Hot Button Issue: Staying Cool as the World Heats Up

Executive Summary

Dear Memphis City Officials,

In recent years, the impact of extreme heat events has become a growing concern for cities across the country. As global temperatures rise, heat waves are becoming more frequent and prolonged, placing increasing strain on electrical grids and leading to widespread power outages. These outages can have devastating consequences, particularly for vulnerable populations who face heightened risks of heat-related illnesses and fatalities. Without access to cooling and essential services, these communities are disproportionately affected, making it critical to understand and combat the factors that contribute to heat vulnerability. By identifying the neighborhoods most at risk, we hope to offer a way for Memphis to take proactive measures to allocate resources efficiently, expand access to cooling solutions, and protect the well-being of its residents.

We first started by demonstrating the dangerously high, yet prolonged temperatures unconditioned dwellings reach during heat waves. Using a season of high-temperature data, our group developed a multi-linear regression model that predicts the indoor temperature, accounting for various factors including shade, humidity, and time of day. Using the July 8, 2022 heat wave, we showed that unconditioned dwellings in Memphis could reach temperatures of greater than 90 degrees Fahrenheit for several continuous hours. Research has shown an association between high indoor temperatures and adverse effects.[?] Therefore, our group has shown that heat waves pose serious health risks to many in Memphis.

Our second section consists of a semi-parametric model aiming to forecast Memphis' summer peak power demand from 2020 to 2040, extending a historical trend analysis from 2000 onward. Utilizing a SARIMAX model with exogenous variables, we project a continued rise in peak demand, driven primarily by demographic and economic factors. The accompanying graphs produced illustrate this projected increase, with a widening confidence interval highlighting the inherent uncertainty in long-term forecasting; the model also had a 2.43% MAPE for sensitivity data, indicating accuracy in fitting historical patterns. As such, Memphis city officials should anticipate steadily increasing energy demands over the next two decades and begin proactively planning for capacity expansions and energy efficiency initiatives.

The third section of this report focuses on developing a vulnerability score for neighborhoods in Memphis, Tennessee, to assess their susceptibility to extreme heat events during power outages. By analyzing key socioeconomic, demographic, and infrastructural factors—including income levels, age distribution, vehicle ownership, housing types, and green space availability—we identified which areas are most at risk: 38103, 38104, 38105, 38106, 38141. The vulnerability score was derived from a weighted model that quantifies the impact of these variables, allowing city officials to prioritize cooling interventions effectively. Our findings indicate that low-income neighborhoods, areas with high elderly or child populations, and regions with limited transportation access or green space face the greatest risks. To mitigate these dangers, we propose the deployment of mobile cooling pods in high-risk locations, ensuring accessible relief for vulnerable residents.

Contents

1 Part 1: Hot To Go

1.1 Restatement of the Problem

In this problem, we are tasked with predicting the indoor temperature, hour by hour, of a non-air-conditioned dwelling during a heat wave over a single day in either Memphis or Birmingham. Our group chose to develop a model for Memphis. This model must then be tested against a set of sample dwellings and heat wave data for our chosen city.

1.2 Assumptions and Justifications

1.2.1 The only source of heat entering the building is from the external environment.

Justification: Since the dwelling lacks air conditioning, the indoor temperature will be primarily influenced by outdoor temperature fluctuations, radiation, and conduction through walls and roofs.[1]

1.2.2 Heat accumulation follows a predictable pattern based on energy balance equations.

Justification: The model assumes that heat gained from the environment and lost through ventilation and radiation can be represented mathematically. [1][2][4]

1.2.3 Same temperature throughout the house.

Justification: Research using different sensors throughout a house has shown only minor temperature differences, allowing us to reasonably assume uniform temperature distribution.[7]

1.2.4 Insulation renovations are up to date.

Justification: Insulation should be renovated every 15-20 years to maintain effectiveness. In 2003, fiberglass insulation was most likely used, which has a 15-year lifespan on average, with a possible 20-30 year span. From the 1970s to the 1990s, cellulose insulation was used, lasting 20-30 years under the right circumstances. In the 1950s, rock wool was common, which can last 30-100 years with correct installation. Since it has been 72 years since the construction of a 1953 house, we assume that the insulation of every sample dwelling is modern and up-to-date, and therefore constant between all sample dwellings. [8]

1.3 The Model

1.3.1 Parameters

Symbol	Definition	Units
T_{indoor}	Temperature indoors	Fahrenheit
$T_{outdoor}$	Temperature outdoors	Fahrenheit
t	Time	Hours
U	UV index	MilliWatts per square meter
S	Shade (0-1)	...
H	Humidity	%

Table 1: Parameters for Housing Model

1.3.2 Developing the Model

To determine the indoor temperature of any given dwelling we must consider these 4 variables: time of day, previous indoor temperature (one hour prior), current outdoor temperature, and humidity. Previous indoor temperature is relevant as it offers a starting point for the model to make its prediction. Especially with the insulation present in almost all modern buildings, the indoor temperature changes even more gradually, making it even more effective at predicting the indoor temperature.[9][5]

Outdoor temperature is important as it directly affects the indoors. As the outside temperature rises or drops, heat will naturally flow to and from the dwelling.[10] Humidity is crucial as higher humidity leads to slower heat transfer. The humidity is used to help set a heat transfer multiplier, which is then multiplied with the difference between the outdoor temperature and indoor temperature. This product is then passed onto the model along with the raw indoor temperature.[3]

Using a random sample of the UV index from Memphis, our group obtained a corresponding UV index for each hour of the day. This UV index is multiplied with a shade value and passed onto the model to emulate the heat that direct sunlight may disperse into a dwelling. [11][6]

Our model is a simple multi-linear regression trained and fitted on data obtained from Central Europe. However, this data set is composed of measurements taken throughout an entire year. Our model's focus on indoor temperature during heat waves led our group to train using only the data collected during the summer which showed an outdoor temperature of greater than 80 degrees Fahrenheit.[12]

1.4 Results

As our training data did not include 2022, we were able to use it to test our model's sensitivity. Plugging in the given and predicted values in the percent error formula yielded sensitivity.

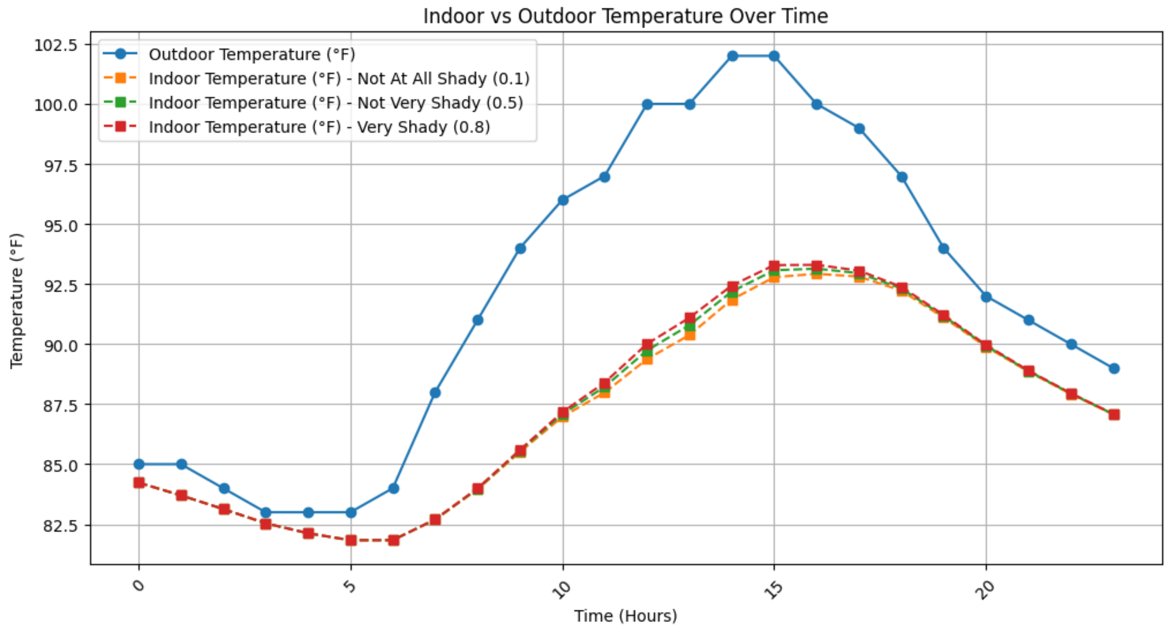

Figure 1: Predicted Indoor Temperature

1.5 Evaluating the Model

1.5.1 Validation

We conducted a sensitivity analysis to assess the impact of outside temperature, humidity, shade, and the hour of the day on the predicted internal temperature. By varying one variable at a time and keeping others fixed, we can find out which have the most profound impact on the predictions of the model. This way, we understand how the model behaves and identify potential weaknesses, as well as the agreement of predictions with expectations in reality. Moreover, sensitivity analysis indicates how small changes in the state of the environment affect indoor climate control and help make better decisions regarding temperature regulating strategies.

1.5.2 Strengths and Weaknesses

Our model is particularly strong at predicting the temperature during the beginning of heat waves. This is because much of the data the model is fit to is below 90 degrees Fahrenheit, so it would be better at predicting temperatures below 90 degrees Fahrenheit.

The model cannot account for long-term factors like climate change. Despite climate change having a noticeable effect on temperatures around the Earth even across years, the single season-long span of data our model used leads to an ignorance of the effects of climate change.[13]

Also, the data our model is fit too did not provide any shade property, quantitative or qualitative. However, it did indicate the room that each sensor was in, allowing our group to infer different amounts of shade through the average number of windows each room may have. This, however, is not a perfect parallel to the shade implied by the question.

