

M3 Challenge 2023:

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

TEAM #16401

March 5th, 2023

Executive Summary

To the head of the U.S. Department of Transportation,

As climate change becomes an increasingly pressing issue, policymakers are looking towards alternate forms of transportation to gas-powered cars. Since August 2022, states such as California, Massachusetts, and New York have passed laws that will ban the sale of gas-engine vehicles by 2035^[1]. As a result, motorists are looking to electric vehicles, which do not require the use of gasoline, as an alternative form of transportation. Although electric cars are a viable EV option, electric bikes (e-bikes) offer a more affordable, flexible, and enjoyable form of transportation^[2]. Along with these consumer benefits, e-bikes are over 20 times more efficient than electric cars at combating climate change^[3]. Thus, it is imperative to comprehensively understand the growing role e-bikes will play in the future of transportation.

We predicted the number of U.S. e-bike sales in 2025 and 2028 using an autoregressive moving average (ARIMA) model. Because we lacked data about historical e-bike sales, we applied an adjustable multiplier to relate e-bike and plug-in hybrid electric vehicle (PHEV) sales. We used monthly sales to represent the seasonality trend found in our data and estimated monthly e-bike sales through a time series from 2018 to 2030. Our model predicts 1.409 million annual e-bike sales in 2025 and 1.912 million annual e-bike sales in 2028.

We then determined the significance of four factors: gas prices, median disposable income, transportation distance, and environmental awareness in e-bike sales using a Granger causality test. We represented each of the four factors as a time series that could be compared to the original time series representing e-bike sales and found whether a factor could reasonably forecast e-bike sales. We used the Granger causality test to find the p-value and F-statistic, which we then used to test for significance with an α -value of 0.05. In addition, we took into account how certain factors may influence e-bike sales in the short and long term by performing the Granger causality test after 1 month, 3 months, and 12 months. We found that both gas prices and environmental awareness only impact e-bike sales in the short run after 1 month, whereas transportation distance only impacts e-bike sales in the long run after 12 months. Changes in median disposable income did not significantly impact e-bike sales at any time.

Finally, we quantified the effects of growing e-bike usage on carbon emissions, traffic congestion, and exercise using a Markov chain model. Taking into account transportation preference trends and potential motivations for switching transportation modes, we created a migration matrix to predict the number of people using the following: gas cars, electric cars, public transportation, electric bikes, and regular bikes in two and five years. Combining this information with our relative effect weighting matrix — which determined how different transportation modes influenced carbon emissions, traffic congestion, and user health — allowed us to calculate an impact score for each of these factors. We found that increased e-bike usage decreased carbon emissions and traffic congestion as well as increased user health. Our model predicts that carbon emissions will decrease by 3.85% in two years and 9.52% in five years, traffic congestion will decrease by 0.644% in two years and 1.60% in five years and health will increase by 0.0891% in two years and 1.60% in five years.

We believe these results will assist policymakers in ensuring e-bikes become an integral part of an efficient and sustainable energy plan in the United States.

Table of Contents

Executive Summary	2
Global Assumptions	4
1 Q1: The Road Ahead	
1.1 Defining the Problem	4
1.2 Assumptions	4
1.3 Variables	5
1.4 The Model	5
1.5 Results	7
1.6 Discussion	8
1.7 Sensitivity Analysis	8
1.8 Strengths & Weaknesses	8
2 Q2: Shifting Gears	
2.1 Defining the Problem	8
2.2 Assumptions	9
2.3 Variables	9
2.4 The Model	10
2.5 Results	11
2.6 Discussion	12
2.7 Strengths & Weaknesses	12
3 Q3: Off the Chain	
3.1 Defining the Problem	13
3.2 Assumptions	13
3.3 Variables	14
3.4 The Model	14
3.5 Results	17
3.6 Discussion	18
3.7 Sensitivity Analysis	18
3.8 Strengths & Weaknesses	19
Conclusion	19
References	20
Code Appendix	22

Global Assumptions:

G-1. For our purposes, all hybrid bikes will be considered electric bikes.

- **Justification:** Hybrid bikes are defined as a hybrid between electric mountain bikes and electric road bikes. All hybrids have the same motor and power system as e-bikes, so we will assume they have the same functionality^[4].

Q1: The Road Ahead

1.1 Defining the Problem

The problem asks us to develop a model to predict U.S. e-bike sales in 2025 and 2028. Our model will take into consideration previous sales data from the U.S. in related markets.

1.2 Assumptions

1-1. The growth of plug-in hybrid electric vehicle (PHEV) sales and e-bike sales will be proportional to each other by an adjustable multiplier.

- **Justification:** Although limited data was provided for electric bike sales by the Mathworks Math Modeling Challenge^[9], it was given that electric bike sales are outpacing electric car sales. Therefore, we are assuming there is a trend for the yearly multiplier between e-bike and e-car sales; this can then be used to extrapolate e-bike sale data from e-car sale data.

1-2. Electric bike sales and PHEV sales both peak during the summer months on average every year.

- **Justification:** E-bike sales peak during the summer months, while PHEV sales peak during two periods: from March-May and September-December, giving a bimodal distribution in the number of sales^[5]. Averaging these two peaks gives an average peak during the summer seasons, similar to e-bike sales throughout the year^[6]. This suggests that electric bike sales and PHEV sales follow similar trends and can be related to each other.

1-3. COVID-19 will have no effect on the e-bike market after 2023.

- **Justification:** The economic effects of COVID-19 have died down in the United States as lockdown restrictions will be completely lifted by the end of 2023^[7]. Hence, our model will ignore the previous impacts of COVID-19 in order to make a reasonable prediction for the future.

1-4. Current e-bike manufacturers will not undergo any drastic changes in the next decade.

- **Justification:** It is difficult to predict the business decisions of any companies or manufacturers; thus, for simplicity's sake, we will choose to exclude these interventions from our model. We will also assume that companies will pursue growth in the next decade.

1-5. Annual e-bike sales from 2018 to 2022 in the U.S. is sufficient to determine annual e-bike sales in the future.

