Predicting Power Outages

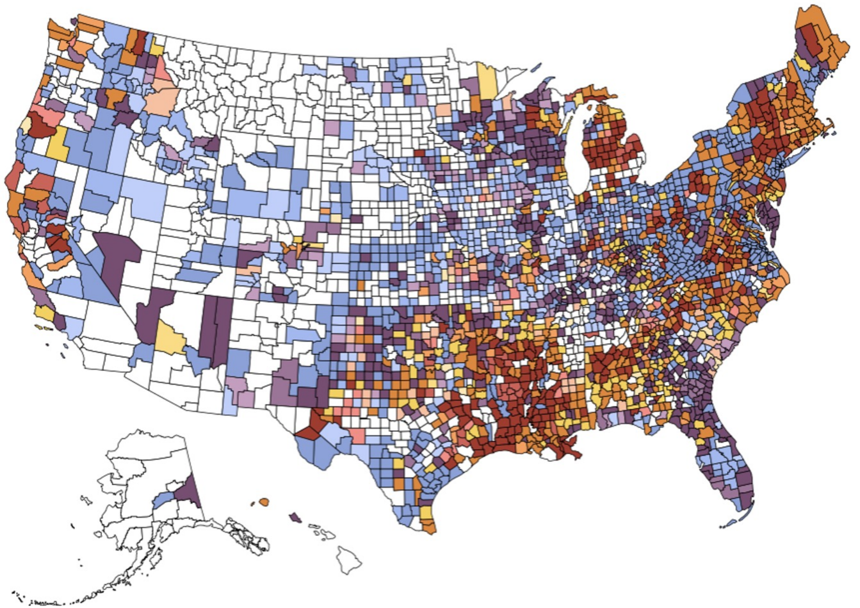
Julio Caceres, Evan Morris, Aaron Weinberg, Anna Zuckerman

Image: NASA

Motivation

Modern life runs on access to electricity!

Outages impact livelihoods + the economy:



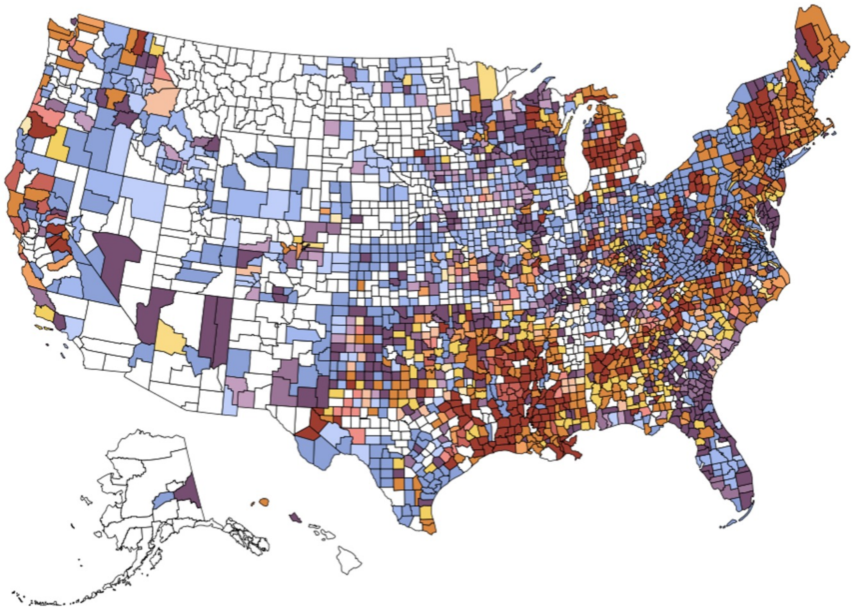
Heatmap of US outages 2018-21 [8]

Motivation

Modern life runs on access to electricity!

Outages impact livelihoods + the economy:

$ ~$44 billion cost to US annually [5]



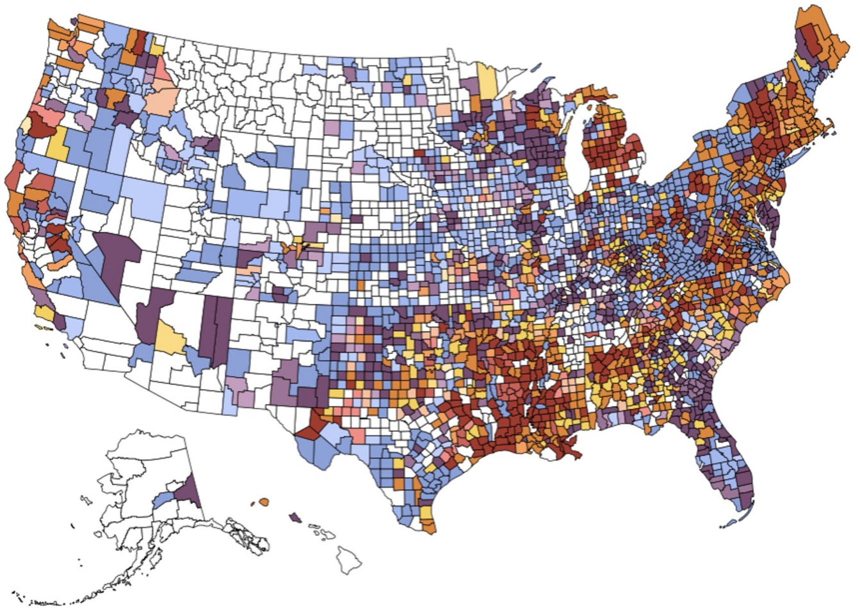
Heatmap of US outages 2018-21 [8]

Motivation

Modern life runs on access to electricity!

