HEAT VS POWER:

TEMPERATURE VS. ELECTRICITY USE IN U.S. STATES

Data Analytics using Python

CS 2704 – April 9, 2025

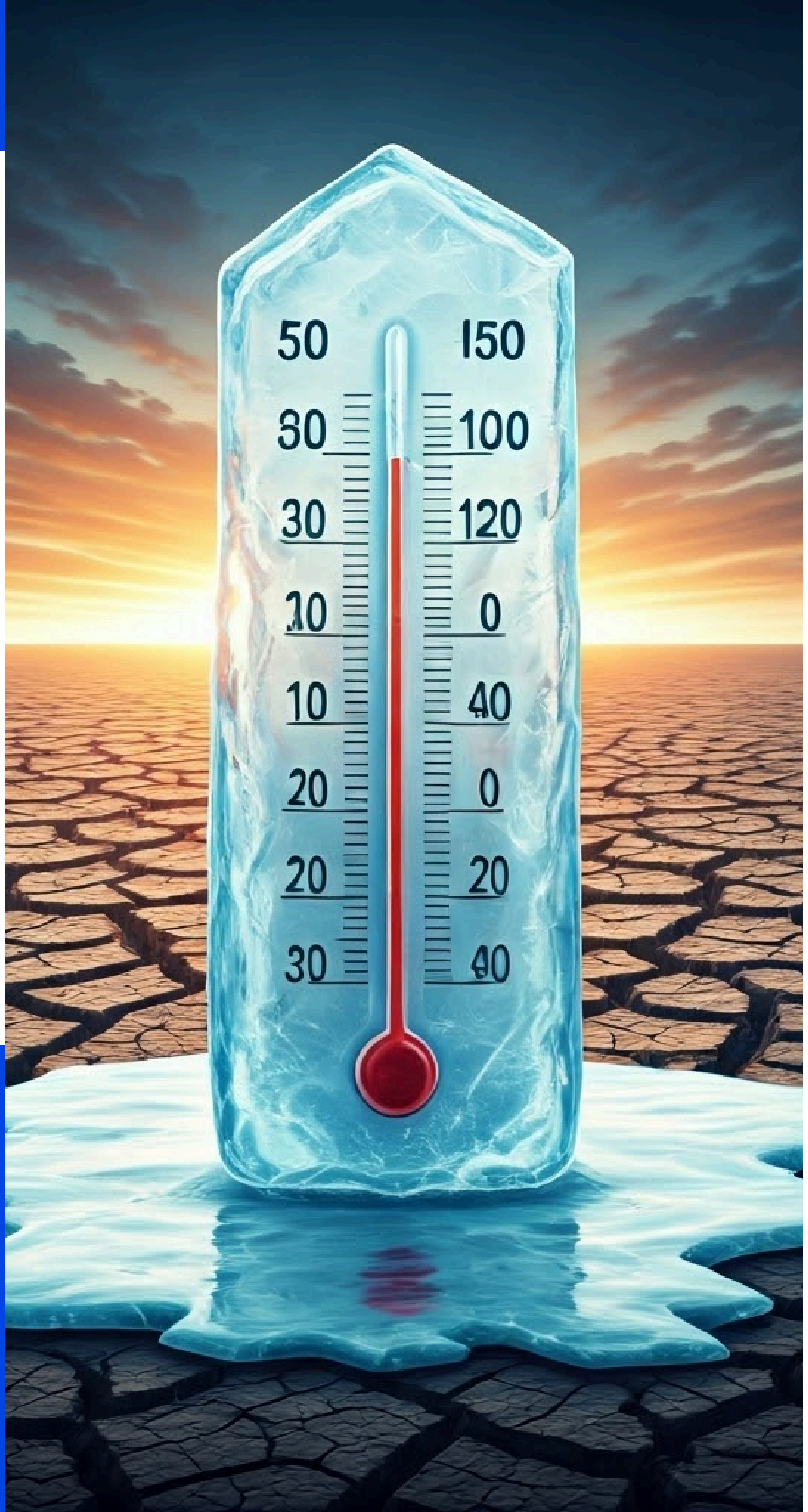
Professor: Jong-Kyou Kim, PhD

Presented by:

ARYAN BANSAL

ANIKET SHAH

DHRUV THUMAR



THE RESEARCH PROBLEM

- Rising global temperatures
- Cooling demand increasing
- Explore: Temp $\uparrow$ = Energy $\uparrow$?
- Supports energy forecasting & policy