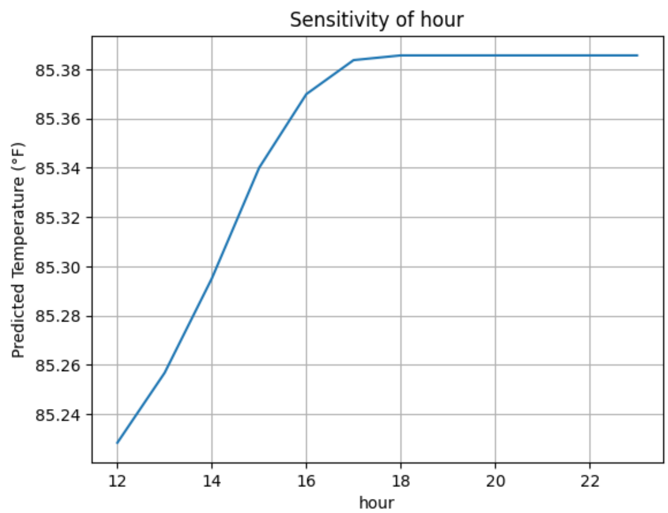

Figure 2: Sensitivity Analysis (hour): 0.18%

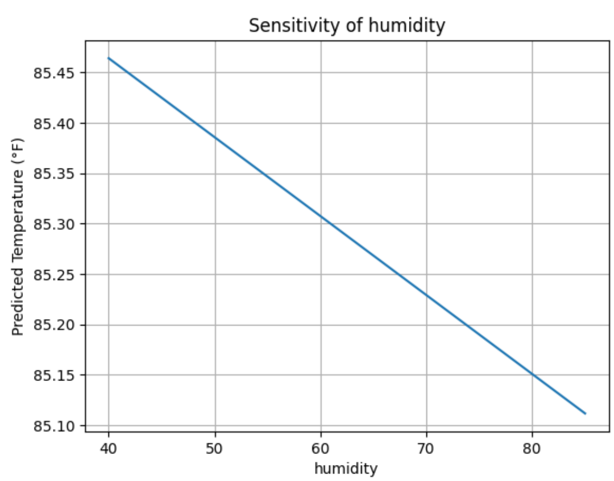

Figure 3: Sensitivity Analysis (humidity): 0.41%

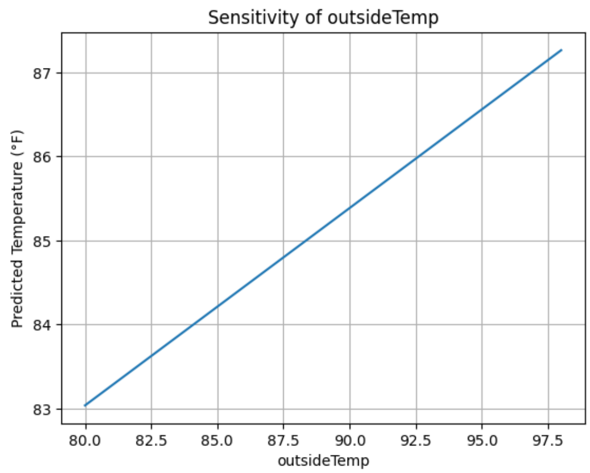

Figure 4: Sensitivity Analysis (outsideTemp): 5.09%

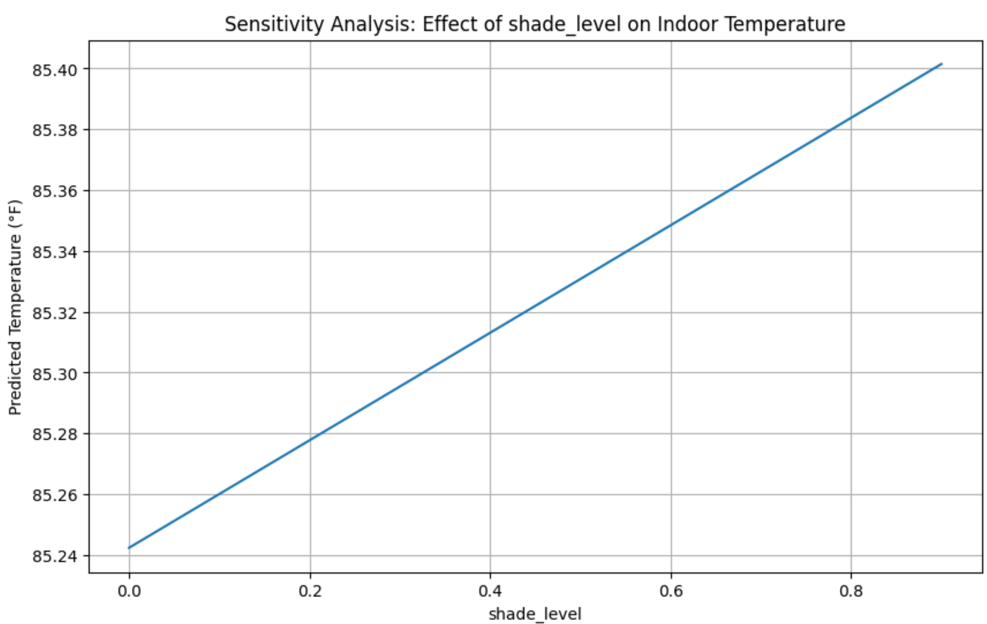

Figure 5: Sensitivity Analysis (shade): 0.19%

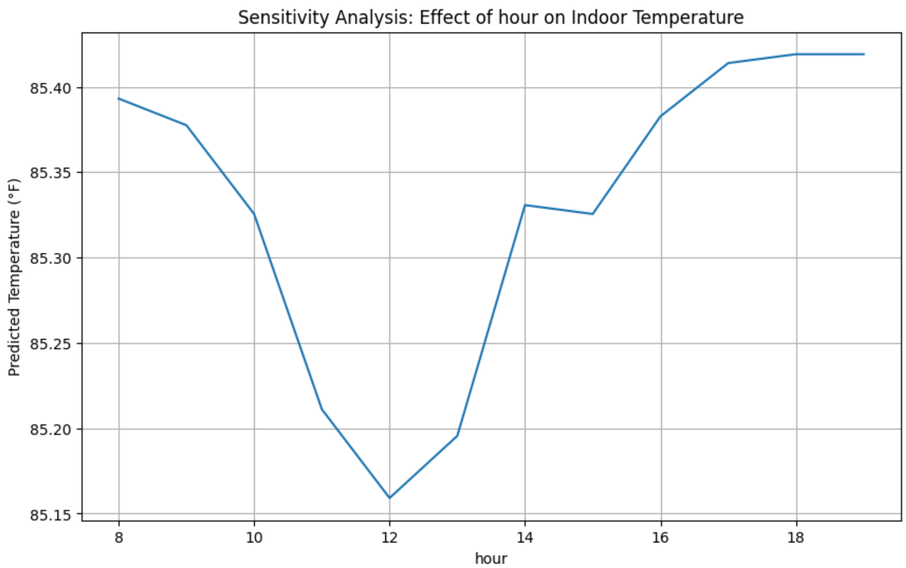

Figure 6: Sensitivity Analysis (insideTemp): 0.31%

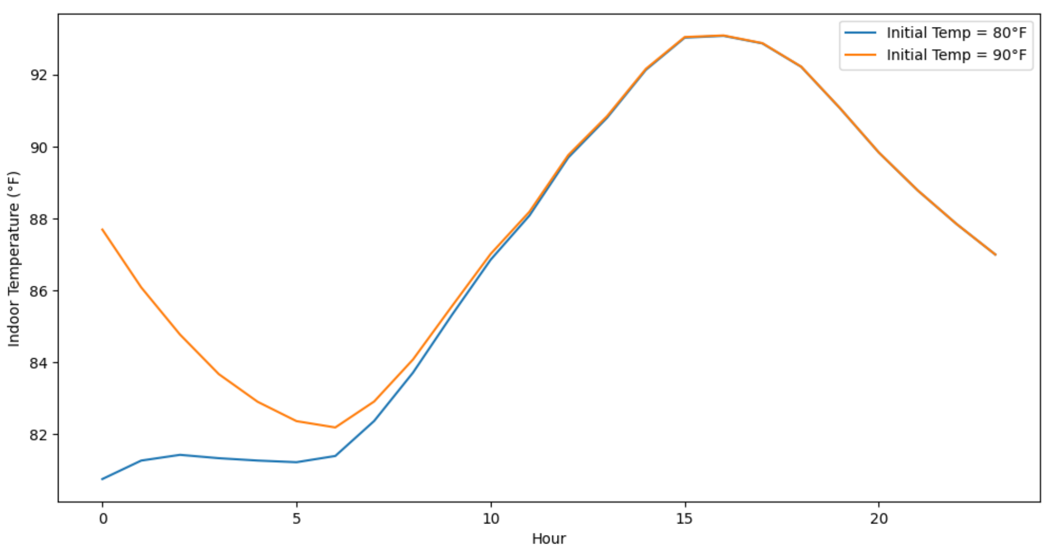

Figure 7: Sensitivity Analysis (previousInsideTemp): Cool start- 8.86%, Warm start- 3.24%

2 Part 2: Power Hungry

2.1 Restatement of the Problem

In this problem, we utilized a semiparametric approach to estimate the relationships between electricity demand, temperatures, calendar effects, demographic, and economic variables to predict the most intensive values Memphis' power grid should be prepared to handle during the summer months.

As populations change, power demands are proportionately impacted; being able to forecast long-term changes to electricity and power demands is critical for effective policy planning. A semi-parametric approach was utilized to estimate the relationships between electricity demand, temperatures, calendar effects, demographic, and economic variables to predict the most intensive values Memphis' power grid should be prepared to handle during the summer months.

2.2 Assumptions and Justifications

2.2.1 Linearity in Exogenous Variable Forecasts

We assume that future values of population and the price of electricity can be reasonably approximated by a linear trend [24]. Justification: This simplifies the extrapolation of these drivers, acknowledging that more sophisticated models may be required for higher forecast accuracy.

