M3 Challenge 2025:

Hot Button Issue: *Staying Cool as the World Heats Up*

Team #17621

March 3rd, 2025

1 Executive Summary

Dear Memphis Policymakers:

As heat temperatures rise across the globe due to the rapid onset of global warming, new records are being set[1] in every corner of the Earth. Climate change has posed new challenges for those without the proper infrastructure to withstand the year's heatwaves, which are ever-increasing in intensity. To help mitigate this problem, our team of researchers has been driven to analyze numerous energy and heat trends across the city of Memphis, a city that has been hit particularly hard by these issues, to deliver to you the best analysis of which areas of the city need the most assistance and why.

In our first endeavor, we sought to predict indoor temperatures for homes without air conditioning in Memphis and the surrounding areas. We first modeled outdoor temperatures as a function of time and realized that the graph resembled a periodic function. We then accounted for numerous other factors, such as heat transfer, home size, proximity to green cover, number of stories, number of people per room, the amount of shade nearby, and the age of the home. Using all of these factors, we were able to create a highly accurate model that can predict the indoor temperature of an at-risk house without air conditioning, many of which reached dangerously high temperatures of 110° F. We also realized that poorer houses and neighborhoods were disproportionately affected by such heatwaves, and then realized how important it was for us to address these issues.

We then moved on to our second question, which asked us to predict energy consumption across the city during peak consumption seasons, which we argued was during heat waves in the summer. We considered factors such as declines in the current population and the energy consumption per resident, outside factors such as climate and temperature effects, and, more importantly, the rising demand for electricity caused by electric vehicles (EVs). Due to the rapid growth in the popularity of EVs, we modeled their growth using a logistic model, with a carrying capacity of nearly 100% adoption to be achieved after around 2060. We discovered that our power grids are ready for the next 20 years, as we predict power consumption to go down by 24% primarily due to a decreasing population and consumption trend in Memphis. From these factors, we observed that the energy consumption during summer months was set to decline. This clearly shows that while investing government assets into increasing power grid capacities might seem like an optimal idea, it is wiser to spend these resources to bridge the gap between Shelby County's economic inequalities.

This brings us to our final question, where we will present our overarching solution. We were asked to address each neighborhood's risk level using a vulnerability score, which would allow Memphis to better allocate resources to its neighborhoods. We chose to focus on six important factors: income level, proportion of households with adults over the age of 65, indoor temperature, access to health resources, working population, and households with vehicles. We created a model that takes into account each of these factors with its own weightage, which we assigned based on their relative importance. We found that East Memphis - Colonial Yorkshire, Coro Lake / White Haven, and South Memphis were the most at-risk neighborhoods, as they had the top three highest vulnerability scores. We recommend that these areas get the most support from the Memphis city government, as their combination of senior citizen populations, lack of access to health resources, and high indoor temperatures make them uniquely susceptible to heat waves. Furthermore, we recommend that you allocate 10% of the local monetary resources into generators that can provide electricity in the case of a power outage. Additionally, because we found that dangerously high indoor temperatures were highly correlated with a neighborhood's vulnerability, we strongly believe that the remaining resources should address this issue and should be invested into community cooling centers that can serve the low-income and the elderly as well as green spaces with trees and plants that can reduce the temperature.

With all of these factors considered, we have high hopes for Memphis and its ability to serve its citizens accurately and efficiently. We humbly plead with you to take into account our ideas and make your vibrant and wonderful city even better and poised to take on the challenges of the future.

Contents

2 Part I: Hot to Go

2.1 Defining the Problem

The problem asks us to develop a model to predict the indoor temperature of non-air-conditioned dwellings during a heat wave over a 24-hour period. We selected Memphis, Tennessee, as our city. Our model will use the provided dataset containing sample dwellings and specific heat wave data for Memphis.

2.2 Assumptions

Assumption 1: The indoor temperature follows a sinusoidal pattern.

- *We argue this because outside temperature also follows such a pattern since the sun comes out in the morning and goes away at night. Additionally, because we are predicting the temperature of a house without air conditioning, the temperature will accumulate over time and will lag behind the outdoor temperature, since heat transfer takes time.*

Assumption 2: Older homes will on average have higher temperatures.

- *We argue this because older homes will most likely have inferior insulation*[2]*, which cannot protect its residents from the heat.*

Assumption 3: July 2022 is sufficient data to model a heat wave in Memphis, Tennessee.

- *We argue this because July 2022 is fairly recent, and it can reasonably predict what a heatwave would look like if it were to happen tomorrow.*

Assumption 4: Indoor temperature is some function of outdoor temperature.

- *We argue this because by the First Law of Thermodynamics, energy, including heat, is conserved, so the outdoors heat will inevitably seep inside. However, it lags behind the outdoor temperatures because of the reasons we have stated above. We also have consulted sources*[3] *that back this theory up*

Assumption 5: Outdoor temperature is uniformly distributed across the city of Memphis.

- *We argue this because over a long period of time, the heat will reach a steady state and will be approximately constant throughout the city.*

Assumption 6: Each "shade level" contributes to a 1.35°F increase in the indoor temperature.

- *We argue this because we are given three shade levels in the provided data (the fourth one is implied). Based on sources*[4]*, we see that less shade outside the residence generally increases indoor temperature by 1.3 to 1.4°F.*

Assumption 7: Green cover doesn't affect temperature by more than $2°F$.

- *We argue this because of research papers*[5] *arguing this assumption.*

Assumption 8: The Age of a home is modeled by a sigmoid function due to developments in infrastructure.

- *We argue this because there was huge development in 1980 for housing infrastructure and indoor temperature changed significantly*[6] *from the new development. Therefore, we are using a sigmoid function to represent the difference before and after 1980. We use 5°F increase for a house built before the 1980s.*

Assumption 9: Dew point and humidity do not affect the actual indoor temperature of the home.

- *We argue this based on research from our sources*[7]*. This makes sense because these two variables primarily affect how hot it feels, but not the actual temperature itself.*

Assumption 10: Wind speed does not affect the indoor temperature of the home.

- *We argue this because we assume windows are closed and that wind effects are negligible. However, some may argue that we should consider when windows are open. If we did consider this, we would only be pushing around the same temperature of air, which would only allow us to reach our max temperature sooner as air is being spread at increased rates, but it would not actually change the peak temperature, which is what the problem is asking.*

Assumption 10: We can generalize Newton's Law of Heat Transfer for these dwellings, allowing us to use the size of the home to affect the indoor temperature.

- *Because the residences are not air-conditioned, the heat transfer is limited to heat seeping in from outside. Therefore, we can generalize Newton's Law, which is often used for smaller objects, to a larger residence.*

2.3 Model Development

We chose a sinusoidal function to model the relationship between the time of day and the outdoor temperature. Because the sun rises in the morning and sets at night, the outdoor temperature accumulates over time. Though the sun reaches its highest point at around 12:00 PM, the hottest time of day is around 2:00 to 3:00 PM, which is reflected by the data.

Some may argue that other factors such as humidity and dew point will affect the outdoors temperature, but we have found this to be untrue, as these factors are a measurement of the moisture in the air. Because the question has asked us to predict the indoor temperature, we will not be taking these factors into account for Question 1. However, they will become increasingly important for Question 3, as humidity can greatly affect how people "feel" during a heatwave.

The general form of a sinusoidal function is as follows:

$$f(x) = A\sin(bx + c) + d$$

Table 1: Variables

Symbol	Variable	Temperature Analogue
A	Amplitude	Fluctuation from Avg. Temp.
B	Period	24 Hour Repetition
C	Phase Shift	Accumulation of Heat Factor
D	Midline	Median Temperature

We can use the *scipy.optimize* module and its *curve_fit* function to fit our data to this general form, which will allow us to get our coefficients and constants for the sinusoidal function.

Though we once considered polynomial regression, we realized that the data is periodic because the temperature will approximately repeat every 24 hours. Therefore, a sinusoidal regression is very helpful in modeling such a function, since we can add in a period of 24 hours inside our sinusoidal regression.

The below images show the scatterplot and fitted curve of the given data. We achieved an $R\hat{2}$ value of 0.9718 and a Root Mean Square Error of 1.09, both of which suggest a very strong correlation between the sine of the time of day and the temperature.

