MathWorks Math Modeling Challenge 2023

FW Buchholz High School

Team #16629, Gainesville, Florida

Coach: Ziwei Lu

Students: Himal Bamzai-Wokhlu, Nicholas Dang, Nolan Gao, Melissa Li, Nathan Wei

M3 Challenge FINALIST—$5,000 Team Prize

JUDGE COMMENTS

Specifically for Team # 16629 —Submitted at the Close of Triage Judging:

COMMENT 1: Good recognition of model limitation for Q3, important points identified in overall conclusion

COMMENT 2: Well written paper. Excellent summary that showcased your results and findings. The log transformation is used to convert the exponential behavior into a linear one in modeling growth in the first part without accounting for a limiting capacity for the growth of e-bike users. All figures and graphs had a clear title, a label, a caption, and the axes were labeled. Even your tables did not miss a title, a header, a label, and a caption. In developing your model for each section, all variables and parameters were clearly stated and defined, and the provenance of their numerical values should be clear with their units. You provided an articulation of your assumptions with justification based on your reasoning and prior research.

Monte Carlo simulation was given to consider multiple factors together in a comprehensive way in second part to discern the most prominent factors contributed to growth in e-bike popularity. In particular, the team assessed the relative importance of transportation costs, maintenance costs, and commute costs in both rural and urban settings. Nice work. The sensitivity analysis could have been improved in each part. For the last part your model clearly forecasted a 7.16 million ton reduction of CO2 output, an 8% decrease in traffic, and a net decrease of 648 deaths. You had fresh ideas and novel thinking in integrating upkeep costs which also points to some assumptions such as the battery costs for ebikes and their upkeep may not remain fixed. The model in part 2 have potential to adopt in different geographical locations as it pulls the data regarding the proportion of consumers who worked from home, commuted by car, and commuted via public transportation based on location.

Excellent work and reporting.

COMMENT 3: The team showed a very good understanding of model development. The report was very well written and explained. In the future, the team will benefit from carrying out sensitivity analysis on each model to predict its robustness and validity. Great job!

COMMENT 4: The paper presented solid solutions to all three problems given. The assumptions are very well established, and the analysis are detailed, particularly for the problem 2 and 3. The assumption “Pandemic era growth in e-bike sales will continue” used for the problem 1 could have been more thoughtful. This assumption could have led to inaccurate result as the majority of the history data used for the regressions do not belong to the pandemic era. In any case, all over, very well-presented paper.

COMMENT 5: Good ES with model motivation and results. Solution to Q3 is extensive but could still be improved by considering commuting distance.

Executive Summary

TO SECRETARY BUTTIGIEG:

We are writing to report to you the status of e-bikes as emerging transportation assets for the United States. By combining speed, flexibility, and affordability, the e-bike occupies a unique niche that is set to capitalize on the zeitgeist of our fast-paced, interconnected modern world. In this report, our team has developed a three-part model to predict the spread of e-bikes throughout the US in the coming years, identify the most salient causes behind the growth of e-bikes, and quantify the direct effects of an increase in e-bike usage for Americans.

The first section of our report models the growth in electric bike sales to predict the number of e-bikes that will be sold in the US 2 years and 5 years in the future. We first obtained data on the number of e-bikes sold in the past few years in the US. The data was fit to a linear regression by graphing the natural logarithm of e-bike sales versus year. Using this model, we predict there will be approximately 1.545 million e-bikes sold in 2025 and 2.75 million e-bikes sold in 2028, as compared to 1.053 million in 2023. Our model shows exponential growth in the number of e-bikes sold, demonstrating that e-bikes are poised to become a vital form of transportation in the US.

Our second model utilizes a robust Monte Carlo simulation to discern the most prominent underlying factors that have contributed to growth in e-bike popularity. In particular, we assessed the relative importance of transportation costs, maintenance costs, and commute costs in both rural and urban settings. Our findings suggest that car maintenance costs have been a driving factor in both urban and rural communities' transition to e-bikes over the past decade. However, our findings for 2021 also hint that changes in remote workplace dynamics, as a result of recent technological improvements and the COVID-19 pandemic, will be of critical consideration in the future.

The third section of our report seeks to quantify some of the impacts that this growing shift to e-bikes may have on the climate, traffic conditions, and the health and safety of our citizens. To determine the decrease in CO_2 output due to electric bikes sold in a given time period, we obtained data to calculate the effective number of e-bikes and created a linear regression to predict the average biking distance per consumer. In order to determine the potential reduction in traffic congestion, we calculated the number of passengers who would no longer travel in a private vehicle and subsequently considered the mean occupancy of a vehicle. We compared accident rates among cars compared to bikes to model the number of expected deaths due to the increase in electric bikes. Our predictions emphasize the potential of electric bikes to affect positive change; by 2028, our model forecasts a 7.16 million ton reduction of CO_2 output, an 8% decrease in traffic, and a net decrease of 648 deaths.

From the results of our three models, it is apparent that electric bikes are experiencing a boom in popularity and will greatly expand in scope. Their efficiency and affordability make them desirable to the general market, and the positive societal externalities such as decreased pollution, traffic, and accident-related deaths support the continued expansion of the market. Clearly, society has seen the electrifying effects of switching gears and consuming electric bikes.

It is therefore the recommendation of this modeling group that your department consider the enormous utility of our three models in the realization of your future bureaucratic agenda.

Contents

1 Q1: The Road Ahead

1.1 Problem Restatement

In part I of this problem, our team was prompted to create a model to predict the growth in e-bike sales in the United States. We were then asked to apply our model to predict the number of e-bikes that will be sold 2 and 5 years from now, in 2025 and 2028 respectively.

1.2 Assumptions

1. **There will be no drastic changes in consumers' preference for electric bikes in the near future.** Changes in consumers' attitudes towards e-bikes heavily influence their sales. For example, if people begin to heavily prioritize environmental protection, there will likely be a larger increase in e-bike sales. However, such a shift is very difficult to predict, so for the sake of this model we assume this will not happen in the near future. This assumption allows us to assume that current trends in e-bike sales will continue into the future.
2. **Pandemic era growth in e-bike sales will continue in the near future.** There was a noticeable increase in e-bike sales in 2021 and 2022. While this may be a fairly temporary effect, we will likely see this increase persist for several years, as it takes time for markets to readjust. In addition, it is safe to assume that some pandemic-motivated shifts, such as transitions to online-work, will continue indefinitely. Since we are asked to predict the growth after only 5 years, it is reasonable to use data from the pandemic era to model growth in e-bike sales for the short-term future.

1.3 Variables

Variable	**Definition**	**Unit**
T	Time	Years
$B(T)$	Number of e-bikes sold in a given year	Thousands of e-bikes per year
β	Slope of our linear regression curve	Thousands of e-bike sales/Year^2
α	y-intercept of our linear regression curve	Thousands of e-bike sales/Year
R^2	Percentage of the variation of our dependent variable ($B(T)$ or $\ln(B(T))$) that can be explained by our independent variable (T)	—
n	Number of data points	—

1.4 Gathering Additional Data

Through our research, we were able to find additional data on the number of e-bikes sold in the US for the years 2012, 2013, and 2017 [1][3]. However, due to time constraints and the limited availability of reliable data, we were unable to provide data for the years 2014-2016.

