Riding into the Future: Evaluating E-Bikes

1 Executive Summary

From 2018 to 2022, the number of e-bikes sold in the US grew from 369,000 units to 928,000 units- a 251% increase. The emergence of e-bikes as a viable commuting option could change the landscape of transportation in the country. If this trend continues, these vehicles provide the US DoT with an opportunity to take cars off the road as part of a potential sustainable energy plan. To help the DoT decide whether to use the recent success of the e-bike to establish another means of transportation, we analyzed industry trends to forecast the number sold in 2025 and 2028.

Our prediction is that 1,014,000 units will be sold in 2025, while 1,561,000 units will be sold in 2028. An analysis of these values showed that sales would increase a mere 7% from 2022-2025, but this stagnation would be followed by a 53% increase from 2025-2028. Since we used a multivariate linear regression, we found that sales are projected to remain at the current rate until 2024 when lithium production will rebound. As such, the number of sales will grow more from 2025-2028 than it will from 2022-2025.

We then carefully examined the impact of each variable to find the cause of the changing growth rate.

The energy efficiency of a lithium-ion battery was the most significant factor in determining total sales. From 1991-2022, the amount of energy available in a battery per unit of mass increased linearly, which coincided with the decrease in the cost of lithium-ion Batteries for the same time period. Due to lower battery prices, the cost of e-bikes decreased, and the number of units sold increased from 2018-2022. However, the cost of these batteries increased for the first time in 2022 (measured by Gravimetric Energy Density), and the number of sales had a proportional relationship, we conducted a sensitivity analysis and analyzed our regression coefficients. These tests confirmed our conclusion that increases in the number of e-bike sales directly result from advancements in energy efficiency.

The DOT should use our projections and insights as a basis to promote research into Lithium Battery efficiency because growth in the e-bike industry is based on improvements in the energy sector.

Before e-bikes become a transportation staple, we must understand how increased usage will impact carbon emissions, traffic congestion, and health. We found that, on average, 10,000 e-bikers would ride 7,477,880.34 miles per year, saving 1190.4 metric tonnes of CO2, and this decrease in emissions results in 220,224 dollars saved per year in climate change costs. Furthermore, 1.2 million car rides were replaced by e-bike rides. To quantify health impact, we found that 224336410 calories were burned by 10,000 e-bikers.

The combination of stable growth and positive effects on CO2 emissions, traffic congestion, and calories burned make e-bikes a must.

Contents

2 Introduction

In recent years, consumers have been considering alternative modes of transportation, with one particular mode earning significant attention - e-bikes. Sales of e-bikes in the United States have been steadily increasing in recent years, with estimates projecting them to grow at a compound annual growth rate of 7.8% between 2020 and 2027[12]. E-bikes are gaining popularity for many reasons, including cost-effectiveness, health benefits, convenience, and environmental factors. With their ease of use and eco-friendly features, e-bikes have become a popular choice for commuters, recreational cyclists, and delivery services.

Many companies manufacture e-bikes, such as Rad Power Bikes, a Seattle-based company that offers a variety of affordable e-bikes, and Juiced Bikes, which feature a selection of high-performance e-bikes. The average cost of an e-bike ranges from around $500 to over $5000, depending on its qualities and features. E-bikes contain an electric motor that assists with pedaling, and their speeds vary depending on their models and regulations in the area. E-bike batteries, the most common of which are lithium-ion batteries, last anywhere from 20 to 100 miles on a single charge, depending on the model and how much assistance is used.

Many cities across the country implement e-bike share programs to meet the growing demand for alternative modes of transportation. For example, New York City's Citi Bike system recently introduced e-bikes to its fleet, and other cities such as San Francisco and Austin have also implemented e-bike share programs[14][15][16].

The rising usage of e-bikes has many positive consequences, including reduced traffic congestion and improved air quality. Many states, such as California, have made attempts at reducing greenhouse gas emissions due to transportation. The state has reduced emissions through fuel economy standards, but growing population and urbanization have caused greenhouse gas levels to remain relatively high.

In Seattle, Washington, the city's bike share program "Lime" introduced a fleet of e-bikes in 2018[13]. According to a report by the Seattle Department of Transportation, the introduction of e-bikes as part of the Lime program resulted in a significant reduction in car trips and an increase in bike trips. The report found that Lime e-bikes accounted for 21% of all trips taken using the Lime program and that 10% of Lime e-bike riders reported that they would have otherwise driven a car if the e-bike was not available. This suggests that e-bikes are helping to reduce the demand for cars and improve traffic flow in Seattle.

Additionally, the rise in e-bikes can have a significant impact on the economy. The e-bike industry is creating new jobs in the US. E-bike manufacturers, retailers, and service providers are all creating new job opportunities, particularly in urban areas where e-bikes are gaining popularity. E-bikes are a more cost-effective mode of transportation than cars, particularly for short trips. This is helping to reduce transportation costs for individuals and businesses, which can have a positive impact on the overall economy.

As more people recognize the benefits of e-bikes, it is reasonable to predict that they may someday become an integral part of transportation in the future.

3 The Road Ahead

3.1 Restatement of the Problem

From 2006 to 2019, E-Bike usage increased annually in Europe[18]. The US experienced a similar trend, as the number of e-bikes sold in the US more than doubled from 2018 to 2022[17]. In light of this growth, e-bikes could support a sustainable energy plan because greater e-bike usage would decrease automobile usage, thereby decreasing the industry's carbon footprint. To discern the role that e-bikes play in transportation, we must create a model that predicts the number of e-bikes sold in 2025 and 2028.

