Ride Like the Wind Without Getting Winded: The Growth of E-Bike Use

Team #16978

Summary

As the policymakers of the United Kingdom see the rising number of e-bike sales and wonder if we can make energy spent on transportation more sustainable by getting cars off the roads, we seek to answer their important questions. These questions involve the forecast for e-bike sales 2 and 5 years from now, understanding significant contributing factors to the growth in e-bike sales, and the quantification of impacts these sales have had on them. This report looks to summarize our findings and illustrate some realizations that were found.

Using MATLAB, we created an exponential regression model to estimate sales in Europe overall to use to predict future sales counts. We then estimated the proportion of that value that the UK would likely be responsible for, using the relative GDPs of the European Union and the UK to properly scale the values.

We then sought to determine whether or not the changes in battery efficiency and gas prices over time had a significant impact on e-bike sales over time. To do this, we created a multivariable function in MATLAB, using regressions of both battery efficiency and gas price relative to sales volume to estimate a 3D function to relate these three variables. Then, we took the partial derivatives of both with respect to time, using the chain rule to help us determine the change in sales with respect to time and to help us prove that the variables were significantly related.

Finally, we analyzed the potential impacts of e-bike usage on air quality and traffic congestion. Our models showed that e-bike usage has the potential to reduce traffic congestion and improve air quality, but additional research may be needed to fully understand the extent of these impacts. Overall, we believe that e-bikes can provide a more sustainable method of transportation for the UK and are definitely worth investing energy towards.

Table of Contents

Introduction

The rise of electric bikes has been astronomical in nature for public consumption. As improvements have been made to their energy efficiency and battery technology, along with more environmental awareness, e-bikes have become an increasingly useful method of transportation. In order to assist the UK Department of Transport, we have analyzed data regarding Europe and the UK.

For the first problem, we were tasked with predicting the growth in e-bike sales. We were then tasked with predicting the number of e-bikes that will be sold two years from now and five years from now.

In the second problem, we were asked to determine both what factors contributed to e-bike sales and how said factors did so.

In the third problem, we were tasked with quantifying any factors that may have been impacted based on an increase in e-bike usage.

1) Q1: The Road Ahead

1.1) Defining the Problem

We were tasked with developing a mathematical model that would accurately estimate the number of e-bikes sold two and five years from now.

1.2) Assumptions

1. E-bike technological advancements will continue at the same rate.
   Justification: Technological advancements are still being made in the e-bike space (specifically in battery technology), and many recent products have lowered the prices of e-bikes or increased their efficiency [1].
2. E-bike sales will continue to grow based on past trends (accounting for COVID-19).
   Justification: Although the COVID-19 pandemic may have impacted E-bike sales, the impacts were not large enough to significantly alter the data and produce outliers [2].

3. European sales can be specified to the UK.
   Justification: E-bike regulations do not greatly differ from country to country and the United Kingdom follows similar regulations to the European Union, so it should be acceptable to compare them regarding their relative GDPs [3].
4. The United Kingdom has a proportional amount of sales relatively similar to its GDP.
   Justification: As the British economy improves, e-bike sales should also increase as citizens will have the funds available to purchase said bikes.

1.3) Analysis

Symbol	Definition	Unit
t	Time	Years
S	E-bikes sold in Europe	E-bikes (in thousands)

Table 1: Definition of variables for e-bike sale regression model

1.4) The Model

We initially attempted to model the past trend of e-bike sales in Europe using a linear regression model. However, we realized that it may not be the best fit due to the low coefficient of determination (or R-squared value).