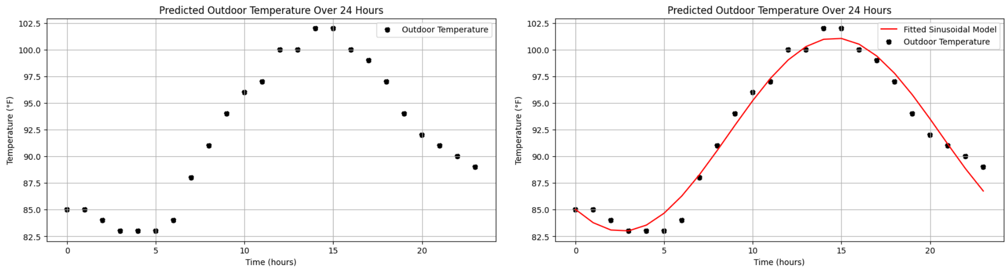


Figure 1: The scatter plot and the fitted curve of our temperature data laid side by side.

The following table shows us the calculated values for each variable. The b-value is 0.2618, which works out perfectly, since the period for a daily repeating function must be 24 hours and $2\pi/24 = 0.261799$.

Table 2: Calculated Variable Values

Symbol	Variable	Value
A	Amplitude	-9.07
B	Period	0.2618
C	Phase Shift	0.88
D	Midline	92.04

2.3.1 Model Development Part II

Now that we have our model for the outdoors temperature, we can use this in tandem with other important variables to calculate our indoors temperature. But first, we can use Newton's Law of Heat Transfer to find the relationship between indoors temperature and outdoors temperature. To restate Assumption 8, because the homes we are looking at are not air-conditioned, there is no external heat transfer other than the heat seeping in from outdoors. Therefore, we can generalize Newton's Law to large residences such as those described in the data. Newton's Law of Heat Transfer argues:

$$\frac{dT}{dt} = -k(T_0 - T_s)$$

where T represents the temperature inside, k is the heat transfer coefficient, T_0 represents the initial temperature, T_s represents the surrounding temperature, and t represents the time of day. To solve this ordinary differential equation, we can separate variables and get our general solution for T:

$$\hat{T_{in}}(t) \propto T_{out}(t) + (T_0 - T_{out}(t)) \cdot e^{-kt}$$

Here, T_{in} represents the indoors temperature, T_{out} represents the outdoors temperature, and all other variables are the same from before. We now observe that we have already experimentally defined $T_{out}(t)$ as the sinusoidal function from before, so we can substitute that into $T_{in}(t)$ to understand how the indoors temperature varies with the outdoors temperature. However, there are many other factors that we must consider when creating our model to predict our indoors temperature. We brainstormed some of the variables that could influence indoors temperature and listed them in the table below:

Table 3: Other Factors that Influence Indoors Temperature

Symbol	Variable	Definition
S	Shade Level	quantitative metric of how much shade exists near a residence
G	Green Cover	calculated metric of how much green space exists near a residence
Y	Age	The year that the home was built in
H	Stories	How high up a residence is
P	People	Number of people in the residence
R	Rooms	Number of rooms in the residence
Z	Size	Size of home in square meters

The last of these factors is the most interesting: the size of a home can determine how fast it warms up, as smaller homes will warm up faster because heat will have less space to fully occupy. Similarly, larger homes will take longer to warm up. Therefore, the size of the home must be taken into consideration when determining the rate at which the indoors temperature of the home warms up. Our new equation is as follows:

$$\hat{T_{in}}(t) \propto T_{out}(t-\tau) + (T_0 - T_{out}(t-\tau)) \cdot e^{\frac{-kt}{z_{scaled}}} = T_{phys}(t-\tau)$$

We introduce the new z_{scaled} factor into our exponent, which is the size of the home in square meters z divided by 100, which we have estimated to be the size of a medium-sized home in Memphis, Tennessee. We also equate our expression with $T_{phys}(t)$ to simplify our equation for later use. Additionally, we also introduce a new phase shift τ into the T_{out} equation, which allows us to account for the indoor temperature taking some time to accumulate heat, therefore predictably lagging behind the patterns of the outdoors temperature. Researchers from the Harlem Heat Project^{source} determined that τ was around 2.5 hours, so we will generalize this value to our project as well.
Shade level S is a variable that is given in the data for each of the example homes. The possibilities are "not at all shady," "not very shady," "shady," and "very shady" (the third one is implied). Each of these possibilities are assigned an integer value from 3 to 0, respectively. To prevent our model from predicting unrealistically high values, we have limited this term to a maximum influence of 4.2° F, which means our coefficient for this term is 1.4.

Green cover G is a variable that depends on the proportion of developed open spaces. We derived a formula to determine green cover based on this proportion, as we found that 20% of developed open spaces and 60% of non-developed open spaces are green^{source}. Then, we scaled this number to values between 0 and 2, as we believe that the amount of green cover near the residence can't influence the temperature more than 2° F.

The age of the residence Y is used in our model because homes with older infrastructure are less energy-efficient and can become hotter than newer homes. We have implemented a sigmoid function because lots of developments have been made in the past half-century or so. That is, a home built 100 years ago is closer in quality to a home built 300 years ago than it is to a home built today. The center of the sigmoid function is established as 1980 as many cities started transitioning to better insulation infrastructure as well as policy changes, as we stated in the assumptions.

The number of stories H affects the indoor temperature because warm air rises. The number of people per room P/R affects the temperature because it shows the population density for the residence, and more people would use the home's facilities more.

2.3.2 Model Execution

Therefore, our final model is as follows:

$$\hat{T_{in}} = T_{phys}(t-\tau) + 1.4S + G + Y + 0.5H + 0.1P/R$$

We then graphed the models for indoor temperature for each of the four example homes given versus time of day, and the graphs are below.

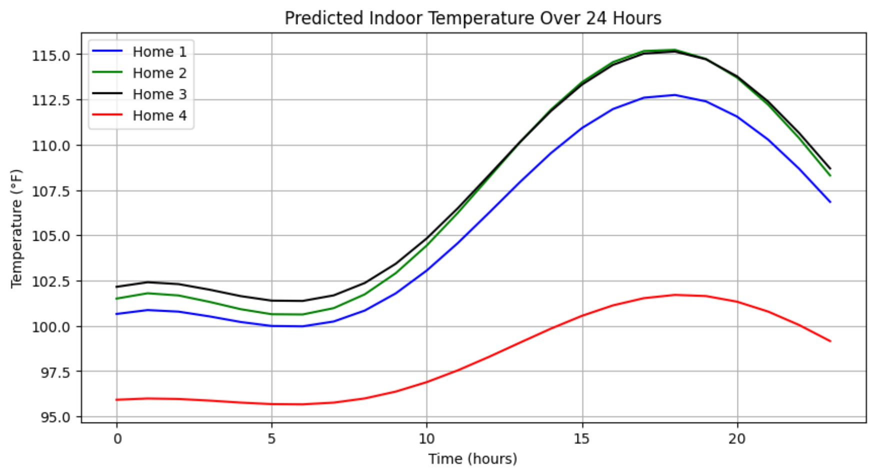


Figure 2: Four graphs for each of the four example homes described in the data

2.4 Results

We can see here that the sinusoidal graphs reach a maximum at around 4 to 5 PM, which predictably lags behind the outdoors temperature, which peaks at around 2 to 3 PM. Additionally, the peak temperature for each house ranges from 102° F to 115° F. Though these temperatures may seem extremely high, it is logical that a home without air conditioning will accumulate heat without dissipating easily, especially in a muggy, urban environment like Memphis. Additionally, when a similar experiment was run in Chicago[6], the indoor temperatures of homes reached a high of 120° F, which shows that our results are not unreasonable. We will now briefly discuss the correlations present in the above graph.

2.5 Discussion

Homes in urban environments and at high elevation are more prone to high indoor temperatures, like Home 3. Home 3 and Home 2 have similar temperatures, but Home 2 had better heat dissipation due to its lower elevation so it reached a lower temperature at night. Finally, Home 4 was by far the lowest, as it was a single-family residence in a suburban area with high square footage. These results show us that poorer areas are disproportionately affected by heatwaves, as urban areas are often poorer and such residences often have lower square footage, which contributes to rapid heat flow. These hot temperatures can be very dangorous, as high temperatures can cause heat strokes and other fatal health threats.[8]

2.6 Sensitivity Analysis

To determine the accuracy of our predictions, we randomly offset each data point by 5% for each Home and then ran the model again to get a new prediction. From there, we calculated the percentage difference between the original prediction and the new altered prediction. We repeated this process 5 times and averaged the differences. This process determined that our prediction of the indoor temperature for all homes had the following average jitter variations.