Outages impact livelihoods + the economy:

- ~$44 billion cost to US annually [5]
- Monthly mortality increases 0.04% per hour of power outage [1]



Heatmap of US outages 2018-21 [8]

Motivation

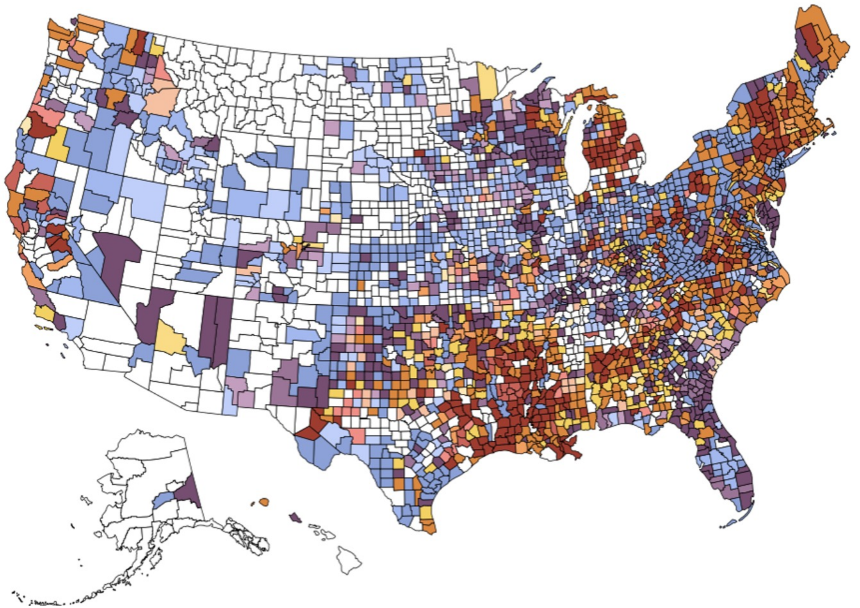
Modern life runs on access to electricity!

Outages impact livelihoods + the economy:

~$44 billion cost to US annually [5]

Monthly mortality increases 0.04% per hour of power outage [1]

Strain emergency services



Heatmap of US outages 2018-21 [8]

Motivation

Modern life runs on access to electricity!

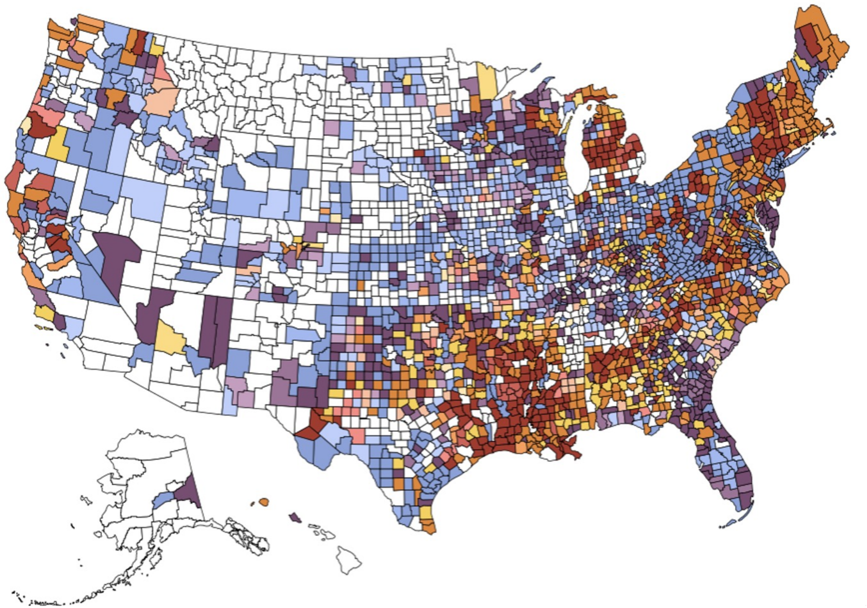
Outages impact livelihoods + the economy:

~$44 billion cost to US annually [5]

Monthly mortality increases 0.04% per hour of power outage [1]

Strain emergency services

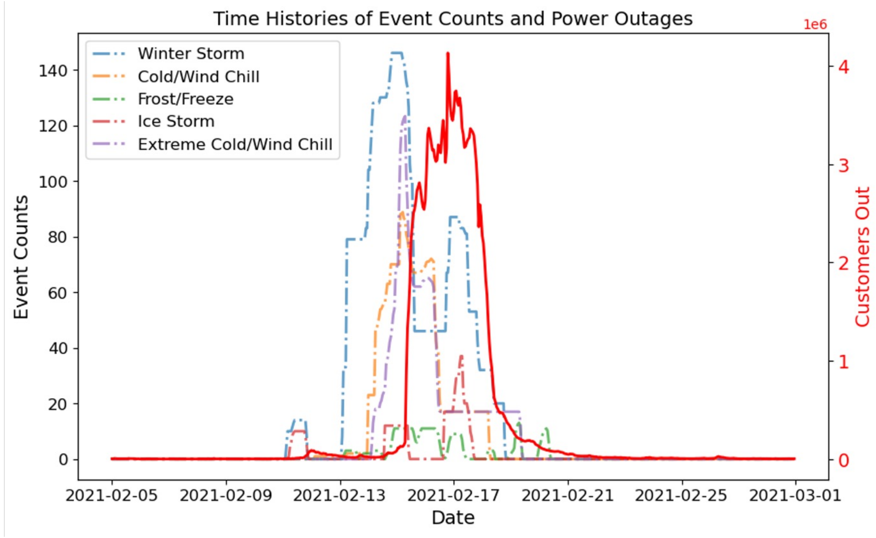
Damage to power infrastructure



Heatmap of US outages 2018-21 [8]

Modeling task

Challenge:
Create a reliable system to accurately predict power outages



Severity of power outage can be associated with extreme weather events

Figure: ThinkOnward, Dynamic Rhythms project introduction

Modeling task

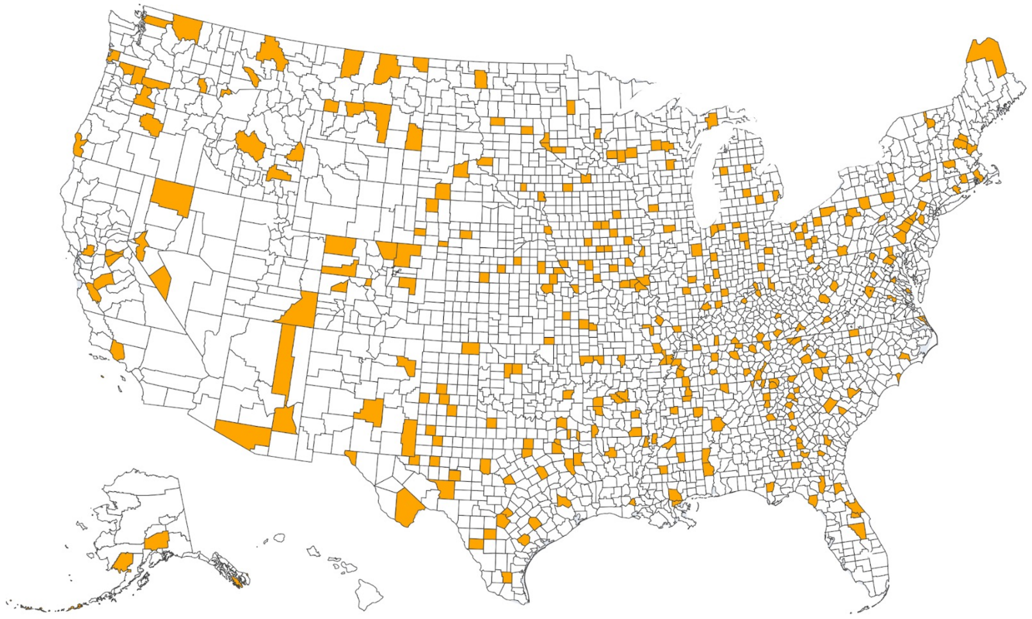
Using **weather data** from the past 5 days, predict the maximum fraction of **people without power** at the county level tomorrow