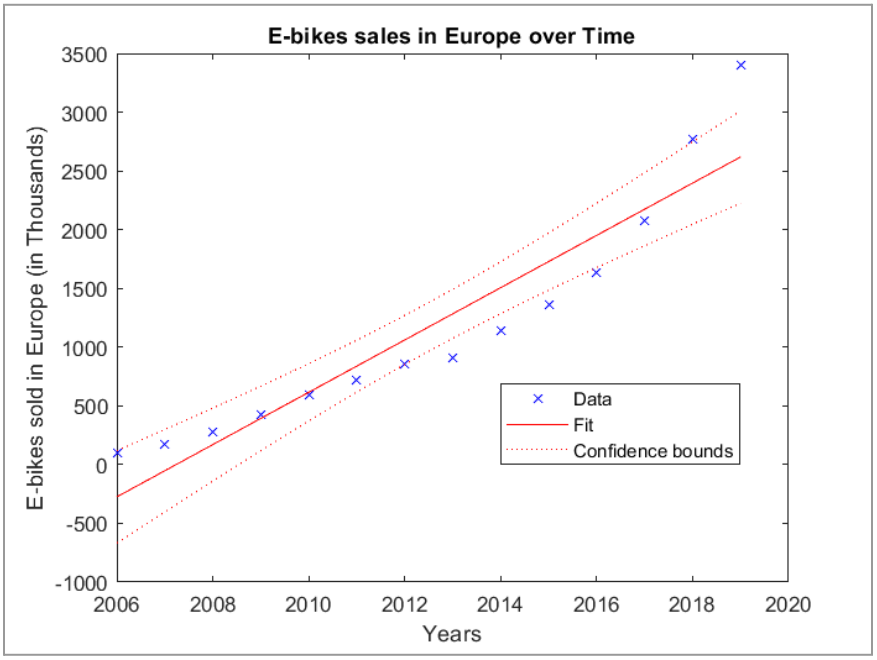

Figure 1: Linear regression model of e-bike sales

Then, we examined the data of e-bikes sold in Europe [4] and created an exponential regression model which fit the data much better than the linear regression model. This line was used to predict the e-bike sales in Europe overall in 2025 and 2028.

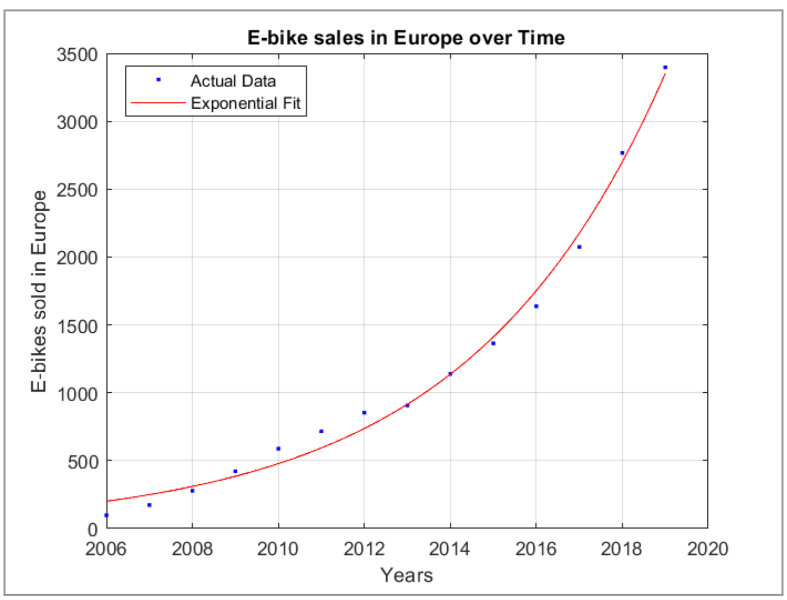

Figure 2: Exponential regression model of e-bike sales

Then, we used the ratio of the GDP values between the UK and Europe to determine an estimate of the e-bikes sold in the UK alone.

1.5) Results

Using MATLAB, we were able to clean the data and remove all non-existent values. Upon removing those values, we constructed a linear regression model (and later an exponential-regression model) that was able to be extrapolated from to obtain additional data points.