Table 4: Average Jittered Variations of Indoor Temperatures in Homes

Home 1	Home 2	Home 3	Home 4
1.93%	1.89%	1.98%	1.07%

Since all of the average jittered varations are low, we are confident that our model is resilient to random error.

2.7 Strengths & Weaknesses

2.7.1 Strengths

- Broad applicability, as the model takes into account numerous factors that can contribute to high indoor temperatures
- The sinusoidal base model fits temperature well as it varies with the time of day in this way
- Our exponential decay factor introduced through Newton's Law of Heat Transfer accounts for heat accumulation
- Our delay factor τ accounts for the lag due to the time it takes for heat to flow

2.7.2 Weaknesses

- Though we achieved plausible values, we were not able to validate our indoor temperatures, as finding public data for this particular situation is extremely difficult
- The model may not account for nonlinear factors, as we have assumed that all of our factors, barring outdoor temperature, linearly affect indoor temperature
- Other human actions like opening windows may have a negligible effect on the indoor temperature in the short term, but they may add up in the long run
- The year that the home was built could oversimplify the truth about the home's building material quality and its thermal insulation mass

3 Part II: Power Hungry

3.1 Defining the Problem

The problem asks us to develop a model that predicts the peak demand of a city grid over the next 20 years, and Memphis, Tennessee, was the city of our choice. Our model is based on the given dataset to project future trends. We turn this problem into several questions: how does our summer climate change? How is the demand for energy changing? How does the effect of population change factor into our model? We design a model that accounts for the variables referenced in these questions and more.

3.2 Assumptions

Assumption 1: There are no main policy changes regarding energy consumption.

- *We believe this because it is infeasible to account for unpredictable policy changes. We only consider current policies regarding energy consumption for the next 50 years. This includes the government's missions to go fully electric by 2060*[9]

Assumption 2: The energy used in kWh per person is the same

- *For the sake of simplicity, we assume that the energy used by every individual is the same. We have gathered this constant using research*[10]*, and will explain the implications later in this section.*

Assumption 3: The energy consumption in Memphis is proportional to the energy consumption in Shelby County.

- *Because of assumption 2, we assume that the amount of energy consumed by a single person is equal among all residents. Thus, it is reasonable to assume that the energy consumption of Memphis, a city inside Shelby County, is proportional to the energy consumption of Shelby County, with the proportionality factor being the difference in populations between Memphis and Shelby County. We get population data from online census data*[11]

Assumption 4: Heat waves will increase intensity by 0.025°F per year for 20 years

- *Due to climate change, the intensity of heat waves increases every year*[12]*. From the 1960s to the 2020s, the intensity has increased by 0.5°F. Since the rate of climate change is also increasing, we believe it will increase by 0.5°F in 40 years rather than 60 as it has before, which is 0.025°F per year.*

Assumption 5: The peak demands on the power grid in the summer months will be during a heat wave

- *Since the highest temperatures are observed during a heat wave, we believe that the peak demand on the power grid will be during a heat wave as, in our model, the higher the temperature, the more energy used.*

Assumption 6: Population trends will follow trends shown from the past 10 years.

- *Since we cannot expect nor predict any major events that would influence the population of Memphis, we model future trends from the population data from the past 10 years.*

Assumption 7: Energy efficiencies are negligible.

- *We find it infeasible to predict how energy efficiency policies may cause total energy consumption to decrease. Since we are trying to find the peak, which is the worst-case scenario, we only consider energy consumption.*

Assumption 9: Electric vehicles are the only significant increase in power consumption.

- *We model all other trends to be linear and accounted for in our regression for current trends. Because trends are hard to predict, and most of the consumer habits will be modeled in our current trends regression line, we group all of those factors together and leave EVs as the only other energy consumption contribution.*

Assumption 9: Electric Vehicles (EV) follow a logistic trend

- *The past 5 years have shown remarkable increases in the acquisition of EVs. Due to Tennessee's ambitious plan for EVs in the next 50 years, as well as potential gasoline vehicle bans in the future, we assume EVs will completely dominate the market in 50 years.*

3.3 Model

3.3.1 Model Development

To model several variables all accumulating to the same variable, we decided to use linear and logistic regression over several different factors to provide us with a total energy consumption during the summer per year. We decided to use linear regression to model population changes, climate changes, and current energy trends, as after graphing our data, we noticed that all of these variables showed linear trends. Regarding EVs and their expected contribution to power consumption, we found that a logistic curve modeled the situation best, as EV owners will eventually reach a max population around 2060, when gas vehicles are projected to be almost completely eradicated from roads[13].

Table 5: Factors and variables that influence energy consumption

Function	Regression
Population	Linear
Climate	Linear
Consumption	Linear
EV	Logistic

To combine our variables, we first discuss how we can relate each variable. We generalize all of the current climate and consumption trends to a usage per person and then multiply by the expected population of

Memphis. We also incorporate a $\frac{\Delta\text{power}}{1^\circ F}$ term. This term, gathered by research in the assumptions, argues that for every $1^\circ F$ change in temperature, the average resident of Memphis will increase their energy consumption by Δpower kWh, primarily through the use of air conditioning and other cooling technology. This gives us an expected annual energy consumption. Since we are only concerned about energy usage during the summer months, we take 30% of this total, as according to the data provided by Mathworks as well as assumption 10, summer consumption is on average 30% of the annual power consumption. We then add the total energy consumption caused by the growth in EVs, which is nonnegligible. Our general function is as follows:

$$\hat{E}(x) = \text{population(t)} \times (\frac{\Delta\text{power}}{1^\circ F}\text{climate(t)+consumption(t)}) + \text{EV(t)}$$

We will now proceed to briefly discuss the derivation of each of our sub models.

Population

This function measures the population of Memphis with respect to time. To determine this value, we analyze Memphis population data from the past 10 years according to census data from the assumptions, and run a linear regression to predict values in the near future. Our data and regression line, generated with Python Scikit, is below:

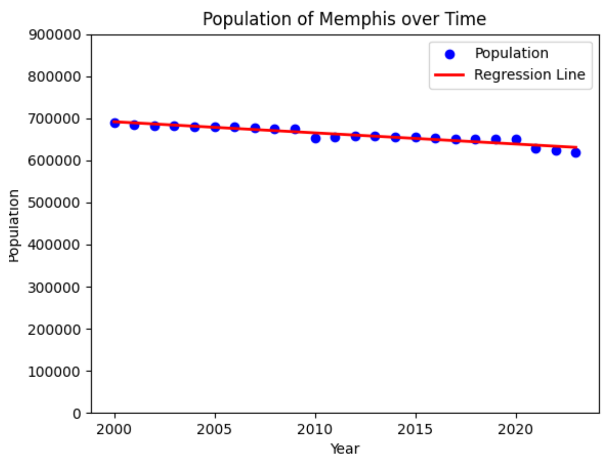


Figure 3: Population of Memphis

$$\text{population(t)} = -2645.0178t + 5981712.2322$$

Climate

We acknowledge that climate includes temperatures that increase over time. To model this, we assumed climate change would increase temperatures at a rate similar to today, approximately 0.1 degrees Fahrenheit a decade, or 0.2 degrees for the entire 20-year period. Additionally, we found that a change of 1 degree Fahrenheit in temperature results in about 3 percent more electricity consumption.

Current Energy Use

The bulk of our terms comes from the current predicted trends of energy consumption. After plotting our data given to use by mathworks[0], we see that, like our population trends, we have yet another linear pattern. We aren't surprised by this; the past decade has led to energy optimization, and without knowledge of changes in energy population or conservation technology, we expect this to continue. Below, we run the same linear regression algorithm we used for the population term. Note that this is per resident; we show the dimensional analysis later on to convert this to kWh during peak usage. Critics may argue that a linear model is inappropriate as it argues that, eventually, energy consumption per resident becomes 0. However, this would only occur in cases of clear extrapolation, as the model becomes too low very far into the future, well beyond the 20-year scope that we require.

