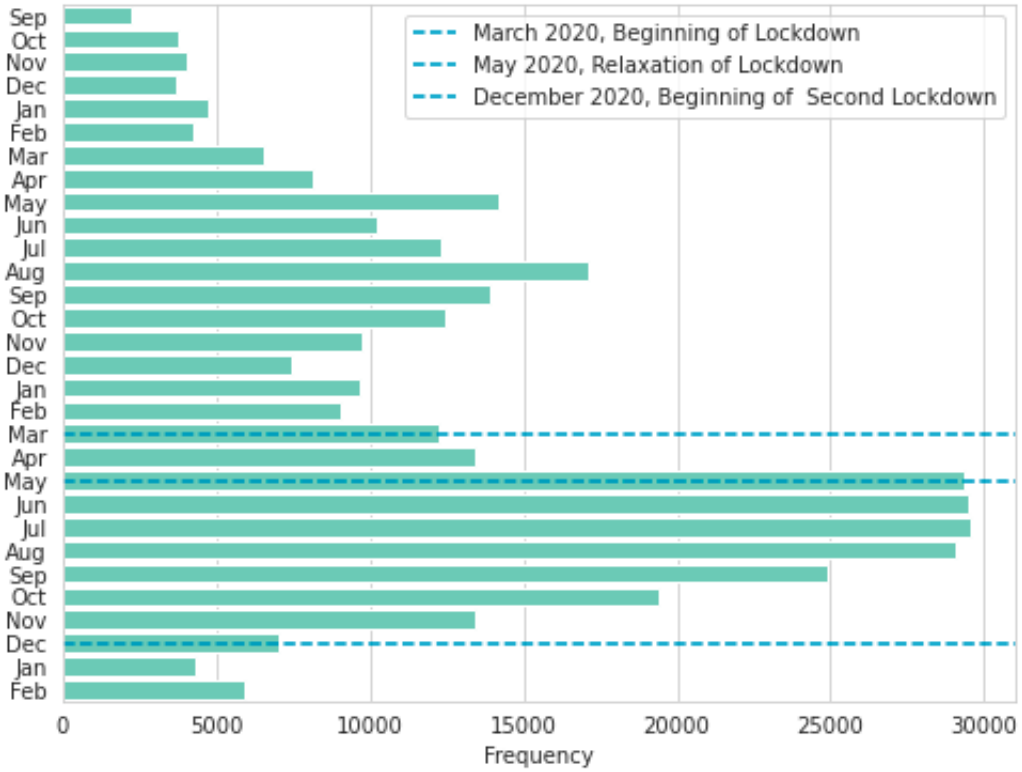

Figure 1: Number of trips per month from September 2018 to February 2021.

As it was also the case that the population was encouraged to work from home wherever possible, the trips made during the lockdown would have largely been trips for leisure rather than commuting. This would have an effect on the number of trips made during the morning rush hour as can be seen in Figures 2 and 3. We see a reduction in the amount of trips made between 7am and 9am. We also see an increase in trips during the late morning and afternoon; a time where usually people are at work or school but now have more free time to travel outdoors.

To see whether these changes in the number of trips made throughout the day was significant, we decided to construct four χ^2 tests on four periods throughout the day. These periods were broken down into early morning (between 5am and 10am), late morning and afternoon (11am to 4pm), evening (5pm to 10pm), and night (11pm to 4am).

We took the average number of trips made at each time period from before the lockdown to be the expected values and the observed values were those from during the lockdown. The values can be seen in Table 1 below.