Modeling task

Target: County-Level maximum fraction customers without power 2014-2023

Modeling task

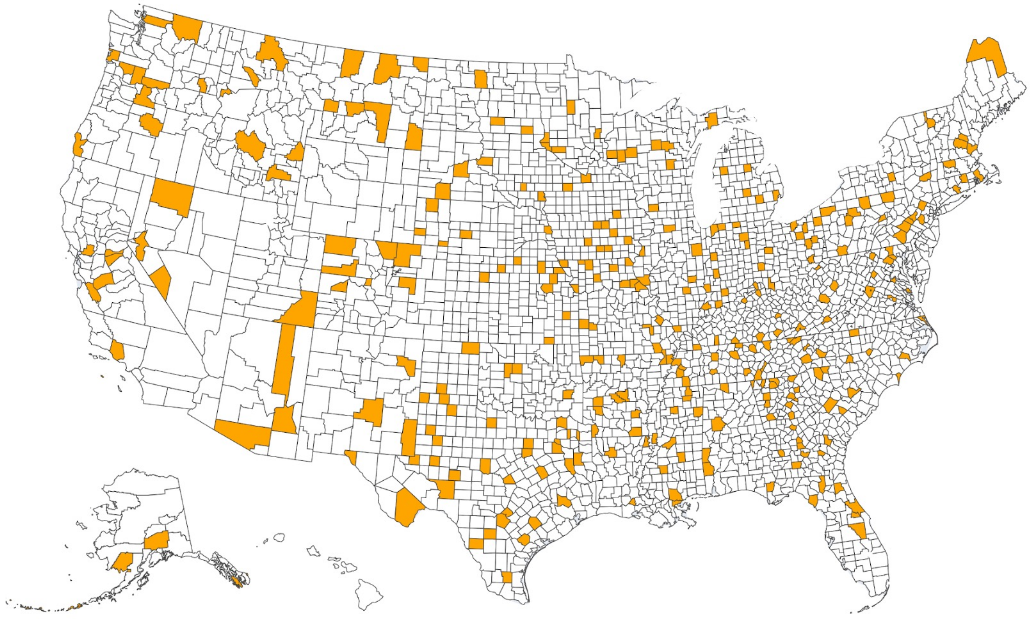
Target: County-Level maximum fraction customers without power 2014-2023