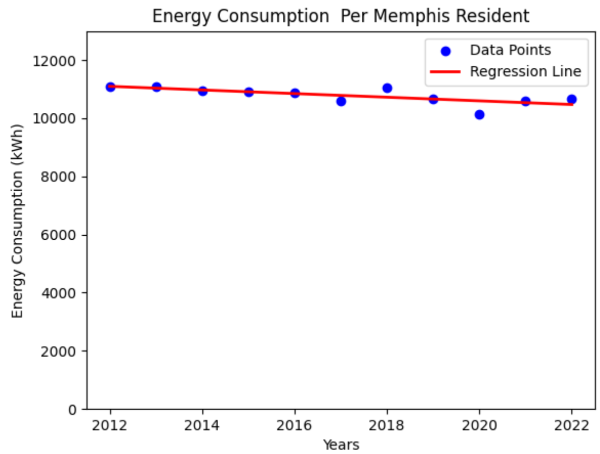
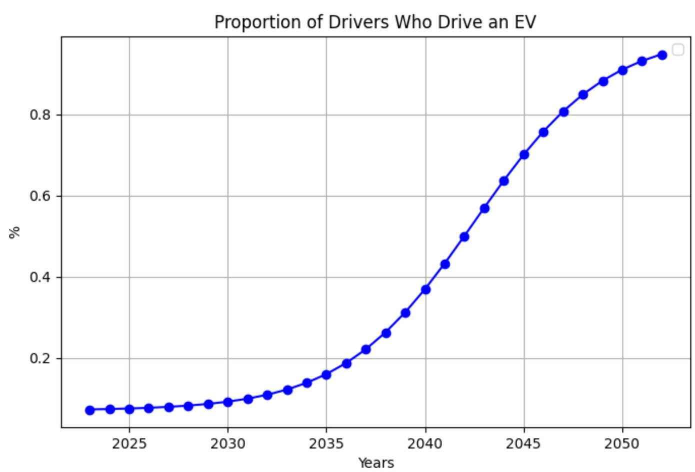


Figure 4: Current kWh per resident trends

$$\text{population(t)} = -62.5868t + 137020.8529$$

EVs
Finally, we argue that EV consumption over time will bring a significant amount of power consumption to power grids, which the city of Memphis should consider. To model this situation, we use a logistic function. We derive this function in the same way from Newton's Law of Heat Transfer in part 1, but we make use of Python's *scikit* and *scipy* libraries and the built-in logistic regression tools. This gives the following differential equation, graph, and particular solution:

$$\frac{dP}{dt} = kP(1 - \frac{P}{L})$$



$$\text{p(t)} = \frac{.93}{(1 + e^{-.298306*(t-2042.5)})} + 0.07$$

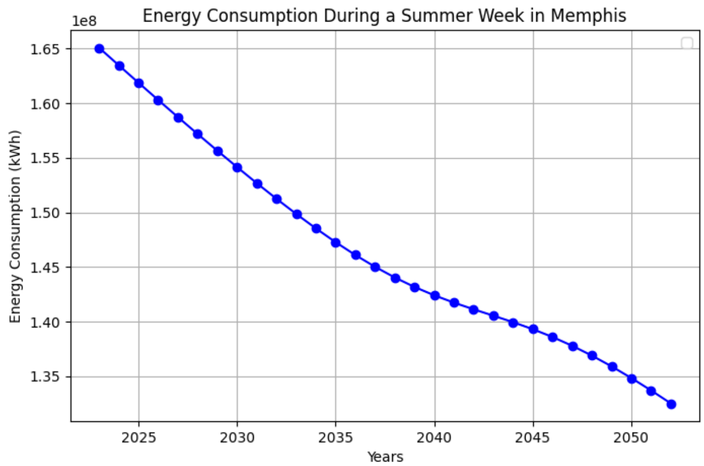
To use this proportion and convert it to energy consumption in kWh, we multiply by the number of drivers in Shelby County^{source} and multiply by the population proportion. This gives us the number of EVs in Memphis. We then multiply by the average number of miles driven^{source} and kWh per mile driven[14]. This builds the following relationship:

$$EV(t) = p(t) \times \frac{575000 \text{ Drivers in Shelby County}}{\text{population(2024)}} \times \text{population(t)} \times \frac{1123 \text{ miles}}{1} \times \frac{.35 \text{ kWh}}{1 \text{ mile}}$$

To summarize, we have determined $\hat{E}(x)$, the expected energy consumption during the summer of a year. To determine the temperature in a given week, we divide by 12 as there are 12 weeks in a summer. This gives a conservative upper bound on the predicted energy consumption during a week in the summer of Memphis.

3.3.2 Model Execution

As we build our model with Python with the form discussed in the previous section, all that is left is to run it with our input variables. We are asked to discuss the maximum demand in a given week 20 years from now. We will graph the next 30 years of predicted energy consumption:



3.4 Results

By plugging in 2045 into our model, we get 139 321 844 kWh. This means that during a hot week in the summer at Memphis, a power grid can expect to draw 139 321 844 kWh of energy. We are also asked to compute current power consumption. We can do this by plugging in 2025 into our model, which gives us 161 872 733 kWh. We can see that energy consumption during the summer months is predicted to decline.

3.5 Discussion

This lines up with our prediction as energy consumption decreases to represent both consumer behavior and population decline. We argue this because this number lines up with the trends shown by the data given in the past 10 years, as well as our expectations of how energy consumption will decline because both average energy consumption per person and the population of Memphis decrease. We also measured the energy use per person, and this number lines up with estimates from other cities as well.

3.6 Sensitivity Analysis

To determine if our model is accurate, we remove the last data point of our model, retrain our model, and examine the error from our predicted value to our real value. We use the following definition for error:

$$\text{Error} = \frac{|\text{Actual} - \text{Predicted}|}{\text{Actual}} * 100\%$$

We remove the data point for 2022. The new predicted value becomes 16 346 678. Our real data set is $.68 \times .3 \times 814\ 024\ 666$, as that is the expected value with the assumptions that energy uses are proportional. We get an error of 0.38%, showing our model is robust and not sensitive to changes.

3.7 Strengths & Weaknesses

3.7.1 Strengths

1. By using multiple linear regressions, we are consistent with trends shown in the past 20 years, and without considering major population-changing events or policies, this seems to align with predictions.

2. Since we describe our function of energy consumption as a combination of other variables, it is extremely easy to "plug and play" different factors to strengthen our model and try them in different situations.

3.7.2 Weaknesses

1. Linear regression doesn't emphasize the effect on heat trends as well.

2. This model assumes population and baseline consumption trends will be the same in the next 20-30 years. This is an assumption that ultimately makes a big difference in the model as this is what drives the model to decrease rather than increase, as some may argue.

3. We may have overestimated the effect that EVs can have on the model. The effect of EVs is just one of many different variables that contribute to the change in energy consumption.

4 Part III: Beat the Heat

4.1 Defining the Problem

The problem asks us to create vulnerability scores for the various neighborhoods of Memphis to express the vulnerability of these areas to heat waves and/or power grid failures. Then, we were to propose a method by which the city could more effectively distribute resources.

4.2 Assumptions

Assumption 1: Any failures regarding the power grid will affect all neighborhoods equally.

- *Based upon Question 2, the city has a uniform power grid, so any change to the grid should affect all neighborhoods in the city.*

Assumption 2: The humidity is the same throughout the city.

- *Humidity is a potential factor for vulnerability, but we find it unlikely to have significant changes between the neighborhoods of the city.*

Assumption 3: Other factors, such as the ability to relocate, are correlated with pre-existing factors like wealth

- *In order to prevent the model from becoming too complicated, we reduced the number of inputs by selecting the most significant factors and assuming them to be the underlying cause behind other plausible factors.*

Assumption 4: The number of hospitals per neighborhood correlates to greater access to general healthcare.

- *The more hospitals are available in a region, the easier individuals can receive medical attention, and we are assuming that the presence of other forms of healthcare is correlated to the presence of hospitals.*

4.3 Model

4.3.1 Model Development

To create a model that generates a vulnerability score for each neighborhood in Memphis, we first isolated 6 variables that have the greatest impact on potential vulnerability to heat waves or power grid shortages: Income, number of households with adults aged over 65, indoor temperature, access to health resources, working population, and households with vehicles.

Age
Age, or more specifically, the proportion of households with adults over the age of 65, was considered to be a key factor in vulnerability because older adults are more susceptible to adverse health conditions. Thus, because of this significant correlation, we assigned a weight of 0.25 to our age variable.

Income
We deemed income to be an essential variable in vulnerability[15] because those less financially well-off may struggle to pay for cooling, backup electricity, and other critical elements. Additionally, lower income has been proven to correlate with increased health complications and deaths during extreme weather conditions.[15] Due to this strong proven correlation, we assigned a weight of 0.2 to income.

