MathWorks Math Modeling Challenge 2023

Adlai E. Stevenson High School

Team #16407, Lincolnshire, Illinois

Coach: Paul Kim

Students: Jack Chen, Collin Fan, Aadit Juneja, Aayush Kashyap, Nathan Ma

M3 Challenge FINALIST—$5,000 Team Prize

JUDGE COMMENTS

Specifically for Team # 16407 —Submitted at the Close of Triage Judging:

COMMENT 1: This was an excellent contribution! The executive summary to the US Secretary of Transportation was informative and concise. The response to the first and second parts were particularly impressive, especially the inclusion and comparison of distinct modeling efforts. Similarly, the response to the third component represented strong critical and analytical thinking skills. The strengths and weaknesses mentioned were also a nice portion of the project, and the conclusion was well-formed. Fantastic job!

COMMENT 2: Great executive summary -- you summarized the problem, your methods, and proposed solutions. For each question a reasonable model is presented, along with a well-reasoned explanation, including derivations of each term. Valid sources are cited. I was especially impressed by how well-written and organized the report is. Team did a great job explicitly stating assumptions and providing reasonable justifications. Good strengths and weaknesses analysis.

COMMENT 3: Very nice paper. Your executive summary was well done and included results, as well as your detailed set of assumptions with justifications. In Part 1, you chose the logistic equation to model sales -- nice choice but it would have been nice to see a graph as well to help the reader "see" the shape. In general, you should consider adding more graphs and charts to help tell your story.

COMMENT 4: Implementation of a Monte Carlo simulation in Part III is really creative! I also like that this team distinguishes bikes that have been sold vs those that are projected to be in use.

Ride Like the Wind: *The Growth of E-bike Use*

Team #16407

March 6, 2023

0 Executive Summary

Dear Secretary Buttigieg,

It's like riding a cloud—you can feel the fresh breeze of the city air, your heart pumping, your mind clearing, all on your commute to work. We're sure you've seen this latest example of the wonders of modern technology: the electric bike. E-bikes provide people with a quick means of transportation that comes without the worries of finding parking or getting stuck in long lines of traffic. As they begin to dominate the streets, e-bikes have become an essential part of urban life in the US. In this report, we predict the number of e-bikes sold in two and in five years from now. Next, we determine the factors that are most relevant to generating e-bike sales. Finally, we quantify the impact of e-bike sales in terms of carbon emissions, traffic congestion, and health.

In Part I, we create a logistic model that predicts the number of e-bikes sold in two years and in five years. We utilized existing data from the market for regular bicycles, and principles of economic demand to find the maximum limit for e-bike sales in a year. Then, we wrote a program to find the optimal logistic growth coefficient. Combining these factors allowed us to determine the final logistic growth equation. From our logistic function, we found that 2,719,755 e-bikes will be sold in 2025 and 5,391,897 e-bikes will be sold in 2028.

In Part II, we find the factors that have been most influential behind the rapidly growing popularity of e-bikes. We considered five factors: the proportion of Americans that already own an e-bike (a "coolness factor"), commute time, affordability, health, and environmental awareness. We synthesized these factors into a function describing the probability of purchasing (adopting) an e-bike for a given individual, mapping this consistently with the logistic model we found in Part I. Then, we took the partial derivative of the rate of adoption with respect to each factor to determine the relative importance of each factor. We found that the coolness factor was most important, followed by financial incentive, commute time, health concerns, and environmental awareness.

In Part III, we quantify the impact that e-bikes will have on the US. As more Americans turn to riding e-bikes, we will observe less carbon emissions and improvements to health. We synthesized our results from Part I and II to find a numerical function for the probability of adoption based on the factors we noted in Part II. Then, we applied these probabilities to a sample population of 10000 individuals, and ran a Monte Carlo simulation to find resulting changes in gas consumption and calories burned. We found that, over the next five years, e-bike adoption will reduce total US carbon emissions by 2,482,912 tons and burn 81.5 calories per day for the average new e-biking American.

Secretary Buttigieg, we wish you and others within the Department of Transportation luck in shaping the future of America. We hope that our findings have convinced you to make e-bikes a central part of that future.

Sincerely,

Team #16407

Contents

1 Part I: The Road Ahead

1.1 Restatement of the Problem

In this part, we develop a model that predicts the number of e-bikes sold in 2025 and 2028 in the US.

1.2 Assumptions

1. *We neglect the volume of consumers leaving the e-bike market.* Since e-bikes are an emerging technology, all products are relatively new (i.e., very few e-bikes are breaking down) and the purchase of e-bikes has not significantly reduced traffic (which would increase the appeal of cars and result in some e-bike drivers switching back to cars).

2. *The number of yearly sales of regular bikes follows the same overall shape as the number of yearly sales of e-bikes.* Regular bikes and e-bikes are different versions of the same product, the only difference between which is the motorized component of an e-bike. Hence, we use regular bike sales to project the long-term trends of e-bikes.

3. *The population change of the US over 5 years will not meaningfully impact the market cap of e-bike users.* The population of the United States has changed by roughly 2.4% from 2018 to 2023. [27] We predict that there likely will be a similarly negligible percent change in the US population from 2023 to 2028 when considering the e-bike market. As such, we will not consider how US population growth impacts the number of e-bike sales.

4. *E-bikes and regular bikes are substitute goods.* Both products serve the same purpose. It is very unlikely an individual will own both a regular bike and a e-bike.

5. *The number of annual e-bike sales after the Covid-19 pandemic must be considered separately from the annual sales before the pandemic.* Before the pandemic struck, e-bikes were primarily used as an easier method of transportation. However, once the pandemic began and many people became restricted to their home and the outdoors, e-bike sales soared due to their recreational and environmental appeal [10].