Environment for the Analysis of Geo-Located Energy Information (EAGLE-I) dataset

Modeling task

Target: County-Level maximum fraction customers without power 2014-2023



Environment for the Analysis of Geo-Located Energy Information (EAGLE-I) dataset

↓

Fill missing data

↓

Combine yearly datasets

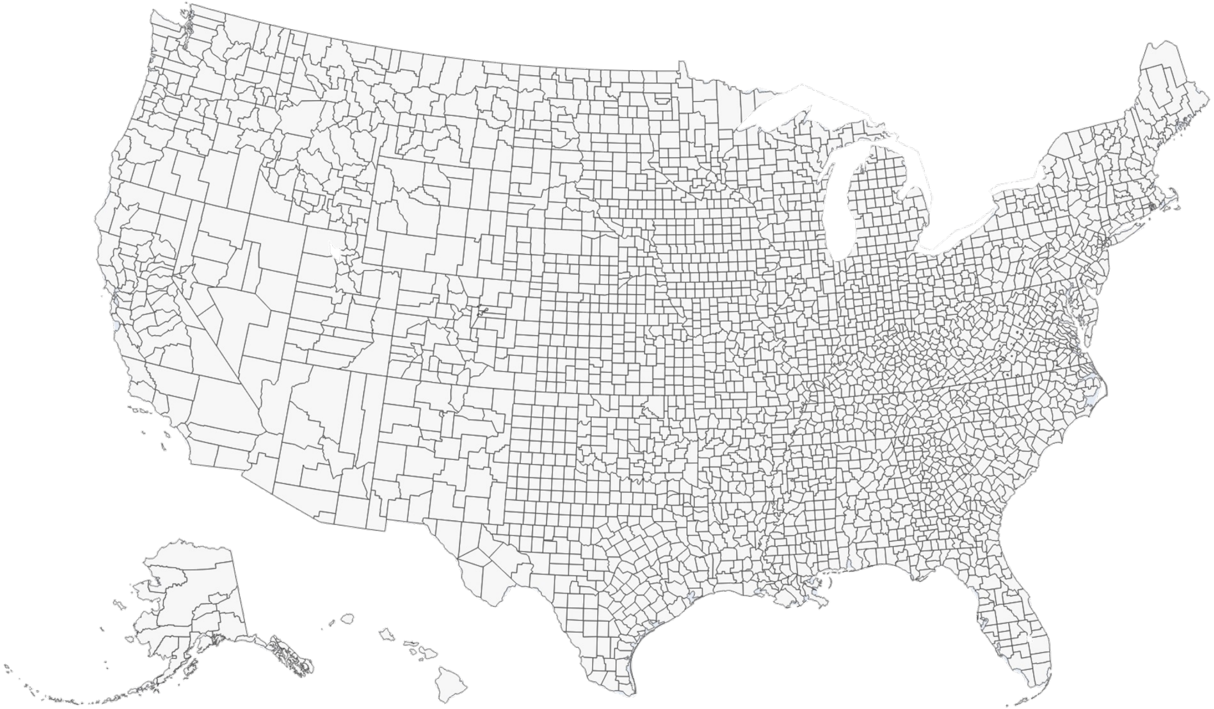
↓

Downsample to 6-hr cadence

↓