1.5 Our Model

Using the data points shown in **Table 1** below, we used linear regression to calculate the Least Squares Regression Line for the graph of e-bike sales versus time.

For a linear regression line, we have $B(T) = \beta \cdot T + \alpha$, where $B(T), T, \beta,$ and α are the variables defined above, and

$$\beta = \frac{n\sum(TB(T)) - \sum T \sum B(T)}{n\sum(T^2) - (\sum T)^2}$$

$$\alpha = \frac{\sum B(T) - \beta \sum T}{n}$$

To expedite this process, we utilized the curve-fitting function in Excel to quickly obtain our regression lines.

T	B	$\ln(B)$
2012	103	4.635
2013	185	5.220
2017	263	5.572
2018	369	5.911
2019	423	6.047
2020	416	6.031
2021	750	6.620
2022	928	6.833

Table 1. Values of T and B. [**20**]

Our original plot (shown in **Figure 1A.**) assumed a linear relationship between B and T. The low R^2 value that was yielded suggested that a linear model did not properly predict the growth in e-bike sales.

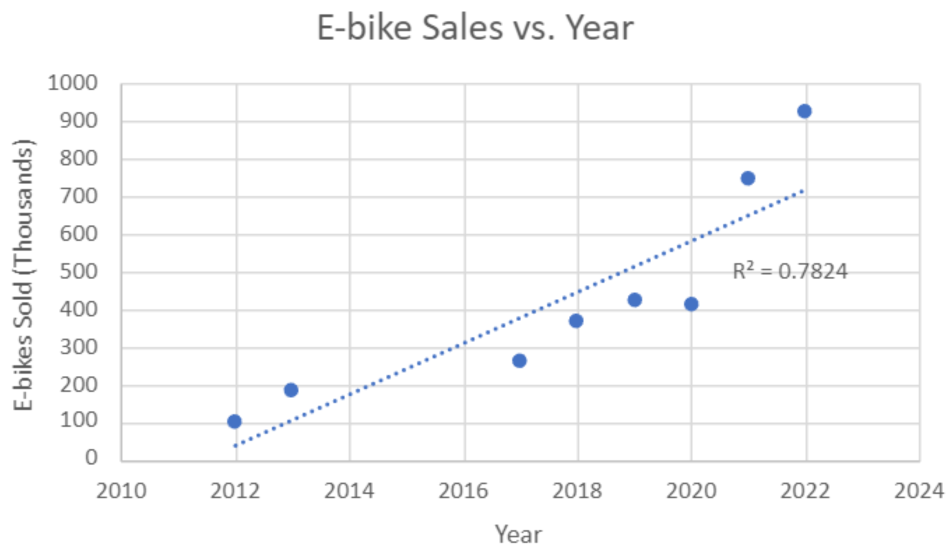
Because the e-bike sales data shows that e-bikes are selling more rapidly as time passes, we decided that an exponential curve would be logical. Thus, to linearize the data, we graphed the natural logarithm of B, $\ln(B)$, against T.

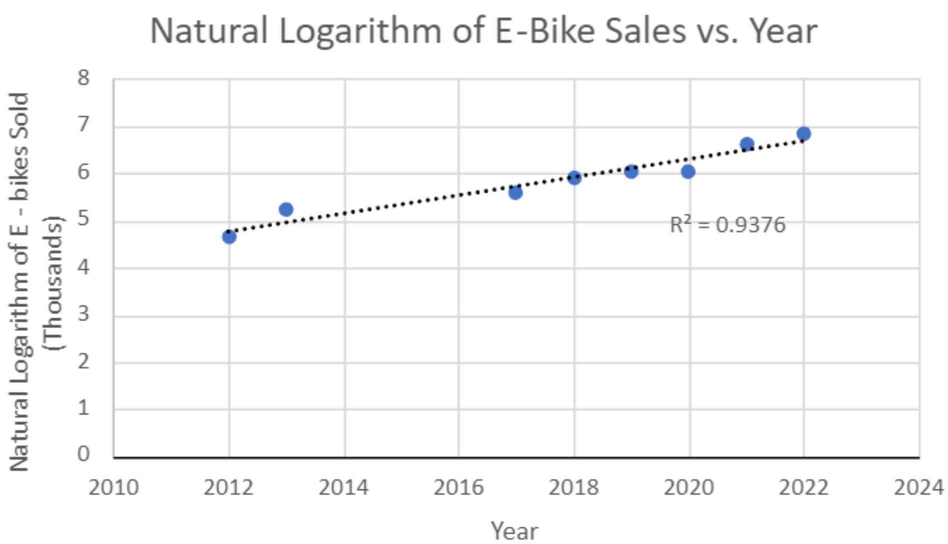
To obtain our predicted value for the number of electric bikes sold in future years, we plugged the new year into our equation to get our value of $\ln(B)$. To convert this to B, we used the following exponential property:

$$e^{\ln(B)} = B$$

The above equation produces the predicted value for the number of e-bikes sold in a certain year. Both the results are displayed in the Results section under **Figure 1**.

1.6 Results





	β	α	R^2
B vs T	68.527	-137841	0.7824
$\ln(B)$ vs T	0.1917	-380.85	0.9376

Figure 1. Results for Least-Squares Linear Regression for E-Bike Sales versus Years and ln(E-Bike Sales) versus Years.

The graphs of our two linear regressions are shown above, as well as a table of the slope, y-intercept, and R^2 values for both of our regression lines. With a Pearson Correlation Coefficient of .97 and Coefficient of Determination of .94, there is strong evidence of exponential growth in future e-bike sales. The equation of our Least-Squares Regression Line is:

$$ln(B) = 0.1917T - 380.85$$

or, in exponential form,

$$B = 3.97 \cdot 10^{-166} \cdot e^{0.1917T}$$

A visualization of this exponential curve plotted against the data points is shown below:

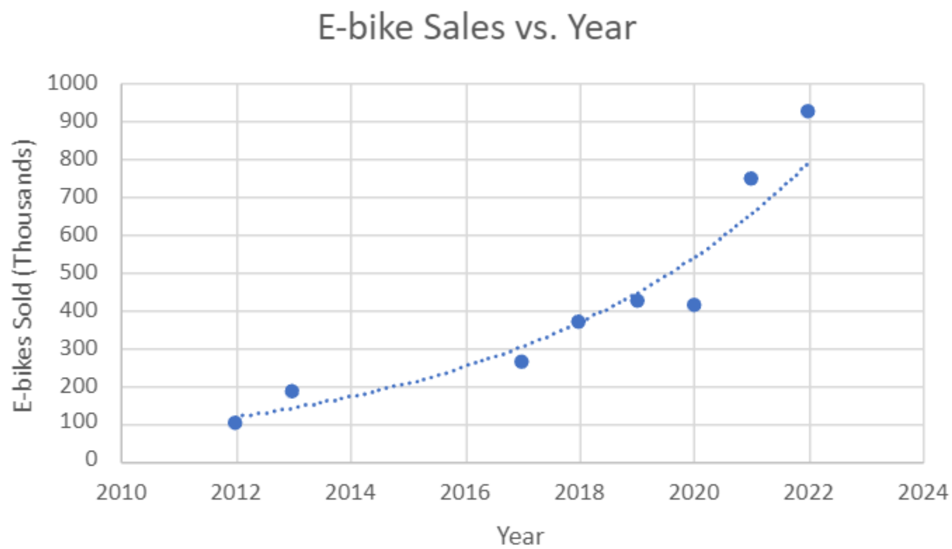


Figure 2. Exponential Regression Model for E-bike Sales over Time

1.7 Discussion and Analysis

The R^2 value of our exponential graph was 0.9376, which was significantly higher than the R^2 value of 0.7824 for our linear model. This high coefficient of determination indicates strong evidence for an exponential growth of the E-bike sale market over the next five to ten years. Using this model, we expect **1.545 million e-bikes** to be sold in 2025 and **2.745 million e-bikes** to be sold in 2028.

Growth in E-bike sales from 2012-2019 was slower than in the last three years. In fact, during the initial phases of the pandemic of 2020, there was actually a slight decrease in the number of e-bikes sold in the US. This can be attributed to the rapid shutdown of the economy, leaving many consumers wary and less willing to purchase an e-bike. As the pandemic continued, people needed to return to work. However, public health concerns discouraged many people from utilizing traditional public transport, such as buses or subways [9]. E-bikes offered a cheap and safe alternative to public transport, causing their popularity to increase exponentially. Compounding this, the fear of public transport also increased the number of cars on the road, causing more traffic and congestion. The increased traffic would also encourage more people to choose alternative transport, such as e-bikes. We expect these trends in e-bike usage to continue.

1.7.1 Strengths

1. Our linear regression model is easy to implement, requiring little processing power or computational storage space.
2. Our model's clear-cut, simple form allows for facile visualization of future trends.
3. The values of Pearson Correlation Coefficient ($R = .97$) and Coefficient of Determination ($R^2 = .9376$) indicate the model is very accurate in predicting the next few years of e-bike sales.

1.7.2 Weaknesses

1. Due to the limited number of data points, our model assumes that post-pandemic growth in e-bike sales will continue in the near future. If e-bike sales were to return to pre-pandemic levels, a linear model may be a better fit for growth in sales.
2. The exponential nature of our model makes it potentially dangerous to extrapolate too far into the future.
3. Similarly, our assumption that consumers' preference for electric bikes will not change in the near future makes our model risky for extended periods of time.

2 Q2: Shifting Gears

2.1 Problem Restatement

In part II of this problem, our team was tasked with evaluating the underlying reasons for the recent rise in e-bike popularity. We identified a number of factors that regulate an individual's decision to purchase an e-bike, of which transportation expenses, vehicular upkeep, value of time, and personal finances were considered. Our team decided on a "bottom-line" monetary metric to quantify the relative importance of each factor to consumers.

2.2 Assumptions

1. **People make rational decisions.** For the sake of our model, it is logical to assume that people make rational decisions. Thus, if using an e-bike provides a net benefit to a consumer, we predict that they will likely purchase an e-bike.
2. **People in rural areas do not use public transit.** The majority of rural areas do not have well-established public transit infrastructure, resulting in the usage of personal vehicles for the vast majority of rural inhabitants. Thus, we can treat the small number of people in rural areas who use public transit as negligible.
3. **We will only observe the population of workers as viable buyers of electric bikes. We assume each worker works 250 days a year.** A typical year has 260 work days, and there are 11 federal holidays, some of which may fall on weekends, resulting in an average of 250 days in which a worker will have to commute to work.
4. **Each hour spent commuting has a value equivalent to a worker's hourly wage.** Time is money! The most practical way to quantify the monetary value of time lost in transit is with a person's wage, as that time could be spent doing other things, such as working.

2.3 Variables

Variable	Definition	Unit
p_u	Proportion of US who live in urban areas	—
p_{fr}	Proportion of workers in US who work fully remote	—
p_{pr}	Proportion of workers in US who work partially remote	—
p_c	Proportion of workers who use a car to commute to work	—
p_{pt}	Proportion of workers who use public transit	—
C_{pt}	Cost of public transportation per hour	Dollars/Hour
U_c	Cost of Car Upkeep	Dollars
U_b	Cost of Bike Upkeep	Dollars
g_p	Price of Gas in US	Dollars/Gallon
C	Commute Time	Hours
$\mathbb{T}(x)$	Yearly Savings in Transportation Costs from E-Bike Usage	Dollars
ΔU	Yearly Savings in Upkeep from E-Bike Usage	Dollars
ΔV	Yearly Savings in Time from E-Bike Usage	Dollars
r	Gallons per Hour used in cars	Used .16 gal/hour
d	Days Worked per Year	Days
D_c	Commute Distance	Miles
s	Speed of E-Bike	Miles Per Hour

2.4 Our Model

2.4.1 Finding Our Parameters

Based on review of data from the Bureau of Labor Statistics [2], savings in transportation costs, upkeep costs, and commute time from an individual's switch to an e-bike all are important contributing factors in increasing the popularity of e-bike usage. We assessed these quantities at three time marks (2005, 2010, and 2021) to evaluate how the relative importance of these factors to consumers may have changed over the years in association with recent increases in e-bike sales.

In a study by Iowa State University [12], the average gas prices between urban and rural areas was found to significantly differ, along with the proportion of remote and partially remote workers. Because the appeal of e-bikes may be very deterministic from transportation availability and other factors, it was decided that our model would stratify American consumers based on their belonging to either rural or urban areas.

2.4.2 Transportation Costs

We determined that whether an individual comes from a rural or metropolitan geographical area would have a major impact on transportation costs for commuting. For example, public transportation (p_{pt}) options that may be readily available in urban communities, are often difficult to find in rural settings. In addition, car usage (p_c), commute times (C), and aggregated transportation costs (ΔT) vary.

Therefore in order to calculate ΔT, the projected yearly savings in transportation costs from switching to e-bike usage, we performed the following calculation for an individual x

$$\Delta\mathbb{T}(x) = \begin{cases} g_p \cdot r \cdot C \cdot d & \text{if } x \text{ commutes via car} \\ C_{pt} & \text{if } x \text{ commutes via public transport} \\ 0 & \text{if } x \text{ works from home} \end{cases} \tag{1}$$

The values in this computation were calculated accordingly at the 2005, 2010, and 2021 timepoints. In addition, the proportion of consumers who worked from home, commuted by car, and commuted via public transportation were determined with geographical location in mind. Further sourcing of data can be found in the appendix under the code for question 2.