1.3 Variables

Variable	**Description**
t_0	Base year for logistic regression (2020)
t	Year under consideration
$y(t)$	Total number of e-bikes in use in year t
$z(t)$	Total number of e-bikes sold in year t
k_{rate}	Relative growth rate coefficient for $z(t)$ regression
k_c	Relative growth rate coefficient for $y(t)$ regression
L_{rate}	Maximum number of e-bikes in use possible in the USA
L_c	Maximum cumulative number of e-bikes in use in USA

Table 1: Variables for Part I

1.4 Model Development

Since we do not desire to utilize raw pre-COVID data, we will split up our model for Part 1 into 3 sections:

- Derive a logistic regression for $z(t)$ using post-COVID data.
- Take the integral of $z(t)$ over 10-year time intervals to determine the cumulative number of e-bikes in use from 2013-2022. Since e-bikes have a life span of 10 years [8], it is valid to take the integral of $z(t)$ to find the cumulative number of e-bikes in use in year t.
- Derive another logistic regression based on our data from the previous subpart to determine the cumulative number of e-bikes in use in year t.

1.4.1 Logistic Regression for $z(t)$

E-bikes are a relatively new product, currently exhibiting high growth within the US. However, this growth rate must decline and stabilize, as there cannot be an infinite amount of people owning an e-bike. From Assumption 2, we know that the growth rate will not become negative in the near future, as this has not yet happened for regular bikes. So, when the annual number of e-bikes sold maximizes, the number of e-bike owners should reach a steady state, similar to what we are already seeing with bikes [4]. This is because the number of e-bikes that stop working each year will be offset by the number of e-bikes sold in the same year.

A logistic regression can model growth of a population given a carrying capacity and constant relative growth rate coefficient. Within a logistic regression, the population initially increases exponentially before decaying exponentially towards the maximum population size.

Since e-bike sales will follow this pattern of high initial growth before receding and stabilizing at a constant value, we use a logistic regression to model annual e-bike sales. Our model will thus take the form:

$$z(t) = \frac{L_{rate}}{1 + e^{-k_{rate}(t-t_0)}} \tag{1}$$

Now, we must derive L_{rate} and k_{rate}.

1.4.2 Deriving $L_{\textbf{rate}}$

Current e-bike sales are still growing and are lower than the true maximum capacity for e-bike sales. Thus, to find the limiting constant for e-bike growth, we derive it from limiting constant for a very similar product: regular bikes. As such an established product, the regular, non-electric bike has already hit its own maximum capacity for growth. Over the past 24 years, there have been an average of 17.4 million bikes sold each year [4].

Furthermore, past research has shown that the market for bikes is perfectly competitive [21]. In a perfectly competitive market, the demand for the product is unit elastic, meaning that any change in price is met with an equally proportional change in quantity demanded. This concept can be used to estimate the maximum number of annual e-bike sales. We know that the maximum number of annual regular bike sales is approximately 17.4 million, the average price of regular bikes is \$200 [2], and the average price of e-bikes is \$3000 [19]. Thus, we have the following expression for L_{rate}:

$(200/3600) \cdot 17.4$, which gives a value of about 1.16 million e-bikes for L_{rate}.

1.4.3 Deriving k_{rate}

While the shape of the number of e-bikes distribution mimics the shape of the number of regular bikes distribution, the change in annual sales will not necessarily be the exact same. Thus, after finding L_{rate} to be 1.16 million, we still must find k_{rate}, the relative growth rate coefficient.

Using existing data from the past three years, we create a Python program that iterates through different values of k, finding the mean squared error for each respective value. This code is found in our Part 1 code appendix. After running this function, we find the error-minimizing value of k_{rate} to be 0.665.

So:

$$z(t) = \frac{1160000}{1 + e^{-.665(t-2020)}} \tag{2}$$

1.4.4 Re-calculating Annual E-bike Sales

Per Assumption 5, we must consider the logistic models for post-Covid e-bike sales and pre-Covid e-bike sales separately. Thus, to predict e-bike owners in future years, we cannot use data on annual sales from before the pandemic. As a result, we must extrapolate data points for the number of annual e-bike sales from 2013 through 2020 using data from post-Covid, in effect, working around the dramatic impact that the pandemic had on the number of e-bike sales. To do this, we use our equation for $z(t)$. Plugging in values of t from 2013 to 2022, we procure the following table:

Year	**Number of e-bikes sold**
2013	10,932
2014	21,070
2015	40,279
2016	75,835
2017	138,886
2018	242,625
2019	393,956
2020	580,000
2021	766,044
2022	917,375

Table 2: Projected number of e-bikes sold in each year from 2013-2022

1.4.5 Logistic Regression for $y(t)$

Now that we have our modified data for the number of e-bikes sold each year, we now desire $y(t)$, which takes the form:

$$y(t) = \frac{L_c}{1 + e^{-k_c(t-2020)}}$$

1.4.6 Deriving L_c

We now will find L_c. Since e-bikes have a 10-year lifespan [8], we can multiply L_{rate} by 10 to find L_c because after $z(t)$ becomes asymptotically close to L_{rate}, then the maximum number of usable e-bikes, and thus L_c, will be $10 \cdot L_{rate}$=11,600,000.

1.4.7 Deriving k_c

We must now find k_c. We use the same Python program that iterates through different values of k and finds the mean squared error for each respective value. This code is once again found in our Part 1 code appendix. After running this function, we find the error-minimizing value of k_c to be 0.5461. So:

$$y(t) = \frac{11,600,000}{1 + e^{-.5461(t-2020)}} \tag{3}$$

1.5 Results

After obtaining the values of k_c and L_c, our logistic regression model can be written as:

$$y(t) = \frac{11,600,000}{1 + e^{-0.5461(t-2020)}} \tag{4}$$

To find the number of e-bikes sold in year t, we take $y(t) - y(t-1)$. As such, we find that approximately 2.7 million e-bikes will be sold in 2025. 5.4 million e-bikes will be sold in 2028.

1.6 Strengths and weaknesses

One strength of our model is that our utilization of a logistic model accounts for one core aspect of the e-bike market: there is a cap on the number of e-bike users. Most other models would allow for the number of e-bike users to grow indefinitely, though this is not logical in practice. By using a logistic model, we account for the finite nature of the e-bike market, making it more practical and applicable.