Models	Equation	Coefficient of Determination (R^2)
Linear-Model	$S(t) = 222.6t - 446810$	0.881
Exponential-Model	$S(t) = (7.077 * 10^{-187})e^{(0.2163t)}$	0.9928

Table 2: Equations and calculations for e-bike sale regression model

Current GDP of the European Union: $14.45*10^{12}$ Euros [5]

Current GDP of the UK: $3.131*10^{12}$ US Dollars [6]

Current Exchange Rate: 1 Euro per every 1.06 USD

Current GDP of the UK in Euros: $2.95377*10^{12}$ Euros

Ratio of GDP of UK to GDP of European Union: 0.204413 = P

$$P * S(2025) = 2424.3696$$

$$P * S(2028) = 4638.8734$$

We predict there will be about 2,424,370 e-bikes sold in the UK in 2025.

We predict there will be about 4,638,873 e-bikes sold in the UK in 2028.

1.6) Discussion

This model can be tested for accuracy by obtaining the actual data for e-bike sales in the UK specifically and assessing the difference between the true and the estimated values. The stability of this model is very low since it was solely based on the number of sales in the past few years. Both the accuracy and sensitivity to assumptions can be tested by regenerating the equation using all but the most recent data point, and checking whether the model accurately predicts the unincluded value. If so, then the assumptions made still result in an accurate prediction of the e-bike sales.

Strength:

- The model displayed a very strong coefficient of determination, indicating that there was an association to be determined between the independent and dependent variables.

Weakness:

- We may have overfitted the regression model to the data (which was hard to avoid due to the low volume of available data), so it may not necessarily be a good predictor for data points not included in the set used to build the regression.

1.7) Technical Computing

The first issue that needed to be addressed was the lack of data for some entries. MATLAB was used as the primary software tool to clean and remove the data. We also used MATLAB to construct the regression fits to extrapolate our data from. Within MATLAB, we specifically used the curve-fit app which gave us the appropriate model and suggested coefficients for us to use. We were also able to generate the R^2 values for these models as well and these were computed by the MATLAB software.

2) Q2: Shifting Gears

2.1) Defining the Problem

We were tasked with considering one or more possible contributing factors to the growth of e-bike usage and creating mathematical models to justify whether or not those factors were significant reasons for the growth in e-bike sales.

2.2) Assumptions

1. Both variables considered in the model affected the number of e-bike sales.
   Justification: The increasing price of gas would cause people to more frequently utilize vehicles that do not need gas (like e-bikes), which can increase the number of sales. The increasing battery energy densities make batteries last longer for cheaper, allowing for more efficient batteries for e-bikes, which can also increase their sales volume.
2. The linear regression models for gravimetric energy density and gas price were accurate representations of their values over time.
   Justification: The respective coefficients of determination for gravimetric energy density and gas price were 0.989 and 0.869, meaning that they are both generally accurate models of these variables.

2.3) Analysis

Variables/Constants	**Values**
A	881.5
B	$-1.384 * 10^{15}$
C	$1.384 * 10^{15}$
D	$1.532 * 10^{15}$

E	$-\ 5.415 * 10^{15}$
F	$3.883\ *\ 10^{15}$
P(t)	$3.226t\ -\ 6380$
G(t)	$7.431t\ -\ (1.453\ *\ 10^{4})$

Table 3: Variables and constants for the multivariable model

2.4) The Model

Upon selecting the price of gas and the gravimetric energy densities as being significant factors influencing the rise of e-bike sales due to their logical relationship with it, we decided to construct a multivariable-regression model. The regression model generates 6 coefficients, one for each term in the polynomial. We set the functions P(t) and G(t) equal to x and y respectively to illustrate how the sales were a function of the price of gas and gravimetric energy densities. Those were, in turn, functions of time. Taking the partial derivatives of S concerning P, G, and later t using the chain rule, we were able to see the effects of these values on sales.