Indoor Temperature
The temperature inside of houses, we determined, was the next most significant factor as the indoor temperature is what individuals within the neighborhood experience. Thus, as higher indoor temperatures greatly increase the risk to people's health, we assigned the indoor temperature variable a 0.2 as well.

Healthcare Facilities
The initial three variables dealt primarily with the risk to one's health they posed, but our fourth variable, the availability of healthcare facilities, deals with the ability a neighborhood has to provide care for health issues. An inability of a neighborhood to do so will naturally result in problems for its inhabitants, making them more vulnerable to heat waves, so we assigned this variable a weight of 0.15. We use hospital data from online research[15]

Vehicles
Heat stroke and related health issues arise more commonly amongst those who take public transportation, as public transportation is generally hotter than personal transportation. Thus, our fifth variable was the proportion of households in any given neighborhood that owned vehicles of their own, and this we assigned a weight of 0.1.

Work
Our final variable was the proportion of residents within a neighborhood that worked for themselves. Individuals who work for themselves are more likely to be able to support themselves and thus less vulnerable to heat waves. However, as this factor is not quite as significant as the others, we gave it a weight of 0.1.

This led to our final equation:

$$Vulnerability = (0.2 \times \text{Income Score}) + (0.25 \times \text{Age Score}) + (0.2 \times \text{Temp Score}) + (0.15 \times \text{Health Score}) + (0.1 \times \text{Vehicle Score}) + (0.1 \times \text{Work Score})$$

To find each individual score, first, we have to find data for the factor. Then, we have to normalize the data using min-max scaling. We chose to min-max scaling since it preserves the relationships among the original data values, reduces the variance, and magnifies the effect of outliers.
Here is the equation for the min-max scaling:

$$\text{Normalized Score} = \frac{\text{Value} - \text{Min Value}}{\text{Max Value} - \text{Min Value}} \times 100$$

To calculate the **Vulnerability Score** for each neighborhood in Memphis, Tennessee, we followed these steps:

1. **Income Score** – We used median household income data from MathWorks, applied min-max scaling, and normalized the values.
2. **Age Score** – We used the percentage of residents aged 65+ from MathWorks, found the min and max values, and applied normalization.
3. **Vehicle Score** – We used the number of households with at least one vehicle, applied min-max scaling, and normalized the values.
4. **Work Score** – We used the number of people aged 16+ who work, found the min and max values, and normalized the data.
5. **Health Score** – We calculated the number of hospitals per capita in each neighborhood and took the inverse ($1 - \frac{\text{hospitals}}{\text{population}}$) to reflect higher vulnerability for areas with fewer hospitals. We then normalized the values.
6. **Temp Score** – Using housing data from MathWorks, we computed a raw temperature score and applied min-max scaling to normalize it.

After computing all six scores, we combined them to determine the final **Vulnerability Score** for each neighborhood.

4.3.2 Model Execution

We built our model using the equation in the previous section in Python. The final step in executing the model is to run the code for each neighborhood based on the inputs for each of the 6 factors in each neighborhood to find the vulnerability scores. This graph provides the vulnerability scores for each neighborhood in Memphis:

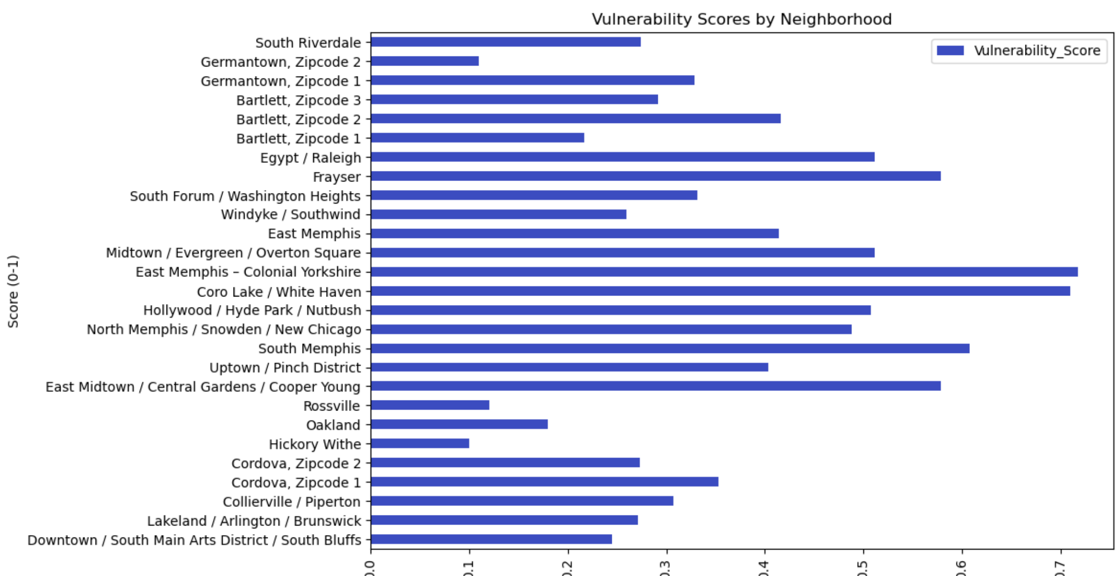


Figure 5: Graph of Vulnerability Scores for each of the neighborhoods in Memphis

To calculate the likelihood of a power grid failure during a heat wave (when there is peak usage), we found the maximum amount of electricity the MLGW power grid can generate per week. The grid has the capacity to handle 280,000,000 kilowatts of demand per week[21]. Since the peak usage that we calculated in the model for Question 2 will only be 165,000,000 kWh, which is significantly less than the capability of the power grid, the likelihood of a power grid failure is low.

4.4 Results

Based off of the vulnerability scores that we calculated in the previous section, the most vulnerable neighborhoods are East Memphis - Colonial Yorkshire and Midtown / Evergreen / Overton Square. The least vulnerable neighborhoods are Hickory Withe, Germantown, Zipcode 2, and Rossville.

4.5 Discussion

Our vulnerability scores for various neighborhoods, calculated with the 6 factors of Age, Income, Indoor Temperature, Healthcare Facilities, Availability of Personal Vehicle, and proportion of people who work, provide a useful scale of which communities are the most in danger during a heat wave or power grid outage.

Table 6: Factors and variables that influence energy consumption

actor	R-value
Median Household Income (USD)	-0.6771
Households with 1+ people 65 yrs and over	0.5271
Indoor Temperatures (Question 1's Model)	0.8237
Number of Hospitals in neighborhood	-0.2092
Population age 16+ years old who work	0.2305
Households with 1+ vehicles	0.3542

Using the correlational analysis between the vulnerability score and each of the 6 factors allows us to see which factors are most strongly correlated with how vulnerable a community is and what resources need to be. Since Indoor Temperature had the r-value with the highest magnitude of 0.8237, focusing on reducing the Indoor Temperature will be the most effective way to reduce the vulnerability of a community to heat waves.

One approach to manage heat waves in Memphis by using these vulnerability scores is to allocate more resources to the communities with the highest vulnerability score. This can be done by providing resources proportionally to each neighborhood's vulnerability score out of the sum of all vulnerability scores. Since we determined that it is very unlikely for there to be a power grid outage, even at peak demand during a heat wave, allocating 10% of the total monetary resources into generators that can provide electricity is sufficient.

The remaining 90% of resources should primarily be used to address the issue of overly high indoor temperatures. We recommend that 20% of resources be invested into community cooling centers, especially in the most vulnerable areas with elderly and low-income populations. 20% of resources should also be used to plant more trees, and vegetation should also be planted to reduce the temperature. Additionally, better building materials, or rather less insulatory ones, should be encouraged to allow houses to avoid building up heat. Finally, cool roofs, roofs made with materials that absorb less sunlight, should be installed on homes with poorer insulation to reduce indoor temperature. For these building materials and cool roofs, we recommend using about 30% of the total monetary resources. The remaining 20% of resources should be used to address any other factors that may be deemed important to reducing the negative impact of heat waves.

4.6 Sensitivity Analysis

To determine if our model is accurate, we decided to go with the Monte Carlo simulation. We chose it because it models real-world uncertainty by randomly varying weights within a realistic range, providing a more robust analysis of how different factors influence vulnerability. It also generates a distribution of outcomes, allowing us to assess sensitivity through standard deviation and identify which neighborhoods have the most stable or fluctuating rankings. We use this over other methods since it considers multiple factors changing at once, capturing complex interactions and providing a full range of possible outcomes rather than a limited set of fixed scenarios. After computing the Monte Carlo Simulation with 1000 iterations, we got our values for the vulnerability mean and standard deviation for each neighborhood, and here is the data on a plot.

