MathWorks Math Modeling Challenge 2023

St. John's School

Team #16573, Houston, Texas

Coach: Dwight Raulston

Students: Ananya Das, Caden Juang, Oliver Lin, Ben Lu, Addison Spiegel

M3 Challenge RUNNER UP—$15,000 Team Prize

JUDGE COMMENTS

Specifically for Team # 16573 —Submitted at the Close of Triage Judging:

COMMENT 1: The paper is well written, providing a clear and detailed background alongside comprehensive model descriptions. Notably, the strengths and weaknesses of the models are explicitly stated and practical in nature. The author(s) have made great use of previous research and information, justifying the proposed models through reasonable assumptions. Throughout the paper, the data tables effectively demonstrate the necessity of the dataset for modeling purposes. Additionally, the model results are reasonably well explained and quantified through the use of tables.

COMMENT 2: Your Executive Summary was well written! However, it failed to stay within the 1 page limit. / Does it seem reasonable that the carrying capacity for e-bikes is the entire population of a country? That is, are they a viable transportation option for each individual? / You had a nice transition between Q1 and Q2. Well done! I also enjoyed the heat maps you presented in Q2. They added a nice readability to your work in a compact way. / Do you know of any way to test for causation between variables? Remember, there may be confounding variables between two ideas, it might not be causation! / It seems slightly strange that you were unable to locate a reference for Assumption 5 in 3.2. Did you try to find the average commute length for people? On the flip-side of this, It was nice to see the split of e-bike usage by age, as not all people's primary use of e-bikes is for commuting. / Overall, well done! Your conclusion really highlights this as a report to someone in the transportation department. It was a nice way to end your discussion.

COMMENT 3: Your summary is well-written and you do a good job of presenting your conclusions, but perhaps a few too many details for a summary concerning your methods. Nice job explaining your first model and you find the desired predictions. I particularly liked your inclusion of the inflection points of the models, and you might have commented on that in terms of what may be an interesting difference in future paths of the two countries. The graphics were appreciated but you should have selected a different vertical scale for your regression pictures as the points are grouped at the very bottom. Good explanation of your second model and well done to realize and state the limitations of correlation when it comes to making conclusions about causation. While you explain the process well it would have been nice to see some examples of the data you used for each factor. Nice graphics. It was fun to see many different model types employed throughout. Your third model had some strong assumptions built into it and again I would have liked to have seem more information about the data you ""gathered"", but you do a nice job of explaining the process and pointing out some potential weaknesses of your approach.

COMMENT 4: This is an excellent essay, congratulations! You provide plenty of sources to sustain your arguments and assumption. The structure and presentation are very sharp. Few remarks for improvement: you could achieve a higher precision in Q1 by distinguishing between rural and urban areas, also by taking part of the population rather than total population. In Q2 you seem to focus on values but do not refer to the statistical significance of each variable. In Q3 you could have included a sensitivity analysis.

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

Executive Summary

In 1993, the first electronic bicycle was sold by Yamaha Motor Company to Japanese consumers as a solution to the high price of gasoline in Japan. Over the next 30 years, the production and capabilities of e-bikes soared, surpassing cars in affordability, speed in short-distance commutes, and popularity due to their low-emission capabilities. As a result, by the end of 2023, there will be an estimated 300 million e-bikes around the world. [1]

We first created a logistic model of the sales of e-bikes in the UK and US to determine the projected future sales of e-bikes. We applied this model to past annual sales data from both nations to predict their respective annual e-bike sales within two and five years. To determine the carrying capacity of both models we divided the total population of both countries by the average lifespan of an electric bike, a figure we assumed to remain constant over time. Applying the model, we predicted that in 2025, $332,824$ new e-bikes would be purchased in the United Kingdom while $563,780$ new e-bikes would be purchased in the United States. Additionally, in 2028, we found that an estimated $1,907,964$ and $3,796,118$ e-bikes would be sold in the United Kingdom and United States respectively. Next, we ran a sensitivity model of our logistic coefficient and determined that our sale predictions in the UK had an average variation of 0.266 percent, and our sale predictions in the US had an average variation of 0.018 percent. This analysis demonstrated the accuracy of our projection of future e-bike sales.

To determine what variables were the most significant in influencing our projected growth of annual e-bike sales, we compiled seven factors prevalent in academic literature and established a theoretical framework for their effect on e-bike sales. By finding the pairwise linear correlation coefficient for all independent variables, we first isolated the variables with a low coefficient value, and subsequently, variables that most likely lacked a significant causal relationship with e-bike sales. Examples of such variables included US Gas Prices, UK GDP Per Capita and Consumer Spending. Next, we employed a multivariate linear regression of the seven variables on annual e-bike sales to determine the relative strength of causation of each variable. We concluded that in the US, economic factors related to increased spending propensity and decreased price had the most significant effect on e-bike sales. In the UK, the most significant factors were social, such as trendiness or attitude towards health and fitness. Our conclusions were confirmed by sensitivity analysis, providing us with an accurate assessment of what factors US and UK DoT officials should prioritize in order to increase the number of e-bike sales.

Finally, we quantified the impact of the projected rise in US/UK e-bike sales on net carbon emissions and average projected lifespan of the population. We used a Monte Carlo simulation to generate a distribution of e-bike owners based on their riding habits. Using the

range of their journey and their replaced mode of transportation, we calculated the carbon emissions saved by riding an e-bike instead of a car or bus. For health benefits, we observed that people exercise while operating e-bikes. Since exercise-minutes are a significant factor in lifespan prediction, we approximate the amount of time spent riding the e-bike to estimate the increase in life years. We then used our Part I.) to model the environmental and health benefits over a ten-year range. In 2028, we project that the United States will save 143 million kilograms of CO2 per year and the UK will save 29 million per year due to e-bikes. In terms of health benefits, the United States will save 9.7 million life years and the United Kingdom will save 1.4 million life years prior to 2028.

Contents

1 Part I: The Road Ahead

1.1 Restatement of the Problem

E-bike technology, popularity, and sales have skyrocketed since in the early 1990s, especially in Western nations. Understanding their projected growth rate is vital for government organizations to predict changing trends in transportation, which is a critical component of their economies. Officials also ought to prepare e-bike-compatible infrastructure in accordance with the growth of the e-bike industry. In this section, we predicted the growth of e-bike sales that will be sold two and five years from now in the United States and United Kingdom.

1.2 Assumptions

1. **E-bikes are two-wheeled bicycles with electric motors that can assist or replace pedaling.** This is the dictionary's definition of an e-bike.[2]

2. **The maximum number of people in a country who will buy an e-bike is equivalent to the total population of that country.** The US and UK lack regulations on who can own an e-bike, so anybody is capable of owning one. We note that due to a lack of necessity, consumers rarely choose to buy more than one e-bike per person.[4]

3. **The growth of e-bikes follows a logistic curve.** This is the growth pattern that other popular technologies, such as the iPhone, have followed. This is a continuation of our previous assumption: there are a finite number of e-bike sales.[3]

4. **People replace their bikes approximately as quickly as the batteries decay.** We cannot predict developments in e-bike form or function. Therefore, bikes only get replaced when they operate poorly, instead of for any new performative or stylistic advantage.

5. **E-bikes have an average battery lifespan of 5 years.** This is the current lifespan of lithium-ion batteries, which is unlikely to improve substantially within our 5-year growth model. Moreover, other technology companies have historically prioritized increasing performance over the improving lifespan in order to maximize sustained long term profits.[5] Those that start using e-bikes generally continue to use them, according to the general manager of the biking division at Lyft. [15]

6. **The growth of e-bikes is independent of which types of e-bikes are purchased.** Some e-bikes provide pedal assist, while others can pedal the user independently. We assume that consumers will simply buy the e-bike that suits their personal preferences, so we do not distinguish the growth of e-bikes overall from the growth of specific varieties.[7]

1.3 Model Development

To predict the future growth in e-bike sales in the United States and United Kingdom for the next 2 and next 5 years, we first found data providing annual sales in both countries. In both the United States and United Kingdom, our data spans the most recent decade of 2012 – 2022 [22] [23] [24] [25] [26] [27] [28].