2.4.3 Value of Time

Some sources suggest that e-bikes have the potential to reduce commute time for workers. Therefore, we thought it would be important to account for the potential savings in time as a result of transitions to e-bikes. Assuming that the importance of a worker's time can be simply quantified as their hourly wage, we calculated the improvements in personal utility as follows:

$$\Delta V(x) = h \cdot d(\text{Commute Time without E-bike} - \text{Commute Time with E-bike})$$

Recall that C_d is defined in our variable table. In order to calculate Commute Time with E-bike, we used the formula of time $= \frac{\text{distance}}{\text{rate}}$

In turn,

$$\text{Roundtrip Commute Time with E-bike} = \frac{2D_c}{s}$$

2.4.4 Upkeep Costs

In an average year, the money car owners spend on their car include costs of maintenance, depreciation of car value, insurance, license, registration, and tires. In our model, upkeep costs for both cars and bikes were treated as constants.

$$\Delta\mathbb{U}(x) = \left\{ \begin{array}{lr} \text{Car Upkeep} - \text{Bike Upkeep} & \text{if } x \text{ commutes via car} \\ -1 \cdot \text{Bike Upkeep} & \text{if } x \text{ commutes via other transportation} \end{array} \right\} \tag{2}$$

2.4.5 Monte Carlo Simulation

Using 3 Monte Carlo simulations adjusted to reflect the composition of the United States at 2005, 2010, and 2021 years, our model sought to estimate the average ΔT, ΔU, and ΔV in each of the three years. In addition, we used p_u to stratify-sample consumers from both urban and rural regions. For maximal robustness, our model sampled 10,000 individuals in each of three years. Averages of ΔT, ΔU, and ΔV were calculated in urban regions, rural regions, and overall. Aggregate savings were also calculated. Finally, "high-impact factors" – factors that delivered the highest dollar value of savings for an individual data point in the sample – were recorded across the three strata.

2.4.6 Technical Computing and Additional Data

Simulations were implemented in Python; code is available in the appendix. Relevant packages included *numpy* to select individuals randomly from normal and binomial distributions. Additional Data is recorded in comments in the script.

2.5 Results

The results from the Monte Carlo simulation model, as described above, are summarized in the table below.

Projected Yearly Savings from E-bike usage			
	2005	**2010**	**2021**
Project Yearly Savings in Transportation Costs (ΔT)	55.88	65.16	52.61
Projected Yearly Savings in Upkeep Costs (ΔU)	5484.58	5339.91	5110.91
Projected Yearly Savings in Commute Time (ΔV)	-2813.22	-3142.87	-5466.74

The savings aggregated in the table were also graphed according to the strata for rural and urban Americans. These results are visualized in the figure below.

Projected Savings from Switching to E-Bike Usage in Urban versus Rural areas in 2005, 2010, and 2021

Figure 3. Proportion of American consumers whose highest impact factors were transportation (blue), upkeep (red), and commute time (yellow) in 2005, 2010, and 2021.

It can be seen from the above table that generally, upkeep costs offered the greatest projected yearly savings for individuals in both rural and urban regions. Savings in upkeep costs were followed by transportation costs, and finally by commute time savings. This is consistent with the data visualized in the bar chart below, where it can be seen that vehicular upkeep has dramatically more impact on consumer decision-making involving e-bike purchases than transportation and commute time value.

Highest Impact Factors for E-Bike Favorability in 2005, 2010, and 2021

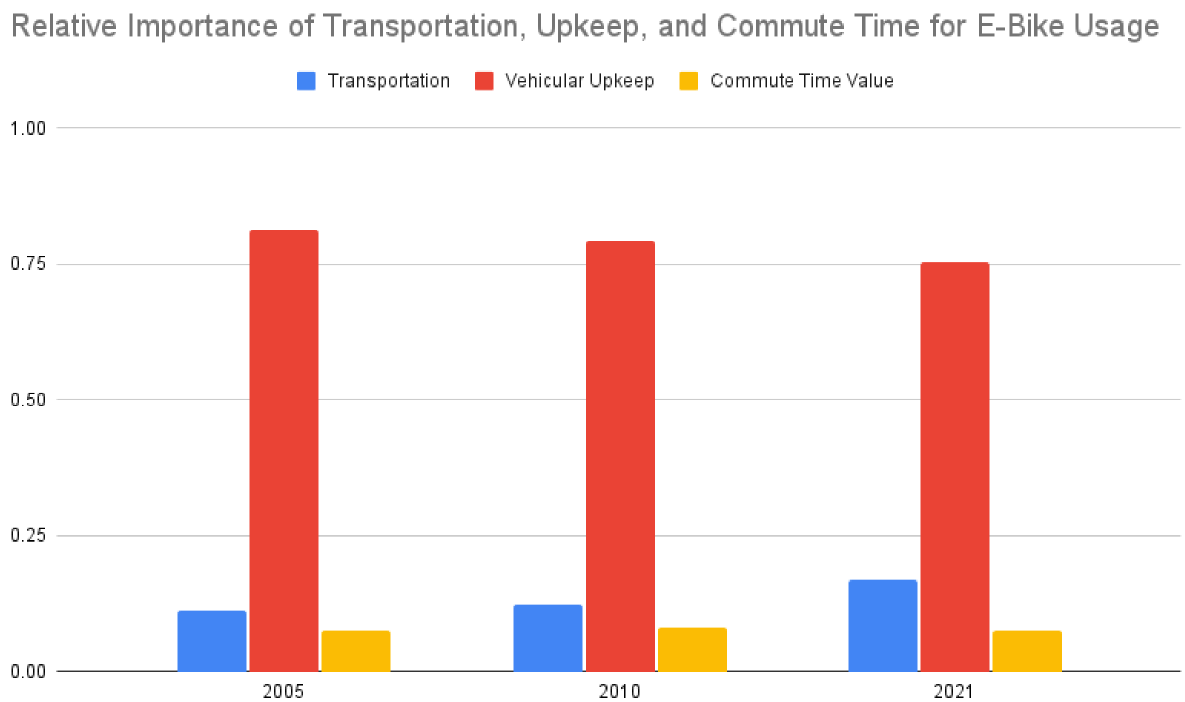


Figure 4. Proportion of American consumers whose highest impact factors were transportation (blue), upkeep (red), and commute time (yellow) in 2005, 2010, and 2021.

2.6 Discussion and Analysis

Our model accurately predicts the relative contributions of transportation costs, commute costs, and upkeep costs to the net benefit of a consumer using an electric bike. We determined vehicular upkeep as the most important factor for the increase in e-bike growth. Of the factors we considered, commute costs were actually worse for e-bike users, reflecting the longer duration of trips by e-bikes in comparison to alternative methods of transportation.

The decrease in transportation costs and upkeep costs as well as the increase in commute costs can be attributed to an increase in remote and hybrid workers resulting from COVID-19. With more remote workers, there are fewer commuters to work and therefore a decrease in transportation costs. The decrease in upkeep costs resulted from a decrease in general travel and the necessity for repairs.