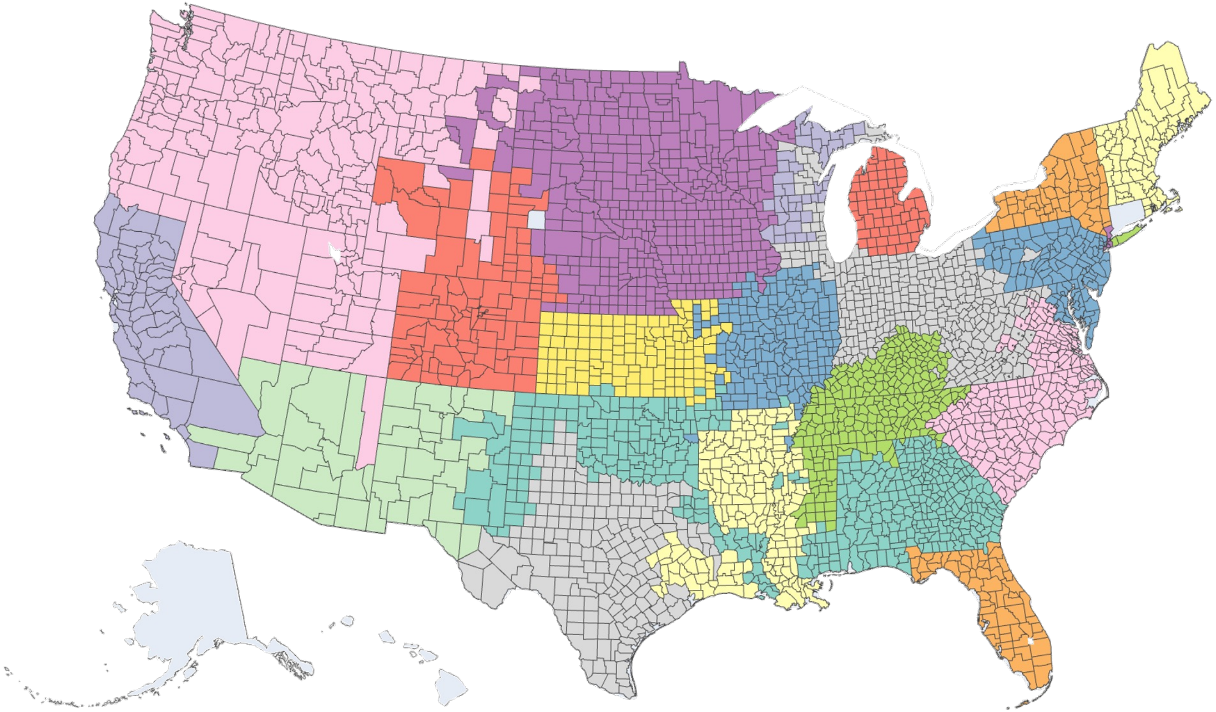
Take maximum over each day

Predictors: County-Level Data



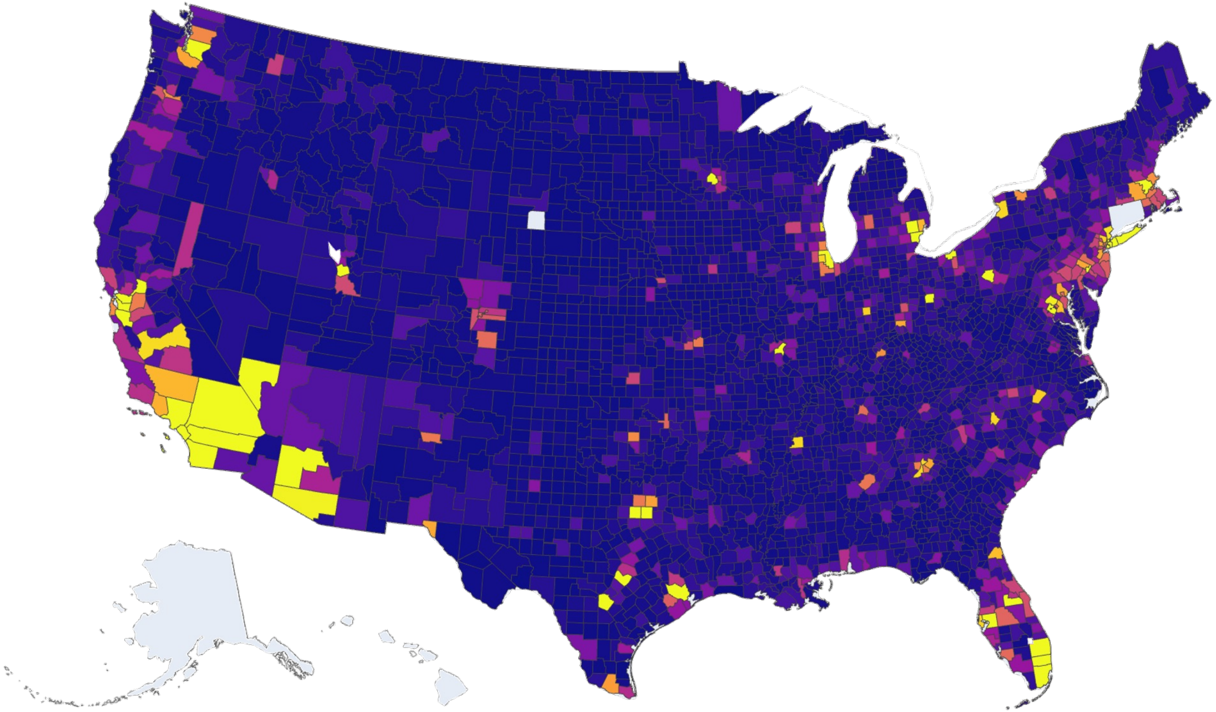
- Census Shapefiles

Predictors: County-Level Data



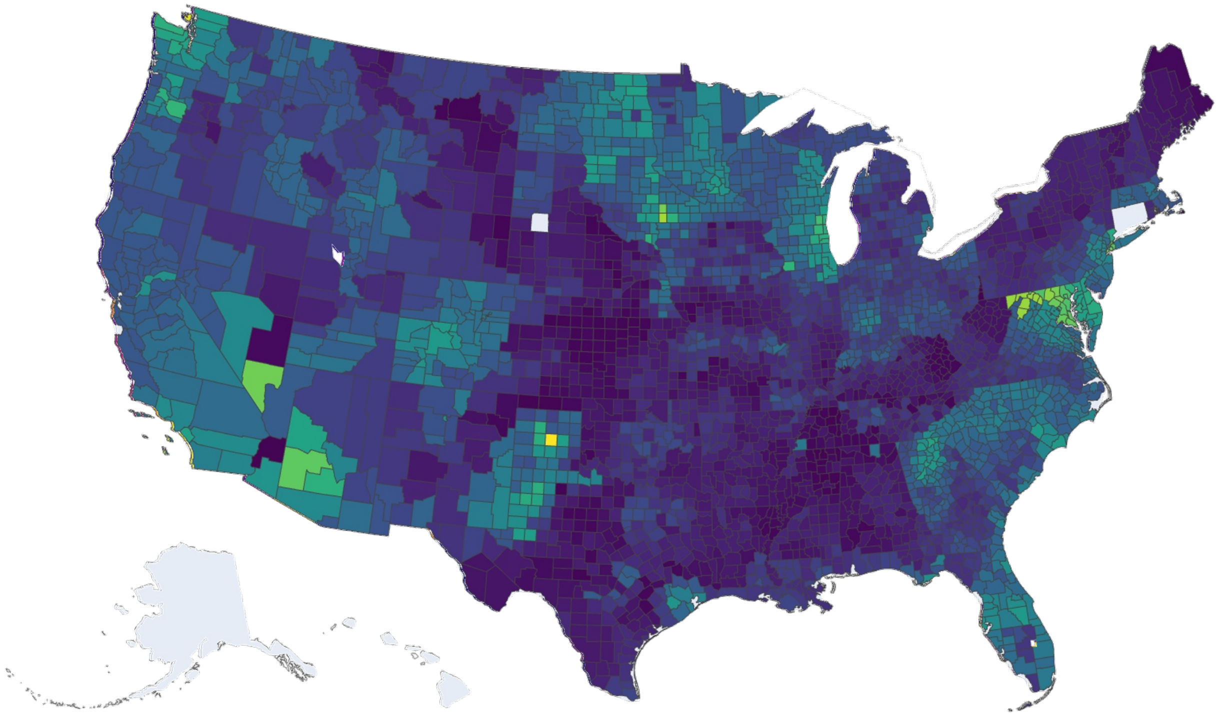
- Census Shapefiles
- EIA Power Grid

Predictors: County-Level Data



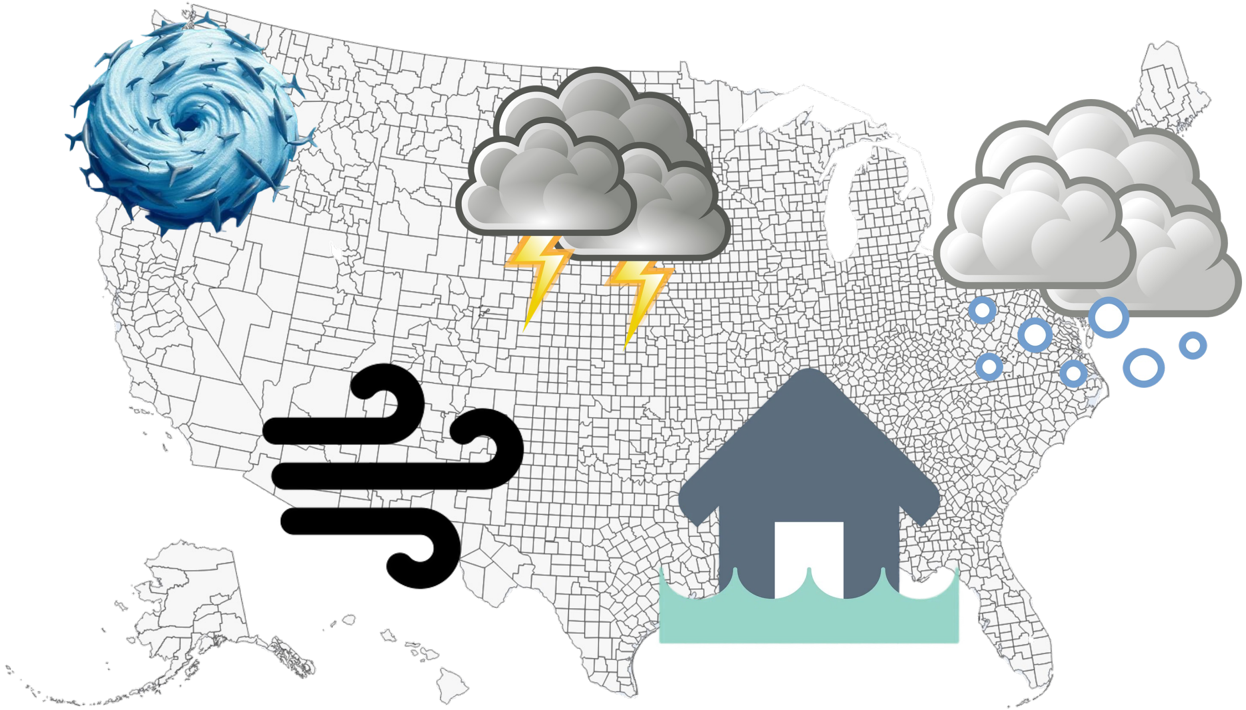
- Census Shapefiles
- EIA Power Grid
- FEMA Pop & Area