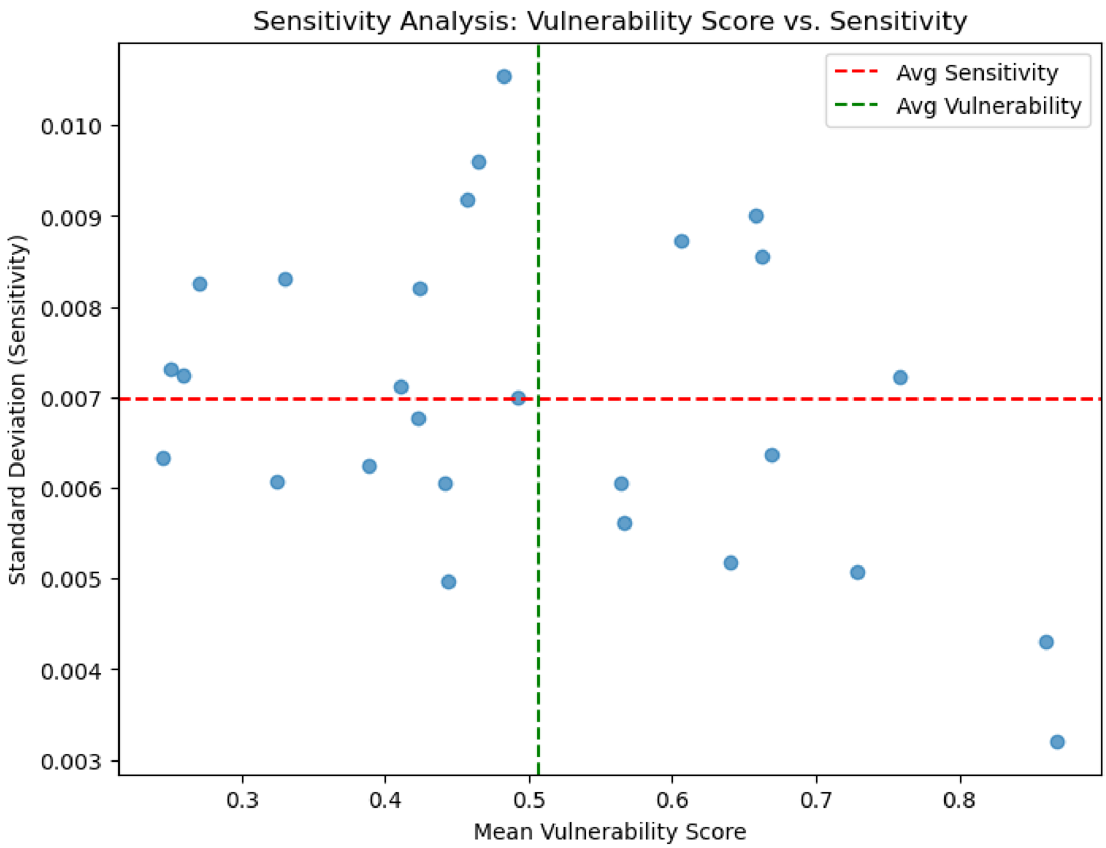


Figure 6: Sensitivity Analysis: Vulnerability Score vs. Sensitivity for each neighborhood in Memphis

Since our average vulnerability and average sensitivity is very low, we are confident that our model is resilient to random error.

4.7 Strengths & Weaknesses

4.7.1 Strengths

1. The correlational analysis between the individual factors and the vulnerability score allows us to be able to easily find what factors are likely to have more impact on vulnerability.
2. Additional factors can be added relatively easily as they just need to be assigned a weight-age while simply being added onto the existing equation.

4.7.2 Weaknesses

1. The amount that each factor is weighted in the model may be somewhat inaccurate.
2. The amount of correlation in a factor does not guarantee that the factor has impact on vulnerability, so the effect of factors such as indoor temperature and income may be overemphasized.
3. This data isn't easy communicable to the general population. We considered using a heatmap to represent how different locations can get affected differently, but due to time constraints we were unable to implement this.

5 Conclusion

5.1 Summarizing Results

From our work today, we have thoroughly examined the effects of numerous important factors on heat and energy consumption in the city of Memphis, Tennessee. Our work has demonstrated that heat and energy consumption are crucial issues that must be further studied. From our study of Problem 1, we used a sinusoidal regression model as well as an exponential decay factor derived from Newton's Law of Heat Transfer to address the relationship between outdoor temperatures and indoor temperatures. In addition to that, we have added on other factors such as home size, green cover, shade level, number of stories, and the number of people per room. Problem 1 showed us that without proper air conditioning, powered by electricity, indoor temperatures can skyrocket to staggering amounts that can induce heat stroke in unsuspecting residents.

From our work in Problem 2, we used a combination of linear regression models and logistic regression models to find the relationship between energy consumption and important factors like population, EV adoption, and more. Problem 2 showed us that energy consumption trends are on the decline due to a decline in the population of the city, demonstrating that it would be more efficient to invest government spending into other areas to bridge the gap between housing and income inequality.

Finally, from our work in Problem 3, we have designed a way to derive the vulnerability score of a neighborhood using linear regression due to factors such as proportion of people who work for themselves, the proportion of senior citizens, access to health resources, number of vehicles, and more. We used correlational analysis to find which factor impacted the vulnerability score the most, and we determined that it was indoor temperature based on our model from question 1.

We also devised a potential solution to reduce vulnerability of neighborhoods in Memphis, and we hope that this can bring attention to the clear impact that these factors have on the safety and well-being of our citizens.

6 References

1. Hot Button Issue, MathWorks Math Modeling Challenge 2025, curated data, https://m3challenge.siam.org/897bjhb54cgfc/.
2. https://www.noaa.gov/news/2024-was-worlds-warmest-year-on-record
3. https://news.arizona.edu/employee-news/ new-conversation-why-older-homes-feel-warmer-thermostat-suggests-and-what-do-about-it
4. https://www.aceee.org/files/proceedings/1988/data/papers/1988V10017.pdf
5. https://www.sciencedirect.com/science/article/pii/S2352484724002452
6. https://www.npr.org/2019/09/03/754044732/as-rising-heat-bakes-u-s-cities-the-poor-often-feel-it-most
7. https://snuggpro.com/blog/item/many-homes-built-prior-to-1980-were-built-without-insulation-in-the-walls
8. https://www.weather.gov/lmk/humidity
9. https://www.elevatenp.org/wp-content/uploads/Summer-Indoor-Thermal-Conditions-1.pdf
10. https://pmc.ncbi.nlm.nih.gov/articles/PMC7413138/sec16
11. https://www.iea.org/reports/global-ev-outlook-2023/prospects-for-electric-vehicle-deployment
12. https://www.energyhub.com/news/how-much-is-one-degree-worth
13. https://datacommons.org/exploreq=population
14. https://apnews.com/article/heat-wave-climate-change-worsen-hotter-797aae046df8165f5f8be7d3f40a8b74
15. https://www.sciencedirect.com/science/article/pii/S0921344921002317
16. https://www.nstlaw.com/memphis/car-accident-lawyers/shelby-county-car-accident-statistics/: :text=Wit
17. https://www.fhwa.dot.gov/ohim/onh00/bar8.htm
18. https://www.edmunds.com/electric-car/articles/how-much-electricity-does-an-ev-use.html
19. https://www.sciencedirect.com/science/article/abs/pii/S0959378021002259
20. https://tha.com/hospitals/
21. https://www.fox13memphis.com/news/tva-asks-customers-to-voluntarily-reduce-power/article498baa3a-d831-11ef-932f-e78608ac8d1a.html

7 Code Appendix

Problem 1:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit

# Time data (Hours)
time_hours = np.arange(24)

temp_outdoor = np.array([
    85, 85, 84, 83, 83, 83, 84, 88, 91, 94, 96, 97,
    100, 100, 102, 102, 100, 99, 97, 94, 92, 91, 90, 89
])

# Define sinusoidal function with fixed frequency
def sinusoidal_model_fixed(x, A, C, D):
    B = np.pi / 12 # Fixed frequency corresponding to a 24-hour period
    return A * np.sin(B * x + C) + D