Another strength of our model was that our model was that we were able to use the existing market for bikes to predict behavior in the e-bike market. Since there was a lack of data for e-bikes, we were forced to look to another market to predict trends. As bikes and e-bikes are substitute goods, the behavior of the two markets can be expected to be very similar, enabling us to create a reliable model for the relatively new U.S. e-bike market.

One weakness of our model is the lack of data driving our derivation of k_{rate} and k_c. In the provided M3 dataset, we were given the number of e-bike sales in the USA each year from 2018-2022. We decided to only use the datapoints from 2020-2022 because COVID greatly morphed the e-bike market [7]. We decided to derive $z(t)$ since that would generate more datapoints to drive the derivation of $y(t)$, though ultimately the model as a whole is based on a very small dataset, which can hinder accuracy.

2 Part II: Shifting Gears

2.1 Restatement of the Problem

In this part, we seek to find the dependence of e-bike adoption on a variety of factors in order to derive a ranking from the most impactful to least impactful factor. In this problem, we consider: commute time (C), affordability (F), health (H), and environmental awareness (E).

2.2 Assumptions

1. *Advancements in battery technology will have a negligible impact on consumer demand for e-bikes.* The current range of e-bikes is 20-100 miles [9], which is more than sufficient for most commutes.
2. *The only financial barrier towards buying an e-bike is the upfront cost of the bike.* In all US states, electricity is cheaper than gasoline [14]. Thus, we will not consider cost of ownership and maintenance on incentive to buy an e-bike, and only on an individual's ability to afford the direct cost of purchasing an e-bike.
3. *We are only considering the usage of e-bikes within urban areas.* With large distances to cover in suburban and rural areas, personal cars are a preferred choice due to the necessary distance and current limits on battery technology. The urban dominance of e-bikes is shown already in the Chinese e-bike market [22].
4. *The proportion of Americans living in urban areas remains approximately constant over short time intervals.* From 2014 to 2021, a 7-year time span, the proportion of Americans living in urban areas increased by only about 2.6% [26]. Thus, it is reasonable to assume that this proportion will remain approximately constant from 2021 (the last time this proportion was measured) and 2028 (the last year for which we are predicting).

2.3 Variables

Variable	Description
N	Population of adults in the United States
p	Probability of adoption in an arbitrary time unit τ
x	Proportion of Americans that own e-bikes, or $\frac{y}{N}$ from Part 1
C, F, H, E	Factors considered in model (see restatement of problem)
$f(C, F, H, E)$	Functional dependence of probability p on the factors C, F, H, and E
Δt_C	Increase in commute time after switching to e-bike (min)
I	Individual income (USD)
H	Rating of health importance $\in [0, 1]$
E	Rating of environmental importance $\in [0, 1]$

Table 3: Variables for Part II

2.4 Model Development

In this part of our model, we seek to find a function for the probability that an individual will purchase an e-bike in a given time τ. From Part I of our model, we found that a logistic model would best fit the trends for e-bike adoption. In this logistic model, the rate of adoption $\dot{x} = dx/dt$ is based on the number of adoptees x and the total population as so, with a constant of proportionality k:

$$\dot{x} = kx(1 - x) \tag{5}$$

In this part, we analyze this differential equation using the probability of adoption p.

Among the $N(1 - x)$ individuals who have not adopted e-bikes at a given time, they have an average probability p of purchasing an e-bike in time τ. Hence, in time dt, a proportion $p(dt/\tau)$ of the $N(1 - x)$ non-adoptees will convert to e-bikes:

$$dx = \frac{p(1 - x)}{\tau} dt \Rightarrow \dot{x} = \frac{p(1 - x)}{\tau} = \frac{p}{\tau}(1 - x) \tag{6}$$

Combining this with Eq. (5), we find that $p = k\tau x$, so $p \propto x$. This linear dependence of p on x is logical; the greater the number of individuals around a person that have adopted e-bikes, the more "popular" it will seem to them, increasing their inclination to purchase an e-bike. This once again justifies the use of a logistical model.

After finding this linear dependence of p on x, we can now consider the other factors that influence p – namely, commute time, affordability, health, and environmental awareness, which we consider to be a part of the function $f(C, F, H, E)$. In this function, $C, F, H, E \in [0, 1]$ represent probability distributions relating to commute time, affordability, health, and environmental awareness, respectively. We can thus redefine p as

$$p = \beta x f(C, F, H, E) \tag{7}$$

for some new constant of proportionality β. In the rest of this section, we seek to find the influence of each factor on the function $f(C, F, H, E)$. Along with the linear dependence of p on x, we will use this function to rank the influence of each factor on the probability of buying an e-bike.

2.4.1 Commute time

First, we will address the effect of commute time in regard to an individual's probability of adopting an e-bike by finding the function C. If a person's commute time increases because of purchasing an e-bike, their probability of purchase will decrease. We can model this according to an individual's attention span, which is a model of how long they are willing to tolerate a lengthier commute time.

According to a 1999 study of human attention span [1], the proportion of individuals who have attention spans of time t is normally distributed with a standard deviation $\sigma = 10$ minutes. We can map this proportion as equal onto our probability C of adoption as shown:

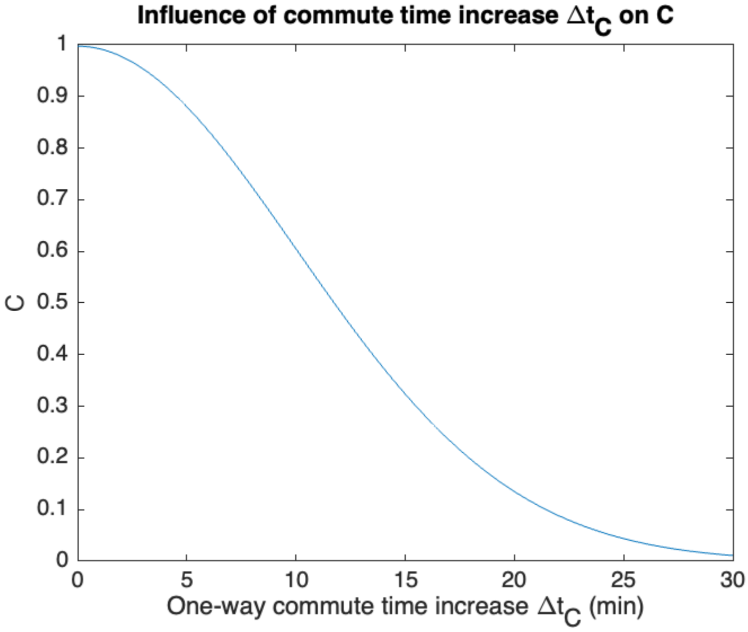

Figure 1: Influence of Δt_C on C

If commute time decreases as a result of purchasing an e-bike, commute time is not an inhibiting factor to purchasing an e-bike ($C = 1$). If commute time increases, we model the probability of purchasing an e-bike as the proportion of individuals that will tolerate such an increase (graph above). Thus, we can define the dependence of C on Δt_C as follows:

$$C = \begin{cases} 1 & \Delta t_C < 0 \\ e^{-0.005 \Delta t_C^2} & \Delta t_C \geq 0 \end{cases} \tag{8}$$

where we have used the definition of a normal model with $\mu = 0$ and $\sigma = 10$, scaling the graph up so the maximal value is 1.

2.4.2 Affordability

Per Assumption 2, we consider only the upfront cost of an e-bike as a financial inhibitor to whether a person purchases an e-bike. According to Marketplace [23], Americans save on average 8-9% of their total income. Thus, using the average price of \$3000 of an e-bike as cited in Part I, an individual needs a minimum yearly income of

$$I_{\text{min}} = \frac{\$3000}{0.085} \approx \$35300 \tag{9}$$

in order to be able to afford an e-bike. However, we do realize that as individual income scales, people will be more willing to spend their finances on an e-bike. The financial comfortable threshold for Americans is \$100000 [24], so we can assume that all individuals with an income above that amount will not have income as an inhibitor to their probability of purchase ($F = 1$). Thus, our probability function F will equal zero at all incomes less than \$35300 equal one at all incomes greater than \$100000. For income levels in between, the income scales directly proportional to the probability, as the proportion of household income spent on an e-bike scales linearly. Thus, for incomes between \$35300 and \$100000, we use a linear model to evaluate the relationship between the income of an individual and their willingness to adopt an e-bike. Our piecewise function for F

can therefore be described as so:

$$F = \begin{cases} 0 & I \leq 35300 \\ \dfrac{I}{64700} - \dfrac{353}{647} & 35300 < I < 100000 \\ 1 & I \geq 100000 \end{cases} \tag{10}$$

2.4.3 Health and Environmental Awareness

With commute time and affordability, we used data from surveys that corresponded to values ranging from 0 to 1 in $E \in [0, 1]$ and $H \in [0, 1]$ (see Section 2.4.5 below for specific correlations).

With commute time and affordability, we realize that if the time increase or financial cost is too large, there is essentially 0 probability that a person will adopt an e-bike; regardless of whether or not they have any health or environmental incentive. However, in the cases of the latter two factors we are considering, having a value of 0 in one of the factors might not skew the total probability to 0. For example, having no health incentive to purchase an e-bike does not necessarily inhibit them from doing so if their environmental incentive is nonzero. Therefore, the dependence of $f(C, F, H, E)$ on H and E cannot be multiplicative as we did with C and F.

Instead, we can combine the influence of H and E into one probability T that is then multiplied to C to yield

$$f(C, F, H, E) = CFT. \tag{11}$$

We can find T via a weighted average by how much Americans prioritize each factor in their daily lives. 68% of Americans greatly value their health [17] and 44% of Americans greatly value the environment [19]. Scaling these to obtain a sum of 100%, we find

$$T = 0.607H + 0.393E \tag{12}$$

Thus, our final function that determines the probability a person will buy an e-bike is

$$p = \beta x f(C, F, H, E) = \beta x CF\,(0.607H + 0.393E) \tag{13}$$

This means the rate of adoption is found as

$$\dot{x} = \frac{p}{\tau}(1 - x) = \frac{\beta x CF\,(0.607H + 0.393E)}{\tau}(1 - x) \tag{14}$$

2.4.4 Ranking methodology

We can find the importance of each factor by finding the magnitudes of their partial derivatives. This is equivalent to finding the possible $\partial \dot{x}$ according to a change in some variable, i.e. ∂x. All of these partial derivatives can be found in the gradient of $\dot{x}$; ranking them will allow us to find the relative importance of each factor.

$$\nabla \dot{x}(x, C, F, H, E) = \left\langle \frac{\partial \dot{x}}{\partial x}, \frac{\partial \dot{x}}{\partial C}, \frac{\partial \dot{x}}{\partial F}, \frac{\partial \dot{x}}{\partial H}, \frac{\partial \dot{x}}{\partial E} \right\rangle \tag{15}$$

To find $\nabla \dot{x}$, we will use current US average values for each of the factors to evaluate each partial derivative. These values can be found in the table below:

Factor	Value	Source
x	0.0145	[19],[11],[12]
F	0.32	[13]
H	0.39	[16]
E	0.25	[19]
C	0.489	[6], [3], [26], [25]

Table 4: Current US average values for each factor

The values for E, H, and C are not readily available online; we provide the derivation below.

2.4.5 Derivation of US average values for E and H

To derive E, we take data from a survey asking respondents how much they cared about the environment [19]. Respondents were given four options: "Great deal," "Fair amount," "Only a little," and "Not at all". We scaled E linearly based on environmental concern. Respondents answering a "Great Deal" were given an E of 1, respondents answering "Fair amount" were given an E of 0.67, respondents answering "Only a little" were given an E of 0.33, and respondents answering "Not at all" were given an E of 0. Multiplying each response value by their respective probability of occurring, we obtain the average E as 0.68.