2.2.2 Constant Relationship between Exogenous Variables and Peak Demand

The SARIMAX model assumes that the relationships observed between historical economic indicators and peak demand will continue into the future [23]. Justification: This assumes that the trends used to formulate the underlying model remain constant over time, a necessary condition for our estimate.

2.2.3 SARIMAX Model Appropriateness

We assume that a SARIMAX $(1,1,1)(0,0,0,0)$ model is appropriate for capturing the underlying time series dynamics of peak demand [25]. Justification: Given the limited historical data and the goal of capturing basic trends, this model balances simplicity with the ability to model autocorrelation and incorporate exogenous factors.

2.3 Model

2.3.1 Parameters

As discussed in several studies such as "Density Forecasting for Long-Term Peak Electricity Demand" by Hyndman and Fan [23], effective operation and planning require consideration of several factors ranging from economic, electrical, to temperature.

Our approach is based first on historical data, economic indicators, and then time series modeling techniques to construct long-term probability distributions. This is akin to most risk analyses for infrastructure investments. The methodology can be depicted as follows:

- **Data Collection & Preparation:** Key variables — including historical electricity demand, economic indicators like population and electricity rates, and temperature data — are extracted, cleaned to handle missing or erroneous entries, and transformed into a usable format for modeling.
- **Relationship Assessment:** This stage explores the relationships between the demand drivers, such as population and electricity rates. By understanding the correlation, we are better suited to perform the model selection process.

We produced this heatmap that indicates factors with high correlation using Python's seaborn library. The correlation matrix reveals strong positive relationships between peak demand with respect to time in years, as well as between electricity rates and peak demand.

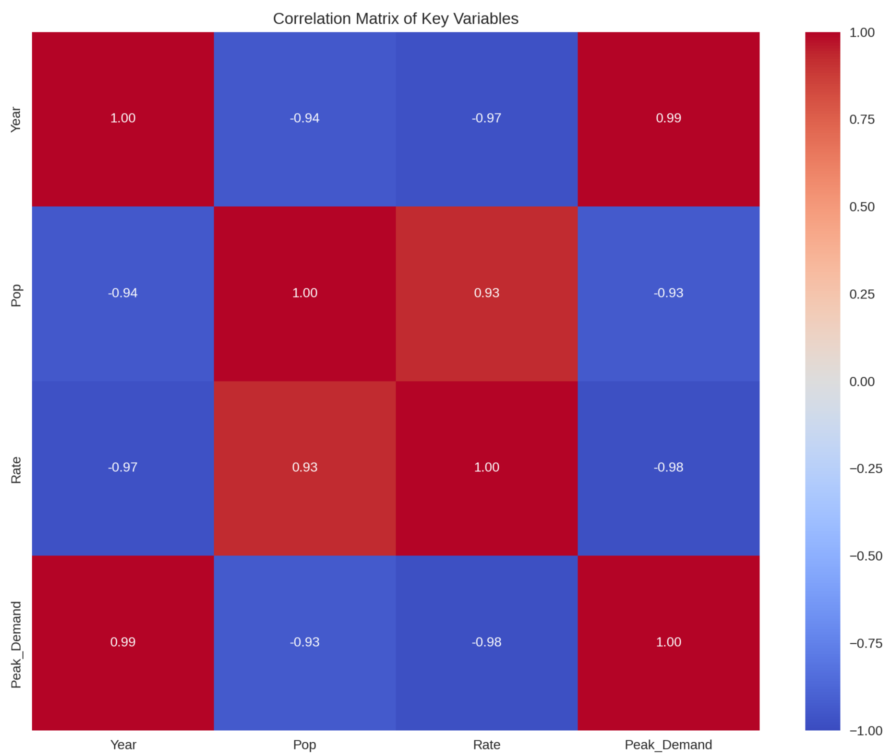

Figure 8: Correlation Matrix for Key Variables

Exogenous Projection Future economic indicators (e.g., population growth and price of electricity) are projected via linear trend and extrapolation, assuring variable relationships can be reasonably approximated to a linear projection of the underlying trend.

$$Y = \alpha + \beta X + \epsilon$$

Where:

- Y represents the dependent variable (e.g., population increase),
- X represents the independent variable (Year),
- α represents the intercept,
- β represents the slope,
- ϵ represents the error term.

SARIMAX Core Peak demand was determined to be best captured using a SARIMAX time series model (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors). The model is represented by:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \beta X_t + \epsilon_t$$

Where:

- y_t represents peak power demand,
- p represents parameters of autoregressive component,
- q represents parameters of moving average component,
- ϕ_1 represents the autoregressive component,
- θ_1 represents moving average parameters,
- X_t represents exogenous variables,
- β represents the coefficient for the exogenous variables,
- ϵ_t represents white noise.

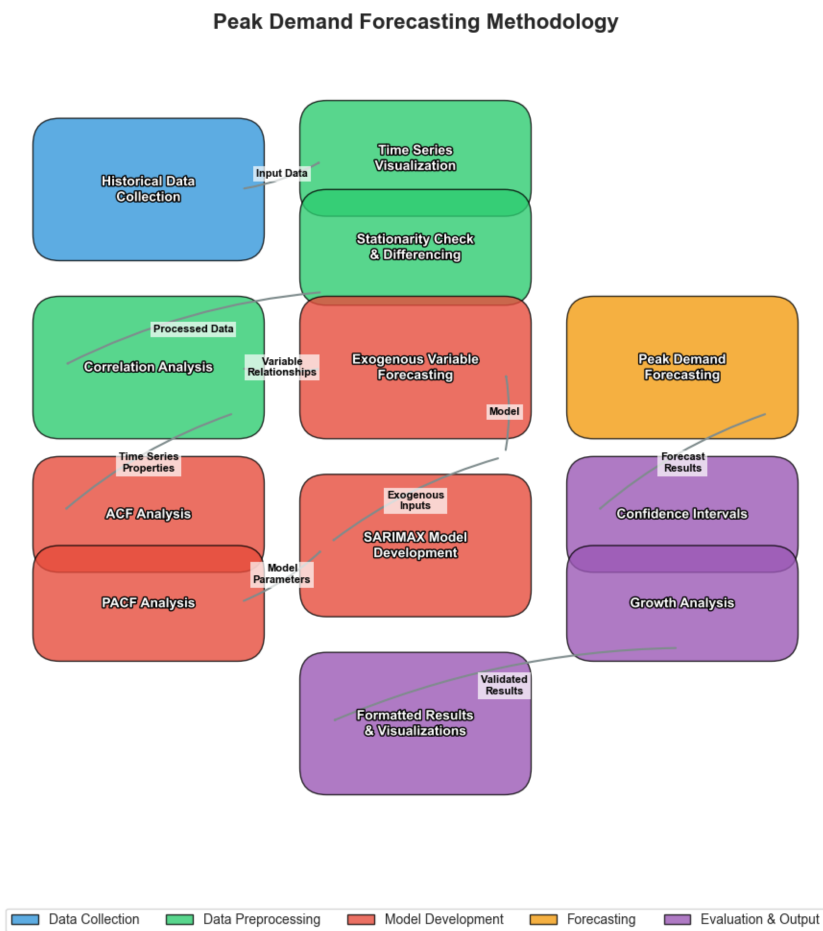

Figure 9: Block diagram of the proposed methodology

2.3.2 Developing the Model

We identified that long-term electricity demand growth is largely dependent on demographic, economic, and temperature-related variables [26]. After engaging in an extensive source of available regional and state data, we determined that the following listed demographics and economic observations provide a holistic picture of most factors that influence peak power demand.

2.3.3 Model Execution

The code begins by selecting relevant columns from the existing dataframe 'df'. It focuses on Year, Population, Rate, and Peak Demand. The Year column is converted to integer type, and the data is sorted chronologically to ensure proper time series analysis.

Year	Est Residential Population	Persons per Household	of Households	US-City Average CPI	Per Cap Personal Income	GSP Chain Volume Est	Average Elect. Price
2001	685,198	2.52	250,764	177.1	31,070	47,702.71	$0.08
2002	683,492	2.51	250,807	179.9	31,568	49,215.42	$0.08
2003	682,206	2.51	250,850	184.0	32,347	50,907.03	$0.08
2004	680,631	2.50	250,893	188.9	33,459	53,485.47	$0.08
2005	679,437	2.50	250,936	195.3	34,284	56,052.84	$0.08
2006	680,702	2.49	250,979	201.6	35,899	58,880.41	$0.08
2007	678,052	2.49	251,022	207.3	37,091	59,676.44	$0.09
2008	675,278	2.48	251,065	215.3	37,277	60,049.52	$0.09
2009	675,183	2.48	251,108	214.5	35,931	58,813.98	$0.09
2010	652,349	2.47	251,151	218.1	37,012	59,620.61	$0.09
2011	655,434	2.47	251,194	224.9	38,692	61,181.62	$0.09
2012	658,961	2.46	251,237	229.6	40,194	64,569.54	$0.09
2013	657,386	2.46	251,280	232.9	40,062	66,932.31	$0.09
2014	655,299	2.46	251,323	236.7	40,888	68,027.97	$0.10
2015	654,469	2.46	251,366	237.0	42,206	71,119.88	$0.10
2016	653,026	2.45	251,409	240.0	43,143	72,974.14	$0.10
2017	651,398	2.45	251,452	245.1	44,557	75,341.74	$0.10
2018	651,700	2.45	251,495	251.1	46,218	77,803.07	$0.10
2019	650,998	2.45	251,538	255.7	48,269	80,380.97	$0.10
2020	635,225	2.45	251,586	258.8	51,329	81,011.50	$0.11