- **Justification:** The data provided by the Mathworks Math Modelling Challenge is only from 2018 to 2022, so we will assume for sake of simplicity that this time frame provides sufficient data for our model.
- 2019 2020 2021 2022 2023 Electric Energy Delivered (millions of kWh) CECNY full service customers 20,579 20,544 20,710 22,547 22,657

1.3 Variables

Symbol	Definition	Units
S_v	Number of plug-in hybrid electric vehicles sold, per year	units/year
S_b	Number of e-bikes sold, per year	units/year
M	Ratio of e-bikes to PHEVs sold, per year (adjustable multiplier)	dimensionless
$S_{v,m}$	Number of PHEVs sold monthly	units/month
$S_{b,m}$	Estimated number of e-bikes sold monthly	units/month

Table 1: Variable definitions for Problem 1

1.4 The Model

1.4.1 Developing the Model

We chose an autoregressive integrated moving average (ARIMA) model to predict the growth of e-bike sales. Because each data point for e-bike sales is associated with a time order, we can employ a time-series modeling approach, such as ARIMA. ARIMA models are often used to forecast future demand as well as financial models^[8], so they can be applied to the e-bikes market. Furthermore, ARIMA models tend to be fairly accurate for short-term predictions, such as two years and five years. Our data exhibits annual seasonality, which is a characteristic property in the majority of ARIMA analysis. Thus, due its applicability in similar fields and the seasonal trends of the data, an ARIMA analysis will be highly effective in predicting electric bike sales in the near future.

ARIMA is characterized by three main parameters: p , d , and q . The p parameter represents the order of the autoregressive component, which models the relationship between the current value and its past values. The q parameter represents the order of the moving average component, which models the relationship between the current value and the past errors (deviations from the predicted value). The d parameter represents the degree of differentiation, which refers to the number of times the time series is different to achieve stationarity.

Our team considered using both linear and polynomial regressions to model the growth of e-bike sales. However, we decided that both were inferior in this situation because they were either too simple or unrealistic. These models would fail to include much of the seasonal business cycles seen in e-bike sales. As stated before, we do not have sufficient data to assume

that e-bike sales will follow the form of a particular model, such as a linear or polynomial function. Using the five data points provided by the Mathworks Math Modeling Challenge for Question 1, a basic linear regression model and a degree two polynomial regression model provided very high R^2 values of 0.849 and 0.949, respectively. These values suggest that the models may potentially overfit the limited data. Furthermore, regression requires independent variable(s) that influence a dependent variable, which is not always found in real-world markets. On the other hand, ARIMA is focused on the behavior of the same variable over time ^[10], which better aligns with our given data on e-bike sales across different years.

1.4.2 Executing the Model

Because seasonal sales are necessary for an ARIMA model, we used estimated monthly sales instead. Due to the lack of data provided for monthly e-bike sales in the US, we chose to relate the number of sales between PHEVs and e-bikes instead. Our team sourced monthly sales data for plug-in electric vehicles from Argonne National Laboratory ^[12]. Since e-bikes are currently outpacing electric cars, we can assume a ratio between e-bike and PHEV sales, which creates an adjustable multiplier for previous years. We then applied this adjustable multiplier to the monthly PHEV sales we found and then applied ARIMA to predict the monthly US e-bike sales in 2025 and 2028.

Year	US annual e-bike sales	US annual PHEV sales	EV adjustable multiplier (M)
2018	369,000	323,912	1.139
2019	423,000	326,000	1.297
2020	416,000	308,000	1.350
2021	750,000	608,000	1.233
2022	928,000	862,000	1.076

Table 2: Annual multipliers between e-bikes and PHEVs

These yearly adjustable multipliers were applied to the monthly number of PHEVs sold in the US in order to obtain an estimated number of e-bikes sold per month from 2018 to 2022. Since we assumed the peak seasons of both products would match, any minute differences in peak seasons would be naturally omitted.

After importing our adjusted data into a Python notebook, our team utilized the autoARIMA function from the *pmdarima* library. This approach offers a significant advantage compared to simpler models as it automatically identifies the optimal values for the p , d , and q constants that are necessary for building an ARIMA model.

We used the entirety of the data, from January 2018 to December 2022, and excluded the remaining data value from January 2023, to serve as training data for a sensitivity analysis. For the ARIMA model, we specified the testing set as the time period of January 2023 to December

2030 on a monthly increment. This gave us a total of $12 \text{ months} \times 8 \text{ years} = 96$ periods to predict via the ARIMA model, giving us a specific dataframe for the predicted sales for each month.

1.5 Results

Below is the graph for our ARIMA predictions of electric bike sales for the next eight years and a table consisting of estimated e-bikes sold per month at the beginning of fiscal quarters.

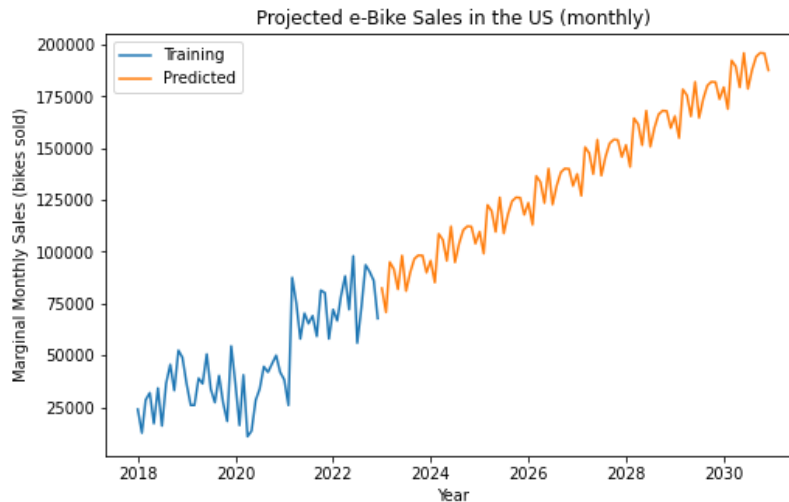


Figure 1: Graph of training and predicted values of e-bike sales through 2030

US projected e-bike sales by month in 2025 and 2028*

Year	Month	Estimated e-bikes sold
2025	January	109,700
	April	119,800
	July	108,900
	October	126,300
	Total	1,409,000
2028	January	151,600
	April	161,700
	July	150,800
	October	168,200
	Total	1,911,500

*Values are rounded to the nearest hundred.

Table 3: Estimation of e-bike sales for 2 and 5 years into the future

1.6 Discussion

In summary, our model predicts that there will be 1.409 and 1.912 million e-bikes sold in the U.S. in 2025 and 2028, respectively. From this, we can conclude that e-bike sales will continue to grow in the future. Our team also broke down our sales into 3-month increments in order to analyze trends in seasonal sales, which is a key factor in ARIMA analysis. The most e-bikes are sold in April and December in both years, which is consistent with peaks in PHEV sales. This makes sense, as e-bikes also tend to be cheaper during these months. This is due to changes in market inventory and demand^[38].