Our overall results are consistent with Electrek's analysis of important factors, primarily emphasizing the value of practicality to the consumer [19]. Since we quantified each factor with a monetary value, we allowed for an accurate comparison between otherwise ambiguous variables.

2.6.1 Strengths

1. Our model's simple methodology of quantifying factors as their projected monetary utility allows comparison of unalike decision-making factors. In addition, it allows our model the flexibility to incorporate more factors in the future.
2. Our model yields insights about the unique impact that e-bike popularity may differ in urban versus rural environments. This allows us greater accuracy in predicting the decision-making of the U.S. as a whole.
3. By accounting for the deviations between remote, hybrid, and daily commuter transportation habits, our model considers the efficacy of the recent COVID-19 pandemic and future changes in American job status.

2.6.2 Weaknesses

1. Our model fails to account for factors that influence increases in e-bike usage on the supply side. With more time, we would like to expand our model to represent the rise in bikeshare companies and an increasing number of courier services utilizing e-bikes [16].
2. Our model does not evaluate the worth of the health benefit or "coolness" factor of e-bikes because of the time constraint and the complexity of these factors. These may contribute additionally to the change in consumer habits, making our model not completely comprehensive.

3 Q3: Off the Chain

3.1 Problem Restatement

In Part III, we were tasked with quantifying the impact that the growth in e-bike usage will have on various aspects of our society. We chose to analyze the impact of this growth on carbon emissions, traffic congestion, and changes in deaths due to car and biking accidents. Our team decided to quantify these factors by crafting various formulas to calculate the net impact of these factors from the period 2023 to 2028.

3.2 Assumptions

1. **The carbon emissions per car can be estimated from the average carbon emissions of a car manufactured in 2021.** We do not have a way to predict the age of the car that will be substituted for by the e-bike. Choosing a more recent year, such as 2021, gives us a reasonable lower bound for the amount of carbon emissions saved. Additionally, carbon emissions per mile for cars has not decreased significantly in the past 10 years [4], making our estimate valid.
2. **Every e-bike purchase is for a new e-bike user.** While many people own regular bikes, averaging 2 bikes per person, e-bikes are much more expensive than a regular bike, and it doesn't make economical sense for a person to own multiple. Additionally, while e-bikes may be damaged due to wearing out or accidents, we are considering this over short time periods, so the number of these incidents is negligible. Thus, the number of e-bikes purchased is equal to the number of new e-bike riders.

3. **E-bike riders travel the same average distance per year as regular bike riders.** Most people purchase an e-bike as a substitute for their regular bike. It stands to reason that these people will still travel the same distances that they used to.
4. **The average distance travelled on a bike in a year remains constant.** Since we are only modeling this for a short period of time, there would not be enough time to drastically change the average time travelled by bike.
5. **Whether a person uses public transportation or not does not affect traffic congestion.** Public transit operates whether any given person is present or not, so a person switching from public transit to an e-bike will not affect traffic.
6. **Road capacities will not change in the near future.** Construction of new roads or traffic lanes is the only way to change road capacity. However, such projects are very time-intensive, so changes to road capacity will not affect our model in the short term.
7. **All vehicles in the United States contribute to traffic, and the number of vehicles remains constant at 300 million.** Although the number of cars increases slightly each year, the amount is very small compared to the number of cars, and accounting for this would make the model unnecessarily complicated. The number of owned cars in the United States in 2021 was 282 million [8], and with the slight growth each year, 300 million is a good estimate of what future numbers will be.
8. **Bicyclist and Car Mortality rates will not change significantly in the coming years.** Both accident and mortality rates are dependent on human attitudes towards driving and biking. These attitudes are unlikely to change in the short term. Additionally, our data shows that the biking mortality rate has remained fairly constant over the past 10 years [13], suggesting our assumption is valid.
9. **The fatality rate of regular bicycles is roughly equivalent to that of e-bikes.** E-bikes suffer from many of the same safety issues as regular bikes, such as car crashes. Our data suggests that this is valid, with a mortality rate of 19 and 21 deaths per million rides for e-bikes and regular bicycles, respectively [14].
10. **We assume that the physical health benefits from using an electric bike are negligible compared to the other relative expenses.** Considering that an electric bike consumer would almost always utilize the motor, thereby reducing the efficacy of the exercise, the health benefit of an electric bike would not reduce a consumer's gym attendance or lead to a significant increase in health. Therefore, it is not necessary for us to account for this factor.

3.3 Variables

Variable	Definition	Units
n_e	One-year equivalent number of e-bikes	E-Bikes
p_b	Total number of e-bike riders in the US	People
Δp_b	Increase in the number of bike riders in the US	People
μ_e	Market share of e-bikes	—
m_b	Total miles travelled on bikes	Miles
$\overline{m}$	Average miles travelled on a bike/e-bike	Miles/Bike
m_e	Total miles travelled on e-bikes	Miles
μ_s	Percent of miles travelled on e-bike that would've been travelled by car	—
m_c	Miles saved by using an e-bike over a car	Miles
A	Average occupancy of a private vehicle	People/Vehicle
T_i	Start of the time period we track	Years
T_f	End of the time period we track	Years
E	Carbon emissions per mile travelled by car	grams CO_2/mile
C	Total carbon emissions saved	grams CO_2
R	Ratio of traffic reduction	—
d_b	Mortality rate for bicyclists	Deaths/100,000 people
d_c	Mortality rate for car-drivers	Deaths/100,000 people
ΔD	Change in deaths	Deaths

3.4 Our Model

3.4.1 Carbon Emission Reduction

The first part of our model quantifies the carbon emissions saved by people choosing to use their e-bikes rather than drive a car. For simplificity's sake, we chose to evaluate the impacts of these factors, such as carbon emissions and traffic congestions, over the period of one year, from 2023 to 2024. Using results from our model in part I, the number of new electric bikes that will be sold in the coming years was predicted.

Next, since we were asked to quantify the impact of people shifting to e-bikes, we only consider the new e-bikes bought during our time period. A bike bought in year t of our period will save carbon emissions for $T_f - t$ years, the remaining years in our period. To make this simpler, we can represent all the e-bikes bought in this time period instead as the equivalent number of bikes that would be bought in the year T_f. For example, a bike bought in year 4 of a 10 year period would have an effective value of 6 bikes bought in year 10 since the bike will save emissions for 6 years. Thus, the total one-year equivalent number of e-bikes bought in this time period can be written as

$$n_e = \Sigma_{t=T_i}^{T_f}(T_f - t)B(t)$$

with B as defined by our linear regression in part I. This gives us the new number of electric bikes bought in our time period. We were given data on the total number of miles travelled by bicycle in past years. Using this data, we created a linear regression of total miles biked in a year(m_b) in billions of miles as a function of year(T).