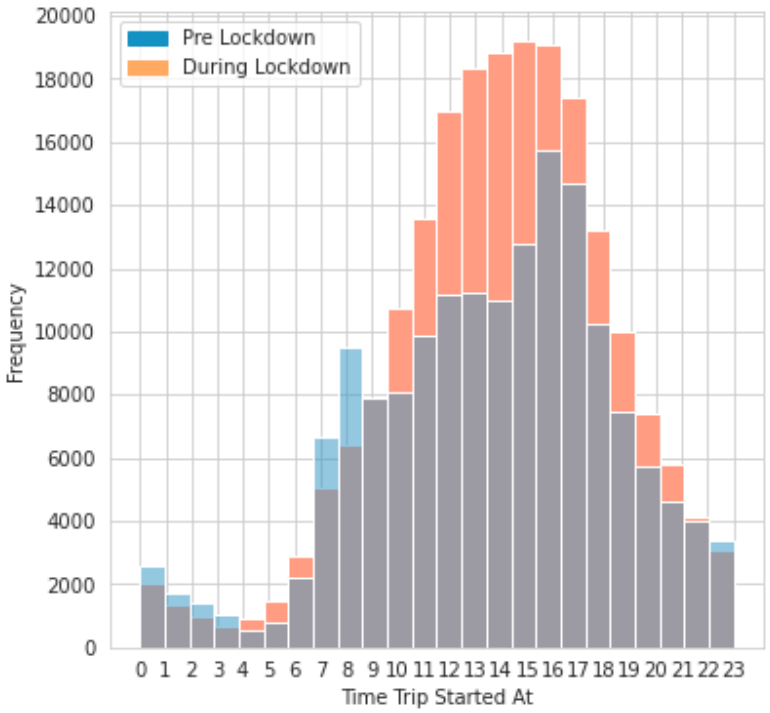

Figure 2: Number of trips made at different times of day from before and during the lockdown.

Our Null Hypothesis is that lockdown had no effect on the average number of trips made using the Just Eat bikes throughout any of the four periods that were previously defined. The Alternate Hypothesis would be that lockdown had an effect on at least one of these periods in the average number of trips made. We conduct this test at the 5% significance level with 5 degrees of freedom for each of the four periods, making the critical value 11.07.

Of the calculated test statistics, the one corresponding to late morning and early afternoon exceeded the critical value. The other test statistics did not.

Early Morning	**5am**	**6am**	**7am**	**8am**	**9am**	**10am**	**Test Statistic**
Before Lockdown	2.09	5.99	18.23	25.88	21.62	22.13	8.6
During Lockdown	4.06	7.94	13.84	17.55	21.59	29.37	$\chi^2 < \chi_{0.05,5}$
Morning/Afternoon	**11am**	**12pm**	**1pm**	**2pm**	**3pm**	**4pm**	**Test Statistic**
Before Lockdown	27.04	30.54	30.67	30.07	34.84	43.04	51.05
During Lockdown	37.2	46.47	50.27	51.57	52.57	52.2	$\chi^2 > \chi_{0.05,5}$
Evening	**5pm**	**6pm**	**7pm**	**8pm**	**9pm**	**10pm**	**Test Statistic**
Before Lockdown	40.21	27.94	20.31	15.72	12.67	10.92	8.41
During Lockdown	47.75	36.25	27.35	20.23	15.81	11.38	$\chi^2 < \chi_{0.05,5}$
Night	**11pm**	**12am**	**1am**	**2am**	**3am**	**4am**	**Test Statistic**
Before Lockdown	9.3	7.02	4.71	3.89	2.78	1.44	2.11
DuringLockdown	8.36	5.55	3.63	2.65	1.78	2.45	$\chi^2 < \chi_{0.05,5}$

Table 1: The average number of trips made at different times of the day from before lockdown and during lockdown. There is also a test statistic column which is the χ^2 value obtained.

This means we can reject the Null Hypothesis as the lockdown had a significant impact on the average number of trips made between 11am and 4pm. Interestingly, there is not a significant reduction in the number of trips made during the morning rush hour of lockdown despite a large number of people now working from home. A possible explanation for this could be due to the free passes that Just Eat Cycles made available to NHS staff in May 2020[11].

We also analysed the destination station of the trips made from before and during the lockdown. Figures 3 and 4 show the stations plotted on a map of Edinburgh with the colour and size of the dot representing the frequency to which that station was visited. From the two figures, it is evident that more trips were made to seaside areas such as Portobello and Cramond Beach during the lockdown.