1.7 Sensitivity Analysis

We performed a sensitivity analysis using a percent error comparing the predicted sales of January 2023 compared to the actual sales of 2023 as per the dataset used. In the month, about 86,000 sales of e-bikes were made. Our ARIMA model predicted the sales of Jan. 2023 to be 82,374.74 units, or 82400 rounded to the nearest 100 sales.

$$P.E = \frac{|predicted - actual|}{actual} \times 100\%$$

Using this formula, we get a percent error of **4.1%**.

This means that our ARIMA model is close to the actual sales from a few months before, but it cannot necessarily guarantee accurate predictions for future sales.

1.8 Strengths & Weaknesses

The use of an ARIMA model takes advantage of the seasonal time series expected of electric vehicle sales. ARIMA combats the issue of auto-correlation, which arises when observations rely on factors other than time. Additionally, given that our data exhibits a clear increasing trend and that the problem only asks for short-term predictions of two years and five years, ARIMA allows us to model the data while also accounting for seasonal fluctuations.

However, our model may not be completely accurate, as it requires assuming a relationship between the PHEV and the electric bike market. Along with this crucial assumption, the ARIMA model has a limited ability to handle any missing data or outliers^[11], which can drastically decrease the effectiveness of our ARIMA model. Our model will also be unable to predict in the long term, as it is based on historical data and parameters that may be affected by human bias. This weakness may have an effect on our results as the COVID years from 2020 to 2022 may be considered outliers in the e-bike market, which can potentially skew our output due to the relative inability of the ARIMA model to mitigate outliers.

Q2: Shifting Gears

2.1 Defining the Problem

The second problem asks us to interpret the significance of several underlying factors that may have contributed to e-bike growth. We chose four factors to analyze against the monthly

e-bike sales in the US by determining which causes could potentially significantly contribute to e-bike growth.

2.2 Assumptions

2-1. Correlations between e-bike sales and a potential underlying factor will be completely independent of other factors.

- **Justification:** When testing one of our factors against e-bike sales, we will have to assume independence among other factors in order to run a Granger causality test. From there, we can determine the significance of any factors that may have a correlation.

2-2. Buying rationale is similar among all consumers in the United States.

- **Justification:** As buying rationale may vary in the United States, it is impossible to accurately represent individual trends without loss of generality. Therefore, for the purpose of this analysis, we assume a constant average for buyer rationale.

2-3. Real disposable personal income per capita can be used as a representative measure of personal finance in the United States.

- **Justification:** Real disposable personal income per capita is calculated by taking income earned from all sources and dividing it by the total U.S. population. It is often used as a benchmark for improvements in the average real living standards of American citizens. Hence, we will utilize this value for our personal finance factor.

2-4. The market of inferior and substitute goods for e-bikes does not have a statistically significant effect on factors contributing to e-bike sales.

- **Justification:** Any external circumstances that may influence someone to buy an electric bike are considered too random to significantly impact overall electric bike sales.

2-5. Sentiments of American adults towards global warming are representative of overall opinions towards environmental awareness.

- **Justification:** It is impossible to analyze personal environment awareness. Using surveys about sentiments toward global warming is an optimal method for quantifying environmental awareness.

2.3 Variables

Symbol	Definition	Units
S_m	Estimated of e-bikes sold, per month	units/month
G_m	Gas price per gallon in the United States, per month	USD
I_m	Real Disposable Personal Income per Capita, per month	USD
V_m	Vehicle miles traveled in billions, per month	miles/month
E_m	Percent of adults that are “very sure” that global warming is happening, per month ^[36]	%

Table 4: Variable definitions for problem 2

2.4 The Model

2.4.1 Developing the Model

We first determined the factors we wanted to consider that may affect e-bike sales. We originally took a financial approach and found data of the gas price per gallon in the United States per month^[33]. Through economic market actions of supply and demand, an increase in gas prices will increase e-bike sales, as cars would be in less demand, increasing demand for e-bikes. After this, we examined the monthly disposable personal income per capita of Americans in relation to e-bike sales^[34]. Our next step in finding data was to assess competition in the transportation industry. Our team found a set of vehicle miles traveled (in billions) per month in the time period, allowing us to make an accurate comparison with the e-bike sales^[35]. The last set we used was one to represent the environmental sentiment of Americans; this was the percentage of adults that believe global warming is occurring in the world^[36]. Because e-bikes have fewer carbon emissions than cars, an increase in this percentage should theoretically yield an increase in e-bike sales.

To assess the significance of factors, our team implemented the Granger causality test. The Granger causality test is a statistical test that helps determine whether a single-time series is useful in forecasting another, providing evidence for causality between the two variables^[37]. It is commonly used in econometrics and finance to investigate causal relationships between variables over time^[40]. The Granger causality test allowed us to find p-values for significance and F-statistic scores. We used the F-statistic to measure significance because the F-test is built for continuous data, which is synonymous with our time-series data sets. The F-statistic was chosen over the Chi-squared test, as our data is not categorical. We will be using an α of 0.05 as a measure of statistical significance, as this is standard practice for non-medical studies.

$$y_i = \alpha_0 + \sum_{j=1}^m \alpha_j y_{i-j} + \sum_{j=1}^m \beta_j x_{i-j} + \varepsilon_i$$

Figure 2: Granger causality general formula

The Granger causality test also includes another parameter known as lag, which is the number of periods it takes for a causality to occur. For example, if the p-value at a lag of 1 is less than the α value, then there is a statistically significant cause-and-effect relationship between the combined dataset. However, if a lag of 12 does not yield a causality (p-value greater than α value), then it can be said that the two sets have a significant correlation in the short term but not in the long term.

2.4.2 Executing the Model

In order to leverage the available datasets, we commenced with the critical step of cleaning the data to align them with the sales parameters that were denominated in months and getting rid of irrelevant data points. Upon ensuring the time period for each set ranging from January 2018 to December 2022 was standardized, we could seamlessly integrate the e-bike sales

dataset with another variable set and proceed to employ the Granger causality methodology in the *statsmodels* Python library as a combined dataframe. We comprehensively tested all the potential factors against the e-bike sales data as reported in Question 1.