# Fit the model to data (only A, C, D are fitted)
def fit_sinusoidal_fixed(x, y):
    params, _ = curve_fit(sinusoidal_model_fixed, x, y, p0=[10, 0, 75])
    return params

A, C, D = fit_sinusoidal_fixed(time_hours, temp_outdoor)
B = np.pi / 12 # fixed frequency

plt.figure(figsize=(10, 5))
plt.plot(time_hours, sinusoidal_model_fixed(time_hours, A, C, D), label="Fitted Sinusoidal Model",
     color="red")
plt.scatter(time_hours, temp_outdoor, label="Outdoor Temperature", linestyle="dashed",
     color="black")
plt.xlabel("Time (hours)")
plt.ylabel("Temperature (F)")
plt.title("Predicted Outdoor Temperature Over 24 Hours")
plt.legend()
plt.grid()
plt.show()

np.save("sinusoidal_params.npy", [A, B, C, D])
print(f"Model parameters: A={A:.2f}, B={B:.4f}, C={C:.2f}, D={D:.2f}")


tau = 2.5
Aout = -9.07 # amplitude
b = 0.2618
c = 0.88
d = 92.04
T0 = 85 # this is the initial value that we were given in the data
k = 0.07

def Tout(t):
  return Aout*np.sin(b*t + c) + d

def Tphys(t, S):
  return Tout(t - tau) + (T0 - Tout(t - tau)) * np.exp(-k*t / S)

```

```python
def Tchange(greencover):
  return 2*(0.7857 - 0.2*greencover + 0.6*(1-greencover))/(0.7857 - 0.5103)

def age(year):
  return 5 / (1 + np.exp(year - 1980))

def Tin(t, S, shadelevel, greencover, year, stories, people, rooms):
  if shadelevel.lower() == "not at all shady":
    shadelevel = 0
  elif shadelevel.lower() == "not very shady":
    shadelevel = 1
  elif shadelevel.lower() == "shady":
    shadelevel = 2
  elif shadelevel.lower() == "very shady":
    shadelevel = 3
  else:
    raise ValueError

  return Tphys(t, S) + 1.4 * shadelevel + Tchange(greencover) + age(year) + 0.5 * stories + 0.1 * people / rooms

plt.figure(figsize=(10, 5))
plt.plot(time_hours, Tin(time_hours, .88, "shady", .4829, 1953, 1, 3, 3), label="Home 1", color="blue")
plt.plot(time_hours, Tin(time_hours, .63, "not very shady", .1047, 1967, 1, 3, 2), label="Home 2", color="green")
plt.plot(time_hours, Tin(time_hours, .74, "not at all shady", .1047, 2003, 15, 2, 1), label="Home 3", color="black")
plt.plot(time_hours, Tin(time_hours, 2.78, "not at all shady", .0451, 1990, 2, 6, 5), label="Home 4", color="red")
plt.xlabel("Time (hours)")
plt.ylabel("Temperature (F)")
plt.title("Predicted Indoor Temperature Over 24 Hours")
plt.legend()
plt.grid()
plt.show()

# Sensitivity Analysis

houses = {
    "Home 1": {"S": 0.88, "shadelevel": "shady",    "greencover": 0.4829, "year": 1953, "stories": 1, "people": 3, "rooms": 3},
    "Home 2": {"S": 0.63, "shadelevel": "not very shady", "greencover": 0.1047, "year": 1967, "stories": 1, "people": 3, "rooms": 2},
    "Home 3": {"S": 0.74, "shadelevel": "not at all shady", "greencover": 0.1047, "year": 2003, "stories": 15, "people": 2, "rooms": 1},
    "Home 4": {"S": 2.78, "shadelevel": "not at all shady", "greencover": 0.0451, "year": 1990, "stories": 2, "people": 6, "rooms": 5}
}

num_iterations = 5 # Number of jitter iterations

print("Average percent difference due to a 5% random jitter for each house:")
for house_name, params in houses.items():
    # Compute the original prediction for the house
    original_prediction = Tin(time_hours, params["S"], params["shadelevel"],
        params["greencover"],
        params["year"],
        params["stories"],
```

```
        params["people"],
        params["rooms"])

    percent_differences = []

    for _ in range(num_iterations):
        # Jitter each numeric parameter by a random factor between 0.95 and 1.05
        jittered_params = {
            "S": params["S"] * np.random.uniform(0.95, 1.05),
            "greencover": params["greencover"] * np.random.uniform(0.95, 1.05),
            "year": params["year"] * np.random.uniform(0.95, 1.05),
            "stories": params["stories"] * np.random.uniform(0.95, 1.05),
            "people": params["people"] * np.random.uniform(0.95, 1.05),
            "rooms": params["rooms"] * np.random.uniform(0.95, 1.05),
            "shadelevel": params["shadelevel"] # remains unchanged
        }

        # Compute jittered prediction
        jittered_prediction = Tin(time_hours, jittered_params["S"], jittered_params["shadelevel"],
        jittered_params["greencover"],
        jittered_params["year"],
        jittered_params["stories"],
        jittered_params["people"],
        jittered_params["rooms"])

        # Calculate the percent difference (averaged over time points)
        perc_diff = np.mean(np.abs(jittered_prediction - original_prediction) /
            np.abs(original_prediction) * 100)
        percent_differences.append(perc_diff)

    avg_percent_difference = np.mean(percent_differences)
    print(f"{house_name}: {avg_percent_difference:.2f}%")
```

Problem 2:

```
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

#determined from data from mathworks
use_per_person = [11097.340449586547, 11073.718272065424, 10941.574701014348, 10925.040055373136,
    10867.72300030933, 10600.545657186545, 11065.241307350008, 10663.471039849586,
    10123.375254923389, 10609.725421769803, 10647.66812003078]

years = range(2012, 2023)

#input data into np arrays for sk
x = np.array(years).reshape((-1, 1))
y = np.array(use_per_person)

model = LinearRegression()
model.fit(x, y)

#lin regression A and B terms
intercept = model.intercept_
slope = model.coef_

print(f"intercept: {intercept}")
print(f"slope: {slope}")

#regression values
y_pred = model.predict(x)
print(f"predicted response:\n{y_pred}")

#Plot data
plt.scatter(x, y, color='blue', label='Data Points')
plt.plot(x, y_pred, color='red', linewidth=2, label='Regression Line')
plt.xlabel('Years')
plt.ylabel('Energy Consumption (kWh)')
plt.title('Energy Consumption Per Memphis Resident')
plt.ylim(0, 13000)

plt.legend()
plt.show()
```

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

#methods that are each of the helper functions we use
def EV(t):
    temp = (.93)/(1+np.exp(-.298306*(t-2042.5))) + .07
    return temp*.68*575000*.35*1123

def curr(t):
    return 137020.8529 - 62.5868 * t

def population(t):
    return 5981712.2322 - 2645.0178*t

def temperature_change(t):
    return (.016+.0125)*(t-2023)*117.98
```

```

def p2calc(t):
    return .3*population(t)*(temperature_change(t)+curr(t)) + EV(t)

#lists used for debugging purposese
next_20_years = []
years = []
currvals = []
tempchange = []
populations = []
evvals = []

for i in range(2023, 2023+30):
    years.append(i)
    currvals.append(curr(i))
    tempchange.append(temperature_change(i))
    populations.append(population(i))
    evvals.append(EV(i))
    next_20_years.append(p2calc(i)/12)


#plotting the code
plt.figure(figsize=(8, 5))
plt.plot(years, currvals, marker='o', linestyle='-', color='b')

plt.xlabel("Years")
plt.ylabel("Energy Consumption (kWh)")
plt.title("Energy Consumption During Summer Months")
plt.legend()
plt.grid(True)

plt.show()
```