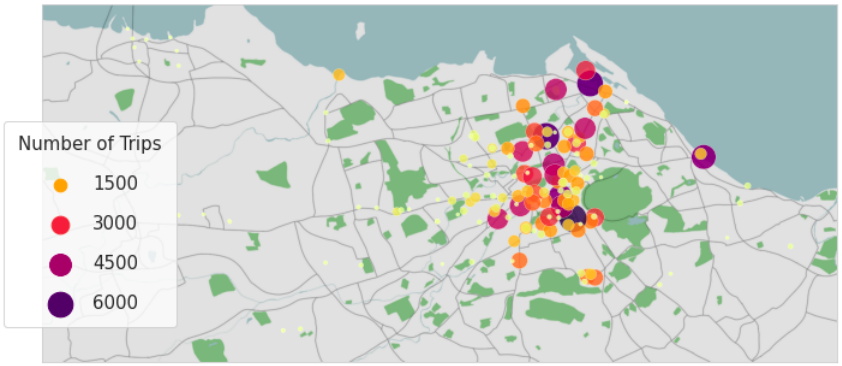

Figure 3: End stations of trips made before lockdown.

We can also see that a larger number of trips were made to the outskirts of the city and less focused on the city center as it was before the pandemic. It is also interesting the note a small increase in trips made to the Queensferry area.

The increase in trips to public spaces in Edinburgh is likely due to the rules of socialisation during the pandemic. For periods of time within the year of lockdowns, it was possible to meet up with people in an outdoor setting. This would influence a larger number of trips to locations where there were more activities to do outside (such as the beach or Leith) or where it would be possible to enjoy the weather during the summer (such as Holyrood Park). All areas that satisfy these criteria saw an increase in visits during the lockdown.

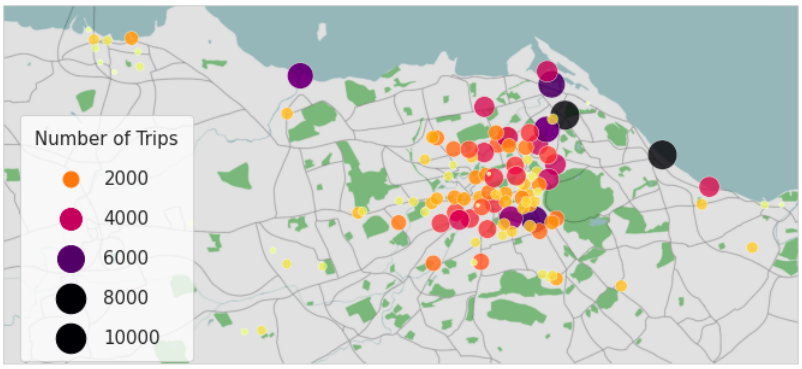

Figure 4: End stations of trips made during lockdown.

Figure 5 shows the frequency of the duration of trips from before the lockdown and during. We see that the means of both time periods appear to be similar but before lockdown there was a larger number of trips of the mean duration. We also see that during lockdown there was an increase in trips of longer time periods, ranging from 30 minutes to 2 hours. From Figure 1 we can see that the number of trips made in lockdown was generally greater than the number made before, so it is interesting to note that the increase in trips is mainly due to trips of longer duration which perhaps supports the hypothesis that more of the trips made were for leisure.

To see if it were actually the case that the change in duration was significant, we decided to construct a confidence interval for the difference in the means of the two populations (before lockdown and during lockdown). Due to the skewness of the data, we also decided that we would initially form a bootstrap sample to analyse in order to normalise it.

The bootstrap sample, β, was formed of 1000 data points $\beta = b_1, b_2, ..., b_{1000}$ where each b_i is the difference in means of two random samples of size 1000 from the pre lockdown data and the during lockdown data with any outliers removed. Let X be the population data of the duration of trips in minutes from before the lockdown and Y be the duration in minutes from during. Then,

$$b_i = \frac{1}{1000}\sum_{i=1}^{1000} x_{k_i} - \frac{1}{1000}\sum_{i=1}^{1000} y_{j_i}$$

where x_i is a data point in X, y_i is a data point in Y, k_i is a uniformly random number from 1 to $\|X\|$, and j_i is a uniformly random number from 1 to $\|Y\|$.

Our Null Hypothesis states that the means of the two samples are the same and so the mean of our bootstrap sample is centered around 0. The Alternate Hypothesis is that the means are different, and so the bootstrap sample mean will be centered around some positive or negative number.