3.2 Assumptions

Assumption: A major recession will not occur within the next five years.

- **Justification:** In 2021, the COVID-19 Pandemic led to a worldwide economic downturn that reduced the consumption of goods. Considering the drop in e-bike sales from 2019 to 2020 (423,000 to 416,000)[19], a similar event would cause e-bike sales to drop. Because of the changes that result from economic downturns, we will assume that the next five years will not be recession years.

Assumption: There will not be any significant policy regulations regarding the e-bike industry within the next five years.

- **Justification:** Major restrictions on e-bike use would cause an unpredictable and severe drop in e-bike sales, so we will assume that US and UK policymakers will not enact e-bike regulations.

Assumption: Citizens of the US and the UK share the same general opinions regarding the benefits and factors of e-bike usage.

- **Justification:** Differing perceptions of the effects of e-bike usage would cause unpredictable variations in e-bike sales between the US and the UK. Since we are using the same contributing factors to predict e-bike sales, we assume that these are the main factors in both the US and the UK. This premise is reasonable because global location should not affect the general needs of citizens.

3.3 Variables

Variable	Definition	Unit
I	Disposable Income	\$ (dollars)
C	Commute Time	minutes
G	Gas Prices	\$ (dollars)
E	Environmental Worry	% of respondants
P	Popularity	sentiment score
D	Gravimetric Energy Density	W-hr/kg
N	Population	persons

3.3.1 Popularity

To find the popularity of e-bikes for each year, we search for public opinion through a dictionary-based sentiment analysis of tweets on the social media website Twitter.

For each tweet, we use a natural language processing technique called Sentiment Analysis which evaluates, from -1 to 1, how positive or how negative a statement is[20]. In calculating Sentiment Scores, each word and lexicon (emoticons, slang, etc.) is assigned a sentiment score based on past data. We then find the sentiment score by aggregating the scores of each word and lexicon, while accounting for other factors like punctuation. For example, given the tweet:

> With a e-bike, you can get most places within a 10 miles radius faster and cheaper, have the ability to stop without worrying or paying for parking & enjoy the ride- something often taken for granted when driving a car.
>
> 9:15 PM · Aug 26, 2018

Words like "faster and cheaper" along with phrases like "enjoy the ride" are conventionally positive terms, which have a high sentiment score. The combination of these terms yields a total Sentiment Score of 0.7579.

To calculate the popularity of e-bikes in a given year, we use the snscrape python module to randomly sample 5000 Tweets written throughout that year containing the term "e-bike." We then calculate the popularity by finding the average of those sentiment scores.

3.4 Model Development

To accurately predict the number of e-bikes sold in the future, we must create a model which analyzes the relationship between our variables in 4.3 and the number of e-bikes sold. To do this, we run a multivariate linear regression with an $r^2 = 0.93$, which calculates the influence of multiple independent variables on a dependent variable. We design the model by creating a function in the form

$$y = b_1x_1 + b_2x_2 + \ldots + b_nx_n + c \tag{1}$$

where b_i represents the weight assigned to each dependent variable.
To account for changes in the state of the independent variables in the years 2025 and 2028, we use a combination of simple linear regressions and distributions, if the change of the variable is independent of time, to predict each variable in 2025 and 2028.

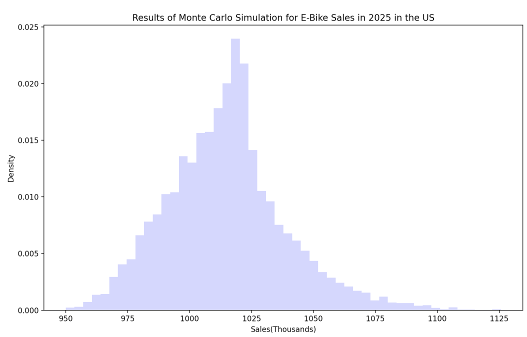
Variable	Method	Parameter	Evaluation
Disposable Income	Linear Regression	$\beta = 877.424, c = -176867$	$r^2 = 0.8103$
Commute Time	Exponentiated Weibull Distribution	$a = 0.1908, c = 3.7915$	KS p-value = 0.908
Gas Price	Nakagami Distribution	$\nu = 0.1991$	KS p-value = 0.525
Environmental	Linear Regression	$\beta = 1.4061, c = -2795.52$	$r^2 = 0.5711$
Popularity	Double Gamma Distribution	$a = 1.8730$	KS p-value =0.972
Gravimetric Energy Density	Exponential Regression	$a = 7.120, b = -14090$	$r^2 = 0.9919$
Population	UN Population Forecasts	—	—

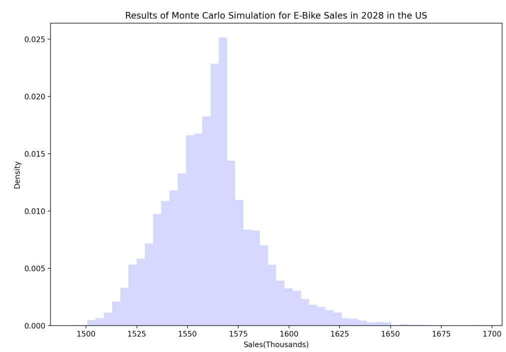
Because there is a variation of possible results for the sales of e-bikes, we use a Monte-Carlo simulation to provide a range of possible sales, one for 2025 and one for 2028. A Monte-Carlo simulation provides us with a clear picture of the model's behavior, rather than a singular forecast. Additionally, the Monte-Carlo will allow us to account for variables that do not necessarily follow a linear relationship, allowing us to account for uncertainty.