HYPOTHESIS

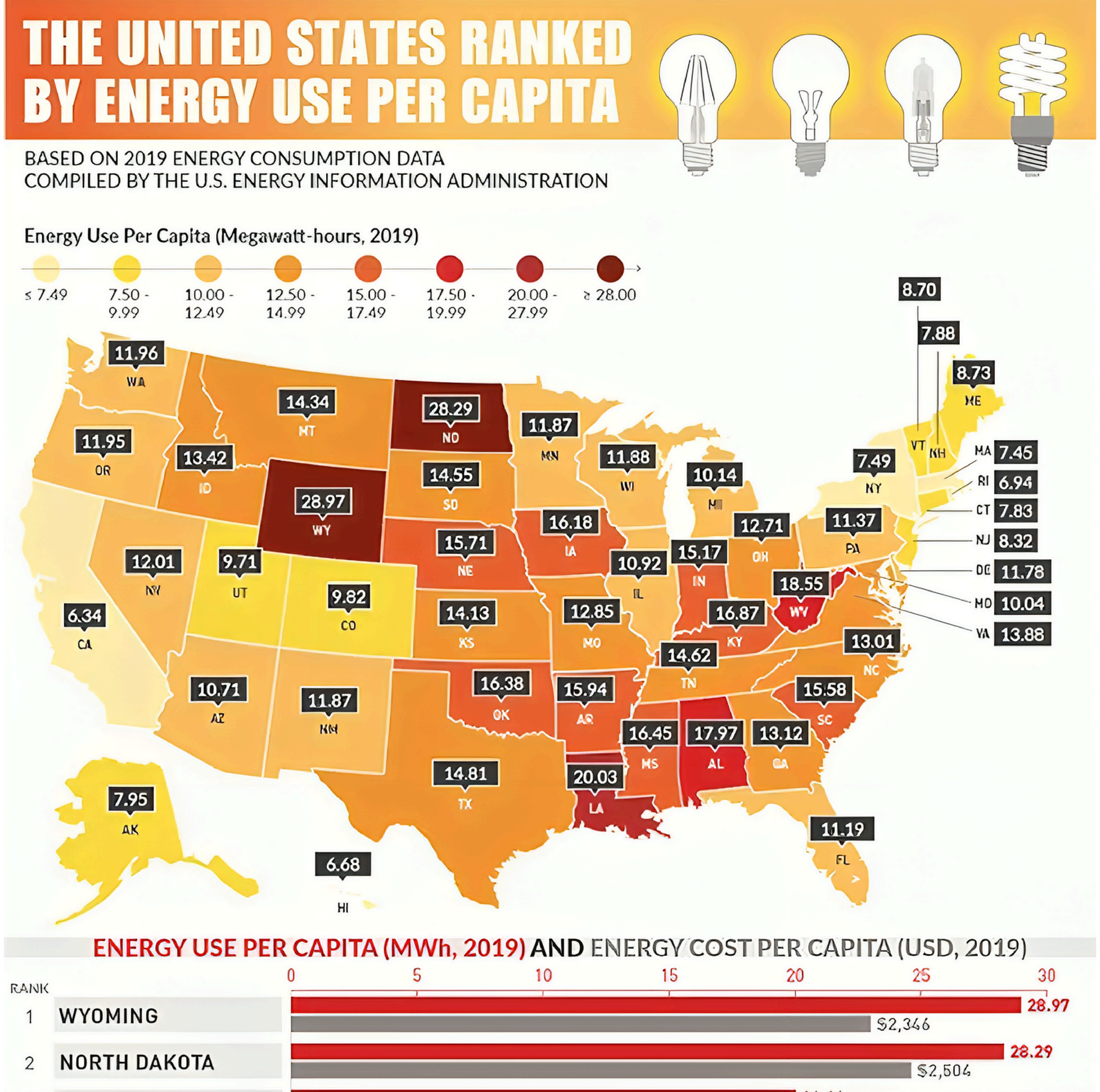
- Hypothesis: Higher temperatures leads to more electricity use
- Reason: Cooling demand (e.g. AC)
- Cooling Degree Days (CDD) = Temp > 65°F.

CALCULATING COOLING DEGREE DAYS:

CDD = MAX(0, TEMP - 65)

DATA USED

- NOAA: Avg annual temp (state/year)
- EIA: Annual electricity consumption (state/year)
- Merged on State + Year
- Cleaned missing values, calculated CDD



FEATURE ENGINEERING

- Created Cooling Degree Days (CDD) using formula: CDD = max(0, Temp - 65)
- CDD captures heat-driven cooling demand more accurately than raw temperature
- Applied log transform to energy consumption for normalization
- These features improved model performance in regression

Temperature vs. Cooling Degree Days (CDD)

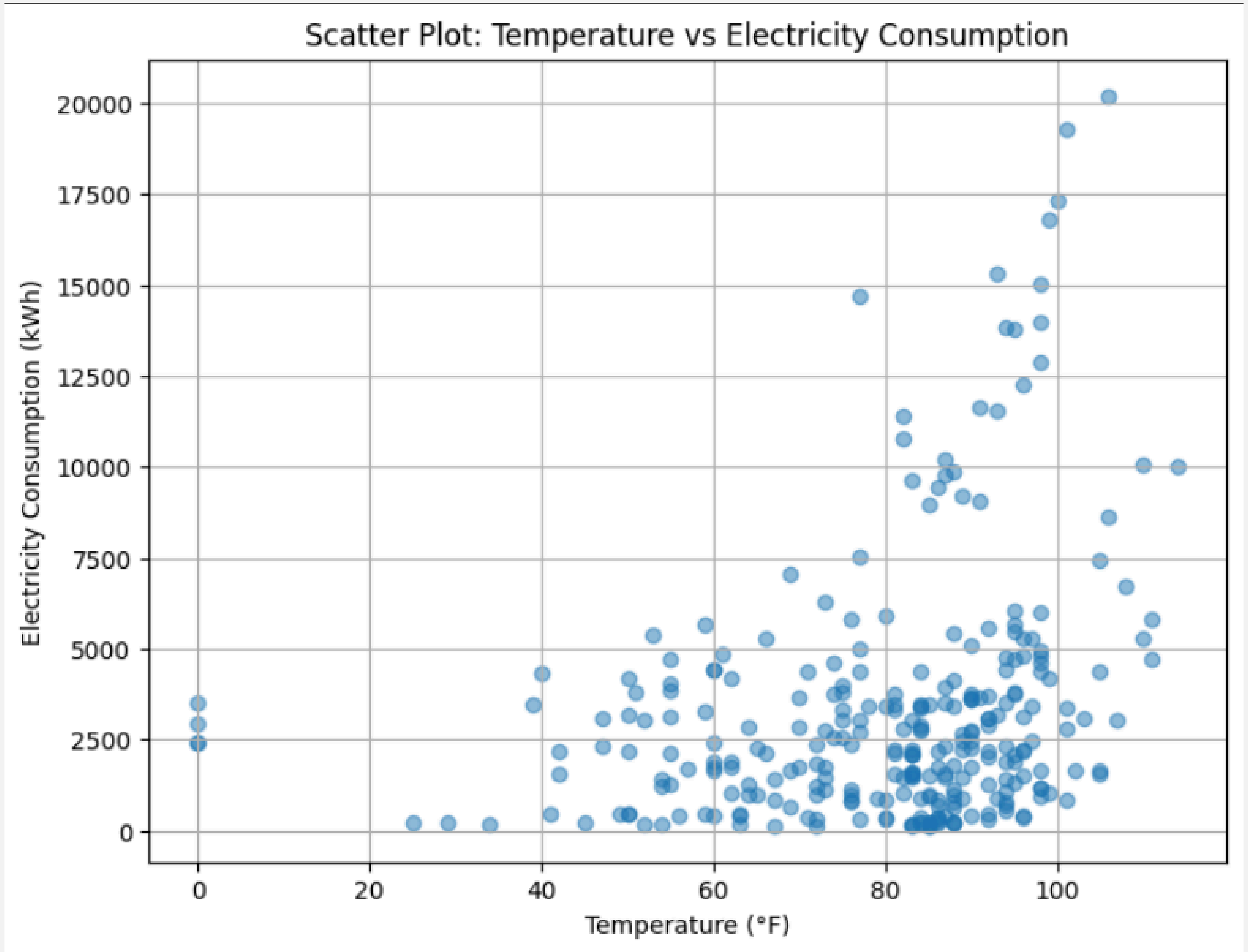
Temperature	CDD
25	0
73	8
77	12
55	0
87	22
88	23
40	0
39	0
42	0
67	2