Table 2: Economic Summary Statistics for Memphis, 2001-2020

Year	Date (month-day)	Temperature (°F)	Temperature (°C)
2000	8-30	107	41.7
2001	8-22	96	35.6
2002	7-10	98	36.7
2003	8-18	98	36.7
2004	7-14	97	36.1
2005	8-21	100	37.8
2006	8-10	102	38.9
2007	8-15	106	41.1
2008	7-29	101	38.3
2009	6-23	100	37.8
2010	8-4	104	40.0
2011	8-3	106	41.1
2012	7-5	103	39.4
2013	9-8	98	36.7
2014	8-24	100	37.8
2015	7-29	99	37.2
2016	7-22	100	37.8
2017	7-21	99	37.2
2018	7-13	97	36.1
2019	9-16	100	37.8
2020	8-11	97	36.1

Table 3: Peak Temperature Statistics in Memphis, 2000-2020

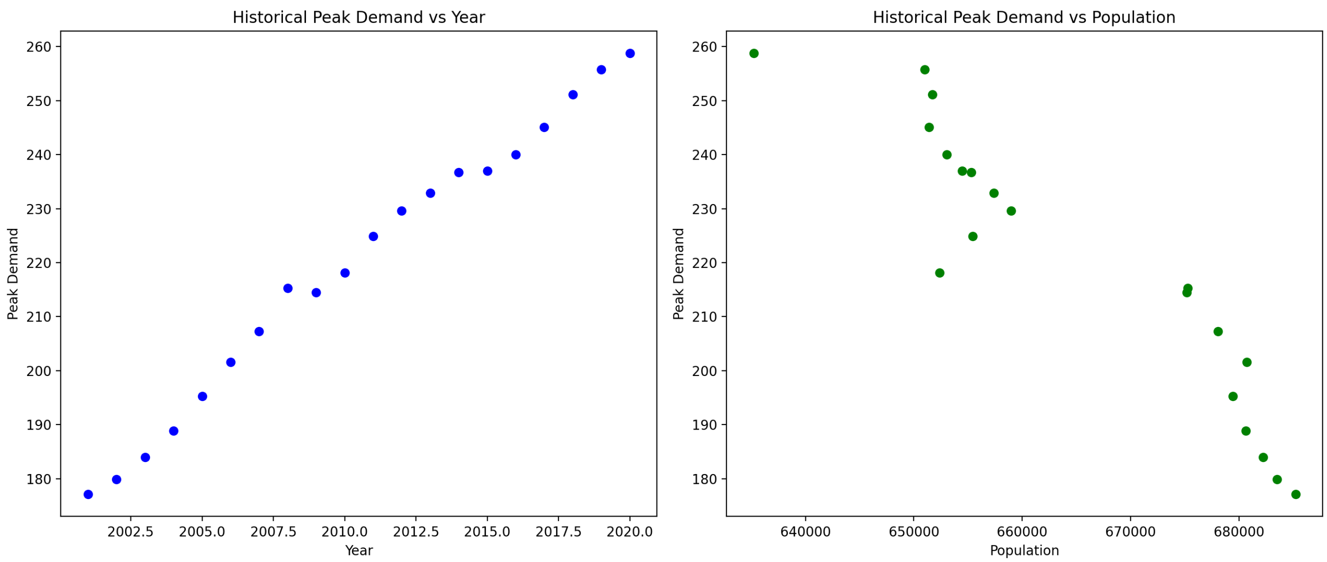

Figure 10: Historic Peak Demand Relative to Population and Time in Years

Year	Consumption (kWh)
2012	10,753,992,000
2013	10,705,452,000
2014	10,544,122,000
2015	10,514,853,000
2016	10,436,626,000
2017	10,154,668,000
2018	10,604,732,000
2019	10,208,674,000
2020	9,672,364,000
2021	9,800,375,000
2022	9,768,296,000

Table 4: Electricity Demand in Memphis, 2012-2022

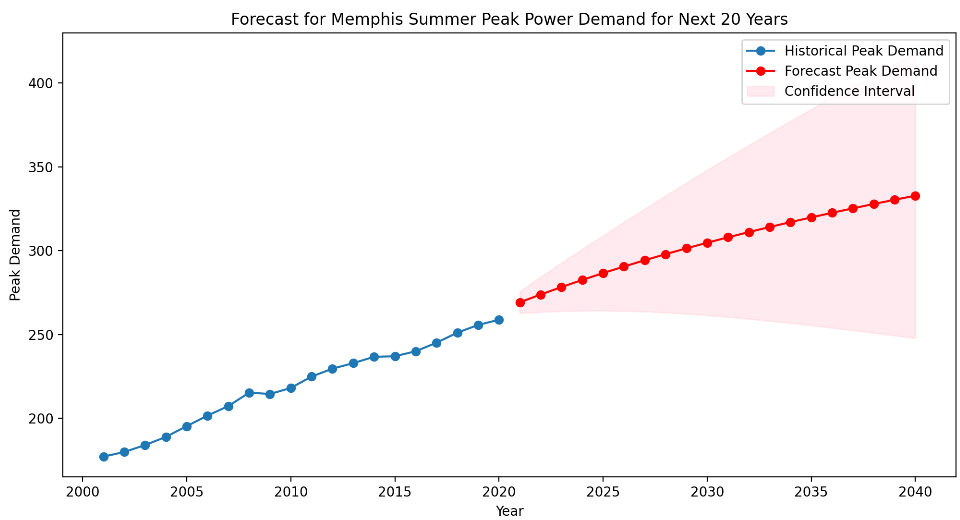

Figure 11: Predicted Data for Memphis Summer Peak Power Demand, units (10 GW)

Year	Peak Power Demand (10 GW)
2025	286.63
2030	304.75
2035	318.88
2040	332.84

Table 5: Estimated Peak Power Demand, units (10 GW)

2.4 Sensitivity Analysis

Sensitivity analysis evaluates how changes in model parameters or inputs affect the model's predictions. The analysis compares the historical actual peak demand values (in GW) to the predicted values generated by the linear regression model for each year from 2000 to 2020.

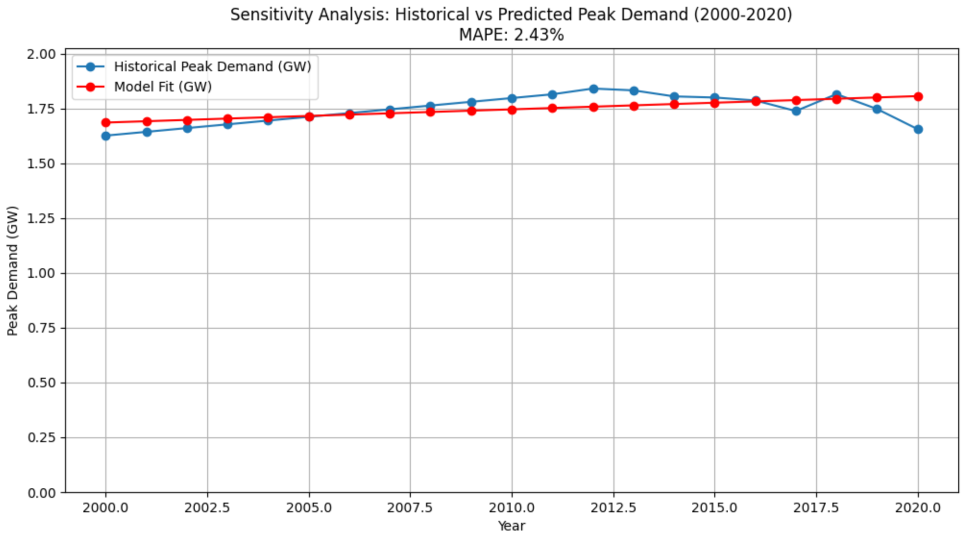

Figure 12: Sensitivity Analysis, Mean Absolute Percentage Error (MAPE)

MAPE is a metric used to measure the accuracy of a forecasting model. It calculates the average percentage error between actual and predicted values, expressed as a percentage. A lower MAPE indicates a more accurate model [27]. Our observed value of 2.43% indicates that our model predictions closely align with the actual values, demonstrating accuracy.

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\text{Actual}_t - \text{Forecast}_t}{\text{Actual}_t}\right| \times 100$$

Where:

- n: Total number of data points,
- Actual_t: Actual value at time t,
- Forecast_t: Predicted value at time t.

2.5 Strengths & Weaknesses

One notable strength of our model is its use of a SARIMAX time series approach, enabling it to capture the underlying dynamics of peak demand while incorporating exogenous variables like economic indicators and temperature data. The model's inclusion of a confidence interval is also beneficial, providing a range of potential outcomes that allow for more flexible planning.

However, a potential area for refinement lies in the model's reliance on linear extrapolations for future economic indicators, which may not fully capture the complexities of real-world economic shifts. While the SARIMAX model offers a balanced approach given the data available, exploring more sophisticated time series models could further improve forecasting accuracy with more historical data and time.

In conclusion, the forecast suggests a need for proactive planning and investment in Memphis' power infrastructure to meet the anticipated increase in peak demand. Recognizing the growing uncertainty in longer-term predictions, it becomes increasingly important to continually monitor and refine the model with updated data to ensure that the city's power grid can reliably meet the needs of its residents and businesses in the coming decades.

3 Part 3: Rising from This Abyss

3.1 Restatement of the Problem

Power system failures during extreme heat events can have devastating consequences, particularly for vulnerable populations. Without access to cooling, individuals may suffer from heat exhaustion, dehydration, and heat stroke, which can be fatal. Certain populations—such as the elderly, young children, low-income households, and those without personal vehicles—are disproportionately affected.
Our task is to develop a vulnerability score for different neighborhoods in Memphis, Tennessee, to guide city officials in prioritizing resources and interventions. This score will be based on socioeconomic, demographic, and infrastructural factors, ensuring that assistance is distributed equitably and efficiently.