Year	United States: E-Bikes/Yr	United Kingdom: E-Bikes/Yr
2012	70	20
2013	159	25
2014	193	50
2015	130	40
2016	152	75
2017	220	55
2018	369	70
2019	423	101
2020	416	160
2021	750	160
2022	928	170

Table 1: Annual E-bike Sales in the US and UK (thousands)

Using this data, we created two independent logistic models to determine the projected number of e-bikes sold in both countries. The logistic formula and variables are shown below:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

Symbol	Variable	Unit
L	Carrying Capacity	1000 bikes/year
x	Year	Year
x_0	Inflection Point	Year

Table 2: Variables in Logistic Function

We determined that a logistic model would be the most appropriate predictor of e-bike sales as other popular tech products such as Apple's iPhones have also proven to reach a carrying capacity a few years after its initial spike in popularity. Once enough e-bikes are sold per year, the market reaches complete saturation which we determined is one electric bike per person.

To determine the carrying capacity of both logistic models we employed the following formula and variables:

$$L_c = \frac{P_c}{S}$$

Symbol	Variable	Unit
L_c	Carrying Capacity by Country	1000 bikes/year
P_c	Total Population by Country	People
S	Lifespan of E-bikes	Years

Table 3: Variables in Carrying Capacity Function

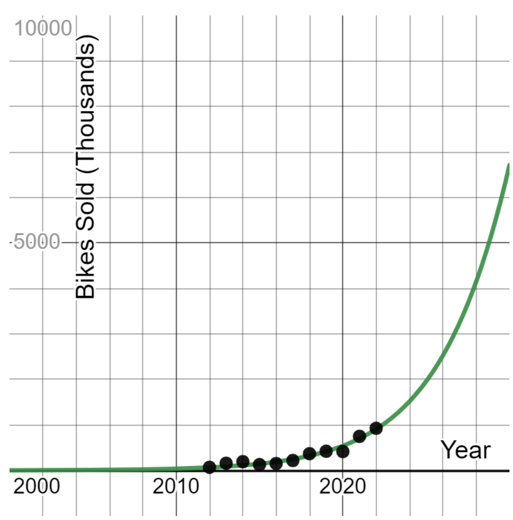
We assume the carrying capacity is equal to the total population of the respective country divided by the average lifespan of an electric bike. The average lifespan for e-bikes has remained at around 5 years historically. [6] This equation outputs a prediction of maximum sales of e-bikes per year – our carrying capacity – for the near future. Thus, we calculate:

$$L_{uk} = \frac{67,000,000}{5} = 13,400,000 bikes/year$$

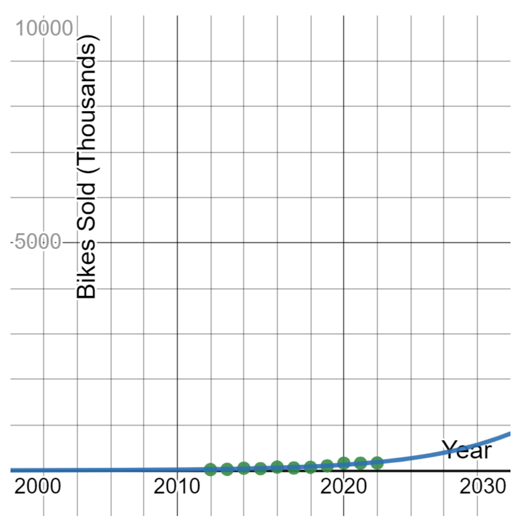
$$L_{us} = \frac{330,000,000}{5} = 66,000,000 bikes/year$$

1.4 Results

Using the predictive logistic model established in the previous section, we were able to forecast the sales of E-bikes in the United States and United Kingdom for 2 and 5 years into the future on Desmos.



(a) US Logistic Regression



(b) UK Logistic Regression

Our carrying capacity L was determined as the total population divided by the lifespan of a bike (5 years). k is a relative growth coefficient, and is calculated to fit the data. x_o is the year at which half the maximum number of sales is reached. This is the point at which half the population is now purchasing an E-bike every five years, which makes sense since approximately half the population in the United States has a commute of fewer than ten miles, the perfect distance for e-biking. [21] The rate of sale growth will slow since the rest of the population does not have an immediate need for an e-bike. Not having the appropriate commute distance, willingness to adapt to alternative transportation methods, and/or fitness to adequately operate a bicycle among other possibilities, this segment of the population will slowly purchase E-bikes, possibly as a device of leisure or to experiment.

The regressions output the following values of these variables:

Variable	US Value	UK Value
L_c	66,000,000	13,400,000
k_c	-0.261688	-0.192052
x_o	2038	2044

Table 4: Logistic Regression Calculated Constants

Thus, we get the following equations projecting future e-bike sale growth and give us the following values of projected bike sales in both countries:

$$f(x)_{us} = \frac{13,400,000}{1 + e^{.261688(x-2038)}}$$

$$f(x)_{uk} = \frac{66,000,000}{1 + e^{.192052(x-2044)}}$$

	2025	2028
United Kingdom	322.824	563.78
United States	1907.964	3796.118

Table 5: Projected Sale of E-Bikes (thousands)

1.5 Sensitivity Analysis

To determine the accuracy of the k-coefficient we calculated, we jittered each data point by a small amount of Gaussian noise and recalculated the resulting k factor, keeping L_c and L_o constant. We then plugged that k value into the original logistic equation to make a new prediction and calculated the percent difference between the original prediction and the new prediction. We repeated this process five times and averaged the percent differences. This process determined that our sale prediction in the UK had an average jittered variation of

0.266 percent, and our sale predictions in the US had an average jittered variation of 0.018 percent. Because these values are relatively low, we can be relatively confident in our model's resilience to random error.

1.6 Strengths and Weaknesses

1.6.1 Strengths

1. **Simplicity.** The model builds upon a basic exponential model, so it can easily be understood and expanded to consider additional data.
2. **Broadness.** The model encapsulates a plethora of auxiliary factors. Nearly any trend, from private investment to social prominence, is reflected by the general development of e-bike sales.
3. **Thrift.** The model is accurate with smaller datasets. A logistic regression provides an accurate prediction with sparse datasets. Since long-term e-bike sale data is difficult to find, logistic regressions are more reliable than alternative, more complex models.

1.6.2 Weaknesses

1. **Generalization.** Without the ability to weigh specific variables, the model assumes that the parameters which affected growth during 2012-2022 remain constant. For example, the increase of positive attitudes towards e-bikes might slow down, prematurely limiting growth of sales in a way not represented by our model.
2. **Carrying Capacity.** Our carrying capacity value is based on the total population and does not account for personal preferences or conditions that would prevent people from buying an e-bike. This limits the accuracy of our logistic curve.
3. **Specificity.** Our model can only demarcate the overall trends in the market. External disruptions like recession or shifts in public sentiment cannot be isolated from the overall trend.

Our model did not consider the factors that contribute towards increasing sales, such as the trendiness of e-biking or attitudes towards health and wellness. We investigate the impacts of these factors and more in the next section.

2 Part II: Shifting Gears

2.1 Restatement of the Problem

Since the sale of e-bikes is projected to continue its meteoric rise in both the United States and the United Kingdom, it is critical that officials in both countries understand the underlying causes propelling the growth. This knowledge will allow government organizations effectively craft legislation and stimulate the necessary factors in order to achieve their respective environmental and economic goals. In this section, we evaluated the relative importance of factors that influence the number of sales of e-bikes: gas prices, environmental consciousness, GDP per capita, consumer spending, e-bike trendiness, attitude towards health and fitness, and median age.

2.2 Assumptions

1. **The price of e-bikes is proportional to the price of lithium batteries.** Lithium batteries are both a primary cost of production and a consistent percentage of production cost. The price of any product is proportional to the price of its inputs and production.[9]
2. **The number of gym memberships within a population is a proxy for the attitudes towards health and wellness of that population.** Because exercise is commonly known to improve physical and mental well-being, a population that cares more about their health is more likely to purchase a gym membership.
3. **There is no significant correlation between the regressed independent variables.** We live in a dynamic world where every factor is probably interrelated in some way that is too complex to be predicted. We chose variables to be as distinct as possible to minimize the cross-talk between factors.
4. **All examined historical trends continue for the next five years.** While it is possible that some unprecedented event alters one of these trends, such an occurrence is impossible to model or predict.
5. **Google search trends are reflective of the popularity of a certain subject within a certain population.** Google is the world's most popular search engine and constitutes approximately 85% of all internet traffic. When people are interested in something, they investigate it on Google as a trustworthy source.[10]
6. **There are no other factors besides those considered that significantly affect the growth of e-bike sales.** Although this is not plausible, we cannot comprehensively include all parameters within the time allotted. p
7. **The data for variables considered is homoscedastic.** The error variance is constant or equal across the levels of independent variables.