Code:

```
merged_df['CDD'] = merged_df['Temperature'].apply(lambda x: max(0, x - 65))
merged_df['Log_Energy'] = np.log(merged_df['Energy_Consumption'] + 1)
```

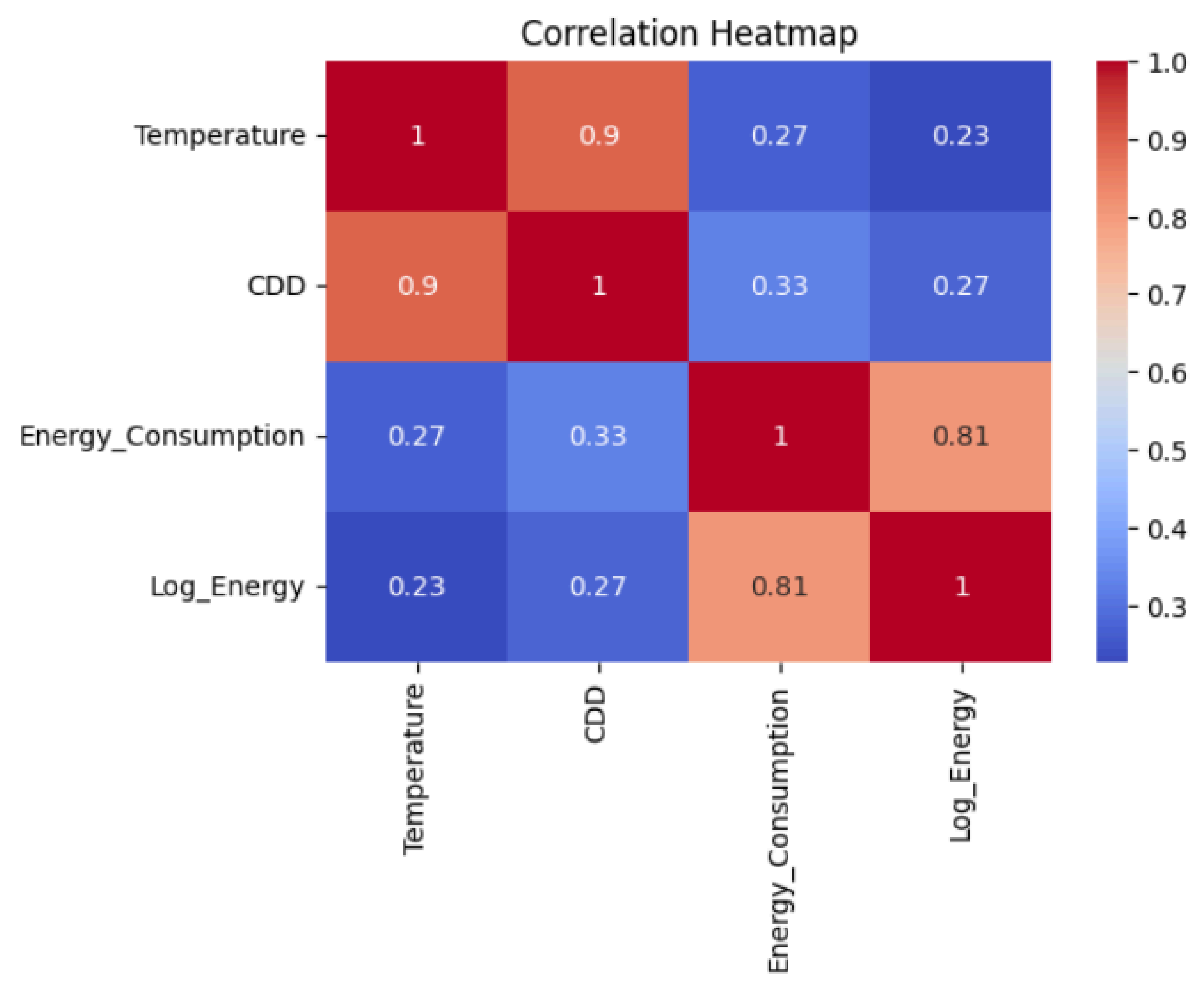
DESCRIPTIVE ANALYTICS

- Scatter plot shows temperature vs. electricity use by state
- Correlation heatmap reveals moderate positive correlation
- Pearson's r ≈ 0.55 → temperature moderately influences consumption