We construct a 95% confidence interval for the mean. As the size of the bootstrap sample is quite large, we can use the Central Limit Theorem to assume that the sample approximates a normal distribution.

$$CI = \left(\bar{x} - Z\frac{s}{\sqrt{n}}, \bar{x} - Z\frac{s}{\sqrt{n}}\right)$$

where $\bar{x}$ is the sample mean, Z is the value from the standard normal distribution, s is the sample standard deviation, and n is the sample size. Therefore, we have a confidence interval of $(-12.26, -9.15)$ to two decimal places. We can be fairly certain that the difference in the means is not 0 and so we can reject the Null Hypothesis. We can also conclude that the duration of trips during the lockdown is, on average, larger than the trips made before the lockdown.

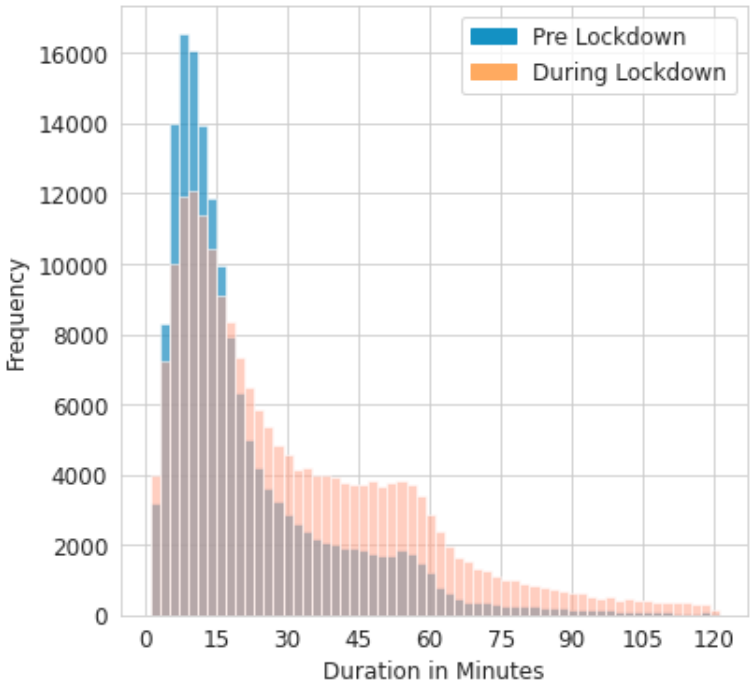

Figure 5: Histogram of the duration of trips made before and during lockdown.

5 Discussion and conclusions

Summary of findings We can see a considerable number of differences between people's habits from before the pandemic and during it with regards to the use of the Just Eat Bikes. During the lockdown and social distancing rules, people spent time outside more frequently and for prolonged periods. They were visiting areas outside of Edinburgh city centre more, such as the beaches in Portobello and Cramond, and visiting the greener spaces within the city too, such as Holyrood Park. The time of day in which these trips were made also saw a change in the late morning and early afternoon as significantly more trips were made in this time during the pandemic than before.

We can attribute a lot of these changes to the effect that the lockdown had on people's lifestyle. The majority of people are now working from home or furloughed from their jobs, and so had the opportunity to make trips outside at around midday during work hours. The restrictions on meeting with other people and leaving your home also would prompt a rise in visits to more scenic locations so that you may take a nice walk or meet up with a friend. Bikes are an incredibly convenient way to get to these destinations as you can also enjoy the outdoors on the way there, as opposed to using a car during the summer. These scenic locations are likely to be further away from where most people live, so we expect longer trips to be made.

It is also possible that lockdown has affected the awareness of people and their effect on the environment. During the first lockdown in March 2020, we saw a significant decrease in air pollution in cities[12] from less traffic on the roads and in the air. Studies have also shown that we have become more environmentally friendly in our consumer habits[13]. These new-found behaviours of ours are also likely to cause an increase in more eco-friendly modes of travel, such as bikes, and so an increase in Just Eat bike usage may have been expected.

Evaluation of own work: strengths and limitations This report took our team a significant amount of work. The major limitation that we found that our original extension of predicting the traffic/cycling speed did not work out due to the limits of multiple linear regression with a higher number of independent variables and how the expected distance was calculated. Due to this, we had to redo our extension. We also found that the Just Eat NHS free pass scheme may have impacted our study such that it is not representative of most cities in the affected months. Our analysis is visually intriguing and contains statistical analysis to support the conclusions that are drawn from the visualisations. This report also explore a number of ways in which the COVID-19 pandemic has affected the habits of those living in Edinburgh.