8. **Subsidies have a negligible impact on sales.** There are no nationwide e-bike subsidies in the US or UK due to partisan tensions, making it impossible to predict when such legislation will pass in the future.[12][11] On the local level, only very few states or municipalities currently have significant e-bike subsidies.[13]

9. **The independent variables in the regression have a causal relationship with annual bike sales in both countries.** This allows us to use the coefficients to determine the relative importance of specific factors in influencing e-bike sales. We provide logical or observed justification for each variable, implying a causal relationship.

10. **Staggering data values by year is negligible.** Although causal effects typically take some lag time to manifest, all variables most likely cause an increase in e-bike sales within the same year the variable changes. Since our data is discretized by year, the time lag will be trivial.

2.3 Model Development

To determine the relative importance of specific factors on the growth of e-bike sales in the United States and the United Kingdom, we created a multivariate linear regression, regressing the variables of gas price, environmental consciousness, GDP per capita, consumer spending, trendiness, attitude toward health/fitness, and median age on the same annual e-bike sales data from Part I. Our data for all independent variables span the years 2012 to 2022, interpolating any missing values. In order to extrapolate causality, we provide a high-level explanation for the causal link.

The following bullets provide the justification for causality for each independent variable:

1. **Gas Prices:** If the price of gas increases, the cost of typical transportation, such as cars and buses, will increase. In response, consumers will seek out alternative, cheaper forms of transportation such as e-bikes. This is because the price of substitutable goods are a determinant of demand.

2. **Environmental Consciousness:** According to a survey by Simsekoglu and Klöckner 18' people care about the environment and might use electric bicycles to avoid pollution from cars, buses, ride-share, or other transportation methods. [14]

3. **GDP per Capita:** An increase in GDP per capita implies that a country's production capacity has increased, purchasing power has increased, and electric bikes are easier to access. People will have the necessary disposable funds to purchase a bicycle when they were unable to before.

4. **Consumer Spending:** In a highly capitalist society, general consumer spending corresponds to a propensity to spend in any specific area of life. Therefore, as consumers spend more, they are more likely to spend on goods like e-bikes even if they are without an urgent need.

5. **Trendiness:** Bicycles have achieved increasing social prominence over the last decade. The COVID-19 pandemic initiated the resurgence of bicycling. Electric bike-sharing initiatives have rapidly popularized in urban locations, and even stationary biking has become a common form of exercise. The general perception of e-bikes has improved accordingly, thereby increasing the probability that someone is going to purchase an electric bicycle. [14]

6. **Attitude towards Health/Fitness:** If people become more conscious about their health and wellness, they might have a higher propensity to purchase e-bikes so that they are able to exercise on their commute. Many that use e-bikes do so because they think it is good for their health or because it promoted physical activity. [14]

7. **Median Age:** Surveys about electric bicycles indicate that e-bikes are more favorable among older populations [14]. As the general population continues to become older due to an imbalance in new births, e-bikes might become a more attractive option to the typical citizen.

8. **Lithium Prices:** Lithium is the largest input cost in e-bike production. As the cost of lithium decreases, the cost of producing e-bikes will decrease, which in turn decreases the price of e-bikes sold on the market and increases the propensity for a consumer to purchase an e-bike due to the law of supply. [8]

We then chose multivariate linear regression because it incorporated a large array of factors and could easily extract which variables were most significant. We used the following equation, regressing for the respective values of α_i for each independent variable:

$$E = \beta + \alpha_1(GP) + \alpha_2(EC) + \alpha_3(GDPPC) + \alpha_4(CS) + \alpha_5(T) + \alpha_6(HF) + \alpha_7(MA) + \alpha_8(Li)$$

Symbol	Heat map Key	Variable	Unit
α_i	N/A	Coefficient	N/A
β	N/A	Error Term	N/A
E	BIKE SALES	Annual E-bike Sales	1000 Bikes/Yr
GP	GAS PRICE	Gas Prices	Dollars/Gallon
EC	ENV	Environmental Consciousness	N/A
GDPPC	GDP PC	GDP Per Capita	Dollars/Person
CS	SPENDING	Consumer Spending	Dollars/Yr
T	TRENDINESS	Trendiness	N/A
HF	GYM MEMB	Attitude toward Health/Fitness	N/A
MA	MED AGE	Median Age	Years
Li	LITHIUM PRICE	Lithium Battery Price	Dollars

Table 6: Variables in Multivariate Annual E-Bike Sales Regression

Using this data we performed multivariate linear regression using sklearn to find the α_i values (correlation coefficients). We then plotted the correlation coefficient between all variables in a heat map using matplotlib and seaborn, allowing us to easily visualize the variables which may be decreasing the precision of our regression. Additionally, the correlation coefficients between the independent variables and the dependent variable (number of e-bikes sold) would help us determine which variable has the largest impact on number of bikes sold. Although correlation does not always imply causation, we demonstrate a correlation between variables and previously demonstrated the rationale behind the presence of a possible causal relationship.

2.4 Results

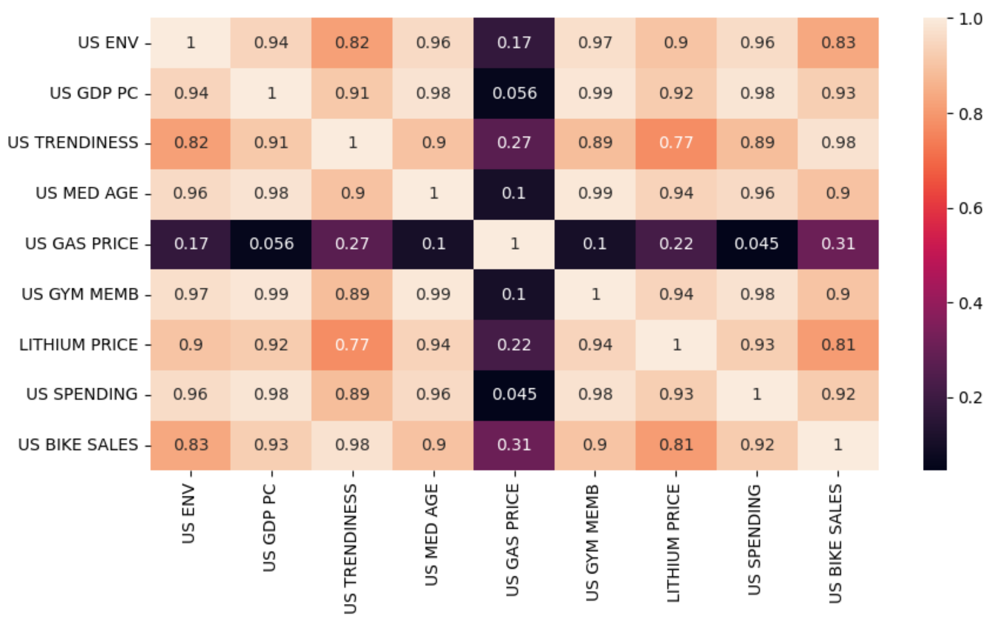


Figure. 2: US Variable Correlation Heat Map

Many of our variables displayed high degrees of correlation for the reasons outlined during model development. Because causal factors must be related to each other, correlation is a prerequisite for causation. We created heat maps of both US and UK data in order to determine which variables were best correlated with bike sales. Variables with low correlation scores, such as gas prices, are not significant causes of bike sales. This makes sense because gas prices are subject to random geopolitical events while bike sales follow a relatively stable trend. On the other hand, variables with high correlation scores, such as trendiness have the potential for a causal relationship as well.