To derive H, we use a similar approach, taking data from a survey on American health [16] respondents either reported that they met both fitness standards, only met one fitness standard, and met no fitness standards. Respondents meeting both fitness standards were given an H of 1, respondents meeting one standard were given an H of 0.5, and respondents meeting no standards were given an H of 0. Multiplying each response value by their respective probability of occurring, we obtain the average H as 0.39.

2.4.6 Derivation of US average value for C

Note that not all states allow Class 3 e-bikes, which can reach a maximum speed of 28 miles per hour. Thus, the safer option is to generalize the e-bikes to go at Class 2 speed, or a maximum of 20 miles per hour. That is very similar to the maximum speed of a normal bike. Therefore, we can assume that e-bikes travel at the same speed as regular bikes in crowded urban sidewalks.

Normally, in congested cities, cars travel around 19 miles per hour [6] while bikes travel around 14 miles per hour [3]. Since we have already established that a normal urban commute takes around 28.1 minutes [25], the additional commute time in using an e-bike is:

$$\Delta t_C = 28.1 \cdot \frac{19-14}{14} = 10.04 \text{ minutes} \tag{16}$$

Plugging this in to our piecewise equation in Section 2.4.1, we get $C_0 = 0.604$. However, per Assumption 3, this change is only applied to individuals living in urban areas. Thus, to scale this to the entire US, we must take a weighted average; i.e. we must multiply C_0 by the proportion of Americans that live in cities, which is 0.813 [26]. Thus, we find $C = C_0 \cdot 0.81 = 0.489$.

2.5 Results

Evaluating each partial derivative using the values from Table 11, we can find their values in the table below (up to the numerical prefactor β/τ):

Factor	Partial Derivative	Value
$\partial\dot{x}/\partial x$	$(1-2x)CF(0.607H+0.393E)$	0.051
$\partial\dot{x}/\partial C$	$(x-x^2)F(0.607H+0.393E)$	0.0015
$\partial\dot{x}/\partial F$	$(x-x^2)C(0.607H+0.393E)$	0.0023
$\partial\dot{x}/\partial H$	$0.607(x-x^2)CF$	0.0014
$\partial\dot{x}/\partial E$	$0.393(x-x^2)CF$	0.00088

Table 5: Partial derivative values for each factor

Hence, **the most significant drivers of e-bike sales are the number of current adoptees (i.e. "coolness factor"), affordability, commute time, health concerns, and environmental concerns, in that order.**

The influence of the first factor (x) is an order of magnitude greater than the others; this is reasonable, since the proportion of current adoptees is so small ($x = 0.0145$) that the visibility of e-bikes to a given consumer is the dominant effect on whether they purchase a bike.

2.6 Strengths and weaknesses

One strength of our model is that we managed to model logical decision-making via piecewise functions for C and F. For example, if commute time is reduced when an e-bike is utilized, then commute time is not an inhibiting factor to purchasing an e-bike and thus $C = 1$. The piecewise nature is representative of how humans often have "cutoffs" when making decisions. The cusps in our piecewise graphs represent these cutoffs.

Another strength of our model is that the factors we analyze are comprehensive of an individual's life. Rather than only analyzing factors within the professional realm of one's life, for instance, we looked at factors from the personal realm (health and environment) as well as the professional realm (commute and finances). This wide variety of factors under analysis enables our model to provide an accurate determination of the most important factor.

One weakness of our model is that the coefficients for H and E in Eq. (12) are scaled to sum to 100%, even though those percentages are from different sources. Ideally, we would have data that has individuals who were asked whether they valued health or the environment more, as that is what we want those weights for H and E to represent. The questions asked in the survey are not exactly equivalent to what we want the weights for H and E to represent in Eq. (12).

Given time and data constraints, we were also unable to consider the expansion of e-bike range due to battery improvement technology. While we only considered the urban market, we realize that with the advancement of battery technology, the range of e-bikes can be improved to include suburban and rural consumers in the market. We note that our results are still highly accurate, though, due to the current dominance of urban consumers in the e-bike market [28].

3 Part III: Off the Chain

3.1 Restatement of the Problem

In this part, we quantify the impact of e-bike sales on carbon emissions and Americans' health after a five-year period.

3.2 Assumptions

1. *The average number of calories burned per hour is similar for every biker.* Although different factors may impact an individual's rate of burning calories while riding a bike, these factors will even out the average at around 300 calories per hour [5].

2. *The average mileage per gallon per car is similar.* Different makes of cars may be more or less efficient compared to the nationwide average, we believe the total number will even out at the average mileage.

3.3 Variables

Variable	Description
p	Probability of adoption in a time unit τ
$\tau = 1$ year	Time unit baseline for p (years)
β	Constant of proportionality dependent on τ (years)
m	Average car gas mileage (mpg)
v	Average car speed (mph)
t_i	Commute time for individual i (hrs/month)
g_i	Gas use per month for individual i (gallons)
Δg_j	Change in gas use in Monte Carlo iteration j
t_e	Time spent commuting on an e-bike per month (hours)
c	Calories burned while riding e-bike (kcal/hr)
$\Delta t = 1$ month	Time interval for each iteration of simulation (months)

Table 6: Variables for Part III

3.4 Model Development

In this section, we find the impact of e-bike sales on carbon emissions, traffic congestion, and health by using a Monte Carlo simulation. We will first find the unknown numerical prefactor β in our equation for p in Part II to find a value of the probability of adoption p_i for individuals in a simulated population. According to this probability, we can simulate the incorporation of e-bikes into the population and their effect on carbon emissions, traffic congestion, and health.

We use a Monte Carlo simulation instead of simply considering averages for each factor (i.e. average reduction in gas use according to average probability and average commute time) because the factors are not independent of one another. For example, a person with a very long commute time is highly unlikely to convert to an e-bike, but could have a significant impact on the reduction in gas emissions (in the highly unlikely case they do convert). Therefore, we must consider all members of a population on an individual basis; only a simulation can accomplish this.