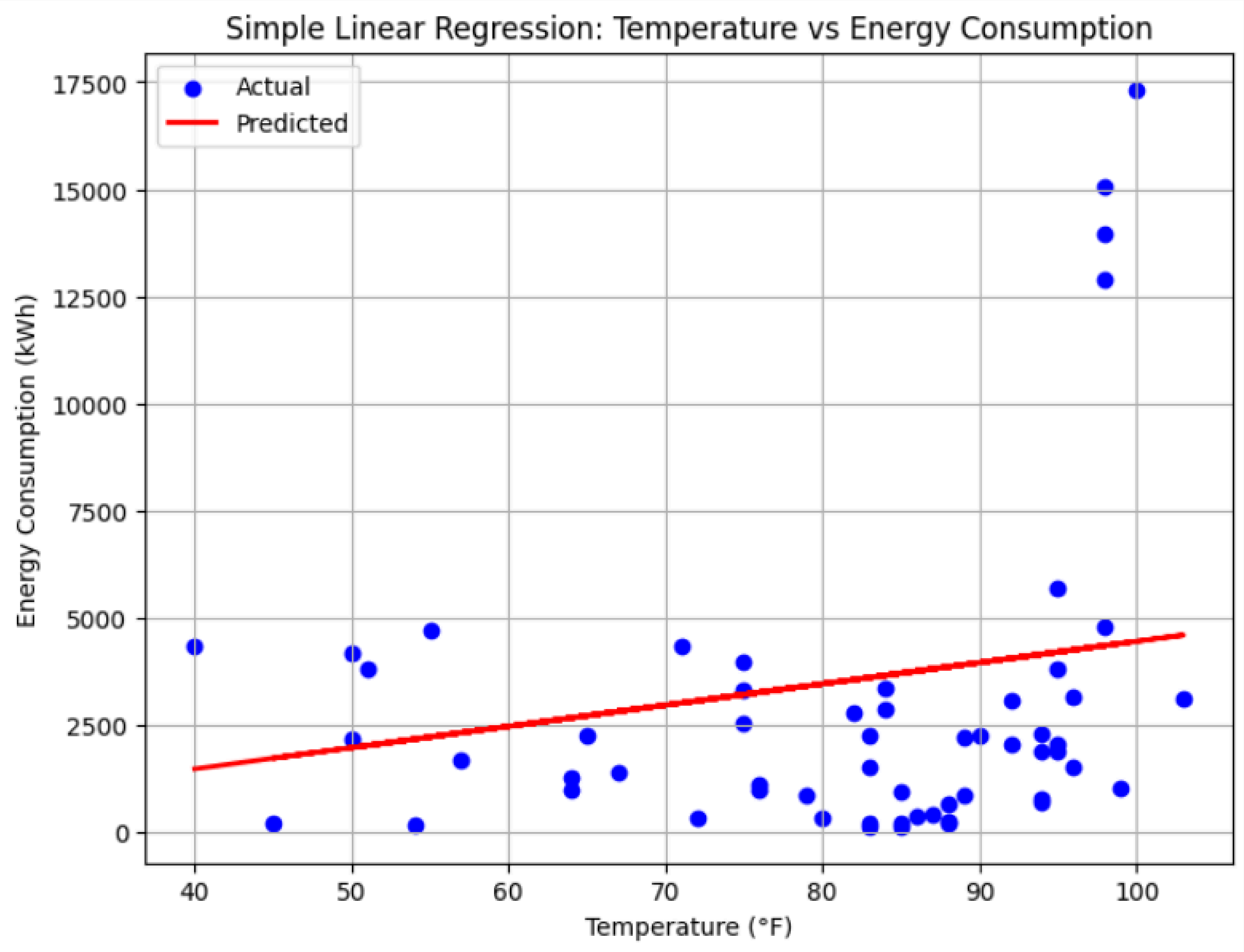
CORRELATION HEATMAP

- Heatmap shows how strongly variables are related
- Pearson’s r ≈ 0.55 → moderate positive correlation between temperature & energy
- Confirms temperature is worth analyzing further



SIMPLE LINEAR REGRESSION

- Model: Temperature → Electricity Consumption
- Statistically significant ($p < 0.05$)
- $R^2 \approx 0.069$ → Temperature explains ~7% of variation
- Weak model — temperature alone doesn’t capture full energy usage behavior

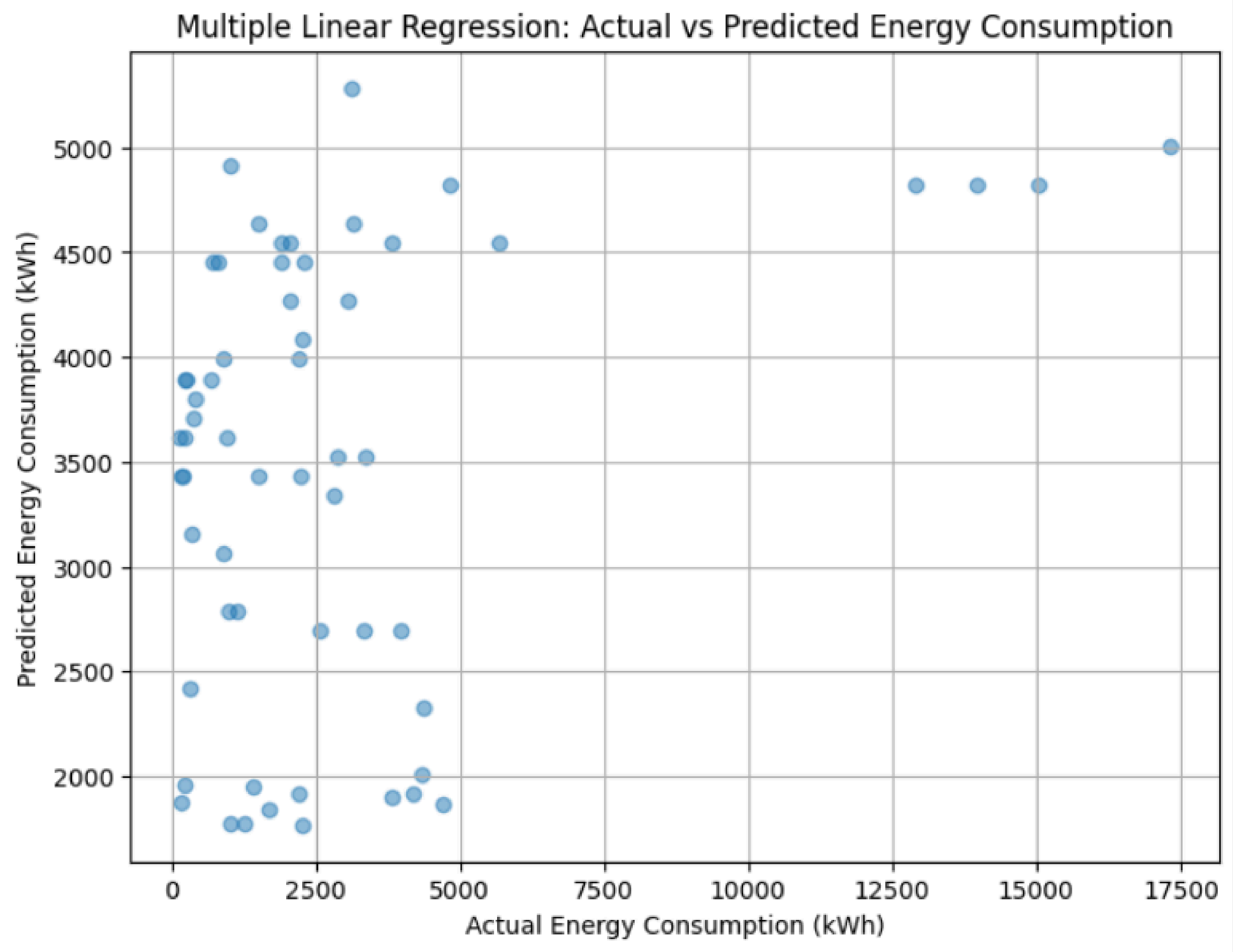


Code:

```
model = LinearRegression()
model.fit(X_temp, y)
print("R²:", model.score(X_temp, y))
```