When dealing with large gaps in data, it can be difficult to accurately represent the true values that may have existed during that time. In order to address this issue, we used linear regression to predict the missing values in the dataset for the environmental awareness variable. By fitting a line between data points with large monthly gaps, we were able to estimate the values that would have likely fallen within the gaps in the data. This allowed us to have a more complete dataset to work with and to better analyze the relationship between environmental awareness and e-bike sales on a monthly basis.

The general formula to determine Granger causality between gas prices and e-bike sales is presented below, where A is the nonempty dataset we gathered for each test and I denotes the information available at the time:

$$P[S_m(t+1) \in A \mid I_{(t)}] \neq P[S_m(t+1) \in A \mid I_{G_m(t)}]$$

Once we obtained F-statistics for different lag values, namely one, three, and twelve (months), our model conducted F-statistic tests and recorded corresponding p-values. These p-values were then compared with our predetermined significance level, α . Variables that showed significance at specific lag levels provided compelling evidence that there was a statistically significant association between that variable and the e-bike sales time series, with respect to the corresponding lag time.

2.5 Results

Below is a table that consists of each factor tested, three different lag values, their F-statistics, respective p-values, and whether they passed our significance level. We found that both gas prices and environmental awareness only impact e-bike sales after 1 month, whereas transportation distance only impacts sales after 12 months. Changes in median disposable income did not ever significantly impact e-bike sales.

Factor	Lag (months)	F-statistic	p-value	Significance ($< \alpha^*$)
Gas prices	1	8.0838	0.0062	Yes
	3	0.9598	0.4190	No
	12	1.3898	0.2398	No
Median Disposable	1	0.1125	0.7386	No

Income	3	0.8456	0.4755	No
	12	1.3689	0.2493	No
Transportation distance (vehicle miles traveled)	1	2.5286	0.1174	No
	3	2.6276	0.0604	No
	12	2.8609	0.0146	Yes
Environmental Awareness	1	4.5847	0.0366	Yes
	3	0.7568	0.5237	No
	12	1.6390	0.1490	No

Table 5. Significance of factors with 1, 3, and 12 months of lag.

**a will be considered as 0.05*

2.6 Discussion

Our Granger causality tests have provided insights into the factors that significantly affect e-bike sales, which include gas prices, transportation distance, and environmental awareness. Specifically, our analysis shows that gas prices and environmental awareness have a short-term impact on e-bike sales, whereas transportation distance has a long-term impact. Interestingly, our results indicate that disposable income does not have any significant impact on e-bike sales, as no tested lag values were found to be significant. Therefore, our findings suggest that e-bike manufacturers and marketers should focus on promoting the environmental benefits of e-bikes and adjust their pricing strategies according to gas prices, rather than targeting consumers based on their disposable income.

2.7 Strengths and Weaknesses

Given that electric bike sales represent a time series, the Granger causality model is a flexible choice in accounting for non-stationary and non-linear data. In its capacity to compare various time series with respect to the different factors we want to analyze, the Granger causality model is effective in being able to compare the significance of different variables, especially in the scope of this problem.

However, while significance is interpreted as causality in this problem, correlation does not necessarily equate to causality, meaning that the factors analyzed may not be linked to the electric bike sales in a strictly cause-and-effect relationship. Therefore, it is important to note that it is impossible to fully determine causation from this data despite the tests our team has run. Furthermore, the data we used for environmental awareness lacked data for certain months, so we fit a line between the nearest data points to predict these values.

Q3: Off the Chain

3.1 Defining the Problem

The third problem asks us to evaluate the impacts of people switching to e-bikes from alternative modes of transportation on carbon emissions, traffic congestion, and health and wellness. We will quantify these impacts by assigning a relative score in each of the categories over time based on the distribution of primary modes of transportation.

3.2 Assumptions

3-1. Society's values and emphasis on the environment will remain constant in the future.

- **Justification:** Society, especially in the West, is currently concerned with the impacts of human activity on the environment. This concern will remain relatively steady until at least 2050, meaning that the focus in the transportation sector, specifically in first-world countries, is to progress in eco-friendly modes of transportation.

3-2. Cars, public transportation, and bicycles make up all primary modes of transportation.

- **Justification:** The Markov chain implements changes in people's primary mode of transportation. We used the three most popular primary forms of transportation in 2021 and 2022, which were cars (76%), public transportation (11%), and bicycles (10%) ^[39]. These made up 97% of all transportation, which is reasonably close to 100%, so we can reasonably assume that they make up all primary forms of transportation.

3-3 The percentage of people who switch from using any type of bike to another mode of transportation is negligible.

- **Justification:** Due to the increasing concerns about the environment, more people are switching to biking. In particular, electric bikes have become more popular for their sustainability and practicality. Taking this into account, as well as the scope of this problem, we will assume that there would be no transition from bikes to other vehicles.

3-4. The larger category of electric vehicles consists only of electric cars and electric bikes.

- **Justification:** This is to simplify our model, as these are the two main modes of electric transportation. We can assume other electronic modes, such as segways and cyber trucks, to be negligible in this case.

3-5. The probability of someone switching from electric vehicles to public transportation & bikes will be the same as someone switching from gas vehicles to public transportation & bikes.

- **Justification:** We assume that the reasons someone will switch from using a car to public transportation or bikes are the same, regardless of whether they drive gas or electric. These reasons include a reduced need to travel long distances, cost, and improved traffic congestion.

3-6. Gas car users would only switch to electric bikes if they moved from a rural area to an urban area.

- **Justification:** Electric bikes are generally only viable in urban areas, as the physical distance between facilities in rural areas makes electric bikes an unfeasible everyday form of transportation. Given that 83% of Americans live in an urban area^[29] and around 10% of Americans move every year^[30], we estimate a maximum of 8.3% of Americans move from an urban area to a rural area annually.

3-7. Currently, no one uses e-bikes to commute to work.

- **Justification:** There is virtually no data on anything related to the number of people that use e-bikes to go to work. Additionally, this assumption simplifies our model and helps better observe the growth in e-bikes in 2025 and 2028.

3-8. Calories burned while using transportation are directly correlated to a user's health and wellness.

- **Justification:** Physical activity plays an important role in ensuring better physical and mental health. Our exercise factor will therefore be measured proportionally using calories burned per hour for each mode of transportation.