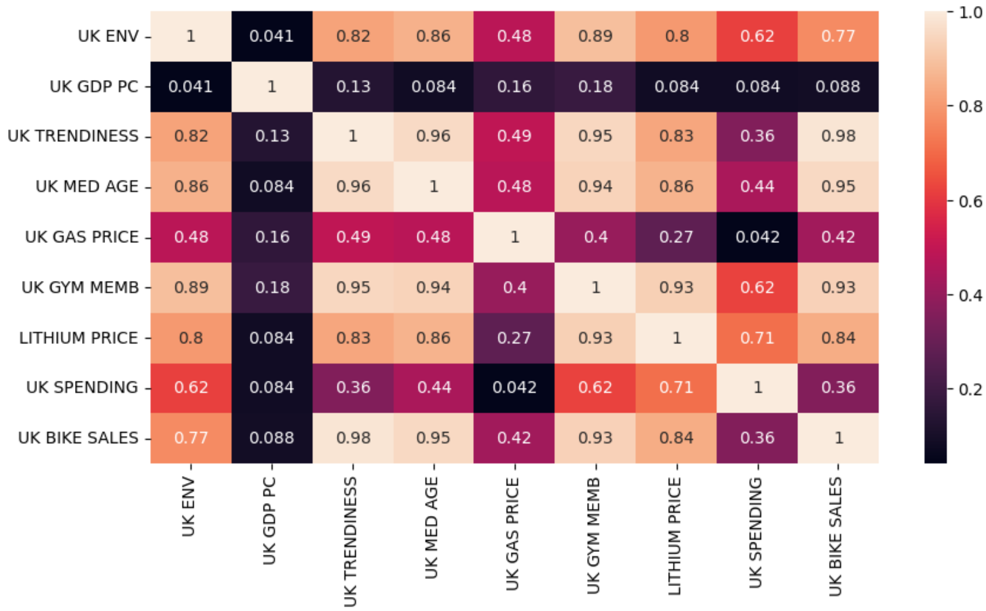


Figure. 3: UK Variable Correlation Heat Map

In order to determine the strength of causation, We then created a bar graph, plotting the absolute value of each correlation coefficient on the y-axis (α_i). Higher correlation coefficients imply that a variable has higher significance in determining bike sales. From the graph, we can see that US GDP per capita is extremely important, while trendiness is particularly important in the UK. This phenomenon can best be explained by a stronger consumer mindset in the US, in which people feel more inclined to spend when the economy is undergoing an overall growth trend, an idea which has been empirically proven by the economy's consistent boom and bust cycles. In the UK, people are frugal in spite of economic growth but make purchases based on those around them.

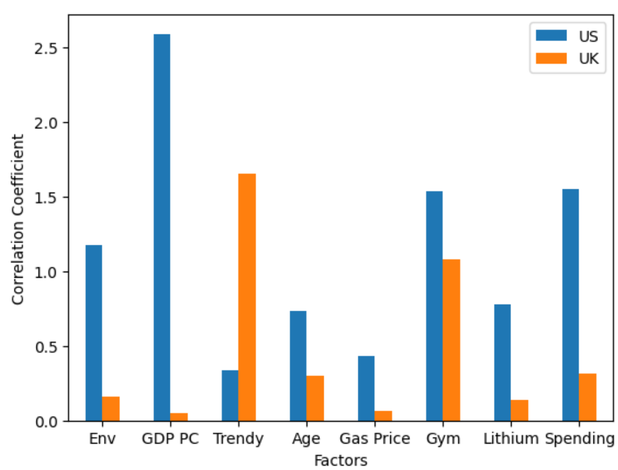


Figure. 4: UK/US Variable Correlation Bar Graph

2.5 Sensitivity Analysis

In order to assess the accuracy of the coefficients in our multivariate linear regression, we jittered the values of all of our data points between a 5% decrease and a 5% increase and plugged the new values into the prior regression. With the added noise, we determined which of the effects are the most susceptible to small changes in data. Notably, the correlation coefficient related to US economic variables seemed to change the most, which is consistent with the aforementioned principle that the United States has a more sensitive consumer mindset. Importantly, other factor coefficients remained relatively similar, allowing us to be confident in our model's overall accuracy and support of our prior analysis on the relative significance of specific factors related to e-bike sales.

2.6 Strengths and Weaknesses

2.6.1 Strengths

1. **Accessibility.** Unlike machine learning models, regressive models are transparent and explain results in the form of regressed coefficients. Our correlation heat map also provides an understandable and intuitive visualization of the relationships between all variables.
2. **Predictive Power.** While machine learning and other statistical methods require large amounts of data to accurately train, a linear regression uses empirical and logical reasoning about causal links to produce better predictions. By strategically choosing variables that are logically related to bike sales, we reduce confounding variables and scored relatively high in accuracy.
3. **Easily Extendable.** Linear regression can incorporate any type of continuous variable and can flexibly expand to account for additional variables. This enabled our model to easily calculate the correlations between all of our data.

2.6.2 Weaknesses

1. **Causality.** Since a linear regression only explains correlation, the model does not explain why variables are related; rather, it provides insight into the existence and strength of a trend. We circumvent this issue by providing a high-level causal inference for each independent/dependent variable pair.
2. **Collinearity.** Although we assumed that our factors were all independent of each other, many of the variables are likely corollaries of overall economic growth. As a result, our model struggles to identify unique influences on the proliferation of e-bike purchases.

3. **Outlier Onset.** Linear regression models are highly sensitive to outliers or high leverage points. If any of these variables undergo substantial changes that develop an outlier data point, it would disrupt the efficacy of the model.

3 Part III: Off The Chain

3.1 Restatement of the Problem

Understanding the projected rise of e-bike purchases and the factors that influence this growth is not enough. For government officials to best allocate resources for creating infrastructure and legislation in this area, it is necessary to evaluate the overall societal impact of e-bike's rise in popularity. In this section, we quantified the impacts of increased e-bike sales in the US and UK on net carbon emissions and projected average lifespan in each country.

3.2 Assumptions

1. **The number of bike users is equal to the number of bikes sold.** Despite the prevalence of e-bike sharing systems, most e-bikes do not lend themselves to consistent sharing, and one person likely does not benefit from owning multiple bikes. Additionally, once a person purchases an e-bike, they will continue to ride an e-bike for the rest of their life. This is a broad assumption that is a necessity for the life-span model to work; however, it is not that far-fetched because e-bike experts report that once people start using them they become continual users. [15]

2. **Each user only uses their e-bike for one use-case.** We could not find any data on the frequency of auxiliary use-cases. We assume that riders have one main use for their e-bike and that any others are trivial.

3. **The average car and e-bike emission per mile is representative of each rider's usage.** The average emission is likely to produce an estimate close to any individual's actual value as car and e-bike emissions are normally distributed and have a low standard deviation.

4. **The production emission of all e-bikes is .16 tons, or 145 kg, and remains constant over time.** This value represents the emissions required to create a standard lithium-ion battery, which is overwhelmingly the most resource-intensive component of e-bike production.[17][18]

5. **The average commute of people that switch to e-bikes is 10 miles.** This value is the current average commute for bikers. We assume that the commute distance for bikers and e-bikers is the same as there are similar threshold conditions needed for someone to bike to work as for someone to an e-bike. Similarly, we assume that the

average distance of a local or recreational trip is 5 miles, and other short-hop journeys are 2 miles long.

6. **All e-bike riding is considered moderate exercise.** Most e-bikes only provide a small amount of assistance, comparable to a tailwind, and require physical movement from their user. We will assume that this is a requisite for all e-bikes.[19]

7. **Health improvements are directly proportional to increased exercise time.** People who exercise more also live longer. We assume that e-bike users are not over-exercising to an extent that counteracts these positive health benefits because such behavior would require an extremely high amount of time e-biking. [16]

8. **Net traffic changes are zero.** Regardless of evolving ratios of e-bikers and car drivers, we assume that the traffic will stay the same. E-bike usage is similarly likely to increase traffic, being susceptible to road accidents and often moving under urban speed limits, or decrease traffic, due to its smaller size.