$$S\ =\ A\ +\ Bx\ +\ Cy\ +\ Dx^2\ +\ Exy\ +\ Fy^2$$

$$Let\ x\ =\ price\ of\ gas(pence/litre)\ =\ P(t)$$

$$Let\ y\ =\ Gravimetric\ Energy\ Densities\ (Wh/kg)\ =\ G(t)$$

$$\frac{dS}{dt} = \frac{\delta S}{\delta P}\frac{dP}{dt} + \frac{\delta S}{\delta G}\frac{dG}{dt}$$

$$\frac{\delta S}{\delta P} = B\ +\ 2Dx\ +\ Ey$$

$$\frac{\delta S}{\delta G} = C\ +\ Ex\ +\ 2Fy$$

Figure 3: Polynomial regression and partial derivative equations

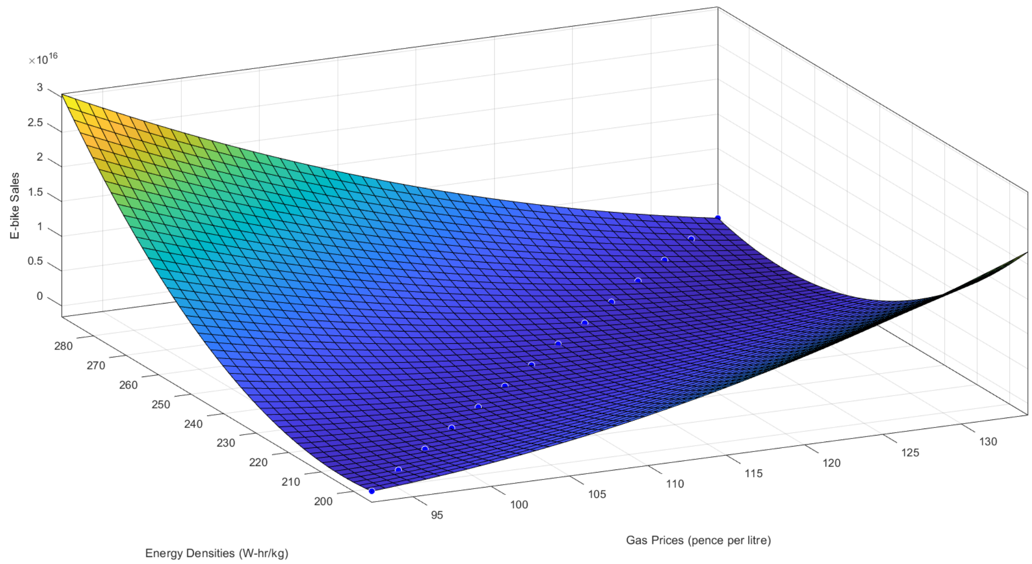

Figure 4: Graphical representation of multivariable model

2.5) Results

This model returned an R^2 coefficient of 0.9905, indicating that the model does accurately model the parameter in question. Upon calculating the partial derivatives of the sales with respect to time, we found that both the partial derivatives for the price of gas and the gravimetric energy densities were positive across the interval we were focusing on. This meant that both factors had a positive relationship with e-bike sales in the UK.

2.6) Discussion

We did the necessary backtesting for this model to ensure its validity. To do this, we entered sample data into the model and the results matched the actual data. To account for overfitting, the regression model can only be extrapolated to limited time in the future. Additionally, the sensitivity to assumptions can be tested by measuring the impact that other variables have on the total e-bike sales since this model does not account for those at all.

Strengths:

- This model allows us to incorporate the impact of more than a single variable on the total UK e-bike sales.
- The 3D model allows us to generate a relatively accurate estimate for any combination of gas price and gravimetric energy density values.

Weakness:

- The data that the model was built on was poor. This can be seen as there were not enough entries to build a proper regression model. Furthermore, some of the data was missing and/or removed altogether.