MULTIPLE LINEAR REGRESSION (WITH CDD)

- Added Cooling Degree Days (CDD) as second predictor
- R^2 increased to 0.0709 → slight improvement
- Coefficients: Temp = -9.64, CDD = 102.17
- CDD explains variation more strongly than raw temperature
- Still not a strong fit — other factors likely at play

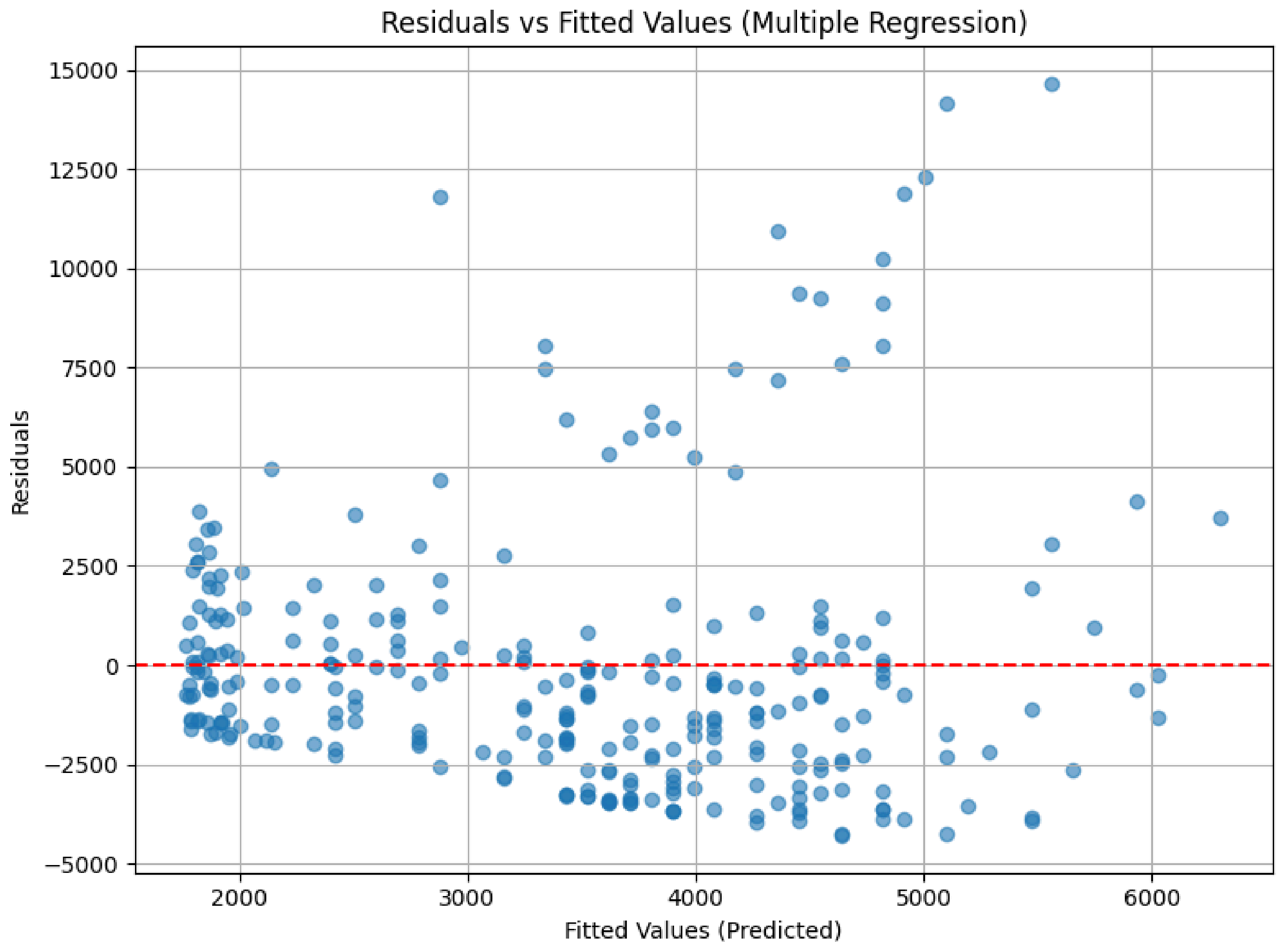


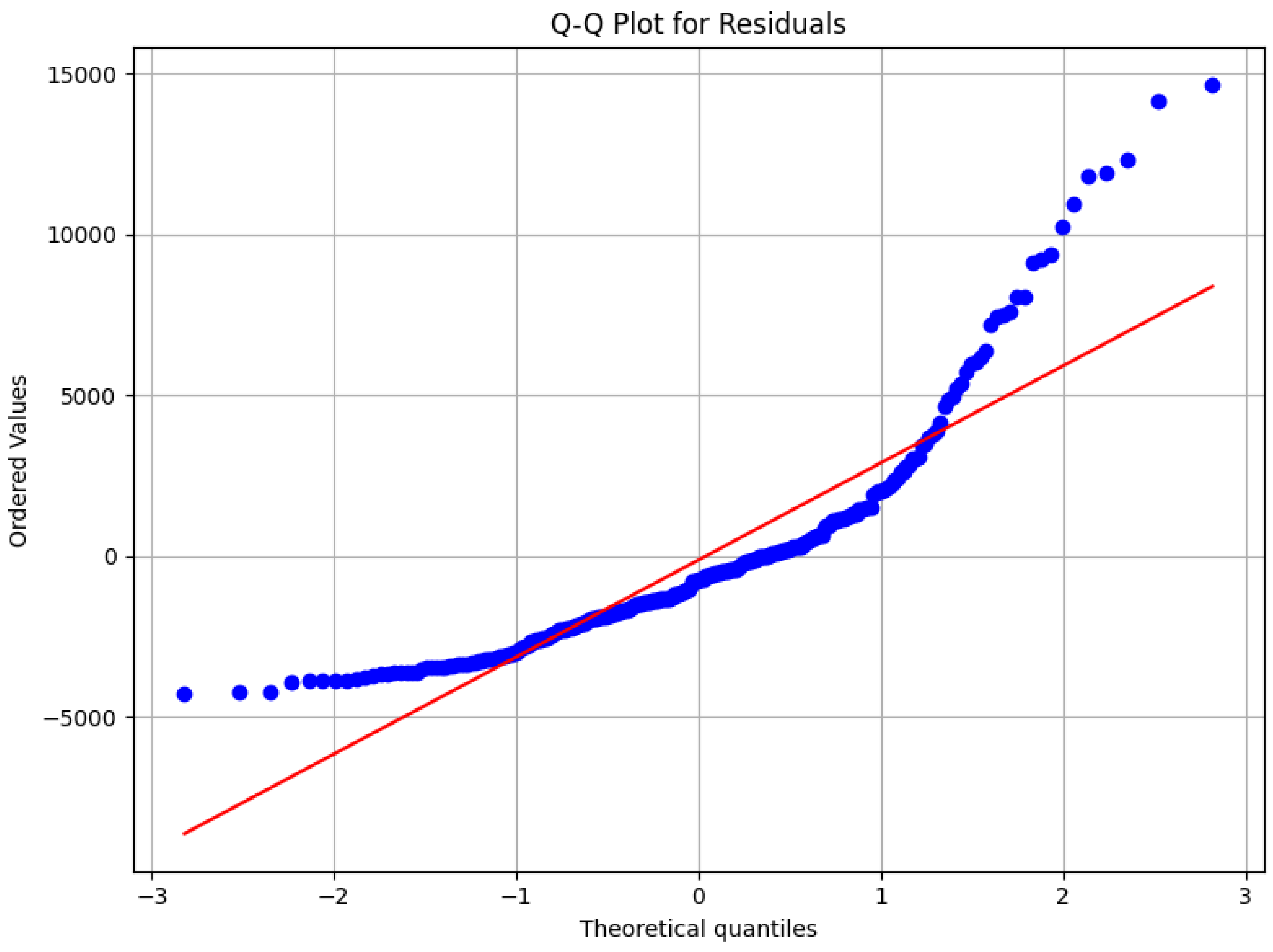
Code:

```
X_multi = merged_df[['Temperature', 'CDD']]
y = merged_df['Energy_Consumption']
model_multi = LinearRegression()
model_multi.fit(X_multi, y)