3.3 Model Development

Variable	Name	Unit
Y_m	Increase in Lifespan	Years per minute per week
P_i	Likelihood of of Use-case i	N/A
E_i	Net Emissions of Use-case i	kg of CO_2
D_i	Distance of Use-case i	miles

Table 7: Variables

We separated e-bike users into two age ranges: older than 55 years old and younger than 55 years old. Approximately 83% of riders are under 55 and the remaining 17% are older than 55. Data showed that these age groups used their e-bikes for different purposes, with younger users being more likely to commute while older riders often made recreational and neighborhood trips. We then gathered data for the specific propensities of each use-case. [14] [29]

Age	Commute	Recreational	Local	Other
Under 55	0.58	0.21	0.09	0.12
Over 55	0.30	0.29	0.31	0.10

Table 8: Use-case Distribution by Age

Based on their use-case, we gathered data on what mode of transportation was replaced by e-biking. In some cases, the rider would never have made the trip without an e-bike, which has been labeled "new trip."[20] Each use-case also had a set number of miles traveled as discussed in the assumptions, which we use to calculate both environmental and health benefits.

Use-case	Commute	Recreational	Local	Other
Miles Traveled	10	5	5	2

Table 9: Distance Traveled per Use-case

Replaced Mode of Transport	Commute	Recreational	Local	Other
Car	0.37	0.47	0.09	0.10
Bus	0.32	0.29	0.19	0.09
Walk/Bike	0.01	0.07	0.48	0.16
New Trip	0.30	0.17	0.23	0.65

Table 10: Replaced Mode of Transport by Use-case

We can use the replaced mode of transport in order to compute the carbon dioxide emissions saved by e-biking.

$$E_i = E_{gross} - E_{e-bike}$$

The carbon dioxide emissions for each mode of transport are shown below in grams of CO_2 per mile. [32] [33] [34]

Mode of Transport	Car	Bus	Walk/Bike	Bike
Net Emissions	348	299	0	8

Table 11: Carbon Dioxide Emissions Per Use-case (g of CO_2)

To calculate the final emissions of a person. We multiply the distance of their journey by the emissions of the replaced mode of transport, and then scale it to a year.

$$E_{total} = 365 \times E_i \times D_i$$

In order to quantify the health benefits of biking, we first turn to the increase in lifespan due to exercise. A 75-minute rise in exercise per week has been empirically shown to increase lifespan by 1.8 years, which we extrapolate to about 0.024 years per exercise-minute per week. [30]

$$Y_m = \frac{1.8 \text{ years per minute per week}}{75 \text{ minutes per week}}$$

Since e-bikes travel at approximately 20 mph, each mile traveled is 3 exercise-minutes per day. [31] We multiply that value by 5 for the number of weekdays in a week and find that each mile of exercise by riding an e-bike adds approximately 0.36 years to the rider's lifespan. Thus, we multiply:

$$Y_{total} = Y_m \times 3 \text{ minutes per mile} \times 5 \text{ days per week} \times D_i$$

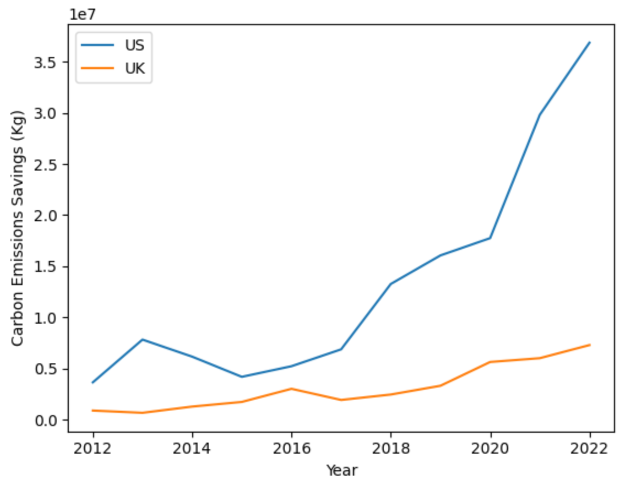
3.4 Results

We used a Monte Carlo simulation to model the entire distribution of electric bike users. Each simulated rider was given an age based on the actual age distribution, chose a use-case based on their age, and then chose a replaced mode of transportation based on use-case. Using the distance of the use case and replaced mode of transportation, we calculate the final environmental savings in 10^7 kgs of CO_2 per year and health benefits in thousands of life years. Note that health benefits assume that people continue to ride their bikes for their entire lives, so the final numbers represent projected life years saved. This assumption is validated by the fact that most riders do not backtrack on their active lifestyles. Limiting carbon emissions is important for a nation because greenhouse gases can increase the temperature of the earth and have adverse affects on the environment, limiting the earth's future potential for prosperity. Additional life years are beneficial for a nation because it allows people to spend more time with their families and output more value.

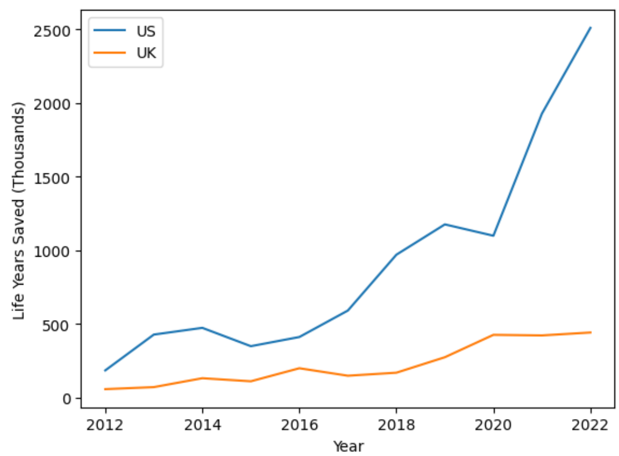
	2025	2028
UK Environmental Savings	1.486	2.951
UK Health Benefits	805.2	1405.07
US Environmental Savings	6.349	14.335
US Health Benefits	4870.8	9792

Table 12: Final Savings Calculated

As a final analysis, we used our data from Part One to project environmental savings and health benefits over a 10-year range and graphed the results.



(a) Environmental Savings



(b) Health Benefits

These numbers are well within expected bounds. Since the United States emits a total of 1 million metric tonnes per year, our projections deem e-bikes an important pillar of climate-aware transportation. Though a large number of life years saved is a cumulative value, the increased exercise from e-bikes demonstrates tremendous potential for both the economy and human fulfillment.

3.5 Strengths and Weaknesses

3.5.1 Strengths

1. **Easily Extendable.** Our Monte Carlo simulation can be easily expanded to account for more models by adding new factors to our rider's choices.
2. **Specificity.** By randomly creating e-bike owners based on known probabilities, we can accurately generate a model for the situation and decisions of an e-bike owner.

3.5.2 Weaknesses

1. **Emission Generalization.** The probability variable for emissions does not consider individual cars, but rather average emissions per mile. The type of car that a person owns might affect their willingness to switch to an e-bike. For example, large trucks make up a larger portion of emissions, but their drivers are unlikely to switch to e-bikes.
2. **Simplistic Riders.** Most e-bike owners likely use their bikes for multiple use-cases. By limiting each owner to one use-case, we lose specificity about other potential uses. Older riders might also use their bikes less frequently than younger ones, which our model does not account for.

Conclusion

Many say that electric vehicles are our future, yet the prevalence and relevance of e-bikes often goes overlooked. To prepare governmental agencies for the ongoing rise of these vehicles, we modeled the spread of e-bikes in the US and UK. After projecting their growth, we determined our prediction's causal factors and finally evaluated their societal impacts.

In both the US and the UK, we determined that annual e-bike sales would continue to increase year over year. However, in the US, overall economic growth accounted for this rise in purchases, while in the UK, e-bikes are popularizing due to societal trends. Although e-bikes are commonly considered a vehicle of those looking to reduce their carbon footprint, we found that environmental awareness was not one of the three most significant causal factors in either nation. Nevertheless, our final model found e-bike growth to decrease both nations' carbon dioxide emissions while offering positive health benefits to e-bike users.

We aim to refine our models in subsequent analyses. Our general projection of e-bike growth cannot be tailored to unpredictable shifts in the status quo, which we would remedy by integrating additional factors into the regression. To address our second model's issues with co-linearity, we would analyze our variables' variance inflation factors or linearly combine some of our most clearly-correlated independent variables. Since certain types of automobile usage are more difficult to transition to e-bike usage, we would finally determine an individual's propensity to transition to an e-bike based on their model of car and daily car usage.