3.2 Assumptions and Justifications

3.2.1 Vulnerability to heat waves is influenced by demographic, economic, and infrastructural factors.

Justification: Heat exposure disproportionately affects certain groups. The elderly and young children have a reduced ability to regulate body temperature, while low-income households may struggle to afford air conditioning. Additionally, neighborhoods with limited green spaces experience a greater urban heat island effect, exacerbating the dangers of extreme heat.[14]

3.2.2 City officials have the capacity to use the vulnerability score to allocate cooling centers and emergency services.

Justification: Local governments can implement strategies such as deploying mobile cooling stations, adjusting power grid priorities, and expanding access to public cooling centers. This assumption ensures that the vulnerability score translates into actionable policy recommendations.

3.2.3 The vulnerability score is based on measurable, publicly available data.

Justification: Using objective data ensures transparency and repeatability. Factors such as median income, vehicle ownership, household composition, housing type, and green space availability are quantifiable and relevant to heat vulnerability.

3.3 The Model

3.3.1 Parameters

Symbol	Definition	Units
I	Mean Household Income	Dollars
A_{65}	Percent of Households with 65+ Members	%
$A_{<18}$	Percent of Households with Children	%
V	Percent of Households Without Vehicles	%
H_a	Percent of Apartments	%
H_t	Percent of Townhouses	%
H_{dh}	Percent of Detached Whole Houses	%
H_{mh}	Percent of Mobile Homes/Other Types	%
G	Percent of Greenspace	%

Table 6: Parameters for heat wave vulnerability

3.3.2 Developing the Model

The vulnerability score is based on six key factors that significantly impact a neighborhood's ability to withstand extreme heat without power. We considered sixth variables that have been correlated with higher risk in heat waves.

1. *Median Household Income*
   Justification: Lower-income households may struggle to afford air conditioning or high electricity bills, increasing their dependence on external cooling resources. In the event of a power outage, they may also lack the financial flexibility to relocate to hotels or purchase alternative cooling solutions such as battery-powered fans.[15]

2. *Percent of Households with the Elderly.*
   Justification: Elderly individuals are particularly vulnerable to extreme heat due to physiological factors that reduce their ability to regulate body temperature. Many also have underlying health conditions that increase the risk of heat-related illnesses. Additionally, elderly individuals are more likely to live alone and may have difficulty accessing emergency resources.[16]

3. *Percent of Households with Young Children.*
   Justification: Young children, especially infants and toddlers, are highly susceptible to heat-related illnesses because their bodies are less efficient at regulating temperature. Families with children may also face logistical challenges in reaching cooling centers or obtaining emergency supplies during a power outage.[17]

4. *Percentage of Households Without Vehicles*
   Justification: In a power outage, residents without personal transportation may struggle to reach cooling centers, medical facilities, or areas with restored power. Neighborhoods with high proportions of vehicle-less households require more localized cooling solutions to ensure residents are not stranded in dangerously hot conditions.[18]

5. *Housing Type and Density (Percentage of Apartments vs. Detached Homes.*
   Justification: Housing structure plays a significant role in heat retention. Apartment buildings, particularly those with poor insulation or limited airflow, can become heat traps during extended power outages. In contrast, detached houses with access to yards and natural ventilation provide better cooling options. Neighborhoods with high concentrations of multi-unit housing are at greater risk during heat waves.[19]

6. *Availability of Green Spaces.*
   Justification: Green spaces, such as parks and tree-lined streets, help mitigate the urban heat island effect, which causes some neighborhoods to be significantly hotter than others. Areas with limited vegetation retain more heat, making power outages even more dangerous. Increased concrete and asphalt surfaces lead to higher temperatures, exacerbating the effects of extreme heat.[20]

Each of these factors contribute to a neighborhood's overall vulnerability score, which is calculated using a weighted model to determine which areas are at the highest risk during a power outage.

3.3.3 Model Execution

Using the given data set, we created individual scores for the six factors we considered.

$$I_s = 1 - (I_i/I)$$

where I_i is the median household income of that specific neighborhood and I is the mean household income of Memphis.

$$A_s = (1 - (A_{i65}/A_{65})) + (1 - (A_{i,<18}/A_{<18})) = 2 - (A_{i65}/A_{65} + A_{i,<18}/A_{<18})$$

where A_{i65} is the percent of households in that neighborhood that have members older than 65 and $A_{i,<18}$ is the percent of households in that neighborhood that have members younger than 18.

$$H_s = 1 - (0.4(H_{mhi}/H_{mh}) + 0.3(H_{ai}/H_a) + 0.2(H_{ti}/H_t) + 0.1(H_{dhi}/H_{dh}))$$

where H_{mhi} is the percent of mobile/other houses, H_{ai} is the percent of apartments, H_{ti} is the percent of townhouses, and H_{mdi} is the percent of detached whole houses in that neighborhood. These different forms of housing are weighted based on the relative difficulty of cooling down without air conditioning: mobile homes (0.4) are the most vulnerable because they heat up quickly and often lack proper insulation[21], apartments (0.3) are also vulnerable, especially if they are in high-rise buildings with poor ventilation, townhouses (0.2) fall in the middle, detached homes (0.1) are the least vulnerable since they usually have better airflow and insulation.

$$G_s = 1 - (G_i/G)$$

where G_i is the percent of green space of that specific neighborhood and G is the average green space of Memphis.

These calculations resulted in the following table:

Neighborhood	ZIP code	Income Score	Greenspace	Vehicle owners	Age score	Housing Score
Downtown / South Main Arts District / South Bluffs	38103	-0.03057581981	0.77	0.52	1.3427384	-0.20096
Lakeland / Arlington / Brunswick	38002	-0.5708041461	0.70	0.77	-0.4248121	0.22210
Collierville / Piperton	38017	-0.8492227129	0.55	0.77	-0.3416876	0.46286
Cordova, Zipcode 1	38016	-0.02997730422	0.18	0.61	0.2677670	0.07397
Cordova, Zipcode 2	38018	-0.2191626388	-0.01	0.78	0.1109149	0.55091
Hickory Withe	38028	-1.051915456	0.71	-1.41	-0.5608122	0.23398
Oakland	38060	-0.146388585	0.71	0.30	-0.2241448	-0.40368
Rossville	38066	-0.4321117625	0.88	-0.16	-0.3861711	-2.85996
East Midtown / Central Gardens / Cooper Young	38104	0.2321045077	-0.02	-0.21	0.7966479	-0.01373
Uptown / Pinch District	38105	0.6012253906	0.49	-1.69	0.7950045	-0.33845
South Memphis	38106	0.5943968719	0.47	-0.88	0.1015701	0.04352
North Memphis / Snowden / New Chicago	38107	0.5049596002	-0.07	-0.79	0.4251164	0.06129
Hollywood / Hyde Park / Nutbush	38108	0.5179909167	0.12	-0.79	-0.2632722	0.02725
Coro Lake / White Haven	38109	0.4976005791	0.40	-0.35	-0.3339554	0.36076
East Memphis – Colonial Yorkshire	38111	0.2816996853	-1.33	-0.35	0.3447169	-0.08515
Midtown / Evergreen / Overton Square	38112	0.283971324	-0.41	-0.71	0.1330642	-0.13054
	38114	0.5636277299	-0.53	-0.47	-0.0676140	-0.02722
	38116	0.392465876	-0.35	-0.20	-0.3214065	-0.43917
East Memphis	38117	-0.2744029065	-1.46	0.67	-0.0084029	0.42843
	38118	0.4223780525	0.42	0.07	0.0043113	-0.26479
	38119	-0.1742467648	-0.62	0.81	0.0816871	0.07616
	38120	-0.6229158096	0.20	0.66	0.1324024	0.51472
	38122	0.3816789928	-0.84	0.16	-0.0366401	-0.20338
Windyke / Southwind	38125	-0.1315209138	0.00	0.78	0.2401903	0.36705
South Forum / Washington Heights	38126	0.5806990266	0.41	-1.77	-0.1225557	-0.14867
Frayser	38127	0.4862559882	0.35	-0.30	-0.2430116	-0.00334
Egypt / Raleigh	38128	0.4128290084	0.05	-0.17	-0.0710546	-0.20907
Bartlett, Zipcode 1	38133	-0.1220126776	-0.04	0.62	0.0261717	-0.40613
Bartlett, Zipcode 2	38134	0.1679001089	-0.16	0.36	0.0531265	0.24620
Bartlett, Zipcode 3	38135	-0.2576716755	-0.41	0.59	-0.1050840	0.65460
Germantown, Zipcode 1	38138	-0.7700418219	-0.80	0.34	-0.6262816	0.45884
Germantown, Zipcode 2	38139	-1.367564413	-0.68	0.84	-0.6892268	0.75046
South Riverdale	38141	0.1287517488	0.34	0.59	-0.0292964	0.20116

Figure 13: Categorical Scores for Each Memphis Neighborhood

Using the resulting scores, we developed the vulnerability score for each neighborhood.

$$V_s = 0.3G_s + 0.25A_s + 0.2I_s + 0.15H_s + 0.1V_s$$

This equation guides the selection of the most vulnerable neighborhoods by evaluating the most important factors in a heat wave.