2.7) Technical Computing

We used MATLAB to generate this model. To begin, we first developed linear regression models to represent gas prices over time and battery power over time. After this, the model was used to fill in missing data points, and we developed a matrix that contained the years 2006-2019, the filled-out gas prices, the filled-out battery efficiency metrics, and the completed sales data from Q1. Then, we sent that matrix to the curve fit application to build the multivariable model. Using the multivariable model, we calculated the partial derivatives to see how each factor contributed to the growth of the sales price over time.

3) Q3: Off the Chain

3.1) Defining the Problem

We were tasked with examining factors that may have been impacted by increases in e-bike usage in the UK, such as carbon emissions and traffic, then quantifying any results.

3.2) Assumptions

1. An increase in e-bike sales results in a decrease in the use of other vehicles.
   Justification: Many people buying e-bikes previously used other methods of transportation that they would most likely use less after obtaining an e-bike.
2. E-bikes have lower carbon emissions than other vehicles.
   Justification: E-bikes use batteries while most other common vehicles use gas. Battery usage will have a lower carbon footprint than gas usage.

3.3) Analysis

To determine whether the increase in e-bike sales had an impact on the CO_2 emissions from vehicles, we had to analyze the provided data for the transport distance of different types of vehicles [4]. We used those numbers in combination with the average passengers per vehicle type and the average CO_2 emissions per vehicle type from each year to calculate the average CO_2 emissions from each vehicle type each year in the UK [7, 8, 9, 10, 11]. By graphing the emissions from each type of vehicle in a stacked area chart with e-bike sales on the x-axis, we can see that the total emissions have significantly decreased as the number of e-bike sales increased.

To determine how e-bikes may have impacted motor vehicle traffic, we found a CSV file on the UK's government website with data for the total numbers of motor vehicles in traffic in various regions of the UK for the years 2000-2020 [12]. We estimated the median total number of motor vehicles across all of the regions for each year, graphed the data in Google Sheets, and took a linear regression to help determine if motor vehicle traffic decreased over time as e-bike sales increased in the UK.