3.3 Variables

Symbol	Definition
\mathbf{x}_0	Initial primary mode of transportation distribution matrix for 2023
$\mathbf{x}_2, \mathbf{x}_5$	Predicted transportation distribution matrices for 2025 and 2028
\mathbf{M}	Probability migration matrix for changes in primary transportation preference
\mathbf{E}	Relative effect weighting matrix (measuring carbon emissions, traffic congestion, and health wellness)
$\mathbf{T}_2, \mathbf{T}_5$	Total effect value matrix in 2025 and 2028
$\mathbf{I}_2, \mathbf{I}_5 (\%)$	Change in impact effect in 2025 and 2028 from 2023

Table 6: Variable definitions for Problem 3

3.4 The Model

3.4.1 Developing the Model

We decided to split this problem into two subsections. First, we will use Markov chain analysis to model the changes in people's preferred modes of transportation. Our migration matrix will take into account changes in car, public transportation, and bike usage. To more accurately measure environmental impacts, we split cars into gas-powered cars and electric cars and bicycles into regular bikes and e-bikes. Starting with an initial state vector representing the current amount of vehicles in circulation for each mode of transportation, we can use the Markov

chain model to predict the changing population over a specific time and then use the final population as a basis to analyze additional factors.

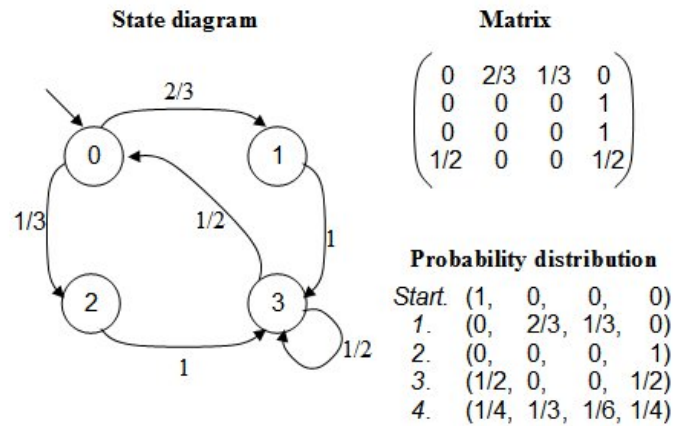


Figure 3: General example of a state diagram and corresponding probability distribution for a Markov Chain

For the migration matrix, we consider the change in usage over five types of vehicles: gas vehicles, electric cars, public transportation, electric bikes, and normal bikes. The majority of commuters do not switch to a new primary mode of transportation. The mode of transportation with the highest proportion of commuters switching to a different mode of transportation is bikes because e-bikes are an increasingly popular substitute for traditional bikes.

We used various data sources and techniques to calculate specific values in the migration matrix. It is known that 25% of people are expected to switch from gas cars to electric cars^[13] and 20% of people switch from electric cars back to gas cars^[14]. Given that people switch to both gas and electric cars every 8 years on average^[26, 27], a total of 3.125% of people can be expected to switch from gas cars to electric cars and 2.5% of people from electric vehicles to gas vehicles annually. Because electric cars make up 37% of the electric vehicles market and around 8.3% of Americans would potentially switch to an electric bike, we estimate 0.1523% of gas car users will switch to an electric bike. Furthermore, 21.3% of people switch from using normal bikes to electric bikes^[15]. Additionally, 0.3% of public transport users switch to bikes on average. As of 2020, around 3% of bikes are electric bikes^[25], which implies that 0.009% of public transport users switch to electric bikes and 0.291% switch to normal bikes. Because an e-bike lasts on average for 10 years^[26], 10% of e-bike users are potentially switching their mode of transportation every year. 11.6% of e-bike users say they would switch to a regular bike^[27], so in a given year 1.16% of e-bike users switch to a regular bike.

Probability Migration Matrix of Switching From					
Switching To	Gas Car	Electric Car	Public Transport	Electric Bike	Regular Bike
Gas Car	0.960797	0.02428	0.01632 [28]	0	0
Electric Car	0.03125	0.89637	0.00068	0	0
Public Transport	0.0057	0.0057	0.98	0	0
Electric Bike	0.001523	0.07292	0.00009	0.9884	0.213
Regular Bike	0.00073	0.00073	0.00291	0.0116	0.787

Table 7: Probability migration matrix for the five transportation units

3.4.2 Executing the Model

We find the number of users for each mode of transportation in 2025 and 2028 by applying the Markov chain to the initial values two and five times, respectively:

$$\mathbf{x}_2 = \mathbf{M}^2 \mathbf{x}_0$$

$$\mathbf{x}_5 = \mathbf{M}^5 \mathbf{x}_0$$

Transportation Mode	Current 2023 Users	Projected 2025 Users (x_2)	Projected 2028 Users (x_5)
Gas Car	250,000,000	231,320,000	207,190,000
Electric Car	2,500,000	16,530,000	30,690,000
Public Transport	6,625,000	9,200,000	12,760,000
Electric Bike	0	1,280,000	6,520,000
Regular Bike	870,000	870,000	950,000

Table 8: Projections for users of the transportation units over time

A matrix of weights for each factor that the migrations will impact is shown below. The impact weighting matrix will contain weights for carbon emissions, traffic congestion, and health. Any category with the largest possible impact will be assigned a maximum score of 1, and all other values will be proportional to that value.

From various sources, we determined the rest of these proportions. Our team found that electric vehicles release half as much greenhouse gas emissions as gas-powered cars on average ^[16]. For every gas car eliminated and replaced with public transportation, a saving of 30% of carbon dioxide emissions can also be realized ^[17]. E-bike power consumption results in an average CO₂ emissions value of 3.5 g/mile ^[18], while the average car may emit 350 g/mile in comparison ^[19]. Regular bikes do not produce any emissions.

Relative Effect Weighting Matrix (E)

Transportation Mode	Carbon Emissions	Traffic Congestion	Health (cal/hr)
Gas vehicles	1	1	0.28 [23]
Electric vehicles	0.5	1	0.28 [23]
Public transportation	0.7	0.68	0.25 [24]
E-bikes	0.01	0.96	0.56 [22]
Regular bikes	0	0.96	1

Table 9: Relative effect weighting matrix for various demand metrics

Gas and electric vehicles contribute to traffic congestion equally, whereas bikes have been reported to reduce congestion in neighborhoods by up to 4% ^[21]. Public transportation, including buses and trains, can reduce up to 32% of traffic in suburban and urban areas ^[20]. Additionally, riding bikes burn the most calories per hour, and the other values are a fraction of that. Since calories burned per hour are assumed to be representative of health benefits, changes in the distribution of primary modes of transportation will impact the health score of a given distribution.

$$\mathbf{T}_2 = \mathbf{E}'\mathbf{x}_2$$

$$\mathbf{T}_5 = \mathbf{E}'\mathbf{x}_5$$

$$\mathbf{I}_2 = \mathbf{T}_2 - \mathbf{T}_0 = \mathbf{E}'\mathbf{x}_2 - \mathbf{E}'\mathbf{x}_0$$

$$\mathbf{I}_5 = \mathbf{T}_5 - \mathbf{T}_0 = \mathbf{E}'\mathbf{x}_5 - \mathbf{E}'\mathbf{x}_0$$

The impact matrix, which shows the impacts on carbon emissions, traffic congestion, and health, is a 3x1 matrix found by multiplying the transpose of the relative effect weighting matrix (E) by the difference in the distributions of people's preferred mode of transportation. We then convert this difference to a percentage.