Problem 3:

```
import pandas as pd

# Income, age, vehicle access spreadsheet
df1 = pd.read_csv("Untitled spreadsheet - Sheet1 (1).csv")

# Work-related data spreadsheet
df2 = pd.read_csv("Untitled spreadsheet - Sheet1.csv")

# Convert relevant columns in df2 (work data) to numeric
numeric_cols2 = [
    "Population age 16+ years old who work",
    "Primary mode of transportation to work (persons aged 16 years+): driving",
    "Primary mode of transportation to work (persons aged 16 years+): walking or public transit",
    "Primary mode of transportation to work (persons aged 16 years+): other and work from home"
]

for col in numeric_cols2:
    df2[col] = pd.to_numeric(df2[col], errors='coerce')

df = pd.merge(df1, df2, on=["Neighborhood", "ZIP code"], how="inner")

def normalize(data, min_val, max_val):
    """Normalize data using min-max scaling."""
    return (data - min_val) / (max_val - min_val) if max_val > min_val else 0

# Define min-max values for normalization
min_max_values['work'] = (
    df["Primary mode of transportation to work (persons aged 16 years+): walking or public transit"].min(),
    df["Primary mode of transportation to work (persons aged 16 years+): walking or public transit"].max()
)

# Apply normalization to Work Score (higher reliance on public transit = higher score)
df["Work_Score"] = df["Primary mode of transportation to work (persons aged 16 years+): walking or public transit"].apply(
    lambda x: normalize(x, *min_max_values['work'])
)
df["Income_Score"] = df["Median household income (in US dollars)"].apply(lambda x: 1 - normalize(x, *min_max_values['income'])) # Inverted (lower income = higher vulnerability)
df["Vehicle_Score"] = df["Households with no vehicles"].apply(lambda x: normalize(x, *min_max_values['vehicle']))
df["Age_Score"] = df["Households with one or more people 65 years and over"].apply(lambda x: normalize(x, *min_max_values['age']))

# Housing data spreadsheet
housing_df = pd.read_csv("queasion3 - Sheet1.csv")

# Calculate total number of homes by summing over construction periods
housing_df['TotalHomes'] = (housing_df["Homes built 2010 or later"] +
                            housing_df["Homes built 1990 to 2009"] +
                            housing_df["Homes built 1970 to 1989"] +
                            housing_df["Homes built 1950 to 1969"] +
                            housing_df["Homes built 1950 or earlier"])

# Compute weighted average construction year.
# Assume approximate mid-year values for each period:
```

```
# 2010 or later: 2015, 1990-2009: 2000, 1970-1989: 1980, 1950-1969: 1955, 1950 or earlier: 1940
housing_df['WeightedMidYear'] = (
    housing_df["Homes built 2010 or later"] * 2015 +
    housing_df["Homes built 1990 to 2009"] * 2000 +
    housing_df["Homes built 1970 to 1989"] * 1980 +
    housing_df["Homes built 1950 to 1969"] * 1955 +
    housing_df["Homes built 1950 or earlier"] * 1940
) / housing_df['TotalHomes']

# Compute building age (assuming current year is 2025)
housing_df['BuildingAge'] = 2025 - housing_df['WeightedMidYear']

# Compute average number of stories.
# Assume weights: Detached whole house = 1, Townhouse = 3, Apartments = 5, Mobile Homes/Other = 1.
housing_df['TotalHousingTypes'] = (housing_df["Detached whole house"] +
                                   housing_df["Townhouse"] +
                                   housing_df["Apartments"] +
                                   housing_df["Mobile Homes/Other"])
housing_df['AvgStories'] = (
    housing_df["Detached whole house"] * 1 +
    housing_df["Townhouse"] * 3 +
    housing_df["Apartments"] * 5 +
    housing_df["Mobile Homes/Other"] * 1
) / housing_df['TotalHousingTypes']

# Estimate greencover as the fraction of homes built 2010 or later (as a proxy for modern, green
    building practices)
housing_df['Greencover'] = housing_df["Homes built 2010 or later"] / housing_df['TotalHomes']

# For S (insulation quality parameter in Tphys), we use the BuildingAge.
# (Assumption: older buildings have lower S, causing higher temperatures.)
housing_df['S'] = housing_df['BuildingAge']

# Compute T_phys for each neighborhood using t and S.
housing_df['Tphys'] = housing_df.apply(lambda row: Tphys(t, row['S']), axis=1)

# Now compute the Temp Score using the provided formula:
# TempScore = Tphys + 1.4*shadelevel + T_change(greencover) + BuildingAge + 0.5*AvgStories +
    0.1*(people/rooms)
housing_df['TempScore'] = (
    housing_df['Tphys'] +
    1.4 * shadelevel +
    T_change(housing_df['Greencover']) +
    housing_df['BuildingAge'] +
    0.5 * housing_df['AvgStories'] +
    0.1 * people_per_room
)

data = {
    "Neighborhood": [
        "Downtown / South Main Arts District / South Bluffs",
        "Lakeland / Arlington / Brunswick",
        "Collierville / Piperton",
        "Cordova, Zipcode 1",
        "Cordova, Zipcode 2",
        "Hickory Withe",
        "Oakland",
        "Rossville",
        "East Midtown / Central Gardens / Cooper Young",
```

```
        "Uptown / Pinch District",
        "South Memphis",
        "North Memphis / Snowden / New Chicago",
        "Hollywood / Hyde Park / Nutbush",
        "Coro Lake / White Haven",
        "East Memphis   Colonial Yorkshire",
        "Midtown / Evergreen / Overton Square",
        "East Memphis",
        "Windyke / Southwind",
        "South Forum / Washington Heights",
        "Frayser",
        "Egypt / Raleigh",
        "Bartlett, Zipcode 1",
        "Bartlett, Zipcode 2",
        "Bartlett, Zipcode 3",
        "Germantown, Zipcode 1",
        "Germantown, Zipcode 2",
        "South Riverdale"
    ],
    # Actual population values for each neighborhood.
    "Population": [11816, 43688, 56225, 44274, 37996, 7699, 12360, 3706, 22121, 4961, 21702,
        14001, 18425, 43639, 42061, 15107, 26258, 42732, 5456, 39404, 43701, 20900, 38849, 30284,
        25171, 16298, 23768],
    # Actual hospital counts for each neighborhood.
    "Hospital_Count": [4,1,0,1,0,0,0,0,3,1,0,1,0,0,0,0,0,2,0,0,2,2,0,0,2,0,0]
}

df3 = pd.DataFrame(data)

# Calculate the hospital ratio: number of hospitals per capita.
df3["Hospital_Ratio"] = df3["Hospital_Count"] / df3["Population"]

# Since higher hospital availability lowers vulnerability, we define a raw health score as the
    inverse.
#   Raw_Health_Score = 1 - Hospital_Ratio
# (Neighborhoods with a lower hospital-to-population ratio will have a higher raw health score.)
df3["Raw_Health_Score"] = 1 - df3["Hospital_Ratio"]

# Define a normalization function to scale values between 0 and 1.
def normalize(series):
    return (series - series.min()) / (series.max() - series.min())

# Normalize the Raw Health Score to get the final Health Score.
df3["Health_Score"] = normalize(df3["Raw_Health_Score"])


# Add everything up to get the Vulnerability Score
df["Vulnerability_Score"] = (
    0.2 * df["Income_Score"] +
    0.25 * df["Age_Score"] +
    0.2 * df["Temp_Score"] +
    0.15 * df3["Health_Score"] +
    0.1 * df["Vehicle_Score"] +
    0.1 * df["Work_Score"]
)


# Sensitivity Analysis Using Monte Carlo Simulation

```

```
# Merge two dataframs into df
df = df.merge(df3[["Neighborhood", "Health_Score"]], on="Neighborhood", how="left")

# Define baseline weights (same order as the columns)
baseline_weights = np.array([0.2, 0.25, 0.2, 0.15, 0.1, 0.1])
weight_names = ["Income_Score", "Age_Score", "Temp_Score", "Health_Score", "Vehicle_Score",
    "Work_Score"]

# Number of Monte Carlo iterations
n_iterations = 1000
vuln_scores_all = np.zeros((n_iterations, len(df)))

# Variability range (10%)
delta = 0.1

# Monte Carlo simulation
for i in range(n_iterations):
    # Generate random weights with 10% variation
    random_weights = np.array([
        np.random.uniform(b * (1 - delta), b * (1 + delta))
        for b in baseline_weights
    ])
    # Normalize to sum to 1
    random_weights /= random_weights.sum()

    # Compute vulnerability score for each neighborhood
    vuln_score_iter = np.zeros(len(df))
    for j, col in enumerate(weight_names):
        vuln_score_iter += random_weights[j] * df[col]

    # Store results
    vuln_scores_all[i, :] = vuln_score_iter

# Compute mean and standard deviation of vulnerability scores
df["Vuln_MC_Mean"] = vuln_scores_all.mean(axis=0)
df["Vuln_MC_Std"] = vuln_scores_all.std(axis=0)

# Display results
print(df[["Neighborhood", "Vuln_MC_Mean", "Vuln_MC_Std"]])
```