Our calculated regression line was

$$m_b = 0.8966T - 1773.6$$

with an R^2 value of 0.9059, showing strong correlation. Using $T = 2023$ to get an estimated value for the total number of miles biked, we obtain $m_b = 40.2218$ billion miles.
n_b is given as 52 million people [5], the total number of bike riders in the US. From this, we can calculate the average mileage travelled in a year by a bike rider,

$$\overline{m} = \frac{m_b}{p_b}$$

Using this, we then calculate the total distance travelled by our e-bike riders,

$$m_e = \overline{m} \cdot n_e$$

Many of these e-bike riders use their e-bikes as a substitute for driving, which is represented by μ_s, the percentage of miles that e-bike owners used their e-bikes that they would have driven their car for. From our research, a reasonable value of μ_s is 0.76. [6]
To calculate m_c, we multiply the total miles travelled on e-bikes by the percent of those which would have been replaced by a car, so

$$m_c = \mu_s \cdot m_e$$

E is reported as 348 grams CO_2 per mile travelled [4]. Multiplying this by m_c therefore gives us our total carbon emissions savings, so

$$C = E \cdot m_c$$

3.4.2 Effect on Traffic Congestion

To evaluate traffic congestion, we decided to use the Volume-to-Capacity Ratio [7][15]. This ratio evaluates the level of congestion by comparing the volume of cars that use the road to the capacity, which is defined as the maximum car flow a road can maintain without interruptions.

The denominator of this metric, capacity, will be constant, as per our assumption. This makes congestion linearly related to volume, or the number of cars that pass through the road. Thus, a decrease in the number of cars on the road will cause a proportional decrease in congestion.

To calculate the reduction in traffic congestion, we wish to find the decrease in the number of vehicles on the road due to the increase in e-bikes. We first need to identify the number of people who no longer contribute to this traffic because of their usage of e-bikes. Although not every consumer of an e-bike will fully replace their car use, treating them as such will establish an upper bound of the traffic reduction.

From our assumption, we can equate the increase in the number of people who use e-bikes with the number of e-bikes sold over the time period. Our model in part I predicts that for a given year t, $B(t)$ is the number of bikes sold that year. Therefore, we can express Δp_b as

$$\Delta p_b = \Sigma_{t=T_i}^{T_f} B(t)$$

Now that we have the number of people who are not contributing to traffic, we seek to find the number of vehicles that this relieves from the roads. According to the given data, 75.6% of people use private vehicles (cars/trucks/vans) to commute [20]. This means that the number of people who are riding bikes instead of contributing to traffic is $0.756\Delta p_b$.

To determine the number of cars that this number of people corresponds to, we needed to calculate the average occupancy of a car. From the given data, we can compute the average occupancy of a car as the expected value of occupants in a car. This can be expressed as

$$A = \frac{0.678 + 2 \cdot 0.059 + 3 \cdot 0.012 + 4 \cdot 0.008}{0.756} = 1.14$$

Thus, the number of cars no longer on the road can be written as

$$\frac{0.756\Delta p_b}{A} = 0.66\Delta p_b$$

To find the ratio of traffic reduction, we will divide this value by the total number of cars, which we have assumed to be 300 million. Therefore, we can express R as

$$R = \frac{0.756\Delta p_b}{300000000} = 2.52 \times 10^{-9}\Delta p_b$$

3.4.3 Effect on Health and Safety

The increasing prevalence of e-bikes will cause the rates of various types of accidents to change. In particular, increasing the number of e-bikes will increase the number of fatalities due to bike crashes, but decrease the number of fatalities due to car crashes.
To calculate the change in deaths due to this shift, we first calculate the increase in death due to increased biking. This is equivalent to the increased number of bike riders times the mortality rate, or

$$n_e \cdot d_b$$

Additionally, we must account for the decrease in deaths due to car accidents, as e-bikes reduce the number of cars on the road. Using n_e as our upper limit for this value, we calculate the decrease in deaths from this factor as

$$n_e \cdot d_c$$

Thus, our total change in deaths is

$$\Delta D = n_e(d_b - d_c)$$

3.5 Results

Using our model, we calculated the subsequent impacts for the period 2023-2028. The table below shows several of our most important values

n_e	m_c	C
31.763 million e-bikes	18.67 billion miles	7.16 million tons CO_2

Table 3. Future Impacts of Increased E-bike Use.

Additionally, we calculated the upper bound for the reduction in traffic congestion by 2028 by setting $T_i = 2023$ and $T_f = 2028$. Using our formula, we calculated R to be

$$R = 2.52 \times 10^{-9} \cdot 3.176 \times 10^{7} = 0.08$$

Thus, we predict an 8% reduction in traffic congestion by 2028.

Finally, we calculated the number of deaths that the increase in e-bike usage would predict using $d_b = 2.3\%$ and $d_c = 0.26$ [11][13]. This gives us a calculated value of $\Delta D = -648$ deaths. From 2023 to 2028, we predict that the switch to e-bikes will save 648 lives due to accidents.

3.6 Discussion and Analysis

The increased use of e-bikes will have a significant effect on the amount of carbon emissions, reducing carbon emissions from cars by 7.16 million tons over our calculated 5-year period. This represents approximately 0.1% of the US's annual CO_2 emissions [17], or 0.38% of the US's annual CO_2 emissions in a year [18]. While e-bike usage currently may represent a very small amount of carbon savings, due to the exponential growth of e-bike use, we can expect larger carbon savings in the future, which will greatly benefit the environment.

As shown, the increased adoption of e-bike usage is predicted to reduce the traffic congestion by at most 8% in the next five years. Even though this is an upper bound, it is a significant amount already and shows the impact that the e-bike is set to have. Furthermore, it is consistent with the predicted doubling of the US electric bike market share from $800 million to $1.6 million. [10]

The shift to e-bikes will also benefit the safety of residents, as we see a slight reduction in fatalities due to bike and car crashes, with a total of 648 lives being saved over a 5-year period. This is largely due to fewer cars, leading to less car accidents. However, our assumption that car usage is perfectly substituted by e-bikes may not be true, meaning the predicted decrease in car use could be less, lowering this effect. Additionally, as stated in the assumptions, the use of e-bikes will have a negligible impact on other health factors such as cardiovascular health due to a reliance on the vehicle's motor. While it may save a few lives, overall, the health and safety benefits of this transition are minimal.

3.6.1 Strengths

1. Our model allows us to predict the reduction in CO_2 output and traffic congestion for any interval of time.
2. Our model predicts a substantial impact on traffic congestion which the corroborates the predicted increase in electric bikes.

3.6.2 Weaknesses

1. Our model for traffic congestion does not have a mechanism to account for the changing quantity of vehicles in the US.
2. Our model assumes that every person who purchases an electric bike will substitute their car usage for the bike. This is unrealistic as there are still scenarios where car usage is necessary. With more time, we would have liked to model how well e-bikes act as substitutes for cars. However, with our limited time we were unable to find data on this. Thus, our models only provide an upper bound, which may not necessarily reflect the true value.