3.4) The Model

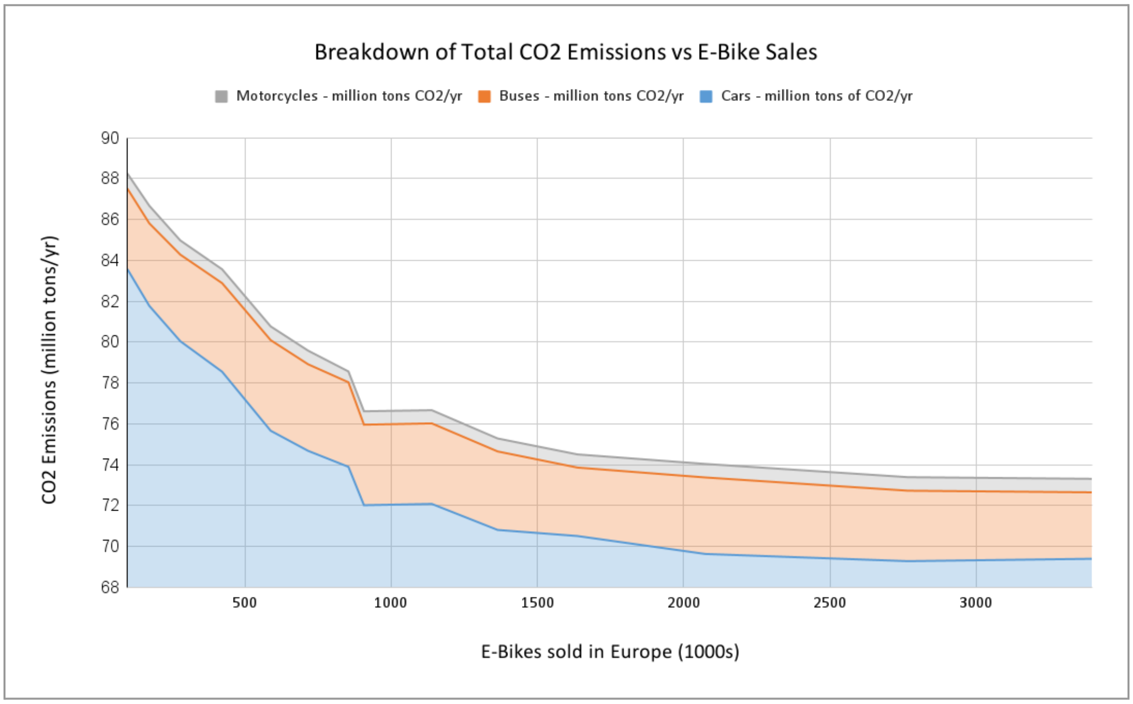

Figure 5: Stacked area chart of CO_2 emissions in Europe

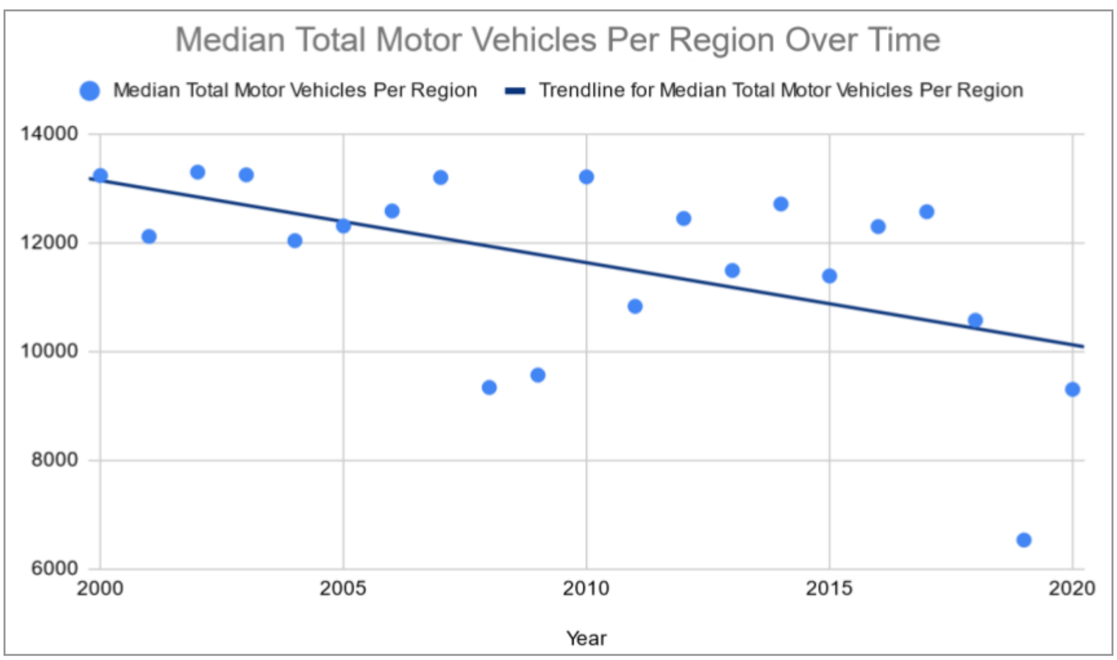

Figure 6: Linear regression model of vehicle counts in the UK

3.5) Results

After graphing each of the vehicle emissions data with respect to e-bike sales, we found that the total carbon emissions of motorcycles, buses, and cars in Europe decreased over time as the number of e-bikes sold in Europe increased.

Additionally, after analyzing the data about traffic counts in the UK over recent years, we found that the median total number of vehicles per region decreased over time.

3.6) Discussion

The carbon emission model can be tested for accuracy for the UK by comparing it to more data specifically regarding carbon emissions and sales in the UK over just extrapolating from Europe. This model could be more stable because there could be other factors affecting the carbon emissions of these vehicles, such as people using electric vehicles.

The vehicle count model can be tested for accuracy by utilizing all of the UK road traffic data instead of just a simple random sample. However, this model is fairly stable due to the large sample size with a large database source.