Comparison with any other related work We first found to have an opposite conclusion to the paper about NYC rides mentioned in Section 2, this may have been due to a number of factors. For example there was the NHS free pass scheme, different travel culture and an alternative approach to the pandemic. We found that our results closely matched conclusions of The Great Bicycle Boom article and the paper going over the change in bike usage across five US cities.

Improvements and extensions For our extension, we will explore how weather and other factors impact the total number of rides each day and if we can use machine learning to predict it effectively.

Our Null Hypothesis, H_0 will be that we cannot predict the number of rides each day based on any of the inputs analysed and our Alternative Hypothesis, H_1 will be that we can predict the number of rides each day.

We first created a heatmap to look at possible correlations with our data. Our analysis immediately noted that both the temperature and lockdown had the most substantial impact on the total number of rides during each day. These factors indicated that they would be good predictors for regression. We see that the average duration has a positive correlation, however we deem this to be too weak for adding it to our regression, similar reasoning applies for the remaining values.

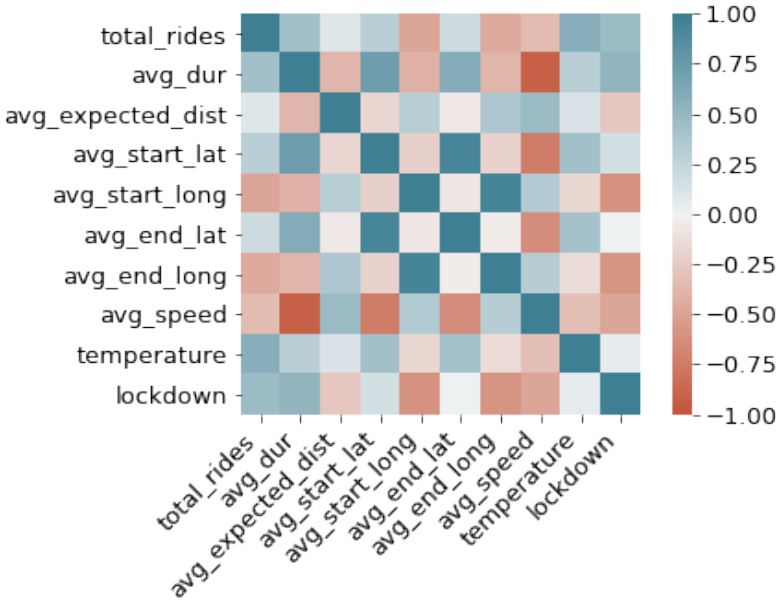

Figure 6: Heatmap of correlations in the Just Eat Cycle data.

We then ran multiple linear regression through two different methods. One which gave us the more practical outlook of the R^2 and allowed us to use test/training samples and another which gave us a more in-depth summary without training.

For the first we used a test size of 0.33; this returned an $R^2 \approx 0.50$, which is a good indication that our predictors are strong. We then ran again, but this time through second method, you can see this result in Table 2.

We can then look at the impact of each of the variables used. The temperature has a coefficient $\approx$ 33, which implies that an increase in temperature causes more bike rides. Lockdown also has a positive coefficient that tells us that lockdown has increased Just Eat bicycle usage. However, to see if these values are a good indicator of the number of cycling rides, we must look at the $P > |t|$, otherwise known as the p-value.

Dep. Variable:	total_rides	**R-squared:**	0.530
Model:	OLS	**Adj. R-squared:**	0.529
Method:	Least Squares	**F-statistic:**	489.6
No. Observations:	871	**Prob (F-statistic):**	4.55e-143

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-65.8330	17.624	-3.736	0.000	-100.423	-31.243
temperature	33.0232	1.381	23.914	0.000	30.313	35.733
lockdown	274.0486	14.487	18.917	0.000	245.615	302.483

Table 2: OLS summary results from statsmodels

This value tells us the probability that our variable has no impact on the dependent variable. Most papers use an alpha of 0.05, and we will use the same here. You can quickly notice that all of our values are much lower than this. Our p values are so small that our summary is unable to display them.