4 Conclusion and Future Directions

Projected increases in electronic bike purchases and use in the United States is a development of great importance for transportation and quality-of-life for Americans. When evaluating the factors that have contributed to trends in e-bike purchases, our limited time frame has prevented us from considering the value of other influential variables including the importance of leisure and exercise as motivating factors.

Our first model predicts the short-term trend of e-bike purchases in the U.S. which is consistent with the data points from a variety of sources. Our model was founded on the basis of national market data, making our model credible in the status quo.

Our second model analyzes the most salient quantifiable motivators of electric bike purchases using a universal monetary standard. Using a Monte Carlo simulation, we considered the variation in hourly wage, gas prices, job status, and commute times as a way to make our findings more applicable to the general population and allow for the comparison of alternative factors given more time.

Our third model utilizes a variety of data to quantify the broad societal impacts of the increased popularity of electric bikes. Among the effects we considered were CO_2 emissions, traffic, and safety. From the linear regression model from Part I, we calculated the one-year effective number of e-bikes that would be accumulated in a period of five years, which, when combined with a derived estimate for the total distance traveled, aided us in calculating the CO_2 offset. Furthermore, we utilized the predicted sales in electric bikes to guide our calculation for the reduction in traffic congestion by incorporating the weighted average for vehicular occupancy. Finally, we examined the difference in death rates for cyclists and passengers to predict the number of lives saved by the shift towards electric bikes. Overall, our models demonstrate the impulses and consequences of the burgeoning electric bike market, one that is sure to have far-reaching ramifications for the future.

After taking the time constraints of this competition into account, we believe there were a number of new directions we could add to our existing models. The growth in e-bike usage will vary between different countries, so examining data from other countries will allow us to develop a better idea of the global trends toward electric bicycles. The presence of international markets, particularly those of Europe and Asia, which have more established e-bike markets, will likely influence the development of the US market and should thus be investigated. These more established markets may also act as a predictor for future conditions in the US market.

Due to the limited time constraints of the competition, we did not examine the supply side of the electric bike market. An important factor contributing to the spread of electric bikes has been the increased accessibility due to the lowered prices as a result of competition among manufacturers. Thus, researching this side of the market could provide insights into additional causes behind the expansion of the electric bike market.

5 References

[1] http://www.ecycleelectric.com/blog/2014/6/25/the-2013-and-2014-usa-electric-bike-market-numbers-explained

[2] https://www.census.gov/content/dam/Census/library/publications/2021/acs/acs-47.pdf

[3] https://www2.deloitte.com/content/dam/insights/us/articles/722835_tmt-predictions-2020/DI_TMT-Prediction-2020.pdf

[4] https://www.energy.gov/eere/vehicles/articles/fotw- 1223-january-31-2022-average-carbon-dioxide- emissions-2021-model-year

[5] https://www.zippia.com/advice/bicycle-industry-statistics/

[6] https://www.bicycling.com/news/a20049844/ebike-study-car-trips/

[7] https://transportation.ky.gov/Congestion-Toolbox/Pages/Congestion-Measures.aspx

[8] https://www.statista.com/statistics/183505/number-of-vehicles-in-the-united-states-since-1990/

[9] https://www.weforum.org/agenda/2021/02/public-transport-covid-data/

[10] https://www.imarcgroup.com/united-states-e-bike-market

[11] https://www.iihs.org/topics/fatality-statistics/detail/state-by-state

[12] https://www.marketplace.org/2022/08/30/in-rural-areas-the-high-cost-of-diesel-is-driving-up-the-cost-of-doing-business/

[13] https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813197

[14] https://www.gallivanlawfirm.com/study-e-bikes-may-cause-more-injuries-than-motercycles-cars/

[15] https://www.oregon.gov/odot/Planning/Documents/V-CRatio_DelayPlanningDesignDecisions_SignalizedIntersections.pdf

[16] https://www.wired.co.uk/article/cargo-bikes-greener-quicker

[17] https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks

[18] https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

[19] https://electrek.co/2022/07/03/electric-bikes-most-popular-attractive-study/