Predictors: County-Level Data



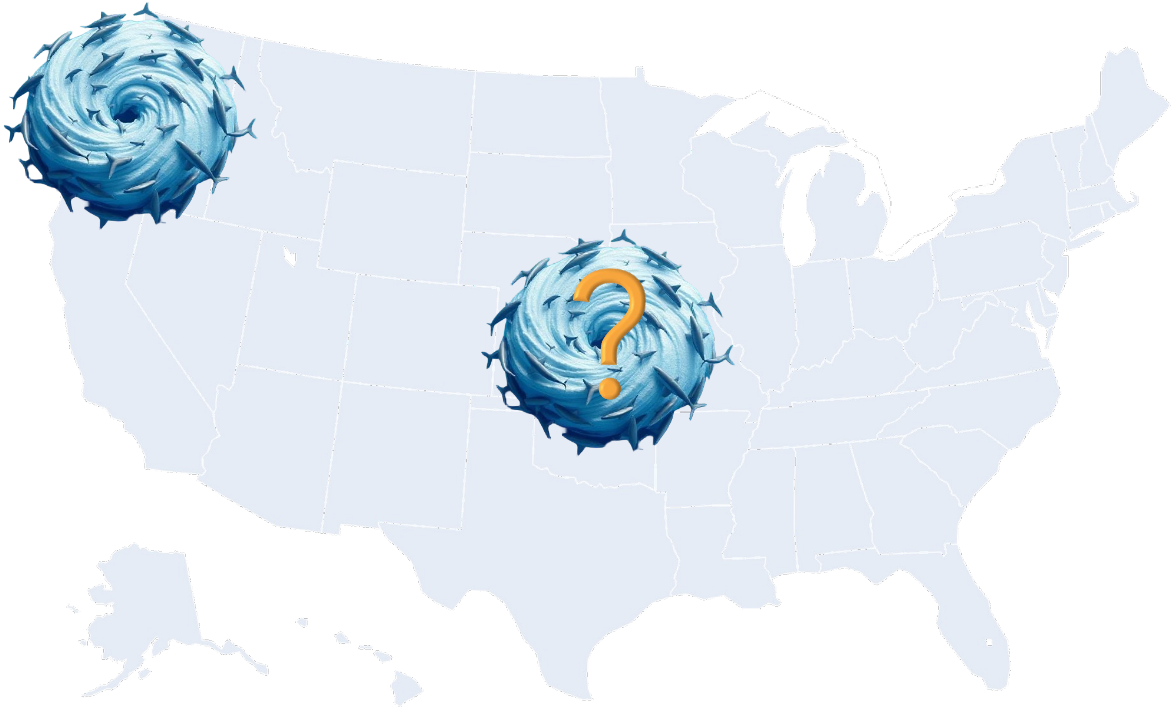
- Census Shapefiles
- EIA Power Grid
- FEMA Pop & Area
- SDC Buried Lines

Predictors: NOAA Extreme Weather



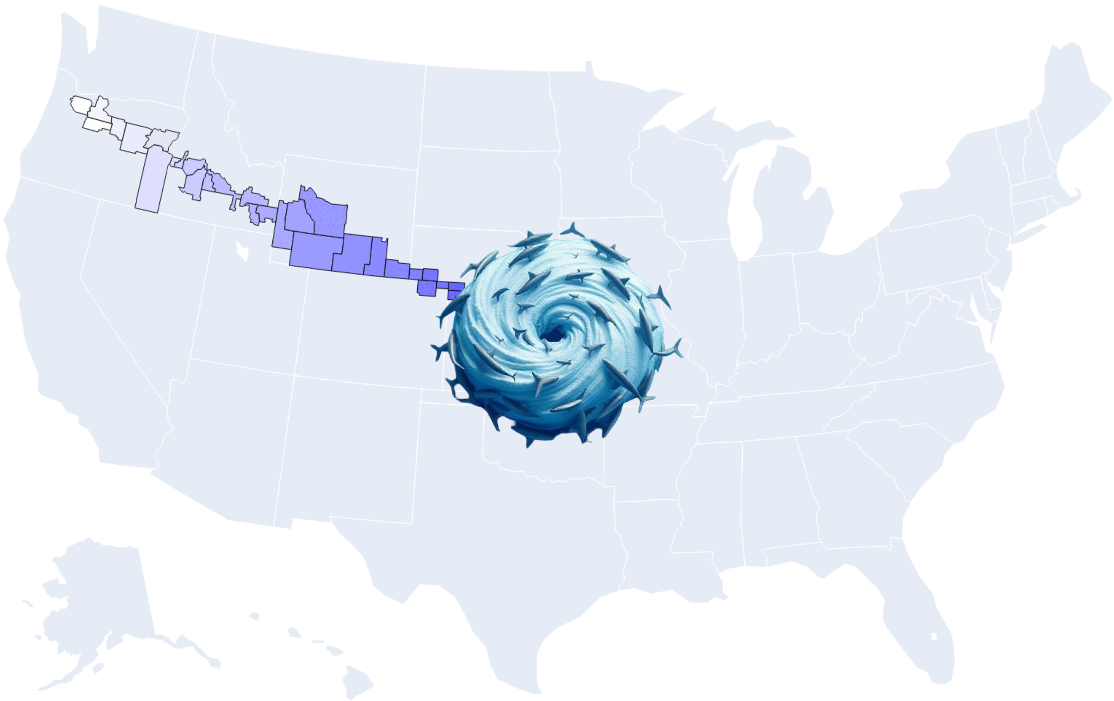
- Begin & End Time
- Narrative
- Location*

Predictors: NOAA Extreme Weather



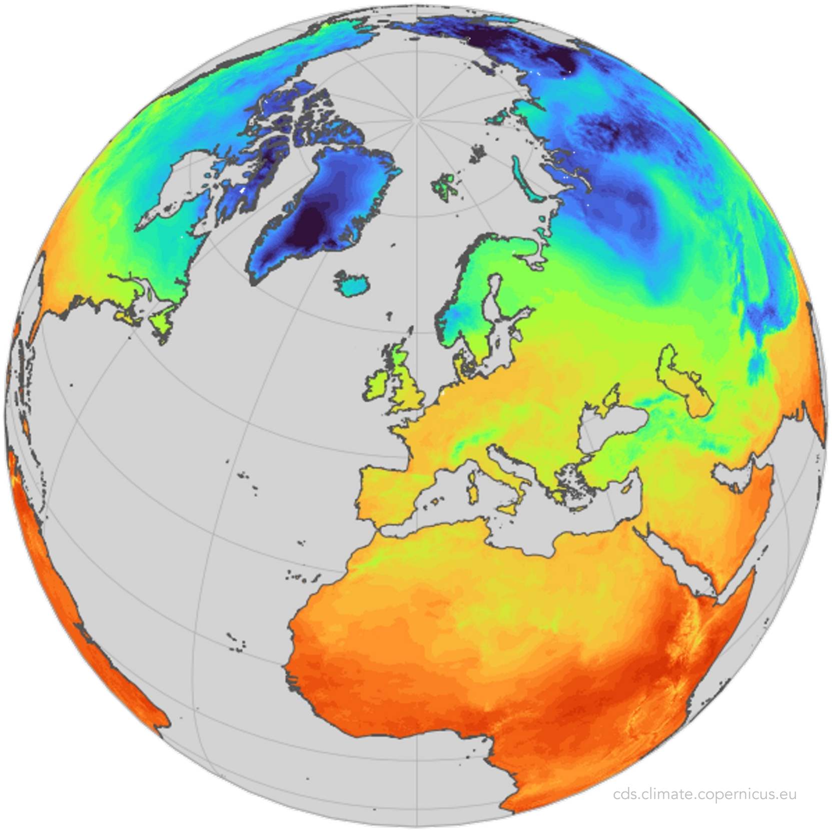
- Classify events
- Identify locations
- Time series