3.4.1 Deriving numerical prefactor β

In Eq. (14) of Part II, we defined the rate of e-bike adoption $\dot{x}$ in terms of a constant of τ, a given time unit, p, the probability of adoption in time τ, and x, the current proportion of adoption. The probability of adoption p depends on τ and a constant of proportionality β as defined in Part II. Note Eq. (14) is the differential equation for a logistic model, where the constant of proportionality k can be found by

$$k = \frac{\beta CF(0.607H + 0.393E)}{\tau} \tag{17}$$

Note that by integration, the following two forms of the logistic model are equivalent:

$$\dot{x} = kx(1-x) \tag{18}$$

$$x = \frac{1}{1-e^{-kx}} \tag{19}$$

Hence, the value of $k = 0.5461$ found in Part I can be applied to Eq. (14) to yield

$$0.5461 = \frac{\beta CF(0.607H + 0.393E)}{\tau} \Rightarrow \beta = 10.42\,\text{years} \tag{20}$$

Since this value of k was found with nationwide averages, we used the data from Table 11 to find β.

3.4.2 Monte Carlo simulations

We will now simulate the adoption of e-bikes in a sample population of 10,000 individuals. For each individual, we will randomly generate a value for F, E, H for the individual, scaled by probability given the survey results of [19][11][12][13][16]. For C, we will use a distribution of original commute time t_c [25] and scale it up to find C as we did in Section 2.4.6. The data we used for our Monte Carlo simulations are listed in Appendix 5.

3.4.3 Carbon emissions

Starting with a Monte Carlo Simulation for carbon emissions, we seek to quantify the reduction in gas usage by US consumers due to transition to e-bikes. In every iteration of our simulation $\Delta t = 1$ month, we consider the probability of each individual of switching to an e-bike $p_i(\Delta t/\tau)$. Given this individual's commute time t_i in hours per month, the amount of gas they use in that month g_i can be found via average car gas mileage m [20] and average urban car velocity v [6].

$$g_i = \frac{\text{distance}}{\text{mileage}} = \frac{vt_i}{m} = 0.785t_i \tag{21}$$

In each iteration of the simulation, the expected value of a given individual's gas savings is $p_i(\Delta t/\tau)g_i$. Thus, the total change in gas saved for the whole population Δg in iteration $j \in 1, 2, 3 \cdots 60$ (for 5 years of 12 months each) can be found by summing this number among all N individuals in the population.

$$(\Delta g)_j = \sum_i^N \frac{p_i g_i \Delta t}{1\,\text{year}} = \sum_i^N \beta^{1/12} x C_i F_i g_i \Delta t(0.607H_i + 0.393E_i) \tag{22}$$

We found β with $\tau = 1$ year; but, given our iterations of units of months (for greater levels of accuracy for our summation), we must take the 12th root of β in our equation. After each iteration, we update the value for the proportion of adoption x in the summation and remove individuals who already adopted e-bikes from the simulation population.

Now, note that individuals that moved off of cars in the first iteration will always contribute to the reduction in gas consumption. Hence, the total gas savings over five years can be found by

$$\Delta g_{\text{total}} = (\Delta g)_1 + [(\Delta g)_1 + (\Delta g)_2] + [(\Delta g)_1 + (\Delta g)_2 + (\Delta g)_3] + \cdots [(\Delta g)_1 + (\Delta g)_2 + (\Delta g)_3 + \cdots (\Delta g)_{60}] \tag{23}$$

Finally, we scale up our results proportionally to the US population from our simulated population of 10,000. This methodology is outlined in our code in Appendix 4.2.

3.4.4 Health benefits

As with health benefits, we approach the situation with the exact same methodology as the previous section, except now considering the time individuals spend commuting on e-bikes t_e rather than gallons of gas saved; as all other. After finding the cumulative total time spent on bikes t_e by individuals of the population, we can multiply our result by the rate of calories burned per hour on an e-bike ride $c = 300$ kcal/hr [5] to find the total number of calories burned for the population. This methodology is outlined in our code in Appendix 4.2.

To derive a more valuable statistic, we can divide this total number of calories burned by our projected total number of adoptees after this 5-year period to find the average number of calories burned over 5-years for the average adoptee. We can also further divide this number to find the average number of calories burned per day.

3.5 Results

After running our simulations, we yield the following results:

Variable	**Value**
Total gas saved (over 5 years)	558,773,834 gallons
Reduction in carbon emissions (over 5 years)	2,482,912 tons
Total calories burned (over 5 years)	333,019,816,920 kcal
Calories burned for average adoptee (over 5 years)	148,796 kcal

Table 7: Cumulative results for Part III after a 5-year period

Thus, over a 5-year period, the total reduction in carbon emissions is **2482912 tons** and the average number of calories burned for the average rider per day is 148796/365/5 = **81.5 kcal**.

3.6 Strengths and weaknesses

One strength of our model was its ability to synthesize the results of the previous two parts to quantify the impacts of e-bike sales, using the constant k generated from Part I and the conversion function found in Part II. By utilizing smaller one month time intervals, we were able to more precisely project the overall adoption of e-bikes and how it would translate towards reducing carbon emissions and burning calories for Americans.

One weakness of our model was that we were not able to consider the carbon emissions of processing electricity (a significant portion of which is still done with fossil fuels). However, as with electric cars, the fossil fuel emissions of running on electricity are much less than a standard combustion engine [15].

References

[1] Enhancing human attention span with hd analytics. URL http://d0a5a114441964e66f5b-0cbc9ea2dd077292aca26784763801f6.r78.cf2.rackcdn.com/2016/01/AttentionEconomicsWhitepaper.pdf.

[2] How much does a bike cost? average bicycle price in 2022, . URL https://www.bikedenver.org/how-much-does-a-bike-cost/.

[3] What is the average speed for urban cycling?, . URL https://sustainabilityinfo.com/transportation/what-is-the-average-speed-for-urban-cycling/.

[4] U.s. bicycle unit sales from 1992 to 2015, . URL https://www.statista.com/statistics/236152/us-unit-sales-of-bicycles/.