Strengths:

- The vehicle count model is based on the median number of vehicles rather than the mean number of vehicles, making our model more resistant to potential outliers from the simple random sample.
- The vehicle emissions model accounted for multiple types of vehicles (cars, buses, motorcycles) instead of just one.

Weaknesses:

- The vehicle count model utilized a simple random sample of the full data, which may give slightly less accurate results due to chance.
- The vehicle emissions model uses a large number of rounded averages in estimating, so its predicted values may slightly vary from the true values.

3.7) Technical Computing

It turns out that the CSV file with the UK traffic data had 33 data columns and over 500,000 rows. When we initially tried to load it into Google Sheets, it exceeded the maximum permitted sheet size. We began writing a Python script (csvCondenser.py) to parse through each row and only select data from the 3 columns we were interested in, which it would then write to a new CSV file. When we tried to run that, the program took an unreasonably long time to run and did not even finish executing before we decided to cut it short. We found that it could process the first 10,000 lines relatively quickly, but 500,000 seemed to be too much. To resolve this, we opted to take a random sample of 10,000 of the lines (following the 10% rule) so we could still get a relatively good representation of the data while not having to process every line.

We uploaded the resulting CSV file to a Google spreadsheet, sorted the data by year, and graphed the data on a scatter plot. We found that graph to not be useful, so we opted to calculate summary statistics using a CSV file version of that spreadsheet. We wrote another Python script (makeSumStats.py), which analyzed the data for each year and calculated annual means and medians, writing the results to yet another CSV file. Then, we uploaded that data into a spreadsheet and took a linear regression of both mean and median over time (though we decided that the median was a more worthwhile indicator to use).

4) References

[1] https://transmitter.ieee.org/technology-and-the-e-bike-boom/[1] Akar, Simay. “Tech That Paved the Way for E-Bike Boom.” IEEE Transmitter, 27 Oct. 2021, transmitter.ieee.org/technology-and-the-e-bike-boom/.

[2] “How Have Supply Chain Shortages Affected the E-Bike Industry? - Features.” *BikeBiz*, 27 Dec. 2021, bikebiz.com/how-have-supply-chain-shortages-affected-the-e-bike-industry/amp/.

[3] McBride, Ryan. “Electric Bike Regulations in the EU and UK | EBike Laws.” *Gadget Review*, 1 Mar. 2023, www.gadgetreview.com/ebike-regulation-in-the-eu-and-uk#:~:text=EBikes%20are%20allowed%20to%20use.

[4] Ride Like the Wind, MathWorks Math Modeling Challenge 2023, https://m3challenge.siam.org/node/596.

[5] “European Union (EU) - Gross Domestic Product (GDP) until 2018 | Statista.” *Statista*, Statista, 2018, www.statista.com/statistics/279447/gross-domestic-product-gdp-in-the-european-union-eu/.

[6] “GDP (Current US$) - United Kingdom | Data.” *Data.worldbank.org*, data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2021&locations=GB&start=2021.

[7] Brown, Colin. *Motorcycle Carbon Emissions*. June 2020.

[8] Carlier, Mathilde. “Average Car and van Occupancy England 2002-2018 Statistic.” *Statista*, 31 Jan. 2023,

www.statista.com/statistics/314719/average-car-and-van-occupancy-in-england/#:~:text=In%202017%2C%20there%20were%201.55%20people%20in%20a.

[9] “Emissions from Bus Travel.” *Www.carbonindependent.org*, 14 Oct. 2022, www.carbonindependent.org/20.html#:~:text=This%20calculates%20the%20average%20CO%202%20emission%20from.

[10] US EPA. “Greenhouse Gas Emissions from a Typical Passenger Vehicle.” *US EPA*, US EPA, 30 June 2022, www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle.