Despite our models' flaws, their predictions are uplifting. They indicate a general inclination towards physical well-being and sustainability, as well as consistent economic growth. To bolster this ongoing upturn, we urge increased e-bike accessibility and safety via the following policy changes:

1. Invest in bike lane infrastructure.
2. Provide rebates for e-bike purchases.
3. Integrate e-bike safety education into drivers license curriculum.
4. Develop secure e-bike parking in public areas.
5. Mandate helmet usage for e-bike riders.

References

[1] 51 Official Ebike Statistics Facts. The RoundUp. https://theroundup.org/ebike-statistics/.

[2] Electric Bike, Merriam Webster. https://www.merriam-webster.com/dictionary/electric%20bike.

[3] The diffusion of iPhones as a learning process. Asymco. http://www.asymco.com/2013/11/06/the-diffusion-of-iphones-as-a-learning-process/

[4] Norman, Paul. BikeRadar. Electric bikes and the law: everything you need to know about ebike regulations. 1 Nov. 2022. https://www.bikeradar.com/advice/buyers-guides/electric-bike-laws/.

[5] Lerma, Andrew, Lithium-Ion vs Lead Acid Battery Life. FluxPower. 14 Mar. 2019. https://www.fluxpower.com/blog/lithium-ion-vs.-lead-acid-battery-life.

[6] Bonk, Lawrence, How Long Do Electric Bikes Last?. Gadget Review. 4 Mar. 2023. https://www.gadgetreview.com/how-long-do-electric-bikes-last: :text=Electric%20bikes%2C%20on%20average%2C%20should,at %20around%20500%20charging%20cycles

[7] How to Choose an Electric Bike. REI. 10 Jan. 2023. https://www.rei.com/learn/expert-advice/how-to-choose-an-ebike.html

[8] Sabastian, Xavier. Price of Lithium-ion Battery Pack through 2010 – 2021. Way. https://www.way.com/blog/lithium-ion-battery-pack-price-through-2010-2021/.

[9] Gross, Sam. Why are electric bikes so expensive?. Electric Bike Report. 1 Feb. 2022. https://electricbikereport.com/why-are-electric-bikes-so-expensive/

[10] Bianchi, Tiago. Worldwide desktop market share of leading search engines from January 2015 to January 2023. Statista. 24 Feb. 2023. https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/.

[11] Hawkins, Andrew. The e-bike tax credit is only mostly dead as supporters plot next steps. The Verge. 9 Aug. 2022. https://www.theverge.com/2022/8/9/23297209/ebike-tax-credit-climate-bill-ev-car-congress.

[12] Laker, Laura. Electric bike subsidy is a no-brainer, so why can't the UK Government see that? Bike Radar. 2 Nov. 2022. https://www.bikeradar.com/features/opinion/electric-bike-subsidy-is-a-no-brainer/: :text=As%20things%20stand%2C%20ebikes%20are,effective %20%E2%80%93%20at%20replacing%20car%20trips.

[13] Toll, Micah. Elektrek. Another US state adds electric bike subsidy, this time with up to $1,700 rebate. 29 Nov 2022. https://electrek.co/2022/11/29/another-us-state-adds-electric-bike-subsidy-this-time-with-up-to-1700-rebate/.

[14] Simsekoglu, Ozlem and Klockner, Christian. Factors related to the intention to buy an e-bike: A survey study from Norway. 11 June 2018. https://www.doi.org/10.1016/j.trf.2018.11.008/.

[15] Surico, John. The Popularity of E-Bikes Isn't Slowing Down. 22 November 2021. https://www.nytimes.com/2021/11/08/business/e-bikes-urban-transit.html.

[16] Moore, Steven. Leisure Time Physical Activity of Moderate to Vigorous Intensity and Mortality: A Large Pooled Cohort Analysis. PLOS Medicine. 6 Nov. 2012. https://doi.org/10.1371/journal.pmed.1001335.

[17] What Does It Cost To Charge An Electric Bike? Hovsco. https://www.hovsco.com/blogs/news/what-does-it-cost-to-charge-an-electric-bike: :text=How%20Much%20Electricity%20Does%20It,cost%20you%205%2D8%20cents.

[18] Crawford, Iris. How much CO2 is emitted by manufacturing batteries? 1 Mar. 2022. https://meche.mit.edu/news-media/how-much-co2-emitted-manufacturing-batteries: :text=The%20Production%20Processtext=Currently%2C%20most%20lithium%20is%20extra

[19] Reynolds, Gretchen. E-Bikes Can Provide a Good Workout. The New York Times. 19 May 2021. https://www.nytimes.com/2021/05/19/well/move/bikes-exercise-workouts.html.

[20] A North American Survey of Electric Bicycle Owners. Portland State University. Mar. 2018. https://www.calbike.org/wp-content/uploads/2019/02/A-North-American-Survey-of-Electric-Bicycle-Owners.pdf.

[21] McCarthy, Neill. How Long Do Americans Commute For? Statista. 21 Apr. 2021. https://www.statista.com/chart/24684/average-duration-of-a-one-way-commute/.

[22] Reed, Betsy. On your e-bike: European craze powers into UK. 12 Oct 2012. https://www.theguardian.com/lifeandstyle/2012/oct/12/e-bike-european-craze-uk.

[23] Statista. Number of electric bicycles sold in the European Union (EU) from 2013 to 2016, by country (in 1,000 units). https://www.statista.com/statistics/397772/electric-bicycle-sales-volume-in-the-european-union-eu-by-country.

[24] Bowden, Alex. UK E-bike sales slow due to cost of living crisis with lack of subsidies a barrier. Feb 4 2023. https://ebiketips.road.cc/content/news/uk-e-bike-sales-slow-due-to-cost-of-living-crisis-with-lack-of-subsidies-a-barrier-4297.

[25] Mintel. On your e-bike: 14% of UK cyclists intend to buy an e-bike in the next year. April 1, 2019. https://www.mintel.com/press-centre/on-your-e-bike-14-of-uk-cyclists-intend-to-buy-an-e-bike-in-the-next-year/.

[26] Bike Europe. UK government statistics reveal huge increase in e-bike imports. https://www.bike-eu.com/39564/uk-government-statistics-reveal-huge-increase-in-e-bike-imports.

[27] Sutton, Mark. A sale every 3 minutes: Electric bikes outsell electric cars in UK during 2020. 24 June 2021. https://cyclingindustry.news/a-sale-every-3-minutes-electric-bikes-outsell-electric-cars-in-uk-020/.

[28] Laker, Laura. UK bike sales take pause, cargo movement powers on. 24 August 2022. https://cyclingindustry.news/uk-bike-sales-cargo-bike-growth/.

[29] Shah, Kineree. Tracing the demographics of e-bike owners in the US. 19 Oct 2022. https://business.yougov.com/content/44110-tracing-demographics-e-bike-owners-us.

[30] Moore et. al 12. Leisure Time Physical Activity of Moderate to Vigorous Intensity and Mortality: A Large Pooled Cohort Analysis. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001335.

[31] Johnson, Gary and Speciale, Henry. How Fast Do Electric Bikes Go? Everything You Should Know!. March 1, 2023. https://biketoworkday.us/how-fast-do-electric-bikes-go/.

[32] Blumgart, Jake. Are Trains or Buses Better for the Environment? Feb. 11, 2022. https://www.governing.com/next/are-trains-or-buses-better-for-the-environment.

[33] Office of ENERGY EFFICIENCY RENEWABLE ENERGY. FOTW 1223, January 31, 2022: Average Carbon Dioxide Emissions for 2021 Model Year Light-Duty Vehicles at an All-time Low. https://www.energy.gov/eere/vehicles/articles/fotw-1223-january-31-2022-average-carbon-dioxide-emissions-2021-model-year.

[34] Bosch. Sustainability at Bosch eBike Systems. https://www.bosch-ebike.com/us/service/sustainability.