3.5 Results

Year	Carbon emission impact score	Traffic congestion impact score	Health & wellness impact score
2023 (T_0)	255,887,500	257,840,200	73,226,250
2025 (T_2)	246,040,000	256,180,000	73,290,000
2028 (T_5)	231,530,000	253,730,000	74,400,000
Change in impact in 2 years (I_2)	-3.85%	-0.644%	+0.0891%
Change in impact after 5 years (I_5)	-9.52%	-1.60%	+1.60%

Table 10: Results of matrix multiplication and predicted change in metrics

3.6 Discussion

As the usage of electric vehicles continues to grow as a whole, carbon emission impact and traffic congestion are set to decrease, while health and wellness are set to increase. In particular, after two years, carbon emissions will decrease by 3.85%, traffic congestion will decrease by 0.644%, and health and wellness will increase by 0.0891%. After five years, carbon emissions will decrease by 9.52%, traffic congestion will decrease by 1.60%, and health and wellness will increase by 1.60%.

Additionally, a decrease in overall carbon emission and traffic congestion impact from an increase in e-bikes intuitively makes sense — e-bikes would generally produce fewer emissions and ease traffic congestion in urban areas. E-bikes also help commuters burn more calories and therefore promote healthier exercise, which explains the slight increase in the health and wellness score.

3.7 Sensitivity Analysis

To test the sensitivity of our Markov chain, we changed the original migration matrix by increasing the probability of a user staying with a gas car by 1% and decreasing the probability that one switched from gas to electric cars by 1%. This ensures that our probability matrix is still column stochastic, as that column will still add up to 1. We then used this migration matrix and the same relative effective weighting matrix to determine the effect of the new migration distribution. Comparing this result to our previous result will determine if our model is adaptable and reasonable.

Altered Probability Migration Matrix for Sensitivity Analysis					
Switching To	Gas Car	Electric Car	Public Transport	Electric Bike	Regular Bike
Gas Car	0.970797 (+1%)	0.02428	0.01632	0	0
Electric Car	0.02125 (-1%)	0.89637	0.00068	0	0
Public Transport	0.0057	0.0057	0.98	0	0
Electric Bike	0.001523	0.07292	0.00009	0.9884	0.213
Regular Bike	0.00073	0.00073	0.00291	0.0116	0.787

Table 11: Altered probability migration matrix with altered values for sensitivity analysis

Year	Carbon emission impact score	Traffic congestion impact score	Health & wellness impact score
Change in impact in 2 years (I_2)	-2.88% (+0.97%)	-0.64%	+0.018%
Change in impact after 5 years (I_5)	-7.19% (+2.33%)	-1.59%	+0.98%

Table 12: Assessment of sensitivity analysis

By decreasing the probability of someone switching to an electric car by 1%, we see that it has a drastic change on the carbon emission impact score over 2 and 5 years. This is because there will be more gas cars overall, which will contribute heavily to the carbon emission impact score for each year (T_2 and T_5). In 2025, there will be a 0.97% increase in carbon emissions due to changes and in 2028 there will be a 2.33% increase in carbon emissions.

3.8 Strengths and Weaknesses

The use of Markov chains allows us to be able to model complex relationships and changes between various modes of transportation. This also allows us to take advantage of past trends and “migrations” found between transport transitions, making our model better suited to account for different changes over time. For example, our model effectively accounts for factors such as changing environmental perceptions and restrictions, sustainability, manufacturing, and the changing urban environment.

However, by this same token, in using the Markov chain model, we assume that the initially observed trend would always be constant, making our model inflexible to more drastic changes that can be found in the long term. Additionally, the more minute relationships between the different modes of transportation, and the various factors that can influence them, may be impossible to be able to account for in our model. Other aspects of our model were also derived from past data — many of which were highly affected due to the COVID-19 pandemic — and may not be entirely accurate to the present day.

Conclusion

In the first question, we employed an ARIMA model to predict electric bike sales in two and five years. Our model predicted 1.409 million annual e-bike sales in 2025 and 1.912 million annual e-bike sales in 2028. We used the Granger causality test to analyze the significance of various factors in the growth of the e-bike market. We found that both gas prices and environmental awareness only impact e-bike sales after 1 month. On the other hand, transportation distance only impacts e-bike sales after 12 months. However, changes in median disposable income did not significantly impact e-bike sales at any time. Finally, we found that as primary transportation methods change, carbon emission will decrease by 3.85%, traffic congestion will decrease by 0.644%, and health and wellness will increase by 0.0891% after two years. After five years, carbon emissions will decrease by 9.52%, traffic congestion will decrease by 1.60%, and health and wellness will increase by 1.60%.

In summary, our findings suggest that the e-bike market is poised for an impressive growth trajectory in the coming years. Furthermore, e-bikes have the potential to create a significant positive impact on society, reducing carbon emissions and traffic congestion while simultaneously promoting health and wellness. These findings and correlations will prove to be useful for policymakers and world leaders to develop the future of transportation: electric bikes.