[5] E-bike-fahren: Der ideale sport für Übergewichtige. URL https://www.focus.de/gesundheit/praxistipps/gelber-bluetenstaub-dicke-luft-problem-fuer-allergiker-folgen-des id_8851576.html.

[6] Gridlocked cities: traffic patterns revealed across 20 major u.s. cities. URL https://www.geotab.com/press-release/traffic-congestion-patterns/.

[7] How this pandemic has affected e-bike demand in the usa, . URL https://slinkerbikes.com/2021/11/22/how-pandemic-affect-the-e-bike-demand-in-usa-2/.

[8] How long do ebikes last?, . URL https://storybicycles.com/blogs/ebike-blog/how-long-do-ebikes-last.

[9] Electric bike range: How far can they go?, . URL https://pedalelectric.com/blogs/journal/electric-bike-range-how-far-can-they-go#:~:text=On%20average%2C%20most%20ebikes%20will,planet%20will%20say%20thank%20you).

[10] Electric bicycles soar in popularity during pandemic, . URL https://spectrumnews1.com/ca/la-west/transportation/2022/01/18/electric-bicycles-popular-during-pandemic.

[11] E-bikes get up to speed in popularity, . URL https://www.washingtonpost.com/health/ebikes-calories-health-exercise/2021/09/24/09035020-17ec-11ec-b976-f4a43b740aeb_story.html.

[12] Estimated sales of electric bicycles in the united states between 2012 and 2016, . URL https://www.statista.com/statistics/326124/us-sales-of-electric-bicycles/.

[13] What is disposable income. URL https://fortune.com/recommends/banking/what-is-disposable-income/.

[14] Driving on electricity is cheaper than gas in all 50 states. URL https://www.pluglesspower.com/learn/driving-electricity-cheaper-gas-50-states/.

[15] Electric vehicle myths. URL https://www.epa.gov/greenvehicles/electric-vehicle-myths.

[16] Only 23% of americans get enough exercise, a new report says, . URL https://time.com/5324940/americans-exercise-physical-activity-guidelines/.

[17] Study: Americans are focusing more on their health as a result of covid-19, . URL https://www.businesswire.com/news/home/20201015005720/en/Study-Americans-Are-Focusing-More-on-Their-Health-as-a-Result-of-COVID-19.

[18] Percentage distribution of household income in the united states in 2021, by ethnicity. URL https://www.statista.com/statistics/203207/percentage-distribution-of-household-income-in-the-us-by-ethnic-group/.

[19] Ride like the wind. URL https://m3challenge.siam.org/node/596.

[20] Average fuel economy by major vehicle category. URL https://afdc.energy.gov/data/10310.

[21] Vosper: Perfect competition, part one: why the bike business has it, and what's so perfect about it. URL https://www.bicycleretailer.com/opinion-analysis/2021/10/13/vosper-perfect-competition-part-one-why-bike-business-has-it-and-what%E2%80%99s.

[22] To e-bike or not to e-bike? a study of the impact of the built environment on commute mode choice in a small chinese city. URL https://www.jtlu.org/index.php/jtlu/article/view/1807/1556.

[23] Americans are saving little — while they spend up a storm. URL https://www.marketplace.org/2022/12/01/americans-are-saving-little-while-they-spend-up-a-storm/.

[24] What salary does it take to be comfortable? here's what americans say. URL https://www.fool.com/the-ascent/personal-finance/articles/what-salary-does-it-take-to-be-comfortable-heres-what-americans-say/#:~:text=A%20new%20report%20from%20Personal,even%20getting%20a%20new%20job.

[25] Travel time to work in the united states: 2019, . URL https://www.census.gov/content/dam/Census/library/publications/2021/acs/acs-47.pdf.

[26] Urban and total populations, . URL https://www.macrotrends.net/.

[27] U.s. population growth rate 1950-2023. URL https://www.macrotrends.net/countries/USA/united-states/population-growth-rate.

[28] Global – are e-bikes the future of urban mobility? URL https://business.yougov.com/content/44158-global-are-e-bikes-future-urban-mobility.

4 Code

4.1 Part I code

```
import math

def logistic(L, k, t):
    return L/(1+math.e**(-k*(t-2020)))

def error_c(L, k):
    sum = 0
    arr = [16000,32000,62000,117000,222000,409000,723000,1207000,1880000,3740000]#calculated via integral
    for i in range(2013, 2023):
        diff = logistic(L, k, i)-arr[i-2013]
        sum+=diff*diff
    return sum


def error_rate(L,k):
    sum = 0
    arr = [416000,750000,928000]#provided M3 data
    for i in range(2013, 2023):
        diff = logistic(L, k, i)-arr[i-2013]
        sum+=diff*diff
    return sum


a = 0.0001
L_rate=1160000 #max number of e-bikes sold in a year
L_c = 11600000 #max number of e-bike in USA
min_error = 9999999999999999


#FIND K_RATE
for i in range(1, 100000):
    cur_error = error_rate(L_rate, i / 10000)
    #print("k: ",k,"Error: ",cur_error)
    if (cur_error < min_error):
        min_error = cur_error
        a = i / 10000
print(a) #display optimal k_rate

#FIND K_C
a=0.0001#manually used optimal k_rate to calculate elements of arr in error_c, and then reset a to 0.0001
min_error = 9999999999999999#reset min_error
for i in range(1, 100000):
    cur_error = error_c(L_c, i / 10000)
    #print("k: ",k,"Error: ",cur_error)
    if (cur_error < min_error):
        min_error = cur_error
        a = i / 10000
print(a)#display optimal k_c
```

```
print(logistic(L_c,a,2025)-logistic(L_c,a,2024))
print(logistic(L_c,a,2028)-logistic(L_c,a,2027))#display sales in 2025 and 2028
```