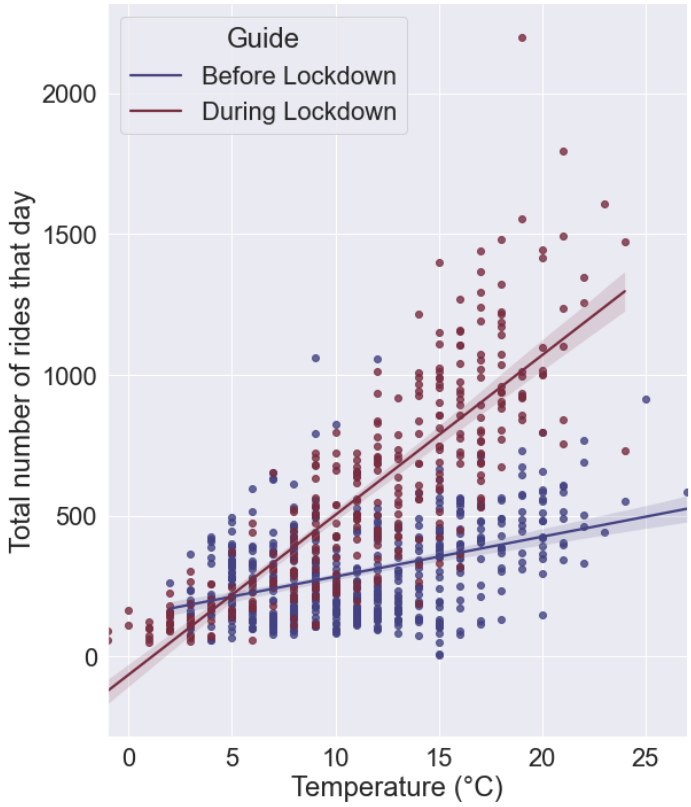

Figure 7: Multiple regression for the total number of rides.

We also created a visual representation for our multiple linear regression to more intuitively understand how it is impacted by our independent variables. There is a clear relation of both lockdown and temperature on the total number of rides a day. We can see that the temperature/weather seems to have a stronger effect during lockdown than before. So in colder temperature the number of people using their bike is reduced during lockdown and in higher temperatures this is increased.

There are numerous possible reasons for why we got the results that we did. The temperature had a strong positive correlation with the number of cycling rides. This relationship is hardly surprising as you would likely expect people to go outside and travel under warmer weather. We also saw that the effect of temperature seems stronger under lockdown. We believe this is due to a decrease in people having to cycle to work or school. Therefore, more people will cycle for recreational purposes. These results closely correlate with the paper going over the impacted of weather when commuting to work by bike.

References

[1] Miguel Lopes. The link between bike sharing and subway use during the covid-19 pandemic: The case-study of new york's citi bike. *https://www.sciencedirect.com/science/article/pii/S2590198220300774*.

[2] BBC. The great bicycle boom of 2020. *https://www.bbc.com/future/bespoke/made-on-earth/the-great-bicycle-boom-of-2020.html*.

[3] Ahmad Ilderim Tokey. Change of bike-share usage in five cities of united states during covid-19. *https://findingspress.org/article/17851-change-of-bike-share-usage-in-five-cities-of-united-states-during-covid-19*.

[4] Brian S Flynn. Weather factor impacts on commuting to work by bicycle. *https://www.researchgate.net/publication/*.

[5] Just Eat Cycles. Historical data. *https://edinburghcyclehire.com/open-data/historical*.

[6] The National Archives. Open government licence. *http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/*.

[7] World Weather Online. Website terms and conditions. *https://www.worldweatheronline.com/terms-and-conditions.aspx*.

[8] Google Maps. Google maps apis terms of service. *https://developers.google.com/maps/terms-20180207*.

[9] Just Eat Cycles. Station map. *https://edinburghcyclehire.com/stations*.

[10] Scottish Government. Timeline of coronavirus in scotland. *https://spice-spotlight.scot/2021/03/12/timeline-of-coronavirus-covid-19-in-scotland/*.

[11] The Edinburgh Reporter. Just eat cycles have free passes for nhs staff. *https://theedinburghreporter.co.uk/2020/05/just-eat-cycles-have-free-passes-for-nhs-staff/*.

[12] The Guardian. Covid-19 lockdowns have improved global air quality, data shows. *https://www.theguardian.com/environment/2020/oct/08/covid-19-lockdowns-global-air-quality-india-london-uk*.

[13] BBC News. Has coronavirus made us more ethical consumers? *https://www.bbc.co.uk/news/business-55630144*.