[11] Yurday, Erin. “Average CO2 Emissions per Car in the UK.” *Www.nimblefins.co.uk*, 24 Apr. 2020, www.nimblefins.co.uk/average-co2-emissions-car-uk.

[12] “Road Traffic Statistics.” *Dft.gov.uk*, 2018, roadtraffic.dft.gov.uk/downloads.

5) Code Appendix

Q1Script.m

```
% Years column from spreadsheet
years = TCP23data(:, 1);
% Europe sales column from spreadsheet
salesInEurope = TCP23data(:, 3);

% clean data: remove last 3 rows because they are NaN
years = years{1:14, :};
salesInEurope = salesInEurope{1:14, :};

% Linear Regression
line = fitlm(years, salesInEurope)

% Exponential Regression (refer to separate function)
model = ExponentialFitEurope(x,y)

% Current year: 2023
% Predict 2 years from now
model(2025)
% Predict 5 years from now
model(2028)
```

ExponentialFitEurope.m

```
function [fitresult, gof] = ExponentialFitEurope(x, y)
%CREATEFIT(X,Y)
%  Create a fit.
%
%  Data for 'ExponentialE-bikesEurope' fit:
%      X Input: x
%      Y Output: y
%  Output:
%      fitresult : a fit object representing the fit.
%      gof : structure with goodness-of fit info.
%
%  See also FIT, CFIT, SFIT.

%  Auto-generated by MATLAB on 03-Mar-2023 08:50:25


%% Fit: 'ExponentialE-bikesEurope'.
[xData, yData] = prepareCurveData( x, y );

% Set up fittype and options.
ft = fittype( 'exp1' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.StartPoint = [2.6946180791496e-178 0.206514653522816];

% Fit model to data.
[fitresult, gof] = fit( xData, yData, ft, opts );

% Plot fit with data.
figure( 'Name', 'ExponentialE-bikesEurope' );
h = plot( fitresult, xData, yData );
legend( h, 'Actual Data', 'Exponential Fit', 'Location', 'NorthEast',
'Interpreter', 'none' );
% Label axes
xlabel( 'Years', 'Interpreter', 'none' );
ylabel( 'E-bikes sold in Europe', 'Interpreter', 'none' );
```

```
title('E-bike sales in Europe over Time')
grid on
```

csvCondenser.py

```
import random

with open('dataToClean.csv') as f:
    lines = f.readlines()

columnTitles = lines[0].split("\",\"")
columnTitles[len(columnTitles) - 1] = "All_motor_vehicles"

titlesToUse = ["Year", "Region_name", "All_motor_vehicles"]
columnNumbers = []
outputString = ""

for title in titlesToUse:
    columnNumbers.append(columnTitles.index(title))
    outputString += "\"" + title + "\","

outputString += "\n"

for i in range(1, 10000):#len(lines)):
    line = lines[random.randint(1, len(lines) - 1)]
    lineArray = line.split("\",\"")
    lineArray[len(lineArray) - 1] = lineArray[len(lineArray) - 1][:-2]
    for index in columnNumbers:
        outputString += "\"" + lineArray[index] + "\","
    outputString = outputString[:-1]
    outputString += "\n"

with open('cleanData.csv', 'w') as f:
    f.write(outputString)
```

makeSumStats.py

```
import statistics

with open('sortedData.csv') as f:
    lines = f.readlines()

# print(lines)

outputString = "\"Year\",\"Mean\",\"Median\"\n"
yearData = []
year = "2000"

for line in lines:
    if line[:4] != year:
        outputInfo = [year, int(statistics.mean(yearData)), int(statistics.median(yearData))]
        for data in outputInfo:
            outputString += "\"" + str(data) + "\","
        outputString = outputString[:-1]
        outputString += "\n"
        yearData = []
        year = line[:4]
    else:
        dataPoint = line[5:-1]
        dataPoint = int(dataPoint)
        yearData.append(dataPoint)

with open('trafficDataSummaryStatistics.csv', 'w') as f:
    f.write(outputString)
```