We than ran 10,000 trials for the Monte-Carlo distribution, with each trial choosing three random variables for commute time, gas prices, and popularity from the three distributions, along with extrapolated data from the linear regressions for the other variables. The simulation then outputs two predictions, one for e-bike sales in 2025 and one for e-bike sales in 2028.

3.5 Results and Discussion

Combining our multivariate linear regression into our Monte Carlo simulations, we arrive at the following skewed-right distributions of possible values for e-bike sales in 2025 and 2028 for the United States.



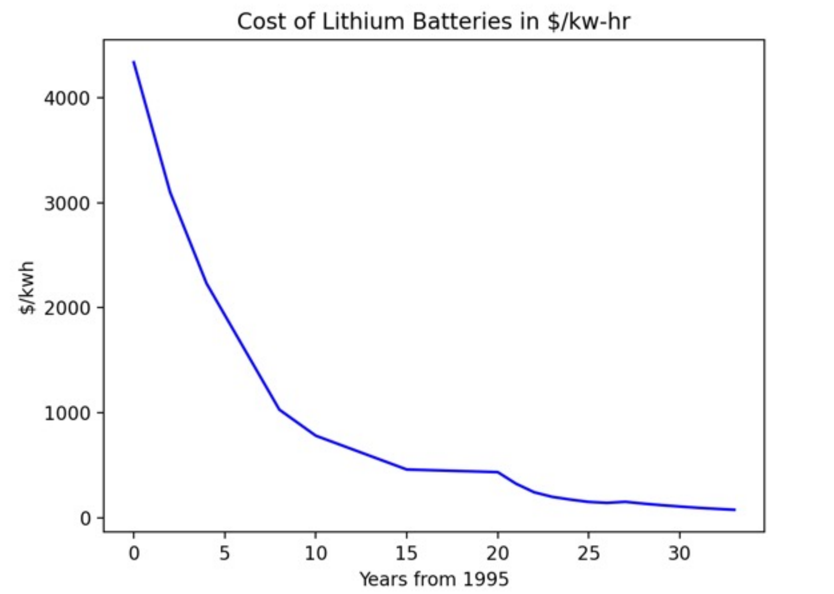


According to the results of our Monte Carlo simulation, we found the average number of e-bike sales to be 1,014,000 in 2025 and 1,561,000 in 2028.

While the 2025 value represents a mere 9% increase from 2022, the 2028 value represents a 54% increase from 2025. Therefore, we find that 2022-2025 will feature slow growth in e-bike sales, but 2025-2028 forecasts growth akin to the early days of e-bikes.

If we examine the historical and modern-day trends in energy efficiency and energy costs, we will find the logic behind these numbers.

The cost of lithium-ion batteries was subject to exponential decay from 1995 to 2021 because of improvements in energy efficiency (represented by D in Section 4.3) during that time period[1][4].



In the graph above, one will see that cost does not follow the exponential decay trend after 2021. In fact, the cost of these batteries rose from \$ 141/kWh to \$151/kWh in 2022 for the first time in years because of the rising price for lithium, nickel, and the other metals[1][4]. This value is expected to hover at \$152/kWh until 2024, when lithium production is expected to rebound. The graph captures this trend, as it remains constant from 2021 to 2024 (x=[26,29]).

The cost is expected to decrease again after 2024, and it will drop below \$100/kWh for the first time in 2026. Make note of this milestone because it explains why e-bike sales increase only 9% from 2022-2025 but jump to a 53% increase from 2025-2028.

The cost stagnated in the first three-year period, so the number of e-bike sales also remained constant, but the return to an annual decrease in cost caused the number of e-bike sales to boom in the second three-year period.

Using our comprehensive analysis detailed above, we recommend that policymakers take note of two key findings:

1. Electric bikes will continue to be a viable transportation option because of the projected growth in sales.

2. To foster the development of the industry, money should be directed to the improvement of energy efficiency. These improvements will lead to a decrease in energy costs and an increase in profit margins for e-bike companies.

3.6 Strengths and Weaknesses

The advantage of a multivariate linear regression is its readability. In the output function, each weight b_i has a meaning that helps us interpret the trends of e-bike sales. For every unit increase in a given dependent variable, the expected change in y is the corresponding weight of the given dependent variable. Because of this feature, we can easily interpret the effect of each dependent variable on the number of e-bike sales.

For example, the dependent variable G representing gas prices has a coefficient of 4.96. With this information, we can conclude that for every dollar increase in gas prices, there will be an additional 5 bike sales.

Another strength of our model is its flexibility, meaning it can consider multiple variables and still produce accurate results. The r^2 for the multivariate linear regression is 0.93, demonstrating a strong relationship between our variables and e-bike sales.

Finally, another strength of our model is its ability to take in factors both tangible and harder to account for. For, example, we used sentiment analysis to account for popularity, a complex variable that is generally difficult to quantify accurately.

One weakness of the model is that there is a small sample size, due to a lack of available data, which produces various statistical issues, like a higher risk of missing relationships between independent and dependent variables, Type I and II errors, and restricted generalizability. This also means that the model is vulnerable to outliers since one data point outside the normal range can significantly skew the linear regression toward one direction.