Predictors: NOAA Extreme Weather

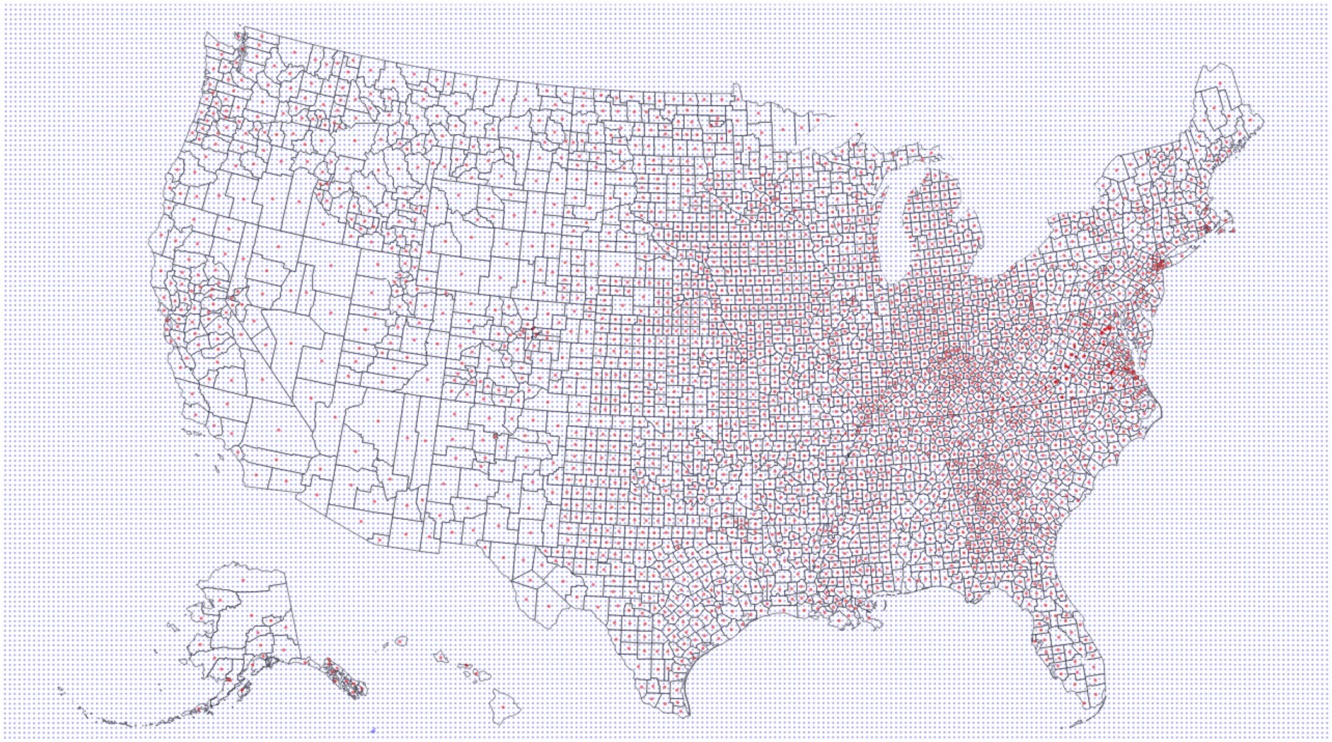


- Classify events
- Identify locations
- Time series
- Identify path
- Compute duration

Predictors: ERA5-Land

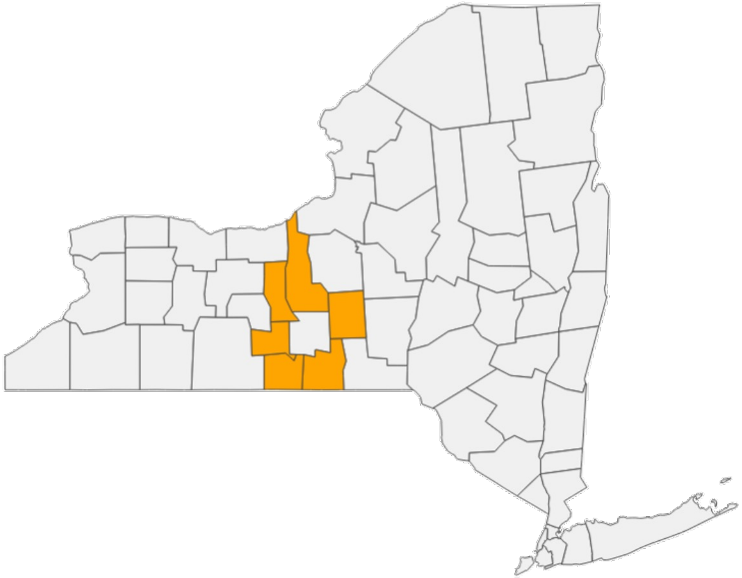


Predictors: ERA5-Land



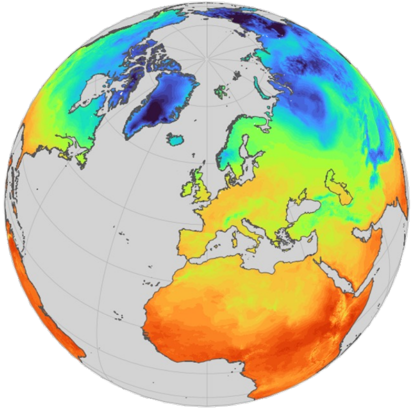
Temperature
Wind Components
Precipitation
Snow Depth
Wind Speed
Cumulative
Precipitation