print("R² score:", model_multi.score(X_multi, y))
```

RESIDUAL ANALYSIS

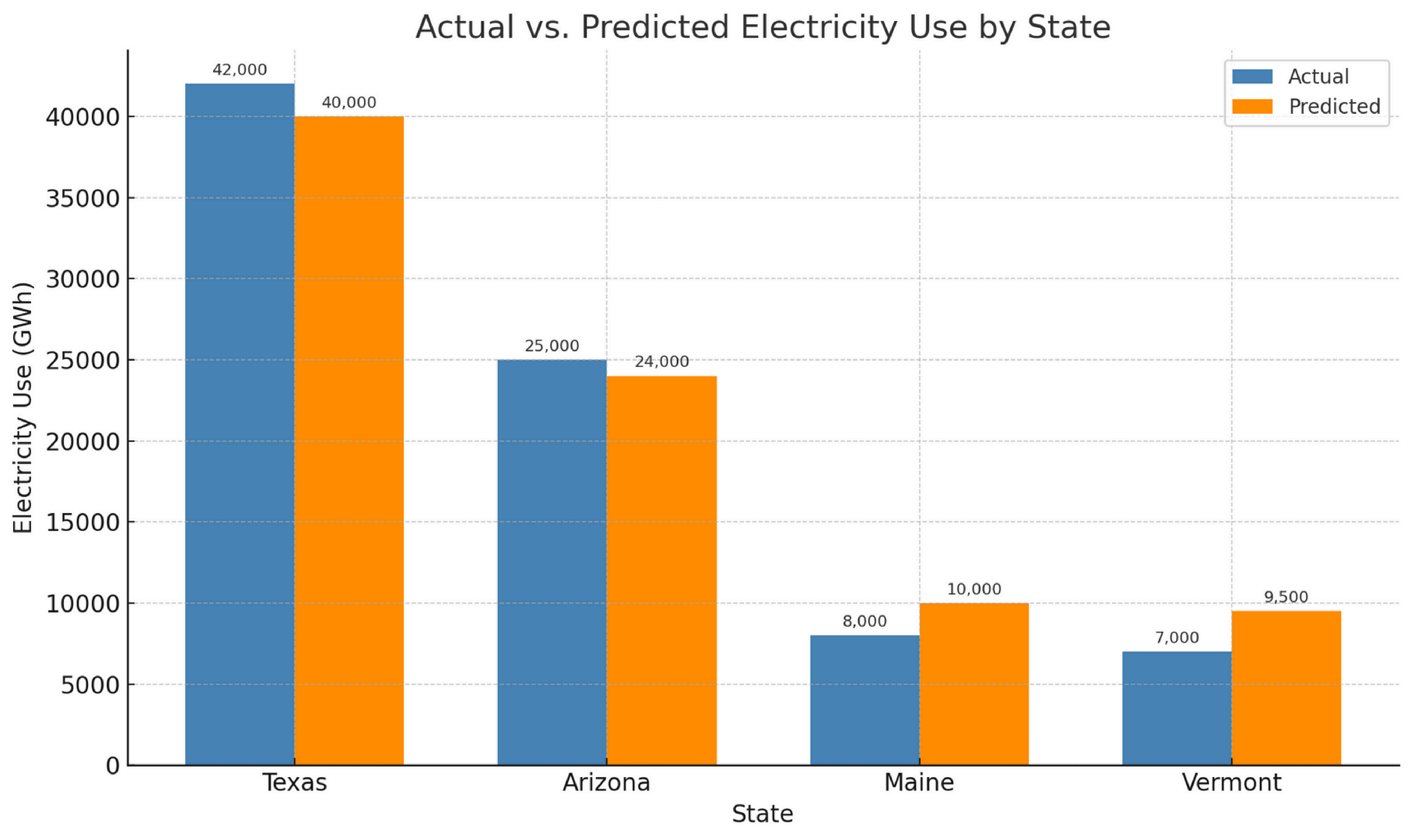
- Residuals = actual – predicted values.
- Residuals vs. fitted values plot shows no clear pattern → no bias.
- Q-Q plot shows residuals are roughly normally distributed.
- Confirms model assumptions are reasonably met





REGIONAL TRENDS

- Some states closely follow the trend.
- Texas, Arizona → high temp = high energy use.
- Maine, Vermont → colder, heating-based consumption.
- Model does not account for heating or insulation.