Another weakness of the model is that we generalized our output to the entire United States. Different regions of the country are bound to have contrasting sales of e-bikes based on factors such as necessity, accessibility, and utility. For example, cities with high population density like New York, with over 6.7 million e-bike rides in 2022[10], will have drastically different e-bike sales than cities in the Midwest and the South, where traffic congestion is much less severe. Because of a lack of e-bike sales data for different regions and states in the United States, we are unable to stratify and account for differences in results by region.

4 Shifting Gears

4.1 Restatement of the Problem

To understand the growth of e-bike sales, we must identify the factors that cause people to buy e-bikes. Likewise, to observe the importance of each factor to consumers, we must create a model that quantifies the influence that each factor has on e-bike sales.

4.2 Assumptions

Assumption: Climate change will not significantly impact the environmental perception of e-bikes within the next five years.

- **Justification:** As climate change grows more severe, e-bikes may grow much more popular as an environment-friendly alternative to more harmful transportation, such as automobiles and trains. This unpredictable surge in popularity would skew the calculated significance of the factors, so we assume that climate change would not cause a sudden increase in the popularity of e-bikes.

Assumption: Gas prices will not fluctuate significantly because of political or economic complications.

- **Justification:** With recent political conflicts, gas prices grew substantially. Because we cannot account for changes in the international landscape during the next five years, our supposition is that the price of gas will change in proportion to historical data.

Assumption: The amount of energy available in a battery per unit of mass will continue to grow at a constant rate.

- **Justification:** According to historical data, the change in gravimetric energy density (D in 4.3) is linear. We assume that this growth rate will remain constant because we are unable to a technological breakthrough in energy efficiency.

4.3 Model Development

We used the multivariate linear regression model and its results to obtain a relationship between the number of bikes sold, the dependent variable, and the contributing factors, the independent variables. We used the same variables as in section 4.3. A linear regression allows us to create a predictive that considers many parameters and outputs a function that minimizes the error from the dataset. Additionally, the scalar aspect of the model allows us to extrapolate data to missing data and the coming years.

Our model finds regression coefficients that relate each of the variables in the model to the output with a regression coefficient, making the model easy to interpret on the modular level. These coefficients serve as weights and measure the significance each factor has on the output. For a regression coefficient b for example, an increase of 1 to the variable b is attached to, will increase the output by b.

4.3.1 β Coefficients in Regression Analysis

From the regression weights, we have determined that based on our variables, gravimetric energy density positively influences the number of sales the greatest, with a coefficient three times greater than the next leading variable. Gravimetric energy density is related directly to efficiency, which would make e-bikes a more viable option. Environment sentiment and gas prices, which both exhibit nearly identical magnitudes of influence, follow. Conversely, commute time has the greatest negative influence on the number of sales, while population, popularity, and disposable income have a negligible influence on the number of sales.

Variable	Weight
Gravimetric Energy Density	$67.463 * 10^0$
Environment	$21.270 * 10^0$
Gas Prices	$19.060 * 10^0$
Commute Time	$-24.969 * 10^0$
Popularity	$-87.452 * 10^{-2}$
Disposable Income	$-12.806 * 10^{-3}$
Population	$23.315 * 10^{-5}$

4.3.2 Sensitivity Analysis

Sensitivity Analysis +5%

Variable	Change in 2025	Change in 2028
I	-0.3%	-0.2%
C	-3%	-2%
G	+0.2%	+.01%
E	+5%	+3.5%
P	0%	0%
D	+13.6%	+9.2%
N	+0.5%	+2.98%

Sensitivity Analysis -5%

Variable	Change in 2025	Change in 2028
I	+0.1%	+0.1%
C	+0.1%	+0.1%
G	-0.2%	-0.01%
E	-0.2%	-0.2%
P	0%	0%
D	-0.5%	-2.9%
N	-2/%	-2.6%

In our sensitivity analysis, we multiplied a single dependent variable x_i by 1.05 or 0.95 (Sensitivity Analysis +5% or Sensitivity Analysis -5%) and held the other variables constant. After running a linear regression with this new equation, we found the percent change in the number of bike sales from our previous equation. Each percent change is listed row by row in 5.3.1.

The results of our sensitivity analysis confirmed our initial analysis of regression weights. According to the analysis, a 5% change in energy efficiency D corresponds to a 13.6% increase in 2025 and a 9.2% increase in 2028 for e-bike sales. Furthermore, a 5% increase in environmental perception led to a 5% increase in 2025 and a 3.5% increase in e-bike sales. These findings confirm our initial statements that gravimetric energy density and environmental perceptions have the greatest influence on the number of sales.

A 5% increase in commute time corresponds with a 3% and 2% decrease because commute time and the number of sales have an inverse relationship as found in our analysis of the regression β coefficients.

4.4 Results and Discussion

Because of our analysis in 4.3.1, it follows that changes in gravimetric energy density are a significant reason for the upward trend in e-bike sales, which is reasonable because improvements in energy efficiency lead to decreases in the cost of lithium-ion Batteries[21]. Because the production of e-bikes relies on these batteries, profit margins will increase if the cost of production goes down. More available capital allows e-bike companies to create more products, and the increased availability leads to increased usage.