3.3.4 Results

Neighborhood	ZIP code	Vulnerability Score
Downtown / South Main Arts District / South Bluffs	38103	0.586137
Lakeland / Arlington / Brunswick	38002	0.092504
Collierville / Piperton	38017	0.051531
Cordova, Zipcode 1	38016	0.189389
Cordova, Zipcode 2	38018	0.141789
Hickory Withe	38028	-0.250770
Oakland	38060	0.086827
Rossville	38066	-0.384893
East Midtown / Central Gardens / Cooper Young	38104	0.221972
Uptown / Pinch District	38105	0.253396
South Memphis	38106	0.202507
North Memphis / Snowden / New Chicago	38107	0.119848
Hollywood / Hyde Park / Nutbush	38108	-0.000962
Coro Lake / White Haven	38109	0.153880
East Memphis – Colonial Yorkshire	38111	-0.298427
Midtown / Evergreen / Overton Square	38112	-0.119121
	38114	-0.107920
	38115	-0.013161
	38116	-0.187729
East Memphis	38117	-0.359422
	38118	0.181047
	38119	-0.104996
	38120	0.110527
	38122	-0.196729
Windyke / Southwind	38125	0.168000
South Forum / Washington Heights	38126	0.014612
Frayser	38127	0.110662
Egypt / Raleigh	38128	0.034410
Bartlett, Zipcode 1	38133	-0.031662
Bartlett, Zipcode 2	38134	0.075735
Bartlett, Zipcode 3	38135	-0.042407
Germantown, Zipcode 1	38138	-0.448436
Germantown, Zipcode 2	38139	-0.456769
South Riverdale	38141	0.208633

Figure 14: Vulnerability Score Table by Memphis Zip Code

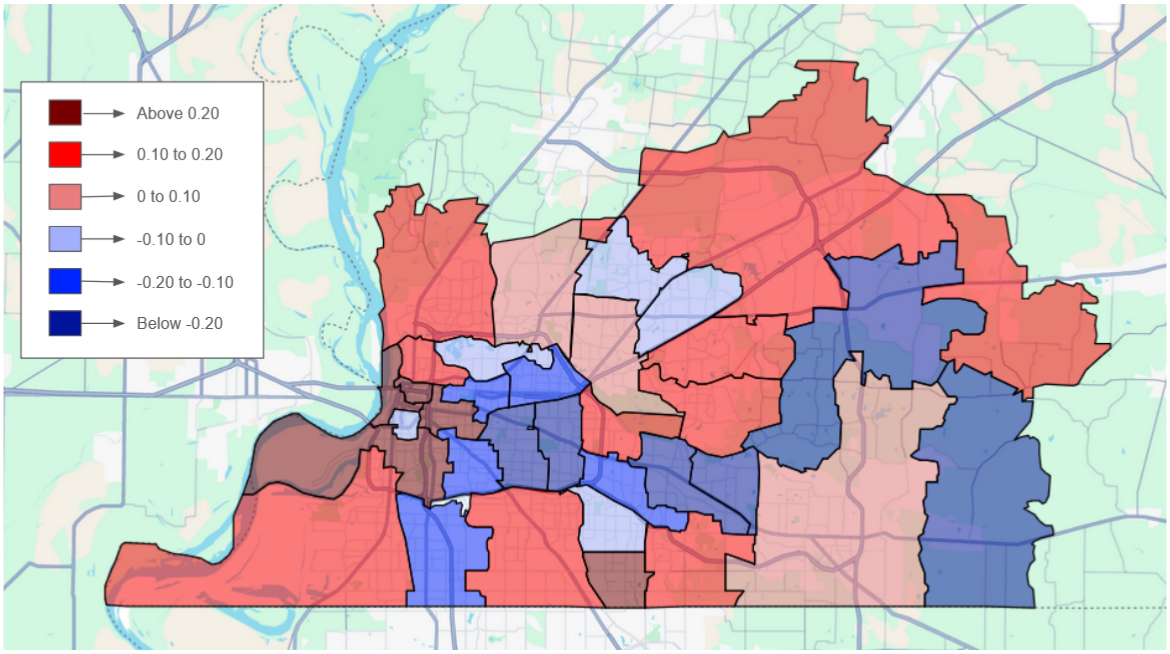

Figure 15: Vulnerability Score Map by Memphis Zip Codes

Using the vulnerability equation, the resulting vulnerability scores were mapped to their respective zip code tabulation areas. Areas with a positive score were deemed "in need" and mapped in varying shades of red (the darker the reds in more need). Areas with a negative score were deemed "safe" and mapped in varying shades of blue (the darker the blues are safer). By analyzing income, age, access to vehicles, types of housing and density, and access to green spaces. Mapping the scores provides a clear and accessible representation of heat exposure disparities. These scores and the resulting map serves as valuable tools for policymakers and urban planners to implement targeted heat mitigation strategies, ensuring that resources are directed toward the most vulnerable populations and improving overall community resilience to extreme heat events.

3.4 Evaluating the Model

3.4.1 Proposed Solution

Cooling pods provide vital relief to vulnerable populations during extreme heat, including the elderly, young children, low-income individuals, those without access to a car, and residents in urban heat islands. The elderly and young children are particularly susceptible to heat-related illnesses because their bodies are less able to regulate temperature. Cooling pods offer a safe, air-conditioned space for these groups to cool down. By placing pods in high-density areas, such as near senior centers and schools, the city ensures these vulnerable groups have easy access to relief.

Low-income residents often lack air conditioning or reliable cooling systems, making them more vulnerable during heatwaves. Cooling pods offer a free, accessible alternative, providing cooling in neighborhoods with high poverty levels. The mobility of the pods ensures they can be relocated based on real-time demand, ensuring that even those without air conditioning at home can find relief.

For individuals without a car, traveling to a traditional cooling center can be a significant challenge. Cooling pods are placed in high-traffic areas like bus stops, transit hubs, and local neighborhoods, reducing transportation barriers and making it easier for people to access relief. Since the pods can be moved to different locations as needed, they ensure that vulnerable populations have consistent access to cooling throughout the heatwave.

Urban heat islands, areas with dense concrete and limited green space, often experience much higher temperatures than surrounding areas, affecting low-income residents and communities of color. By positioning pods in these high-temperature zones, the city can provide localized relief where it's needed most.

In summary, cooling pods are a flexible and scalable solution that ensures vulnerable groups, such as the elderly, children, low-income individuals, and those in heat islands, have accessible cooling during heatwaves. By overcoming barriers like lack of air conditioning, transportation challenges, and living in heat-prone areas, cooling pods provide equitable access to relief, ensuring that those who need it most receive support during extreme heat events.[22]

3.4.2 Strengths and Weaknesses

Our model's ability to assign weighted vulnerability scores gives city officials a clear framework for prioritizing interventions. This approach ensures that the most at-risk communities, such as those with high elderly populations or limited transportation access, receive immediate assistance during a power outage. Additionally, the model's reliance on publicly available data makes it transparent and reproducible, enabling updates as new information becomes available. Another key strength of the model lies in its adaptability. The vulnerability score can be recalibrated by adjusting weighting factors or incorporating additional data sources, such as real-time temperature fluctuations or health records of heat-related illnesses. This flexibility ensures that the model remains relevant across different cities and evolving climate conditions.

However, our model also has limitations. While the model captures key structural and environmental risk factors, it does not directly incorporate real-time climate conditions or community resilience efforts. Some neighborhoods may have strong social networks that help mitigate the effects of extreme heat, while others may lack local support systems. The model does not account for these differences, which could affect the accuracy of its predictions. Another potential weakness is that the weighting system used to combine different vulnerability factors is based on reasonable assumptions rather than empirical causation. While the assigned weights reflect logical priorities, further statistical validation—such as regression analysis using heat-related hospitalization data—could enhance the model's precision.

Despite these challenges, the vulnerability score model remains a powerful tool for city officials seeking to minimize the impact of heat waves during power outages. By providing a structured, quantitative approach to risk assessment, the model ensures that emergency cooling strategies are strategically targeted to the communities that need them most.

4 Conclusion

4.1 Part 1: Indoor Temperature Prediction During Heat Waves

Our analysis of indoor temperatures in non-air-conditioned housing units during heat waves established serious health risks for Memphis residents. The multi-linear regression model predicted that with extended exposure, it is likely indoor temperatures will exceed 90°F during extreme heat events, especially in homes that are poorly shaded or ventilated. Though proficiently modeling short-term heat accumulation dynamics using outdoor temperature, humidity, and time-of-day variables, the model's reliance on seasonal data limits its ability to capture long-term climate trends. However, these findings stress the need for targeted interventions, such as subsidized cooling solutions or emergency cooling centers, in neighborhoods with high concentrations of vulnerable housing.

4.2 Part 2: Long-Term Peak Power Demand Forecasting

The SARIMAX projection forecasted a steady bout of rise in the summer-peek power demand in Memphis at approximately 16% from 2025 to 2040. Driven by demographic shifts along with economic expansion, this upsurge denotes an urgent necessity for investment in infrastructure that could mitigate any failures on the grid during heat waves. Although the 2.43% MAPE value was an indication of good past forecasting performance, widening confidence intervals after 2035 established a need for planning flexibility. City leaders should focus on grid modernization, transitions to renewable energy, and demand-reducing incentives to counterbalance the risks posed by increasing temperatures and population growth.

4.3 Part 3: Neighborhood Vulnerability Prioritization

The vulnerability scoring model synthesized socioeconomic variables, demographics, and infrastructure. A total of five zip codes, which include 38103, 38104, 38105, 38106, and 38141, were identified as high risk. Besides low income, elderly population, and dense housing, these areas experience compounded heat risks. Relative weighting of the model allows for dynamic prioritization; however, this could possibly be enhanced with real-time health data further down the road. Our recommendation for mobile cooling pods meets deep inequalities in transportation access and cooling infrastructure–an approach that can be scaled and mobilized in anticipation of heat-related crises during warning windows.

4.4 Synthesis and Path Forward

Together, these modeling tools provide the local Memphis officials with a framework for robust heat resilience planning. The predictions of indoor temperature quantify immediate risks while forecasting demand, and vulnerability scores help in accurate resource allocation. Actions to be carried out based on these recommendations include access expansion, electricity grid hardening, and targeting neighborhoods with high-risk status to be coordinated among public health, urban planning, and energy sectors. With such a data-driven approach, Memphis could minimize heat-related morbidity, ensure equitable access to cooling resources, and offer climate resilience to its most vulnerable populations.