Appendix

Part I: The Road Ahead

In [1]:
```
import pandas as pd
import janitor as jn
import numpy as np

d = {'year': [2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022],
     'uk_bikes': [20, 25, 50, 40, 75, 55, 70, 101, 160, 160, 170],
     'us_bikes': [70, 159, 193, 130, 152, 220, 369, 423, 416, 750, 928]}
df = pd.DataFrame(data=d)
df
```

Out[1]:

	year	uk_bikes	us_bikes
0	2012	20	70
1	2013	25	159
2	2014	50	193
3	2015	40	130
4	2016	75	152
5	2017	55	220
6	2018	70	369
7	2019	101	423
8	2020	160	416
9	2021	160	750
10	2022	170	928

In [2]:
```
df.jitter("uk_bikes", dest_column_name="uk_jit1", scale=1)
df.jitter("us_bikes", dest_column_name="us_jit1", scale=1)
df.jitter("uk_bikes", dest_column_name="uk_jit2", scale=1)
df.jitter("us_bikes", dest_column_name="us_jit2", scale=1)
df.jitter("uk_bikes", dest_column_name="uk_jit3", scale=1)
df.jitter("us_bikes", dest_column_name="us_jit3", scale=1)
df.jitter("uk_bikes", dest_column_name="uk_jit4", scale=1)
df.jitter("us_bikes", dest_column_name="us_jit4", scale=1)
df.jitter("uk_bikes", dest_column_name="uk_jit5", scale=1)
df.jitter("us_bikes", dest_column_name="us_jit5", scale=1);
df.head()
```

Out[2]:

	year	uk_bikes	us_bikes	uk_jit1	us_jit1	uk_jit2	us_jit2	uk_jit3	us_jit3	uk_jit4	us_jit4	uk_jit5	us_jit5
0	2012	20	70	19.534892	71.087685	20.101495	70.515299	18.248848	68.033886	21.057751	70.714968	21.460196	68.733483
1	2013	25	159	25.784255	159.731954	25.440109	159.691796	24.917498	160.136580	27.140135	159.878496	24.745142	160.996750
2	2014	50	193	50.935359	191.286590	50.959139	192.373690	51.369816	194.439624	49.662531	191.527436	50.146140	193.488267
3	2015	40	130	40.252886	129.277396	38.761249	130.302084	39.635678	130.041924	40.249617	130.411221	40.997599	130.908503
4	2016	75	152	74.724498	149.765040	73.641459	152.027191	74.228529	153.915069	74.889975	153.204309	74.203558	151.679899

In [14]:
```
df['uk_k1']=np.log((13400/df['uk_jit1'])-1)/(df['year']-2044)
df['us_k1']=np.log((66000/df['us_jit1'])-1)/(df['year']-2038)
df['uk_k2']=np.log((13400/df['uk_jit2'])-1)/(df['year']-2044)
df['us_k2']=np.log((66000/df['us_jit2'])-1)/(df['year']-2038)
df['uk_k3']=np.log((13400/df['uk_jit3'])-1)/(df['year']-2044)
df['us_k3']=np.log((66000/df['us_jit3'])-1)/(df['year']-2038)
df['uk_k4']=np.log((13400/df['uk_jit4'])-1)/(df['year']-2044)
df['us_k4']=np.log((66000/df['us_jit4'])-1)/(df['year']-2038)
df['uk_k5']=np.log((13400/df['uk_jit5'])-1)/(df['year']-2044)
df['us_k5']=np.log((66000/df['us_jit5'])-1)/(df['year']-2038)
```

```
In [18]: df['uk_logistic1']=13400/(1+(2.71828**((df['uk_k1'])*(df['year']-2044))))
         df['us_logistic1']=66000/(1+(2.71828**((df['us_k1'])*(df['year']-2038))))
         df['uk_logistic2']=13400/(1+(2.71828**((df['uk_k2'])*(df['year']-2044))))
         df['us_logistic2']=66000/(1+(2.71828**((df['us_k2'])*(df['year']-2038))))
         df['uk_logistic3']=13400/(1+(2.71828**((df['uk_k3'])*(df['year']-2044))))
         df['us_logistic3']=66000/(1+(2.71828**((df['us_k3'])*(df['year']-2038))))
         df['uk_logistic4']=13400/(1+(2.71828**((df['uk_k4'])*(df['year']-2044))))
         df['us_logistic4']=66000/(1+(2.71828**((df['us_k4'])*(df['year']-2038))))
         df['uk_logistic5']=13400/(1+(2.71828**((df['uk_k5'])*(df['year']-2044))))
         df['us_logistic5']=66000/(1+(2.71828**((df['us_k5'])*(df['year']-2038))))
```

```
In [23]: df['uk_logdif_1']=100*(df['uk_logistic1']-df['uk_bikes'])/df['uk_bikes']
         df['us_logdif_1']=100*(df['us_logistic1']-df['us_bikes'])/df['us_bikes']
         df['uk_logdif_2']=100*(df['uk_logistic2']-df['uk_bikes'])/df['uk_bikes']
         df['us_logdif_2']=100*(df['us_logistic2']-df['us_bikes'])/df['us_bikes']
         df['uk_logdif_3']=100*(df['uk_logistic3']-df['uk_bikes'])/df['uk_bikes']
         df['us_logdif_3']=100*(df['us_logistic3']-df['us_bikes'])/df['us_bikes']
         df['uk_logdif_4']=100*(df['uk_logistic4']-df['uk_bikes'])/df['uk_bikes']
         df['us_logdif_4']=100*(df['us_logistic4']-df['us_bikes'])/df['us_bikes']
         df['uk_logdif_5']=100*(df['uk_logistic5']-df['uk_bikes'])/df['uk_bikes']
         df['us_logdif_5']=100*(df['us_logistic5']-df['us_bikes'])/df['us_bikes']
```

```
In [28]: uk_mean=df[['uk_logdif_1','uk_logdif_2','uk_logdif_3','uk_logdif_4','uk_logdif_5']].mean().mean()
         us_mean=df[['us_logdif_1','us_logdif_2','us_logdif_3','us_logdif_4','us_logdif_5']].mean().mean()
         print('Percent Difference for UK:')
         print(uk_mean)
         print('Percent Difference for US:')
         print(us_mean)

         Percent Difference for UK:
         0.26627644461457967
         Percent Difference for US:
         0.017596113417233602
```

Part II: Shifting Gears

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.linear_model import LinearRegression
from sklearn import preprocessing
from sklearn.metrics import mean_squared_error, r2_score
```

```python
data = pd.read_csv("/work/M3 - imputed (4).csv")
data.head()
```

```python
plt.figure(figsize=(10, 5))
sns.heatmap(data.corr().abs(),  annot=True)
```

```python
US = ["US ENV","US GDP PC", "US TRENDINESS", "US MED AGE", "US GAS PRICE", "US GYM MEMB", "LITHIUM PRICE", "US SPENDING", "US BIKE SALES"]
US_DATA = data[US]

UK = ["UK ENV","UK GDP PC", "UK TRENDINESS", "UK MED AGE", "UK GAS PRICE", "UK GYM MEMB", "LITHIUM PRICE", "UK SPENDING", "UK BIKE SALES"]
UK_DATA = data[UK]

st_scaler = preprocessing.StandardScaler()

US_DATA = st_scaler.fit_transform(US_DATA)
US_DATA = pd.DataFrame(US_DATA, columns=US)
UK_DATA = st_scaler.fit_transform(UK_DATA)
UK_DATA = pd.DataFrame(UK_DATA, columns=UK)

plt.figure(figsize=(10, 5))
matrix = np.triu(US_DATA.corr())
sns.heatmap(US_DATA.corr().abs(),  annot=True)#, mask=matrix)

plt.figure(figsize=(10, 5))
matrix = np.triu(UK_DATA.corr())
sns.heatmap(UK_DATA.corr().abs(),  annot=True)#, mask=matrix)
```

```python
US_X_pred = ["US ENV","US GDP PC", "US TRENDINESS", "US MED AGE", "US GAS PRICE", "US GYM MEMB", "LITHIUM PRICE", "US SPENDING"]
US_Y_pred = ["US BIKE SALES"]

X_US = US_DATA[US_X_pred]
Y_US = US_DATA[US_Y_pred]

UK_X_pred = ["UK ENV","UK GDP PC", "UK TRENDINESS", "UK MED AGE", "UK GAS PRICE", "UK GYM MEMB", "LITHIUM PRICE", "UK SPENDING"]
UK_Y_pred = ["UK BIKE SALES"]

X_UK = UK_DATA[UK_X_pred]
Y_UK = UK_DATA[UK_Y_pred]
```

```python
us_model = LinearRegression()
us_model.fit(X_US, Y_US)

uk_model = LinearRegression()
uk_model.fit(X_UK, Y_UK)
```

```
score = us_model.score(X_US,Y_US)
intercept = us_model.intercept_
coef = us_model.coef_

print("--------US--------")
print("Score: ", score)
print("Intercept: ", intercept)
print("Coef: ", coef)
```