Mathematical modeling showed that every 1 unit increase in environmental perception (as measured by the poll provided [4]) accounted for 21 additional sales of an e-bike. Environmental perceptions provide a motive for the purchase of an electronic bike because these vehicles have a reduced carbon footprint when compared with automobiles. If people are attuned to the climate crisis, they will turn to e-bikes when commute time does not interfere with the plausibility of e-bikes as a means of transportation.

4.5 Strengths and Weaknesses

One strength of our model is that it is easy to understand, with clear correlations between independent variables and e-bike sales. Since the model uses coefficients, it clearly expresses the magnitude of the effect that each independent variable has on e-bike sales, allowing for a convenient assessment of the influence of each factor. Additionally, since our model accommodates multiple variables in the same equation, it can effectively gauge the extent of the impact of each factor by comparing the effects of the variables, whether positive or negative.

We can scale this model up using more variables, and given more data, we will be able to tailor the model toward specific locations and circumstances. For example, one change in the environment variable is not equal to a change in the disposable change variable.

Finally, our model is robust, which allows for consideration of outliers and incorporation of new data due to a lack of available data sets. Our linear regression also accounts for multicollinearity, when multiple independent variables are correlated. For example, the popularity of e-bikes and gas prices are likely correlated. This flexibility also allows us to obtain reliable results with a Monte-Carlo simulation.

One weakness is that the weights use different units, so that increase of x in one variable is not necessarily equivalent to an increase in another variable. The weights are significant only when you are changing one variable at a time.

5 Off the Chain

5.1 Restatement of the Problem

Switching to e-bikes as a main method of transportation presents an inherent impact on both the public and the Earth itself. Inherently, implementing e-bikes as a dominant means of transportation would cause fewer people to choose other transportation options. To understand the implications of this change, we must create a model that quantifies the effects of switching to e-bikes on the world. In our case, we look to find the changes as a direct result of riding e-bikes instead of car rides of 10,000 e-bike users in a year.

5.2 Assumptions

Assumption: Carbon emission rates will not dramatically change for all vehicles.

- **Justification:** Fluctuations in carbon emission rates for vehicles would cause unforeseen changes in the effect of altering the number of gas vehicles, distorting the calculated effect of switching to e-bikes on carbon levels. Therefore, we assume that each vehicle continues to affect carbon levels in the same fashion as before.

Assumption: Consumer preferences will not dramatically change.

- **Justification:** If consumers suddenly lose interest in e-bikes, usage of other means of transportation would no longer decrease significantly, nullifying the predicted effects of switching to e-bikes. Our premise that consumers will remain interested in e-bikes ensures that the effects of switching to e-bikes would remain relevant.

Assumption: The price of electricity will remain relatively the same.

- **Justification:** If electricity prices grow significantly, the cost efficiency of owning and using an e-bike will decrease dramatically. E-bike usage will also decrease substantially, causing the effect on the public and environment to be much weaker than calculated. To maintain the accuracy of the model prediction, we assume that the price of electricity remains relatively constant.

5.3 Variables

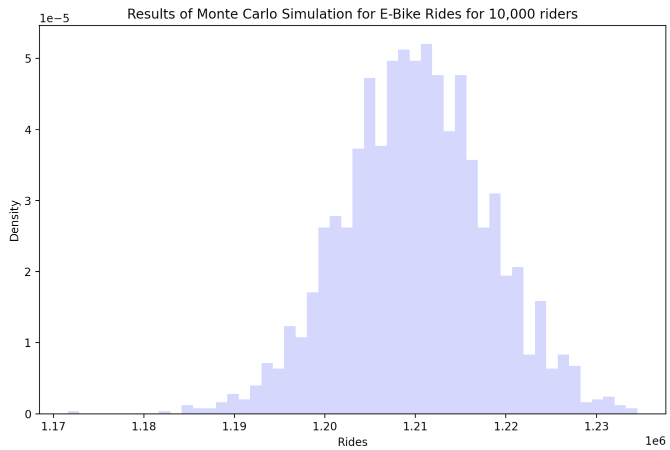
Variable	**Definition**	**Unit**
I	Household Income	dollars($)
S	Household Size	person
D	Population Density	persons/mi^2
T	Count of Daily Trips	trips

5.4 Model Development

To quantify the change in carbon emissions, finance, traffic congestion, and health as a direct result of the usage of e-bikes, we first look to predict the number of bike trips taken, as opposed to car rides, by a random group of 10,000 e-bike riders in a year.

In predicting these values, we use the 2017 NHTS National Survey, which contains data on the variables outlined above as well as the number of bike rides per year [3], to build a Random Forest(RF) Regression Model which would predict the number of e-bike rides taken by the group.

We then generate each variable based on their weights among the US Population(i.e. if 15% of the US Population has a household income between $15-30000, each member of our e-bike rider group has a 15% chance of having that individual income) to find factors for our group of e-bike riders. To account for this variability, we use the law of large numbers and generate a 2000 trial Monte Carlo Simulation to accurately model the number of trips taken by our group of 10,000 riders. The distribution for our simulation can be found below:



Projection	Total Rides
$\mu - \sigma$	1,201,964
μ	1,210,013
$\mu + \sigma$	1,218,062

Next, we will use our previous data on commute time (c) in combination with the average miles per minute speed of an e-bike[2] to find the average distance of a trip (A):

$$A = minutes \times \frac{miles}{minutes} = 6.18$$

We then combine the Monte-Carlo simulation with the average distance to find the average number of miles traveled per year:

$$6.18 \cdot 1,210,013 = 7477880.34 \; mi/yr$$

5.5 Results and Discussion

Using our model, we then find the effect of our group of e-bike riders on carbon emissions, traffic congestion, and personal health.