4.2 Part III code

```
import random
import math
#adoption function
def p_adopt(x, c, f, h, e):
    return 1.22*x*c*f*(0.607*h+0.393*e)
#assigning commute times and their respective probabilities
commute_min = [0]*1000
for i in range(0, 27):
    commute_min[i] = 2.5
for i in range(27, 119):
    commute_min[i] = 7
for i in range(119, 248):
    commute_min[i] = 12
for i in range(248, 397):
    commute_min[i] = 17
for i in range(397, 538):
    commute_min[i] = 22
for i in range(538, 604):
    commute_min[i] = 27
for i in range(604, 743):
    commute_min[i] = 32
for i in range(743, 775):
    commute_min[i] = 37
for i in range(775, 816):
    commute_min[i] = 42
for i in range(816, 901):
    commute_min[i] = 52
for i in range(901, 969):
    commute_min[i] = 75
for i in range(969, 1000):
    commute_min[i] = 90
#assigning health scores
health = [0]*100
for i in (0, 23):
    health[i] = 1
for i in (23, 55):
    health[i] = 0.5
#assigning environmental scores
enviro = [0]*100
for i in (0, 25):
    enviro[i] = 1
for i in (25, 50):
    enviro[i] = 0.67
for i in (50, 75):
    enviro[i] = 0.33
#total urban population
urban_pop = 214000000
#starting e-bike population
start = 3740000
gal_saved = 0
#iterate through 5 years within 60 one month intervals
for i in range(0, 60):
```

```
    new_bikers = 0
    #gallons saved per month
    temp_gal = 0
    for j in range(10000):
        #assigning c, h, e, f
        c = commute_min[random.randint(0, 999)]
        #gallons saved if this specific individual converts
        gal = 19*c*44/(60*24.2)
        c*=5/14
        h = health[random.randint(0, 99)]
        e = enviro[random.randint(0, 99)]
        f = random.randint(0, 5)
        if f==0:
            f = random.randint(0, 27000)
        elif f==1:
            f = random.randint(27000, 52000)
        elif f==2:
            f = random.randint(52000, 85000)
        elif f==3:
            f = random.randint(85000, 141000)
        else:
            f = 141000
        if f<=35300:
            f = 0
        elif f<100000:
            f = f/64700-353/647
        else:
            f = 1
        c = math.e**(-0.005*c*c)*0.81
        prob = p_adopt(start/urban_pop, c, f, h, e)
        if random.random()<=prob:
            new_bikers+=1
            temp_gal+=gal*(60-i)
    #scale gallons saved and new bikers up
    gal_saved += temp_gal*(urban_pop-start)/10000
    start+=(new_bikers)*(urban_pop-start)/10000
print(gal_saved)
```

```
import random
import math
#same general idea of above program but slightly modified to account for calories burned instead of gas saved
def p_adopt(x, c, f, h, e):
    return 1.22*x*c*f*(0.607*h+0.393*e)
commute_min = [0]*1000
for i in range(0, 27):
    commute_min[i] = 2.5
for i in range(27, 119):
    commute_min[i] = 7
for i in range(119, 248):
    commute_min[i] = 12
for i in range(248, 397):
    commute_min[i] = 17
for i in range(397, 538):
```

```
    commute_min[i] = 22
for i in range(538, 604):
    commute_min[i] = 27
for i in range(604, 743):
    commute_min[i] = 32
for i in range(743, 775):
    commute_min[i] = 37
for i in range(775, 816):
    commute_min[i] = 42
for i in range(816, 901):
    commute_min[i] = 52
for i in range(901, 969):
    commute_min[i] = 75
for i in range(969, 100):
    commute_min[i] = 90
health = [0]*100
for i in (0, 23):
    health[i] = 1
for i in (23, 55):
    health[i] = 0.5
enviro = [0]*100
for i in (0, 25):
    enviro[i] = 1
for i in (25, 50):
    enviro[i] = 0.67
for i in (50, 75):
    enviro[i] = 0.33
urban_pop = 214000000
start = 3740000
cal_burned = 0
for i in range(0, 60):
    new_bikers = 0
    temp_cal = 0
    for j in range(10000):
        c = commute_min[random.randint(0, 999)]
        #potential calories burned
        cal = (c*19*300*43)/(14*60)
        c*=5/14
        h = health[random.randint(0, 99)]
        e = enviro[random.randint(0, 99)]
        f = random.randint(0, 5)
        if f==0:
            f = random.randint(0, 27000)
        elif f==1:
            f = random.randint(27000, 52000)
        elif f==2:
            f = random.randint(52000, 85000)
        elif f==3:
            f = random.randint(85000, 141000)
        else:
            f = 141000
        if f<=35300:
            f = 0
```

```
        elif f<100000:
            f = f/64700-353/647
        else:
            f = 1
        c = math.e**(-0.005*c*c)*0.81
        prob = p_adopt(start/urban_pop, c, f, h, e)
        if random.random()<=prob:
            new_bikers+=1
            temp_cal+=cal*(60-i)
    cal_burned += temp_cal*(urban_pop-start)/10000
    start+=(new_bikers)*(urban_pop-start)/10000
print(cal_burned)
```

5 Data

5.1 Data for for Part III Monte Carlo simulations

Factor	Probability	I	Source
Income	20%	0-27k	[18]
	20%	27k-52k	
	20%	52k-85k	
	20%	85k-141k	
	20%	141k+	

Table 8: Generation values of I

Factor	Response	Probability	Value	Source
Health	Meets Fitness Standards	23%	0.23	[16]
	Only Meets One Standard	32%	0.16	
	Meets None	45%	0	

Table 9: Generation values of H

Factor	Response	Probability	Value	Source
Environmental awareness	Care a great deal	44%	0.44	[19]
	Care a fair amount	27%	0.18	
	Care only a little	18%	0.06	
	Does not care at all	10%	0	

Table 10: Generation values of E

Factor	Probability	C	Source
Commute Time	2.7%	0.81	[25]
	9.2%	0.79	
	12.9%	0.74	
	14.9%	0.67	
	14.1%	0.59	
	6.6%	0.51	
	13.9%	0.42	
	3.2%	0.34	
	4.1%	0.26	
	8.5%	0.14	
	6.8%	0.02	
	3.1%	0.005	

Table 11: Generation values of C