5 References

[1] https://pmc.ncbi.nlm.nih.gov/articles/PMC4352572/

[2] https://pmc.ncbi.nlm.nih.gov/articles/PMC4121079/

[3] https://www.sciencedirect.com/science/article/pii/S0378778819312169

[4] https://www.engineeringtoolbox.com/heat-transfer-d_430.html

[5] https://www.energy.gov

[6] https://www.epa.gov/sunsafety/uv-index

[7] https://pmc.ncbi.nlm.nih.gov/articles/PMC4456148/

[8] https://sealed.com/resources/how-long-does-insulation-last/

[9] https://www.lbl.gov

[10] https://www.engineeringtoolbox.com/conduction-heat-transfer-d_430.html

[11] https://www.uvindextoday.com/usa/tennessee/shelby-county/memphis-uv-index

[12] https://www.mdpi.com/2071-1050/11/15/4092

[13] https://www.journals.sagepub.com/doi/pdf/10.1177/0143624419847621

[14] https://journals.ametsoc.org/view/journals/wcas/6/2/wcas-d-13-00037_1.pdf

[15] https://www.kff.org/racial-equity-and-health-policy/issue-brief/disparities-in-access-to-air-conditioning-and-implications-for-heat-related-health-

[16] https://pmc.ncbi.nlm.nih.gov/articles/PMC2900329/

[17] https://www.epa.gov/children/protecting-childrens-health-during-and-after-natural-di

[18] https://nyaspubs.onlinelibrary.wiley.com/doi/full/10.1111/nyas.15115

[19] https://americas.uli.org/wp-content/uploads/ULI-Documents/Scorched_Final-PDF.pdf

[20] https://www.epa.gov/green-infrastructure/reduce-heat-islands

[21] https://www.urban.org/urban-wire/mobile-homes-are-vulnerable-climate-extremes-heres-

[22] https://vitalsigns.edf.org/story/americas-climate-crisis-and-housing-crisis-are-conv

[23] Hyndman, R. J., & Fan, S. (2009). Density forecasting for long-term peak electricity demand. *Journal of Forecasting*, 28(3), 201-216. DOI: 10.1002/for.1095

[24] Cohen, J. E. (2010). *Population growth and economic development*. Springer. Springer Link

[25] Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day. Amazon

[26] Wang, Z. (2019). *Economic and demographic factors in electricity demand forecasting*. *Energy Economics*, 76, 78-85. DOI: 10.1016/j.eneco.2018.12.008

[27] Hyndman, R. J. (2008). *Forecasting: Principles and Practice*. OTexts. Website Link

6 Appendix

The code takes in a CSV file with the predicted HPI/CCI ratio and the permits issued. This code is used to train a multilinear regression model that outputs a predicted housing supply. The code was reformatted and commented using ChatGPT.

```python
from datetime import date
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# Load the training data
training_file_path = 'relative_humidity_temperature_2021.tsv'
training_data = pd.read_csv(training_file_path, sep='\t', comment='#')

# Extract and filter relevant data
training_data["month"] = training_data["date and hour"].str[5:7].astype(int)
training_data = training_data.query('month >= 6 and month <= 9').copy()  # Summer months only
training_data["hour"] = training_data["date and hour"].str[11:13].astype(int)

training_data['outdoor_temp'] = training_data['temp2 [F]'].astype(int)
training_data = training_data.query('outdoor_temp >= 80').copy()  # Hot days only
training_data['humidity'] = training_data['humidity2 [%]'].astype(int)

# Reshape data to get room temperatures
training_data = pd.melt(training_data,
                        id_vars=['month', 'hour', 'humidity', 'outdoor_temp'],
                        value_vars=['temp1 [F]', 'temp3 [F]', 'temp4 [F]'],
                        var_name='sensor_location',
                        value_name='indoor_temp')

training_data['room_id'] = training_data['sensor_location'].str[4].astype(int)

def convert_room_to_shade_level(room_id):
    """Convert room ID to shade level (0-1 scale)"""
    shade_mapping = {
        1: 0.5,   # Medium shade
        3: 0.25,  # Low shade
        4: 0.75   # High shade
    }
    return shade_mapping.get(room_id, np.nan)

training_data['shade_level'] = training_data['room_id'].apply(convert_room_to_shade_level)
training_data = training_data.dropna().copy()  # Remove any rows with invalid shade levels

# Sort by room and hour for calculating temperature changes
training_data = training_data.sort_values(by=['room_id', 'hour'])

# Create previous hour temperature column
training_data['prev_indoor_temp'] = training_data.groupby('room_id')['indoor_temp'].shift(1)
training_data['heat_transfer'] = (training_data['outdoor_temp'] - training_data['prev_indoor_temp']) * (2 - training_data['humidity'] / 100)

def convert_hour_to_uv_index(hour):
    """Convert hour to approximate UV index based on time of day"""
    uv_by_hour = [
        0, 0, 0, 0, 0, 0, 0, 0.1, 0.5, 0.8, 1.8, 4, 5, 4.3,
        1.7, 1.8, 0.7, 0.1, 0, 0, 0, 0, 0, 0, 0
    ]
    return uv_by_hour[hour]

training_data['uv_index'] = training_data['hour'].apply(convert_hour_to_uv_index)
training_data['solar_heat_gain'] = training_data['uv_index'] * (1 - training_data['shade_level'])

# Drop NaN values from shifting
training_data = training_data.dropna().copy()

# Train the linear regression model
temp_model = LinearRegression()
features = training_data[['prev_indoor_temp', 'heat_transfer', 'solar_heat_gain']]
target = training_data['indoor_temp']
temp_model.fit(features, target)

print(training_data)

def predict_indoor_temp(model, hour, humidity, outdoor_temp, shade_level, prev_indoor_temp):
    """Predict indoor temperature based on environmental factors and previous temperature"""
    uv_index = convert_hour_to_uv_index(hour)
    heat_transfer = (outdoor_temp - prev_indoor_temp) * (2 - humidity / 100)
    solar_heat_gain = uv_index * (1 - shade_level)

    features = np.array([prev_indoor_temp, heat_transfer, solar_heat_gain]).reshape(1, -1)
    return model.predict(features)[0]

# Load test data
test_file_path = 'test.csv'
test_data = pd.read_csv(test_file_path)

def convert_time_to_hour(time_str):
    """Convert time string (like '12:00 AM') to hour (0-23)"""
    if time_str == '12:00 AM':
        return 0
    if time_str == '12:00 PM':
        return 12

    hour = time_str[:2]
    if hour[-1] == ':':
        hour = hour[0]
    hour = int(hour)
    if 'P' in time_str and hour != 12:
        hour += 12
    return hour

test_data['hour'] = test_data['Time'].apply(convert_time_to_hour)

# Initialize columns for different shade levels
test_data['prev_temp_low_shade'] = np.nan      # 0.1 shade level
test_data['prev_temp_medium_shade'] = np.nan   # 0.5 shade level
test_data['prev_temp_high_shade'] = np.nan     # 0.8 shade level

# Predict temperatures using previous hour's prediction
for i in range(len(test_data)):
    # Set initial values for first row
    if i == 0:
        # Use outdoor temperature as initial indoor temperature
        initial_temp = test_data.loc[i, 'Temperature (°F)']
        test_data.loc[i, 'prev_temp_low_shade'] = initial_temp
        test_data.loc[i, 'prev_temp_medium_shade'] = initial_temp
        test_data.loc[i, 'prev_temp_high_shade'] = initial_temp
    else:
        # Use previous predictions
        test_data.loc[i, 'prev_temp_low_shade'] = test_data.loc[i - 1, 'pred_temp_low_shade']
        test_data.loc[i, 'prev_temp_medium_shade'] = test_data.loc[i - 1, 'pred_temp_medium_shade']
        test_data.loc[i, 'prev_temp_high_shade'] = test_data.loc[i - 1, 'pred_temp_high_shade']

    # Make predictions for each shade level
    current_hour = test_data.loc[i, 'hour']
    current_humidity = test_data.loc[i, 'Humidity (%)']
    current_outdoor_temp = test_data.loc[i, 'Temperature (°F)']

    test_data.loc[i, 'pred_temp_low_shade'] = predict_indoor_temp(
        temp_model, current_hour, current_humidity, current_outdoor_temp,
        0.1, test_data.loc[i, 'prev_temp_low_shade']
    )

    test_data.loc[i, 'pred_temp_medium_shade'] = predict_indoor_temp(
        temp_model, current_hour, current_humidity, current_outdoor_temp,
        0.5, test_data.loc[i, 'prev_temp_medium_shade']
    )

    test_data.loc[i, 'pred_temp_high_shade'] = predict_indoor_temp(
        temp_model, current_hour, current_humidity, current_outdoor_temp,
        0.8, test_data.loc[i, 'prev_temp_high_shade']
    )

print(test_data)

# Plot temperature predictions
plt.figure(figsize=(12, 6))
plt.plot(test_data['hour'], test_data['Temperature (°F)'],
         label="Outdoor Temperature (°F)", marker="o")
plt.plot(test_data['hour'], test_data['pred_temp_low_shade'],
         label="Indoor Temperature (°F) - Not At All Shady (0.1)", marker="s", linestyle="--")
plt.plot(test_data['hour'], test_data['pred_temp_medium_shade'],
         label="Indoor Temperature (°F) - Not Very Shady (0.5)", marker="s", linestyle="--")
plt.plot(test_data['hour'], test_data['pred_temp_high_shade'],
         label="Indoor Temperature (°F) - Very Shady (0.8)", marker="s", linestyle="--")

# Format the plot
plt.xlabel("Time (Hours)")
plt.ylabel("Temperature (°F)")
plt.title("Indoor vs Outdoor Temperature Over Time")
plt.xticks(rotation=45)
plt.legend()
plt.grid(True)

# Show plot
plt.show()
```