5.5.1 Carbon Emissions

E-bikes have significantly lower carbon emissions per passenger mile (g/pmi) than other more common forms of transportation, such as cars and buses. Taking in all emissions made (e.g., manufacturing, food-related, electricity or gas prices), an e-bike produces around 9.20g CO_2e/pmi (CO_2 equivalent—CO_2, methane, N_2O, etc.) during its entire lifetime. A car that uses gas, in comparison, emits 168.39g CO_2e/pmi during its lifetime, an 18.3x increase compared to e-bikes[5]. According to research done in the UK[6], the upper limit found for reducing carbon emissions by switching from car to e-bike is around 24.4 million metric tonnes of CO_2 per year. In our case, we find that the group of 10,000 e-bike riders will save $(168.39 - 9.20) \cdot 7477880.34 = 1.1904 \times 10^9$ grams or 1190.4 metric tonnes of CO_2e.

In a study[9], the average cost of a metric tonne of CO_2 emissions, which is calculated based on the damages to society, is around \$185 in the US. In the UK, current carbon pricing is £83 as of March 1, 2023, meaning that if the theoretical 24.4 million metric tonnes could be saved, a total of £2.03 billion would be saved per year. Converting our carbon emissions to monetary value, we find that our riders saved $1190.4 \cdot \$185 = \$220,224$ dollars per year.

5.5.2 Traffic Congestion

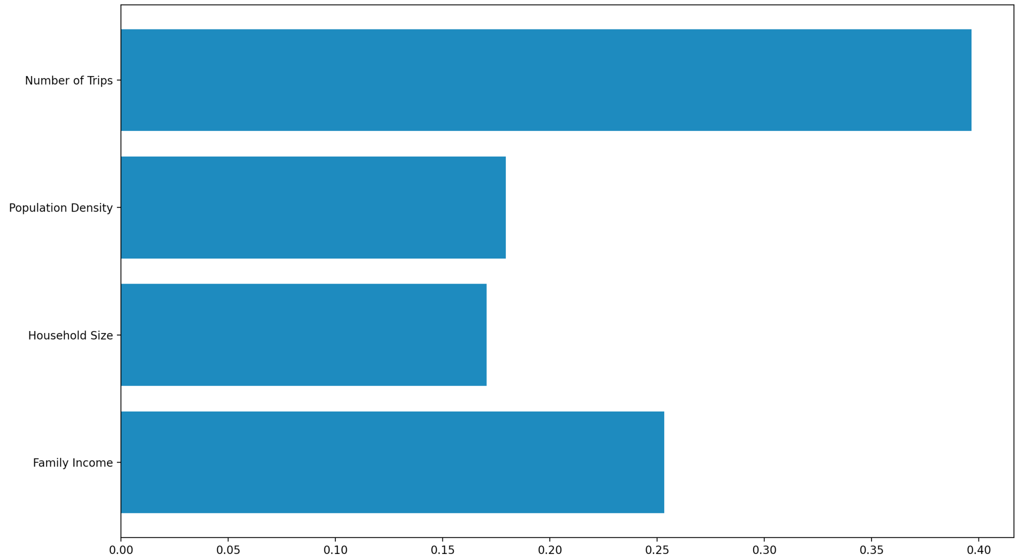
Another major impact of switching to e-bikes is the reduction of overall traffic. Traffic congestion is a major hindrance in the daily life of a worker, easily consuming 30-60 minutes out of a day traveling to and from work. A micromobility ban of e-scooters and e-bikes in Atlanta in 2019[7][8] showed that average commute time increased by around 10%, or 2-5 minutes on average, when e-vehicles were not being used. In our case, based on the Monte Carlo simulation, our e-bike riders took, on average, 1,210,013 e-bike rides instead of car rides, taking their cars off the street and reducing congestion.

5.5.3 Health

Finally, many e-bike owners use their e-bike as a source of recreational exercise. The effect of usage of e-bikes on the health of e-bike users can be found using the approximate equivalence of 1 mile ridden = 30 calories burned[11]. In our case, with over 7 million miles of bike-riding, our group of e-riders burned $30 \cdot 7477880.34 = 224,336,410cal$.

5.6 Feature Importance

The feature importance for our Random Forest Regression can be found below:



In evaluating the feature importance, we found that family income and the number of trips were important in predicting the number of bike rides per year. Considering how e-bike and bike riding in general is a middle to upper-class luxury, it is no surprise that family income is a key predictor in evaluating the number of bike trips. Additionally, the number of trips, in general, was most important because a household that constantly is moving in and out of the house will naturally have more forms of transportation and will tend to use a bike more often.

5.7 Strengths and Weaknesses

A strength of our model is that it samples from a wide plethora of e-bike manufacturers. Since the model considers many top-leading brands of e-bikes in the US, its results are precise, allowing for solid conclusions.

Another strength of our model is its ability to account for variability in a large number of factors. Since the Monte-Carlo simulation samples from differing variables and considers a large number of trials, it can output a reliable average from a distribution, which we can use to find information on the total population.

Also, by using a Random Forest, we are able to consider multiple variables, many of which do not linear relationships, and output a reliable prediction of an average number of trips taken. Because it combines predictions from multiple decision trees, our model is robust and less prone to overfitting. Because the Random Forest regression is scalable and versatile, it can account for different variables and identify which are the most significant.