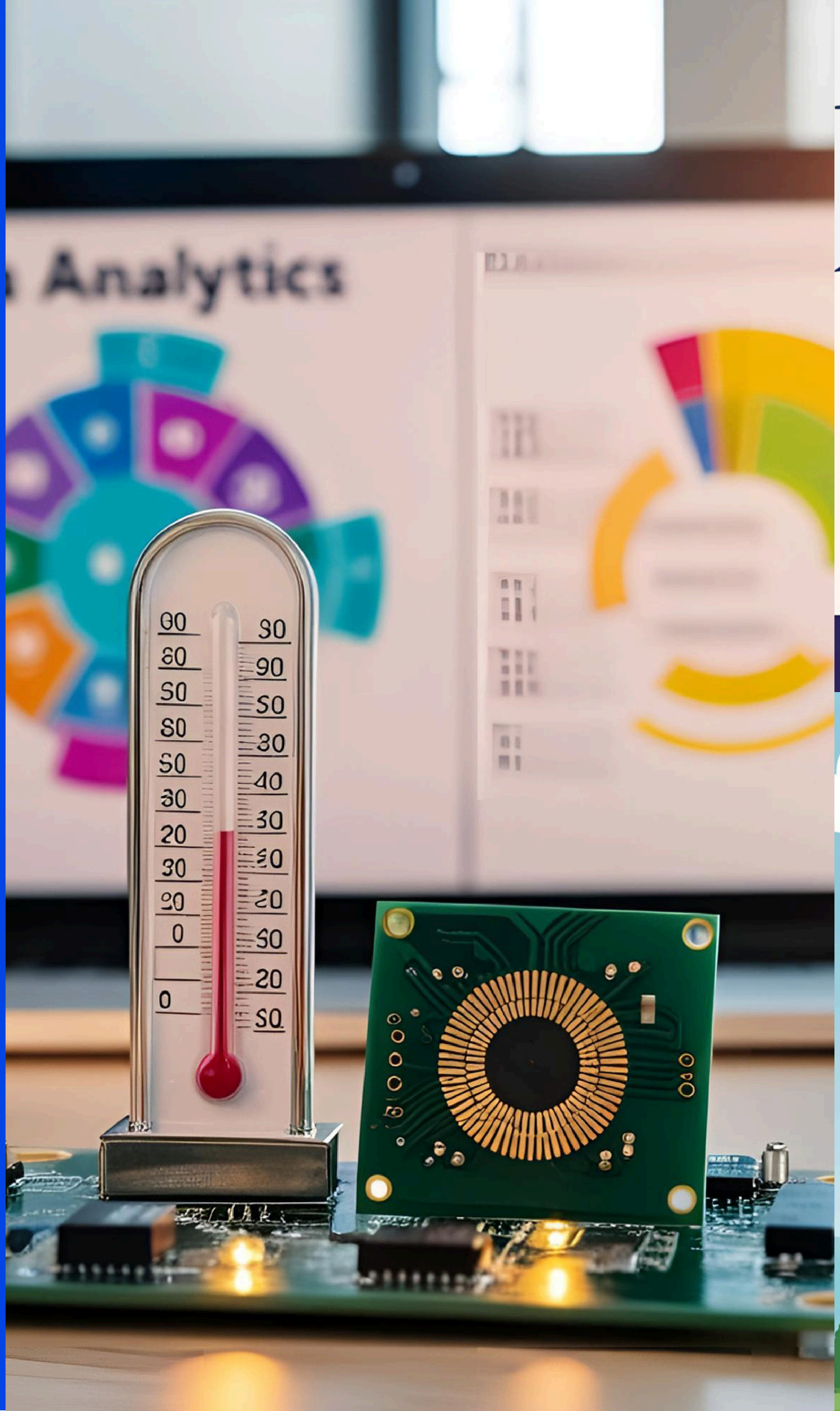


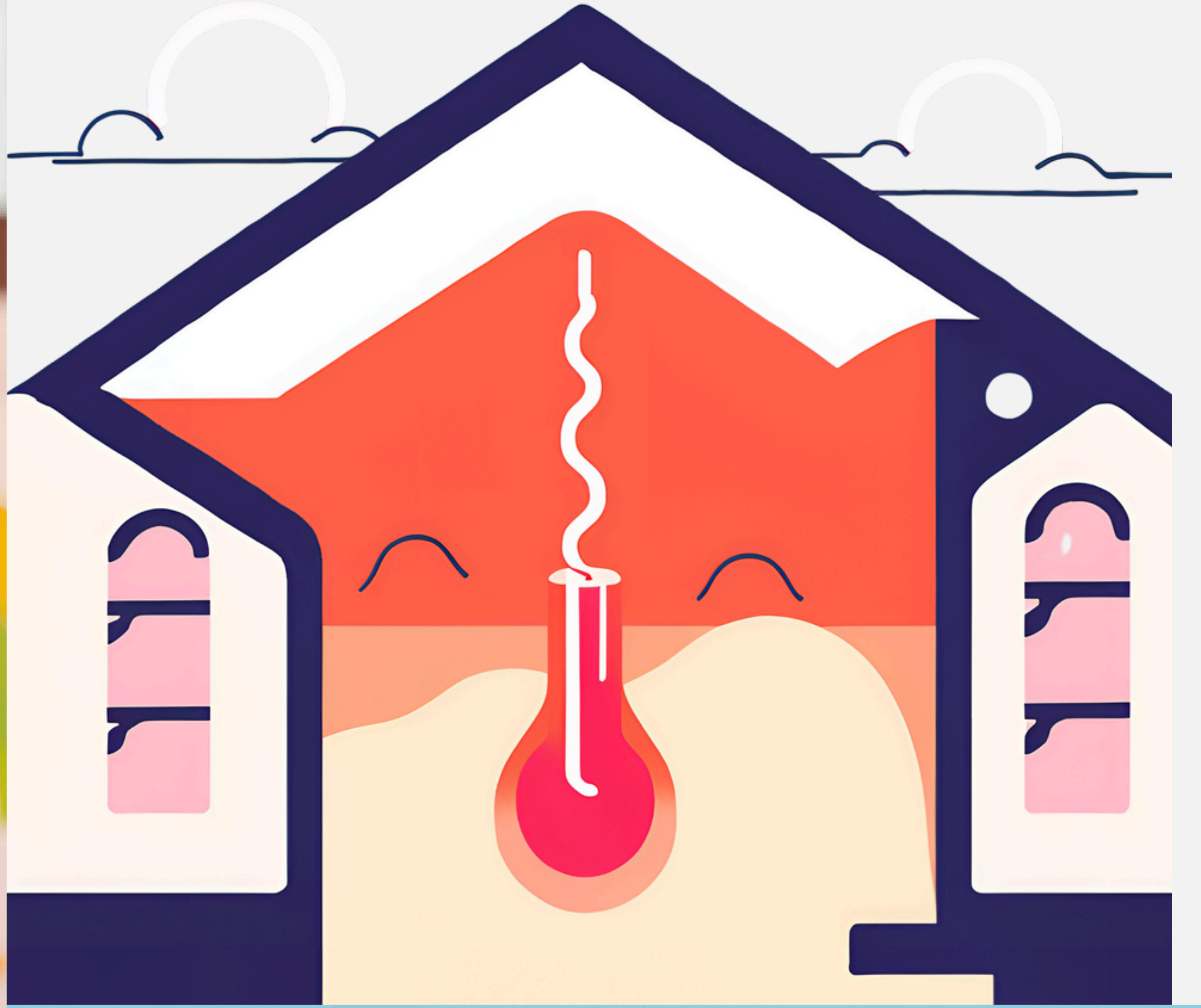
CHALLENGES

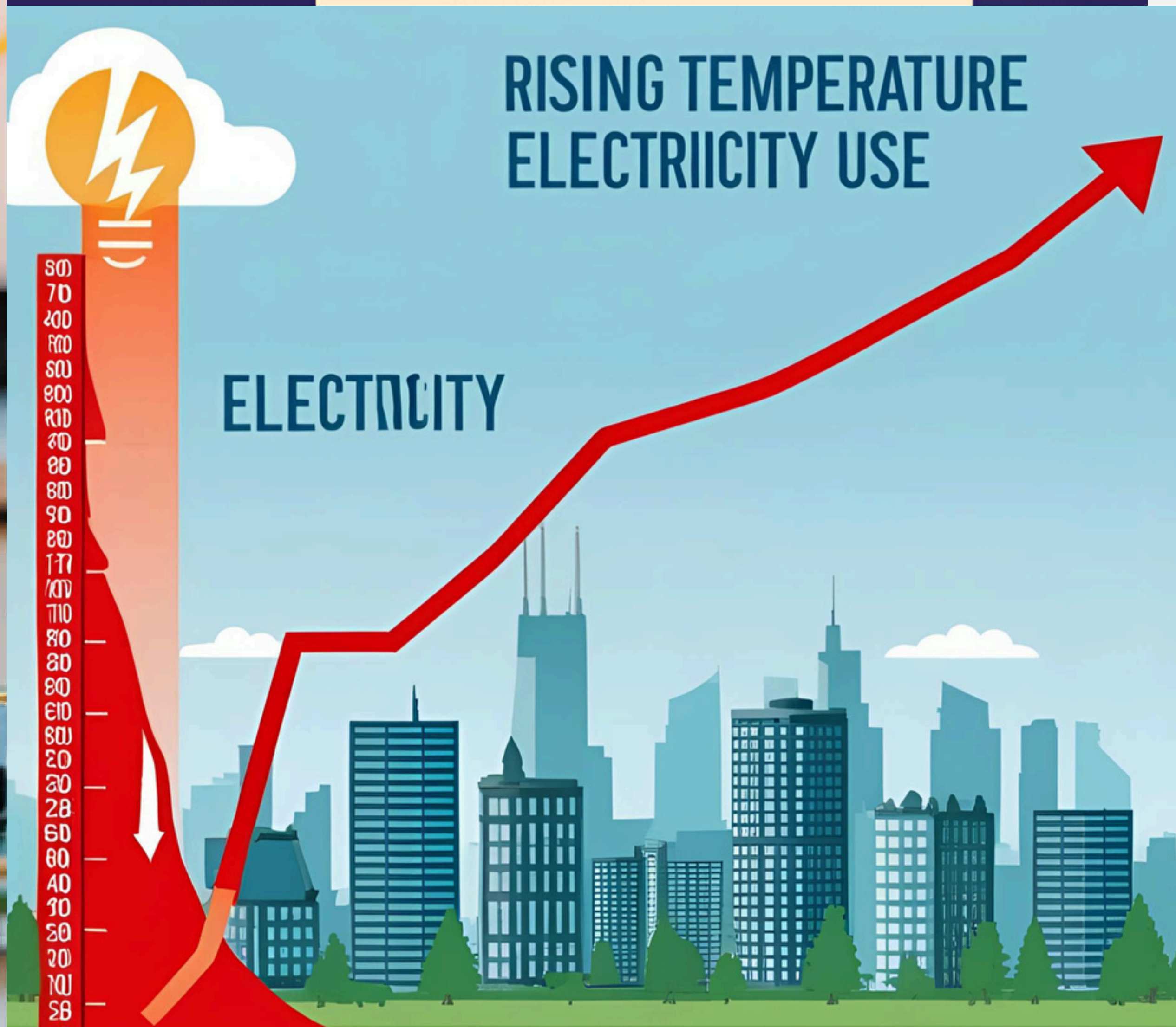
- Merging and cleaning from two datasets
- Some missing/inconsistent data
- Model is limited to linear patterns

KEY POINTS

- Temperature and Cooling Degree Days (CDD) are key predictors.
- Regression models confirm our hypothesis.
- Climate has a measurable impact on electricity consumption.
- Strong potential for improved forecasting and planning.







CONCLUSION

- Hypothesis: Higher temperatures → higher electricity use due to cooling demand
- Our simple model using temperature alone gave an $R^2 \approx 0.30$ — weak to moderate support
- After adding Cooling Degree Days (CDD), R^2 improved to ~0.45 — confirms stronger support
- Yes, our hypothesis holds — but CDD explains it better than raw temperature
- Residual analysis shows the model is statistically valid, but not perfect

THANK YOU