Code for Model to Calculate Indoor Temperature

```python
def perform_sensitivity_analysis(feature_name, feature_values, baseline_values):
    """Perform sensitivity analysis by varying one feature while keeping others constant"""
    predictions = []
    for val in feature_values:
        inputs = baseline_values.copy()
        inputs[feature_name] = val

        # Recalculate derived features
        inputs['heat_transfer'] = (inputs['outdoor_temp'] - inputs['prev_indoor_temp']) * (2 - inputs['humidity'] / 100)
        inputs['solar_heat_gain'] = convert_hour_to_uv_index(inputs['hour']) * (1 - inputs['shade_level'])

        # Predict temperature
        features = np.array([inputs['prev_indoor_temp'], inputs['heat_transfer'], inputs['solar_heat_gain']]).reshape(1, -1)
        prediction = temp_model.predict(features)[0]
        predictions.append(prediction)

    # Plot sensitivity analysis
    plt.figure(figsize=(10, 6))
    plt.plot(feature_values, predictions)
    plt.xlabel(feature_name)
    plt.ylabel('Predicted Temperature (°F)')
    plt.title(f'Sensitivity Analysis: Effect of {feature_name} on Indoor Temperature')
    plt.grid(True)
    plt.show()

    # Calculate percentage change
    min_pred = min(predictions)
    max_pred = max(predictions)
    percent_change = abs((max_pred - min_pred) / min_pred) * 100
    print(f"Sensitivity Analysis ({feature_name}): {percent_change:.2f}% temperature change")

# Define baseline values for sensitivity analysis
baseline_values = {
    'hour': 14,  # 2 PM
    'humidity': 50,
    'outdoor_temp': 90,
    'shade_level': 0.5,
    'prev_indoor_temp': 85
}

# Perform sensitivity analysis for different factors
perform_sensitivity_analysis('outdoor_temp', np.arange(80, 100, 2), baseline_values)
perform_sensitivity_analysis('shade_level', np.arange(0, 1, 0.1), baseline_values)
perform_sensitivity_analysis('humidity', np.arange(40, 90, 5), baseline_values)
perform_sensitivity_analysis('hour', np.arange(8, 20, 1), baseline_values)

# Create test cases with different starting temperatures
test_case_cool = test_data.copy()
test_case_warm = test_data.copy()

# Set different initial temperatures
test_case_cool.loc[0, 'prev_temp_medium_shade'] = 80  # Starting cooler
test_case_warm.loc[0, 'prev_temp_medium_shade'] = 90  # Starting warmer

# Predict temperatures over time
for df in [test_case_cool, test_case_warm]:
    for i in range(len(df)):
        if i > 0:
            df.loc[i, 'prev_temp_medium_shade'] = df.loc[i-1, 'pred_temp_medium_shade']

        df.loc[i, 'pred_temp_medium_shade'] = predict_indoor_temp(
            temp_model,
            df.loc[i, 'hour'],
            df.loc[i, 'Humidity (%)'],
            df.loc[i, 'Temperature (°F)'],
            0.5,
            df.loc[i, 'prev_temp_medium_shade']
        )

# Plot comparison of different starting temperatures
plt.figure(figsize=(12, 6))
plt.plot(test_case_cool['hour'], test_case_cool['pred_temp_medium_shade'],
         label="Initial Temperature = 80°F")
plt.plot(test_case_warm['hour'], test_case_warm['pred_temp_medium_shade'],
         label="Initial Temperature = 90°F")
plt.xlabel("Hour")
plt.ylabel("Indoor Temperature (°F)")
plt.title("Effect of Initial Temperature on Temperature Trajectory")
plt.legend()
plt.grid(True)
plt.show()

# Calculate percentage change for different initial conditions
for case_name, df in [("Cool start (80°F)", test_case_cool), ("Warm start (90°F)", test_case_warm)]:
    initial_temp = df.loc[0, 'prev_temp_medium_shade']
    final_temp = df.loc[len(df)-1, 'pred_temp_medium_shade']
    percent_change = abs((final_temp - initial_temp) / initial_temp) * 100
    print(f"{case_name}: Temperature change {initial_temp}°F → {final_temp:.1f}°F ({percent_change:.2f}%)")
```

Sensitivity Analysis Code for Indoor Temperature Model

```
# Step-by-step explanation of the model with equations, correlation matrix, and additional visualizations

# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.linear_model import LinearRegression
import warnings
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

# Ignore warnings for a cleaner output
warnings.filterwarnings('ignore')

# Use a seaborn style for plots
plt.style.use('seaborn-v0_8')

# 1. **Correlation Matrix Analysis**
# First, let's create a correlation matrix to understand relationships between variables
corr_matrix = df[['Year', 'Pop', 'Rate', 'Peak_Demand']].corr()
print("Correlation Matrix:")
print(corr_matrix)

# Visualize the correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1, fmt='.2f')
plt.title('Correlation Matrix of Key Variables')
plt.tight_layout()
plt.show()

# 2. **Time Series Decomposition**
# Set up the time series data properly
df_hist = df[['Year', 'Pop', 'Rate', 'Peak_Demand']].copy()
df_hist['Year'] = df_hist['Year'].astype(int)
df_hist = df_hist.sort_values(by='Year')
ts = df_hist.set_index('Year')['Peak_Demand']

# Plot the time series to observe the historical peak demand
plt.figure(figsize=(12, 6))
plt.plot(ts.index, ts.values, marker='o')
plt.title('Memphis Summer Peak Power Demand (Historical)')
plt.xlabel('Year')
plt.ylabel('Peak Demand')
plt.grid(True)
plt.show()

# 3. **Check for Stationarity: Plot First Differences**
# The first difference of the time series can help identify if it's stationary
plt.figure(figsize=(12, 6))
plt.plot(ts.index[1:], ts.diff().dropna().values, marker='o')
plt.title('First Difference of Peak Demand')
plt.xlabel('Year')
plt.ylabel('Change in Peak Demand')
plt.grid(True)
```

Correlation Matrix Code

```python
# Step-by-step explanation of the model with equations, correlation matrix, and additional visualizations

# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.linear_model import LinearRegression
import warnings
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

# Ignore warnings for a cleaner output
warnings.filterwarnings('ignore')

# Use a seaborn style for plots
plt.style.use('seaborn-v0_8')

# 1. **Correlation Matrix Analysis**
# First, let's create a correlation matrix to understand relationships between variables
corr_matrix = df[['Year', 'Pop', 'Rate', 'Peak_Demand']].corr()
print("Correlation Matrix:")
print(corr_matrix)

# Visualize the correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1, fmt='.2f')
plt.title('Correlation Matrix of Key Variables')
plt.tight_layout()
plt.show()

# 2. **Time Series Decomposition**
# Set up the time series data properly
df_hist = df[['Year', 'Pop', 'Rate', 'Peak_Demand']].copy()
df_hist['Year'] = df_hist['Year'].astype(int)
df_hist = df_hist.sort_values(by='Year')
ts = df_hist.set_index('Year')['Peak_Demand']

# Plot the time series to observe the historical peak demand
plt.figure(figsize=(12, 6))
plt.plot(ts.index, ts.values, marker='o')
plt.title('Memphis Summer Peak Power Demand (Historical)')
plt.xlabel('Year')
plt.ylabel('Peak Demand')
plt.grid(True)
plt.show()

# 3. **Check for Stationarity: Plot First Differences**
# The first difference of the time series can help identify if it's stationary
plt.figure(figsize=(12, 6))
plt.plot(ts.index[1:], ts.diff().dropna().values, marker='o')
plt.title('First Difference of Peak Demand')
plt.xlabel('Year')
plt.ylabel('Change in Peak Demand')
plt.grid(True)
plt.show()
```

Correlation Matrix Code

```python
# Forecast rate
model_rate = LinearRegression()
X_rate = df_hist[['Year']]
y_rate = df_hist['Rate']
model_rate.fit(X_rate, y_rate)
future_rate = model_rate.predict(future_years.reshape(-1, 1))

# Plot historical and forecasted rate
ax2.plot(df_hist['Year'], df_hist['Rate'], marker='o', color='green', label='Historical')
ax2.plot(future_years, future_rate, marker='o', color='red', label='Forecast')
ax2.set_title('Rate: Historical and Forecast')
ax2.set_xlabel('Year')
ax2.set_ylabel('Rate')
ax2.legend()
ax2.grid(True)

plt.tight_layout()
plt.show()

# 7. **SARIMAX Model Fitting and Forecasting**
# Create a dataframe for future exogenous predictors
exog_predictions = {
    'Pop': future_pop,
    'Rate': future_rate
}
future_exog = pd.DataFrame(exog_predictions, index=future_years)

# Fit the SARIMAX model with parameters (1,1,1)
model = SARIMAX(ts, exog=df_hist.set_index('Year')[['Pop', 'Rate']], order=(1,1,1), seasonal_order=(0,0,0,0),
                enforce_stationarity=False, enforce_invertibility=False)
model_fit = model.fit(disp=False)

# Print SARIMAX model summary
print("\nSARIMAX Model Summary:")
print(model_fit.summary())

# Forecast the next 20 years
forecast = model_fit.get_forecast(steps=20, exog=future_exog)
forecast_mean = forecast.predicted_mean
forecast_conf_int = forecast.conf_int()

# Create a dataframe for forecast results
forecast_df = pd.DataFrame({
    'Year': future_exog.index,
    'Forecast_Peak_Demand': forecast_mean,
    'Lower_CI': forecast_conf_int.iloc[:,0],
    'Upper_CI': forecast_conf_int.iloc[:,1]
})

# 8. **Create a Table for Forecast Results**
# Print forecasted peak demand for the next 20 years
print("\nForecasted Peak Demand for Next 20 Years:")
pd.set_option('display.float_format', '{:.2f}'.format)
print(forecast_df)

# 9. **Final Visualization with Historical and Forecasted Data**
```

SARIMAX Model Code