On the other hand, a weakness of our model is that, due to a lack of available data, it only accounts for commutes to work, disregarding e-bike rides for other activities such as leisure.

This deficiency indicates that the model does not accommodate all e-bike users, which may lead to discrepancies in data.

The model also only accounts for e-bikes and gas-powered cars, excluding cycling and walking. As a result, the health factor does not consider the change in health benefits caused because of the switch from walking or cycling to e-biking.

6 Conclusion

The trend toward e-bikes as an alternative mode of transportation is growing in popularity as noted in our model. This encouraging development decreases the number of gas-powered vehicles on the road, which in turn reduces our carbon footprint and relieves traffic congestion. Given the potential of the e-bike industry, policymakers should seek to implement and fund energy efficiency, which has the greatest effect on e-bike sale growth. Likewise, infrastructure policy and financial incentives will catalyze e-bike sales. Also, along with benefiting the environment and economy, switching to these other modes of transportation benefits the user, since these modes promote activity. Overall, the growth of the e-bike industry brings promising outlooks for the future of the US.

References

[1] Bloomberg. *Trends in Lithium Battery Prices*https://www.bloomberg.com/news/articles/2022-12-06/battery-prices-climb-for-first-time-just-as-more-evs-hit-market?leadSource=uverify%20wall

[2] Portland State University. *Are e-bikes faster than conventional bicycles?*https://trec.pdx.edu/blog/are-e-bikes-faster-conventional-bicycles

[3] National Household Travel Survey. *Compendium of Uses 2017*https://nhts.ornl.gov/2017/pub/Compendium_2017.pdf

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

[5] Stott, Seb. *How green is cycling? Riding, walking, ebikes and driving ranked* https://www.bikeradar.com/features/long-reads/cycling-environmental-impact/

[6] Phillips, et al. *E-bikes and their capability to reduce car CO2 emissions* https://www.sciencedirect.com/science/article/pii/S0967070X21003401?via%3Dihub#bib27

[7] Toll, Micah. *Scientific study shows how much traffic increases when e-bikes and e-scooters are banned* https://electrek.co/2022/10/31/scientific-study-shows-how-much-traffic-increases-when-e-bikes-and-e-scooters-are-banned/

[8] Asensio, et al. *Impacts of micromobility on car displacement with evidence from a natural experiment and geofencing policy* https://www.nature.com/articles/s41560-022-01135-1#Abs1

[9] Rennert, et al. *Comprehensive evidence implies a higher social cost of CO2* https://www.nature.com/articles/s41586-022-05224-9

[10] Cuozzo, Steve. *Beware the E-Bike Menace in NYC*, URL. https://nypost.com/2022/05/14/beware-the-e-bike-menace-in-nyc-they-must-be-banned/

[11] Compendium of Physical Activities *Transportation*, URL. https://sites.google.com/site/compendiumofphysicalactivities/Activity-Categories/transportation-

[12] Grand View Research *Electric Bikes Analysis*, URL.https://www.grandviewresearch.com/industry-analysis/electric-bikes-market

[13] Lime :: Seattle *Free-Floating Bike Share Program Permit Application*, URL.https://www.seattle.gov/documents/Departments/SDOT/BikeProgram/Seattle_2018_Lime_permit_application_final_15Aug18.pdf

[14] CitiBikeNYC *CitiBike*, URL.https://citibikenyc.com/

[15] Lime :: Seattle *Bay Wheels Bike Share Program*, URL.https://mtc.ca.gov/operations/traveler-services/bay-wheels-bike-share-program#:~:text=Bay

[16] MetroBike *AustinBcycle*, URL.https://austin.bcycle.com/

[17] Deloitte *E-bikes merge into the fast lane* , URL.https://www2.deloitte.com/us/en/insights/industry/technology/smart-micromobility-e-bikes.html

[18] Juiced Bikes *E-Bike Facts and Statistics 2020* , URL.https://www.juicedbikes.com/blogs/news/e-bike-facts-and-statistics

[19] Mordor Intelligence *NORTH AMERICA E-BIKE MARKET - SIZE, SHARE, COVID-19 IMPACT FORECASTS UP TO 2029* , URL.https://www.mordorintelligence.com/industry-reports/north-america-e-bike-market

[20] Analytics Vidhya *Sentiment Analysis Using Vader* , URL.https://www.analyticsvidhya.com/blog/2022/10/sentiment-analysis-using-vader/#:~:text=VADER(%20Valence%20Aware%20Dictionary%20for,as%20either%20positive%20or%20negative.)

[21] Office of ENERGY EFFICIENCY RENEWABLE ENERGY *FOTW 1272, January 9, 2023: Electric Vehicle Battery Pack Costs in 2022 Are Nearly 90% Lower than in 2008, according to DOE Estimates*, URL.https://www.energy.gov/eere/vehicles/articles/fotw-1272-january-9-2023-electric-vehicle-battery-pack-costs-2022-are-nearly

Appendix

Sentiment Analysis

```
import snscrape.modules.twitter as sntwitter
import pandas as pd
import os
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
os.system("snscrape --jsonl --progress --max-results 500 --since 2017-01-01 twitter
-search 'e-bike until:2018-01-01' > text-query-tweets.json")
tweets_df = pd.read_json('text-query-tweets.json', lines=True)
print(tweets_df.columns)
analyzer = SentimentIntensityAnalyzer()
scores = 0
for sentence in tweets_df["rawContent"]:
    vs = analyzer.polarity_scores(sentence)
    compound = vs["compound"]
    if(compound > 0.6):
        print(compound)
        print(sentence)
    scores = compound + scores
print(scores/5000)
\end{document}
```