References

1. <https://abcnews.go.com/Business/states-banning-sale-gas-powered-vehicles/story?id=88895372>
2. <https://www.skipeak.net/blog/8-benefits-of-using-electric-bikes>
3. <https://sites.google.com/view/ebikestudy/home?authuser=0>
4. <https://discerningcyclist.com/are-hybrid-bikes-electric/>
5. <https://www.sixthreezero.com/blogs/bike-advice/electric-bikes-on-sale-or-where-and-when-are-e-bikes-on-sale>
6. <https://www.investopedia.com/ask/answers/041715/how-important-are-seasonal-trends-a-utomotive-sector.asp#:~:text=Peak%20Seasons,by%2010%25%20to%2015%25>
7. <https://www.brookings.edu/blog/future-development/2021/05/20/a-long-term-view-of-covid-19s-impact-on-the-rise-of-the-global-consumer-class/>
8. <https://journals.sagepub.com/doi/full/10.1177/1847979018808673>
9. <https://m3challenge.siam.org/node/596>
10. https://medium.com/@lewis_b/next-post-eb6b933256a4
11. <https://towardsdatascience.com/limitations-of-arima-dealing-with-outliers-30cc0c6ddf33>
12. <https://www.anl.gov/esia/light-duty-electric-drive-vehicles-monthly-sales-updates>
13. <https://newsroom.aaa.com/2022/07/americans-reveal-fresh-thoughts-on-electric-vehicles/>
14. <https://www.anthropocenemagazine.org/2021/04/what-makes-people-disloyal-to-their-electric-vehicles/#:~:tAmericans+Reveal+Fresh+Thoughts+on+Electric+Vehicles+|+AAA+Newsroom&text=But%20that%20doesn%27t%20mean%20everyone%20who%20buys%20an,of%20California%20Davis%20decided%20to%20find%20out%20why>
15. <https://www.bicycle-guider.com/bike-facts-stats/>
16. <https://usafacts.org/articles/how-much-emissions-do-electric-cars-produce/>
17. https://www.kcata.org/about_kcata/entries/environmental_benefits_of_public_transit#:~:text=By%20eliminating%20one%20car%20and,dioxide%20emissions%20can%20be%20realized
18. <https://www.bosch-ebike.com/us/service/sustainability#:~:text=If%20you%20only%20consider%20the,depending%20on%20the%20power%20mix>
19. <https://www.energy.gov/eere/vehicles/articles/fotw-1223-january-31-2022-average-carbon-dioxide-emissions-2021-model-year#:~:text=The%20average%20production%20weighted%20carbon,%E2%80%92%20a%20decrease%20of%2049%25>
20. <https://www.bloomberg.com/news/articles/2013-04-01/public-transportation-does-relieve-traffic-congestion-just-not-everywhere>
21. <https://www.itskrs.its.dot.gov/its/benecost.nsf/ID/2bd579fc975fda27852585540063be1a#:~:text=Study%20indicates%20bike%20sharing%20can,as%20much%20as%20four%2039percent.&text=ITS%20Deployment%20Evaluation>

22. <https://www.flyer-bikes.com/en/this-is-how-many-calories-you-burn-when-e-biking#:~:text=Sports%20physicians%20estimate%20that%20an,or%20when%20you%20go%20shopping>.
23. <https://www.bizcalcs.com/calories-burned-driving/#:~:text=How%20many%20calories%20does%20an,to%20250%20calories%20per%20hour>.
24. <https://en.tempo.co/read/1285960/burn-calories-by-riding-public-transport#:~:text=Your%20training%20gets%20more%20effective,100%20kcal%20per%2030%2Dminutes>.
25. <https://www.reuters.com/article/us-health-coronavirus-usa-bikes-idUSKBN21Z1BX>
26. <https://www.thezebra.com/resources/driving/average-length-of-car-ownership/#average-length-of-car-ownership>
27. <https://evelo.com/blogs/learn/a-survey-of-u-s-electric-bike-owners-and-interested-consumers>
28. <https://www.pewresearch.org/fact-tank/2021/12/16/in-2020-fewer-americans-moved-exodus-from-cities-slowed/>
29. <https://css.umich.edu/publications/factsheets/built-environment/us-cities-factsheet>
30. <https://www.shyftmoving.com/blog/moving-industry-statistics>
31. <https://storybicycles.com/blogs/ebike-blog/how-long-do-ebikes-last>
32. <https://wgntv.com/news/how-long-do-electric-cars-last/>
33. https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=emm_epmr_pte_nus_dpg&f=m
34. <https://fred.stlouisfed.org/series/A229RX0>
35. https://www.fhwa.dot.gov/policyinformation/travel_monitoring/tvt.cfm
36. <https://climatecommunication.yale.edu/wp-content/uploads/2022/07/climate-change-american-mind-april-2022.pdf>
37. <https://www.machinelearningplus.com/time-series/granger-causality-test-in-python/>
38. <https://www.bicycling.com/bikes-gear/a20018177/how-to-get-the-best-deal-on-a-bike/>
39. <https://www.statista.com/chart/18208/means-of-transportation-used-by-us-commuters/>
40. http://www.scholarpedia.org/article/Granger_causality

Code Appendix

Question 1: The Road Ahead

```
#Imported Libraries
import numpy as np
import pmdarima as ar
import pandas as pd
import matplotlib as mp
import matplotlib.pyplot as plt
import csv
import math

# Cleaning + Prelim Graph
sales = pd.read_csv('plugin_monthly_sales.csv')
sales = sales.drop("sales_before",axis=1)
sales = sales.drop("sales_next",axis=1)
yearly_multipliers = [1.139198301,1.297546012,1.350649351,1.233552632,1.076566125]

counter = 0
for i in yearly_multipliers:
    for j in range(12):
        sales["sales"][counter] *= i
        counter+=1

for i in range(96):
    sales = sales.append({"year":2023+(i//12), "month": i%12 +
1,'sales':0},ignore_index=True)

sales['date'] = pd.to_datetime(dict(year=sales.year, month=sales.month,day=1))
sales = sales.drop("year",axis=1)
sales = sales.drop("month",axis=1)
sales = sales[["date",'sales']]
print(sales)

sales.set_index("date",inplace=True)
sales.plot()

# Train and Test Value
train = sales[:60]
test = sales[-96:]
plt.plot(train)
plt.plot(test)
```

```
# Arima Model Analysis
arima_model =
ar.auto_arima(train,start_p=0,d=1,start_q=0,max_p=5,max_d=5,max_q=5,start_P=0,D=1,star
t_Q=0,max_P=5,max_D=5,max_Q=5,m=12,
seasonal=True,error_action='warn',trace=True,suppress_warnings=True,stepwise=True,rand
om_state=20,n_fits=50)

arima_model.summary()

# Prediction Table
prediction = pd.DataFrame(arima_model.predict(n_periods=96),index=test.index)
prediction.columns = ['predicted_sales']
print(prediction)

# Values for Year
print(prediction.iloc[[24]])
print(prediction.iloc[[27]])
print(prediction.iloc[[30]])
print(prediction.iloc[[33]])
count_2025 = 0
for i in range(24,36):
    count_2025 += prediction.iloc[i]['predicted_sales']
print(count_2025)
print()
print(prediction.iloc[[60]])
print(prediction.iloc[[63]])
print(prediction.iloc[[66]])
print(prediction.iloc[[69]])
count_2028 = 0
for i in range(60,72):
    count_2028 += prediction.iloc[i]['predicted_sales']
print(count_2028)

# Sensitivity
actualJan2023 = 86000
predictedJan2023 = 82374.740980

percentError = (predictedJan2023-actualJan2023)/actualJan2023 * 100
print(percentError)
```