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

6 Appendix

Code

```
import numpy
from numpy.random import binomial, uniform, normal

transport_total = 0
upkeep_total = 0
timeval_total = 0
saved_total = 0
urban_transport_total = 0
urban_upkeep_total = 0
urban_timeval_total = 0
urban_total = 0
rural_transport_total = 0
rural_upkeep_total = 0
rural_timeval_total = 0
rural_total = 0
days_worked = 250

# Gas Prices Each Year Found from M3 Sheet 2
# Ride Like the Wind, MathWorks Math Modeling Challenge 2023, https://
    m3challenge.siam.org/node/596.
us_gas_prices = {2005: 2.27, 2010: 2.78, 2021: 3.01}

# https://www.statista.com/statistics/678561/urbanization-in-the-united-
    states/
prop_urban = {2005: .808, 2010: .871, 2021: .837}

# https://www.cipd.co.uk/Images/working-from-home-1_tcm18-74230.pdf for
    2005/2010
# https://news.gallup.com/poll/355907/remote-work-persisting-trending-
    permanent.aspx for 2021
prop_fully_remote = {2005: .031, 2010: .035, 2021: .25}
prop_partially_remote = {2005: .22, 2010: .248, 2021: .2}

# used data from 2006 as an approximate for 2005 from US Census Bureau
# https://www.census.gov/content/dam/Census/library/publications/2015/acs/
    acs-32.pdf
# 2021 proportion taken from M3 data
prop_urban_using_car = {2005: .80, 2010: .78, 2021: .732}
prop_rural_using_car = {2005: .90, 2010: .91, 2021: .854}

# Ride Like the Wind, MathWorks Math Modeling Challenge 2023, https://
    m3challenge.siam.org/node/596.
# Assumed that people in rural areas do not use public transit
# Converted by dividing by proportion in urban areas
prop_public_transit = {2005: .01, 2010: .01, 2021: .025}

# https://fred.stlouisfed.org/release/tables?rid=50&eid=6471&od
    =2021-01-01#avg_wages = {2005: 15, 2010: 18, 2021: 24}
# Jan 2021, Jan 2010, Mar 2006 as an approximation for 2005
avg_wages = {2005: 20.04, 2010: 22.41, 2021: 29.92}

# https://thepointsguy.com/guide/monthly-public-transport-costs-worldwide/
```

```
# Used the statistic that "public transport consistently costs 3 percent
    to 4 percent of income"
public_transport_cost = {2005: 15 * .035, 2010: 18 * .035, 2021: 24 *
    .035}

# https://www.aaa.com/autorepair/articles/what-does-it-cost-to-own-and-
    operate-a-car
# Depreciation, insurance, maintenance, license, finance, tires upkeep
    costs were aggregated
car_upkeep = 7293

# https://bikelvr.com/bikes/e-bikes/maintenance-costs-for-an-electric-bike
    /
# "your average maintenance cost per year is about $450 per ear$
bike_upkeep = 450


# Function selects positive value from a normal distribution, ensuring
    only positive values
def positive_from_normal(mu, sigma):
    my_num = 0
    while my_num <= 0:
        my_num = normal(mu, sigma)
    return my_num


# Function that evaluates (1) Change in Transportation Costs fr (2) time
    value from commute time (3) cost of upkeep
"""
Function samples an individual given a year to evaluate
(1) change in transportation costs,
(2) change in opportunity costs (using value of time),
(3) cost of upkeep
"""


def sample_indiv_by_year(year):
    transportation_costs = 0
    is_urban = uniform(0, 1) < prop_urban.get(year)

    # assigns to work from home or not
    random_number = uniform(0, 1)

    if is_urban:
        has_car = uniform(0, 1) < prop_urban_using_car.get(year)
    else:
        has_car = uniform(0, 1) < prop_rural_using_car.get(year)

    if is_urban and not has_car:
        has_public_transport = uniform(0, 1) < prop_public_transit.get(
    year)
    else:
        has_public_transport = False

    # https://www.titlemax.com/discovery-center/money-finance/average-
    commute-time-by-city-and-state/
    # Average Roundtrip Commute Time if fully remote: 0
```

```
    # Average Roundtrip Commute Time if partially remote: Assumed to be half of average American
    # Average Roundtrip Commute Time if brick-and-mortar employee with car: Average of .88 hours/day
    # Average Roundtrip Commute Time if public transit user: 1.1*2 hrs
    if has_public_transport:
        commute_time_hrs = positive_from_normal(60.7 * 2 / 60, .30)
    if random_number < prop_fully_remote.get(year):
        # individual works remotely
        commute_time_hrs = 0
    elif random_number < prop_fully_remote.get(year) + prop_partially_remote.get(year):
        # individual works hybrid
        commute_time_hrs = positive_from_normal(.44, .10)
    else:
        # individual works brick and mortar
        commute_time_hrs = positive_from_normal(.88, .20)

    # Distribution is taken from US Department of Transportation, Bureau of Transportation Stats, 2003
    # Unable to find data for later years
    random_variable = uniform(0, 1)
    if random_variable < .29:
        commute_dist = (1 + 5) / 2
    elif random_variable < .29 + .22:
        commute_dist = (6 + 10) / 2
    elif random_variable < .29 + .22 + .17:
        commute_dist = (11 + 15) / 2
    elif random_variable < .29 + .22 + .17 + .1:
        commute_dist = (16 + 20) / 2
    elif random_variable < .29 + .22 + .17 + .1 + .07:
        commute_dist = (21 + 25) / 2
    elif random_variable < .29 + .22 + .17 + .1 + .07 + .05:
        commute_dist = (26 + 30) / 2
    elif random_variable < .29 + .22 + .17 + .1 + .07 + .05 + .03:
        commute_dist = (31 + 35) / 2
    elif random_variable < .29 + .22 + .17 + .1 + .07 + .05 + .03 + .08:
        commute_dist = 35

    # evaluating changes in gas expenses
    # assuming 250 days worked per year
    gas_price = us_gas_prices.get(year)

    if has_car:
        transportation_costs += gas_price * .16 * commute_time_hrs * days_worked
    elif has_public_transport:
        transportation_costs +=public_transport_cost.get(year)
    else:
        transportation_costs = 0

    hourly_pay = avg_wages.get(year)

    # evaluating changes in traffic as opportunity cost/value of time
    # assumes value of time spent is equal to hourly pay
    # https://www.gazellebikes.com/en-us/how-fast-do-electric-bikes-go#:~:text=So%2C%20how%20fast%20do%20electric,into%20the%20current%20
```

```
    classification%20system.
    # Assumes Average e-bike speed of 25
    delta_timeval = hourly_pay * commute_time_hrs * days_worked - (2 *
    commute_dist / 20) * hourly_pay * days_worked

    # cost of upkeep
    if has_car:
        delta_upkeep = car_upkeep - bike_upkeep
    else:
        delta_upkeep = -1 * bike_upkeep

    # calculating total saved
    total_saved = transportation_costs + delta_upkeep + delta_timeval

    list = [transportation_costs, delta_upkeep, delta_timeval, total_saved
    ]

    return is_urban, list


# 2010 Measurements
urban_max_factors = [0, 0, 0]
rural_max_factors = [0, 0, 0]
total_max_factors = [0, 0, 0]

year = 2021
sample_size = 10000
for i in range(sample_size):
    is_urban, indiv = sample_indiv_by_year(year)

    if is_urban:
        urban_transport_total += indiv[0]
        transport_total += indiv[0]
        urban_upkeep_total += indiv[1]
        upkeep_total += indiv[1]
        urban_timeval_total += indiv[2]
        timeval_total += indiv[2]
        urban_total += indiv[3]
        saved_total += indiv[3]
    else:
        rural_transport_total += indiv[0]
        transport_total += indiv[0]
        rural_upkeep_total += indiv[1]
        upkeep_total += indiv[1]
        rural_timeval_total += indiv[2]
        timeval_total += indiv[2]
        rural_total += indiv[3]
        saved_total += indiv[3]

    # removes total
    indiv.pop()

    max = indiv[0]
    biggest_factor = 0
    for i in range(1, 3):
        if indiv[i] > max:
            max = indiv[i]
```

```
                biggest_factor = i

    if is_urban:
        urban_max_factors[biggest_factor] += 1
    else:
        rural_max_factors[biggest_factor] += 1

    total_max_factors[biggest_factor] += 1

print("Most important Factors Overall ", end="")
print(total_max_factors)
print("Importance of Max Factors in Urban Areas ", end="")
print(urban_max_factors)
print("Importance of Max Factors in Rural Areas ", end="")
print(rural_max_factors)
print("Average Yearly Saved Transport Costs Overall ", end="")
print(transport_total / sample_size)
print("Average Yearly Saved Transport Costs in Urban Areas ", end="")
print(urban_transport_total / sample_size)
print("Average Yearly Saved Transport Costs in Rural Areas ", end="")
print(rural_transport_total / sample_size)

print("Average Yearly Saved Upkeep Costs Overall ", end="")
print(upkeep_total / sample_size)
print("Average Yearly Saved Upkeep Costs in Urban Areas ", end="")
print(urban_upkeep_total / sample_size)
print("Average Yearly Saved Upkeep Costs in Rural Areas ", end="")
print(rural_upkeep_total / sample_size)

print("Average Yearly Saved Timeval Overall ", end="")
print(timeval_total / sample_size)
print("Average Yearly Saved Timeval Costs in Urban Areas ", end="")
print(urban_timeval_total / sample_size)
print("Average Yearly Saved Timeval Costs in Rural Areas ", end="")
print(rural_timeval_total / sample_size)

print("Average Yearly Saved ", end="")
print(saved_total / sample_size)
print("Average Yearly Saved in Urban Areas ", end="")
print(urban_total / sample_size)
print("Average Yearly Saved in Rural Areas ", end="")
print(rural_total / sample_size)
```