Q1 Monte Carlo Simulation

```
import pandas
from sklearn import linear_model
from scipy.stats import exponweib, nakagami, dgamma
import time
monteplusresults2025 = []
monteplusresults2028 = []
monteminusresults2025 = []
monteminusresults2028 = []
df = pandas.read_csv("/Users/lzhou/Downloads/multivariate.csv")
y = df['Sales']
X = df[['Disposable Income','Commute Time','Gas Prices',
'Environment Factors("Great Deal")','Popularity','gravimetric energy densities(watt hour



regr = linear_model.LinearRegression()
regr.fit(X, y)
r_squared = regr.score(X, y)
dfmonte = pandas.read_csv("/Users/lzhou/Downloads/montecarlo.csv")
dfmonte = dfmonte.drop(index=2)
for i in range(100):
    dfmonte['Commute'] = exponweib.rvs(0.19080767212119915, 3.7915627953332645, 25.59999
    dfmonte['Gas'] = nakagami.rvs(0.1991258897808997, 2.1399999999999997, 1.196717266158
    dfmonte['Popularity'] = dgamma.rvs(1.8730282946959174, 0.12971971629881,0.0217292435
    X = dfmonte[['Disposable Income','Commute','Gas','Environment Factors("Great Deal")'
    dfmonte["Prediction"] = regr.predict(X)
    og2025 = dfmonte["Prediction"][1]
    og2028 = dfmonte["Prediction"][0]

    X['Population'] = X['Population'] * 1.05#-.05
    dfmonte["New Prediction+5"] = regr.predict(X)
    new2025 = dfmonte['New Prediction+5'][1]
    new2028 = dfmonte['New Prediction+5'][0]
    diff2025 = (og2025 - new2025)/og2025
    diff2028 = (og2028 - new2028)/og2028
    monteplusresults2025.append(diff2025)
    monteplusresults2028.append(diff2028)
    X['Population'] = X['Population'] * .95#-.05
    dfmonte["New Prediction-5"] = regr.predict(X)
    new2025 = dfmonte['New Prediction-5'][1]
    new2028 = dfmonte['New Prediction-5'][0]
    diff2025 = (og2025 - new2025)/og2025
    diff2028 = (og2028 - new2028)/og2028
```

```
    monteminusresults2025.append(diff2025)
    monteminusresults2028.append(diff2028)
d = {'2025plusresults': monteplusresults2025, '2028plusresults': monteplusresults2028,'2
dfresults = pandas.DataFrame(data = d)
print("2025 Plus: " + str(dfresults['2025plusresults'].mean()))
print("2025 Minus: " + str(dfresults['2025minusresults'].mean()))
print("2028 Plus: " + str(dfresults['2028plusresults'].mean()))
print("2028 Minus: " + str(dfresults['2028minusresults'].mean()))
```

Q2 Monte Carlo

```
import pandas as pd
import numpy as np
import random
import time
df = pd.read_csv("/Users/lzhou/Downloads/csv 2/hhpub.csv")
df = df[['BIKE','HHFAMINC','HHSIZE','HTPPOPDN','CNTTDHH']]
df = df[(df > 0).all(1)]
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
incomeoptions = [5000,12500,20000,30000,40000,60000,85000,110000,135000,175000,250000]
incomeweights = [4.9,5,8.8,9.4,12,17.3,12.9,9.8,5.4,5.4,5.7]
hhsizeoptions = [32.2,42.8,11.7,8.8,3.1,.9,.3,.1,0,0,0,0,0]
hhsizeweights =[1,2,3,4,5,6,7,8,9,10,11,12,13]
popdnoptions = [50,300,750,1500,3000,7000,15000,200000]
popdnweights = [14.7,17.0,9.3,13.2,18.8,20.8,4.6,1.5]


train, test = train_test_split(df, test_size=0.2)
x_train = train.drop(['BIKE'], axis=1)
y_train = train['BIKE']
x_test = test.drop(['BIKE'], axis=1)
y_test = test['BIKE']
regressor = RandomForestRegressor(n_estimators = 100, random_state = 0)
regressor.fit(x_train, y_train)
y_pred = regressor.predict(x_test)
from sklearn.metrics import mean_squared_error
print(mean_squared_error(y_pred, y_test))
from matplotlib import pyplot as plt

montecarloresults = []
for i in range(2000):
    dfincome = pd.DataFrame(random.choices(incomeoptions, weights=incomeweights, k=10000
    dfhhsize = pd.DataFrame(random.choices(hhsizeoptions, weights=hhsizeweights, k=10000
    dfpopdn = pd.DataFrame(random.choices(popdnoptions, weights=popdnweights, k=10000),
```

```
    dfpredict = pd.concat([dfincome,dfhhsize,dfpopdn],axis=1)
    dfpredict['CNTTDHH'] = np.random.randint(0, 95, dfpredict.shape[0])

    newpredict = regressor.predict(dfpredict)
    montecarloresults.append(newpredict.sum())
    print(newpredict.sum())
monteresults = pd.DataFrame()
monteresults["Days"] = montecarloresults
monteresults.to_csv("monteresults1.csv")
```