```
score = uk_model.score(X_UK,Y_UK)
intercept = uk_model.intercept_
coef = uk_model.coef_

print("--------UK--------")
print("Score: ", score)
print("Intercept: ", intercept)
print("Coef: ", coef)
```

```
import statsmodels.api as sm
modelus = sm.OLS(Y_US, X_US).fit()
modelus.summary()
```

```
import statsmodels.api as sm
modeluk = sm.OLS(Y_UK, X_UK).fit()
modeluk.summary()
```

```
import matplotlib.pyplot as plot

US = [1.1738, 2.5911, 0.3355, 0.7360, 0.4316, 1.5388, 0.7765, 1.5585]
UK = [0.1600, 0.0523, 1.6514, 0.3823, 0.0618, 1.0826, 0.1383, 0.3115]
index = ['Env', 'GDP PC', 'Trendy',
         'Age', 'Gas Price', 'Gym', 'Lithium', 'Spending']
df = pd.DataFrame({'US': US, 'UK': UK}, index=index)
df.plot.bar(rot=0)
plot.xlabel("Factors")
plot.ylabel("Correlation Coefficient")
```

Part III: Off The Chain

```python
import matplotlib.pyplot as plt
import numpy as np
import math
```

```python
def get_data(population):
    over_ff_percentage = .17
    under_ff_percentage = .83

    over_ff = math.ceil(population * over_ff_percentage)
    under_ff = population - over_ff

    # For self-reference
    key = ["commute", "local", "recreation", "other"]
    trip_key = ["car", "public", "new", "other"]

    # Transition probability arrays

    commute = [37.26, 32.38, 0.89, 29.47]
    local = [47.04,29.43,6.04,17.49]
    recreation = [9.45,19.36,47.98,23.21]
    other = [9.89,8.79,16.48,64.84]

    under = [58,9,21,12]
    over = [30,31,29,10]

    for arr in [commute, local, recreation, other, under, over]:
        for i in range(len(arr)):
            arr[i] /= 100

    # grams
    health_cost = [3.6,1.8,1.8,0.72]

    car = 348
    bus = 299
    ebike = 8
    bike = 8

    translate = [car, bus, ebike, bike]

    def act(age) :
        # 1 is over 55
        if age == 1 :
            return np.random.choice(4,p=over)
        # 0 is under 55
        if age == 0 :
            return np.random.choice(4,p=under)

    class Person:
        def __init__(self, age, activity):
            self.activity = activity
            self.age = age
            self.start_cost = 145000 / 5
            self.cost = 0

            self.health_cost = health_cost[self.activity]

        def change(self):
            t = 0

            # commute
            if self.activity == 0 :
                new_transport = np.random.choice(4,p=commute)
                t = translate[new_transport]

            if self.activity == 1:
                new_transport = np.random.choice(4,p=local)
                t = translate[new_transport]

            if self.activity == 2:
                new_transport = np.random.choice(4,p=recreation)
                t = translate[new_transport]

            if self.activity == 3:
                new_transport = np.random.choice(4,p=other)
                t = translate[new_transport]

            self.cost -= (t * 365)

        def get_cost(self):
            return self.cost + self.start_cost
```

```python
        def get_cost(self):
            return self.cost + self.start_cost

    saving_sum_arr = []
    health_sum_arr = []

    iteration = range(1)

    for i in iteration :
        # Create population

        population = []

        for i in range (over_ff) :
            population.append(Person(1, act(1)))

        for i in range (under_ff) :
            population.append(Person(0, act(0)))

        saving_sum = 0
        health_sum = 0

        for person in population :
            person.change()
            saving_sum += person.get_cost()
            health_sum += person.health_cost

        saving_sum_arr.append(saving_sum)
        health_sum_arr.append(health_sum)

    return np.array(saving_sum_arr).mean() * -1, np.array(health_sum_arr).mean() * 1
```

```python
year_production_us = [70,159,193,130,152,220,369,423,416,750,928]
year_production_uk = [20,25,50,40,75,55,70,101,160,160,170]
year = [2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022]
```

```python
carbon_us = []
carbon_uk = []

for yr in year_production_us:
    carbon_us.append(get_data(yr)[0])

for yr in year_production_uk:
    carbon_uk.append(get_data(yr)[0])

plt.plot(year, carbon_us, label="US")
plt.plot(year, carbon_uk, label="UK")
plt.xlabel("Year")
plt.ylabel("Carbon Emissions Savings (Kg)")
plt.legend()
```

```python
years_us = []
years_uk = []

for yr in year_production_us:
    years_us.append(get_data(yr)[1])

for yr in year_production_uk:
    years_uk.append(get_data(yr)[1])

plt.plot(year, years_us, label="US")
plt.plot(year, years_uk, label="UK")
plt.xlabel("Year")
plt.ylabel("Life Years Saved (Thousands)")
plt.legend()
```

```python
uk_tf = 322824 // 10000
uk_te = 563780 // 10000

us_tf = 1907964 // 10000
us_te = 3796118 // 10000
```

```
uk_tf_d = get_data(uk_tf)
uk_te_d = get_data(uk_te)

us_tf_d = get_data(us_tf)
us_te_d = get_data(us_te)

print(f"---UK 2025---\nCarbon Emissions Saved: {uk_tf_d[0]/1000}\nLife Years Saved: {uk_tf_d[1]* 10000}\n")
print(f"---UK 2028---\nCarbon Emissions Saved: {uk_te_d[0]/1000}\nLife Years Saved: {uk_te_d[1]* 10000}\n")

print(f"---US 2025---\nCarbon Emissions Saved: {us_tf_d[0]/1000}\nLife Years Saved: {us_tf_d[1]* 10000}\n")
print(f"---US 2028---\nCarbon Emissions Saved: {us_te_d[0]/1000}\nLife Years Saved: {us_te_d[1] * 10000}\n")
```

In [26]:
```python
import pandas as pd
import janitor as jn
import numpy as np

d = {'Activity': ['Commute', 'Local', 'Recreational', 'Other'],
     'Automobile Percent': [45.8, 35, 18.3, 0.9],
     'Active Transportation Percent': [39.8, 21.9, 37.5, 0.8],
     'Would Not Percent': [1.1, 4.5, 92.9, 1.5],
     'Other Percent': [36.2, 13, 44.9, 5.9]}
df = pd.DataFrame(data=d)
```

In [27]:
```python
df["Replaced Automobile Trips Per User"]=(df["Automobile Percent"]*1778/100)/1796
df["Replaced Active Trips Per User"]=(df["Active Transportation Percent"]*1778/100)/1796
df["New Trips Per User"]=(df["Would Not Percent"]*1778/100)/1796
df["Replaced Other Trips Per User"]=(df["Other Percent"]*1778/100)/1796
```

In [30]:
```python
df["Total Trips"]=(df["Replaced Automobile Trips Per User"] + df["Replaced Active Trips Per User"]
+ df["New Trips Per User"] + df["Replaced Other Trips Per User"])
```

In [31]:
```python
df["Percent of Replaced Automobile Trips"]=100*df["Replaced Automobile Trips Per User"]/df["Total Trips"]
df["Percent of Replaced Active Trips"]=100*df["Replaced Active Trips Per User"]/df["Total Trips"]
df["Percent of New Trips"]=100*df["New Trips Per User"]/df["Total Trips"]
df["Percent of Replaced Other Trips"]=100*df["Replaced Other Trips Per User"]/df["Total Trips"]

df[["Activity","Percent of Replaced Automobile Trips","Percent of Replaced Active Trips",
   "Percent of New Trips","Percent of Replaced Other Trips"]]
```

Out[31]:

	Activity	Percent of Replaced Automobile Trips	Percent of Replaced Active Trips	Percent of New Trips	Percent of Replaced Other Trips
0	Commute	37.266070	32.384052	0.895037	29.454841
1	Local	47.043011	29.435484	6.048387	17.473118
2	Recreational	9.452479	19.369835	47.985537	23.192149
3	Other	9.890110	8.791209	16.483516	64.835165