Question 2: Shifting Gears

```
# Imported Libraries
import statsmodels.api as sm
from statsmodels.tsa.stattools import grangercausalitytests
import numpy as np
import pandas as pd
import matplotlib as mp
import matplotlib.pyplot as plt
import csv
import math

# Sales - Cleaning
sales = pd.read_csv('plugin_monthly_sales.csv')
sales = sales.drop("sales_before",axis=1)
sales = sales.drop("sales_next",axis=1)
yearly_multipliers = [1.139198301,1.297546012,1.350649351,1.233552632,1.076566125]

counter = 0
for i in yearly_multipliers:
    for j in range(12):
        sales["sales"][counter] *= i
        counter+=1

sales['date'] = pd.to_datetime(dict(year=sales.year, month=sales.month,day=1))
sales = sales.drop("year",axis=1)
sales = sales.drop("month",axis=1)
sales = sales[["date","sales"]]
print(sales)

sales.set_index("date",inplace=True)
sales.plot()

# Gas Prices - Cleaning
gas = pd.read_csv('gas_prices.csv')

gas['date'] = pd.to_datetime(dict(year=gas.year, month=gas.month,day=1))
gas = gas.drop("year",axis=1)
gas = gas.drop("month",axis=1)
gas = gas[["date","price"]]
print(gas)

gas.set_index("date",inplace=True)
```



```
gas.plot()

# Gas Prices - Granger causality
test = 'ssr_chi2test'
gasCombine = pd.concat([sales, gas], axis=1)
gasCombine.columns = ['sales', 'prices']
gas_test = grangercausalitytests(gasCombine, 12)

# Disposable Income - Cleaning
income = pd.read_csv('income.csv')

income['date'] = pd.to_datetime(dict(year=income.year, month=income.month, day=1))
income = income.drop("year", axis=1)
income = income.drop("month", axis=1)
income = income[["date", 'disp_income']]
print(income)

income.set_index("date", inplace=True)
income.plot()

# Disposable Income - Granger causality
test = 'ssr_chi2test'
incomeCombine = pd.concat([sales, income], axis=1)
incomeCombine.columns = ['sales', 'disp_income']
income_test = grangercausalitytests(incomeCombine, 12)

# Travel - Cleaning
travel = pd.read_csv('travel.csv')

travel['date'] = pd.to_datetime(dict(year=travel.year, month=travel.month, day=1))
travel = travel.drop("year", axis=1)
travel = travel.drop("month", axis=1)
travel = travel[["date", 'dist_traveled']]
travel["dist_traveled"] = travel["dist_traveled"].str.replace(",", "")
travel["dist_traveled"] = travel["dist_traveled"].astype(float)
print(travel)

travel.set_index("date", inplace=True)
travel.plot()

# Travel - Granger causality
test = 'ssr_chi2test'
```

```

travelCombine = pd.concat([sales, travel], axis=1)
travelCombine.columns = ['sales', 'dist_traveled']
travel_test = grangercausalitytests(travelCombine, 16)

# Environment - Cleaning
env = pd.read_csv('environment.csv')

env['date'] = pd.to_datetime(dict(year=env.year, month=env.month, day=1))
env = env.drop("year", axis=1)
env = env.drop("month", axis=1)
env = env[["date", 'env_sent']]
print(env)

env.set_index("date", inplace=True)
env.plot()

# Environment - Granger causality
test = 'ssr_chi2test'
envCombine = pd.concat([sales, env], axis=1)
envCombine.columns = ['sales', 'env_sent']
env_test = grangercausalitytests(envCombine, 12)

```

Question 3: Off the Chain: (MATLAB)

```

%% Statistics/Probability Matrix M and Current distribution matrix x0
% Migration Matrix to change
M = [.960797 .02428 .01632 0 0; .03125 .8964 .00068 0 0; .0057 .0057 .98 0 0;
.0 .07292 .000009 .9884 .213; .00073 .00073 0.000291 0.0116 .787]
% Effect Matrix for CE, TC, and Health/Exercise
E = [1 1 .28; 0.5 1 .28; 0.7 .68, .25; .01 .96 .56; 0 .96 1]
W = E' % Transpose to multiply
% Mode of Transportation Distributions
x0 = [250000000; 2500000; 6625000; 0; 870000] % Current: 2023
x2 = M*M*x0 % 2025
x5 = M*M*M*M*M*x0 % 2028
% Total effect using effect matrix on transportation distribution
T0 = W*x0 % 2023
T2 = W*x2 % 2025
T5 = W*x5 % 2028
% Impact Matrix as percent change in total effect
I0 = T0./T0 .* 100 - 100 % 2023
I2 = T2./T0 .* 100 - 100 % 2025
I5 = T5./T0 .* 100 - 100 % 2028
%% Sensitivity Analysis

```

```
Ms = [.970797 .02428 .01632 0 0; .02125 .8964 .00068 0 0; .0057 .0057 .98 0 0;  
.0 .07292 .000009 .9884 .213; .00073 .00073 0.000291 0.0116 .787]  
M = Ms  
% Effect Matrix for CE, TC, and Health/Exercise  
E = [1 1 .28; 0.5 1 .28; 0.7 .68, .25; .01 .96 .56; 0 .96 1]  
W = E' % Transpose to multiply  
% Mode of Transportation Distributions  
x0 = [250000000; 2500000; 6625000; 0; 870000] % Current: 2023  
x2 = M*M*x0 % 2025  
x5 = M*M*M*M*M*x0 % 2028  
% Total effect using effect matrix on transportation distribution  
T0 = W*x0 % 2023  
T2 = W*x2 % 2025  
T5 = W*x5 % 2028  
% Impact Matrix as percent change in total effect  
I0 = T0./T0 .* 100 - 100 % 2023  
I2 = T2./T0 .* 100 - 100 % 2025  
I5 = T5./T0 .* 100 - 100 % 2028
```