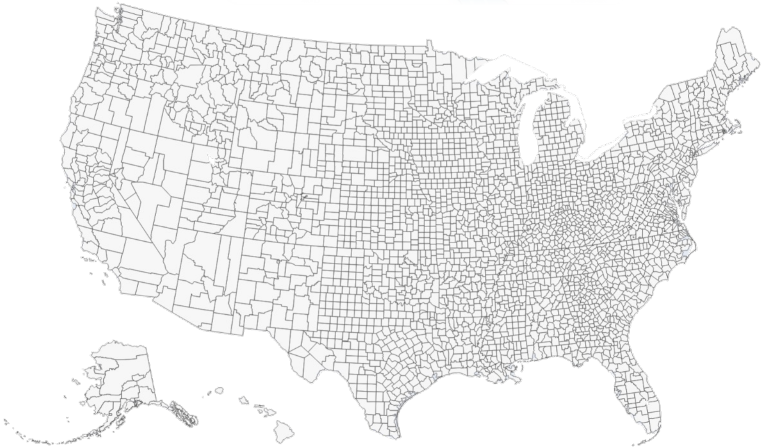
Predictors: ERA5-Land



- Maximum values
- Mean values

Predictors: Merging & Cleaning





Predictors: Merging & Cleaning

FIPS	Datetime	Percent Customers Out	Weather Events	Weather	County Data
23001	2014-11-01 00:00				
	2014-11-01 06:00				
	2014-11-01 12:00				
	2014-11-01 18:00				
	2014-11-02 00:00				
	2014-11-02 06:00				

Modelling approach

Data curation → **Merging and downsampling** → **Feature engineering** → **Fit and predict** → **Compare**

Download:

- EAGLE-I data
- NOAA weather event data
- ERA5-land weather reanalysis
- County-level shapefiles

Downsample all temporal data to a 6-hour cadence and merge by county.

Add weather information from ERA5-land of neighbouring counties

SKTIME

Models used:

- Naive
- Linear regression
- HGBR
- XGBoost

- LSTM neural network

TensorFlow

We fit using 5 day windows and forecast 1 day into the future.

Predictions done at county level, for time periods between 2014 and 2021.

We compute the RMSE for each model at each county and take the mean for each model.

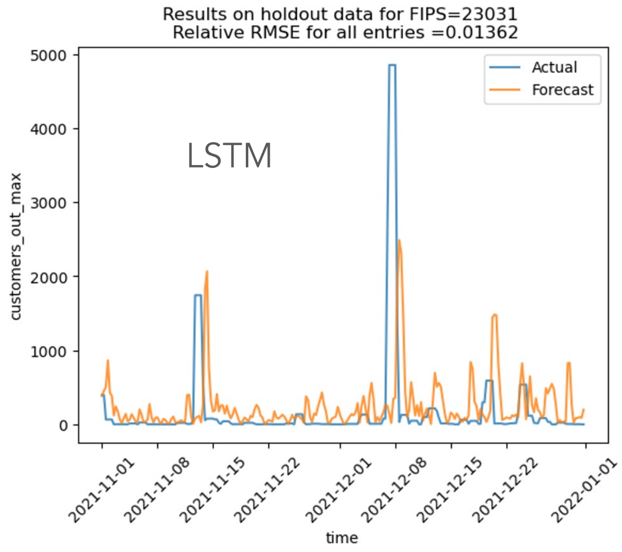
Results on holdout sets

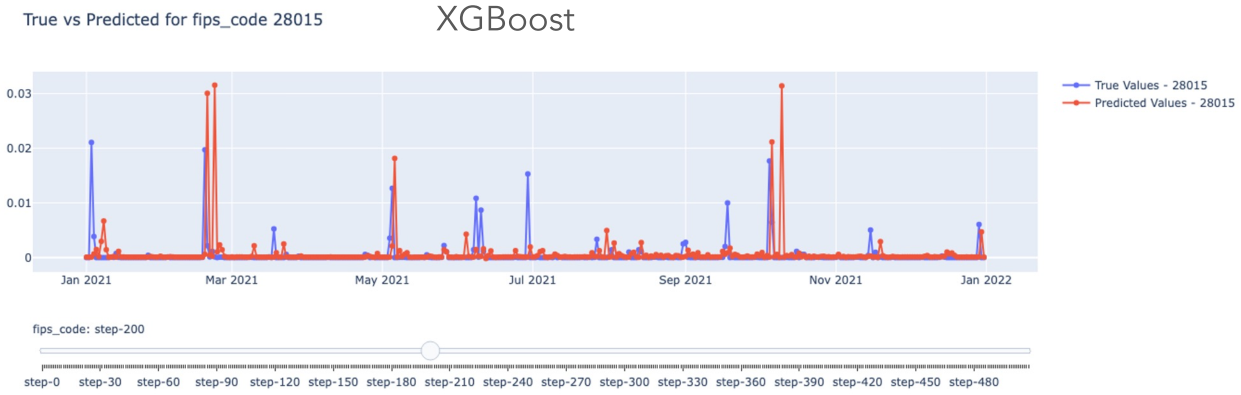
Model	RMSE
Naive	0.003122
Linear Regression	0.003547
HGBR	0.003904
XGBoost	0.004010
LSTM	0.004224

Nothing did much better than the Naive model!

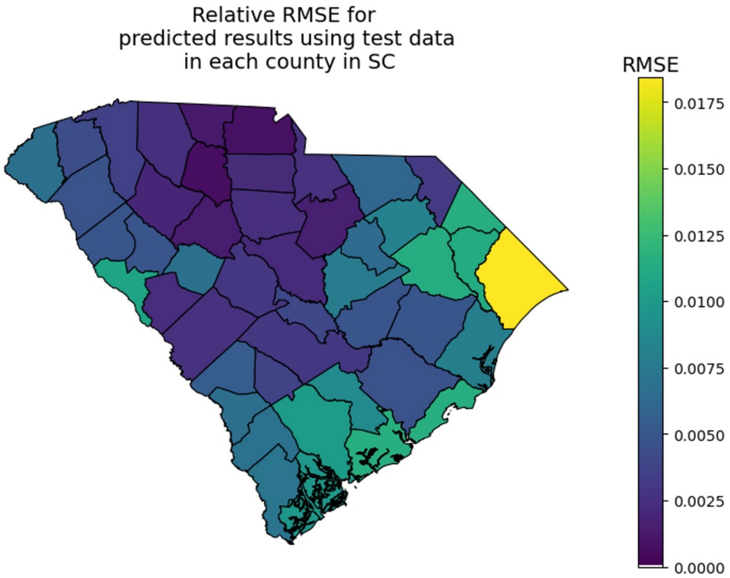
Lagging problem

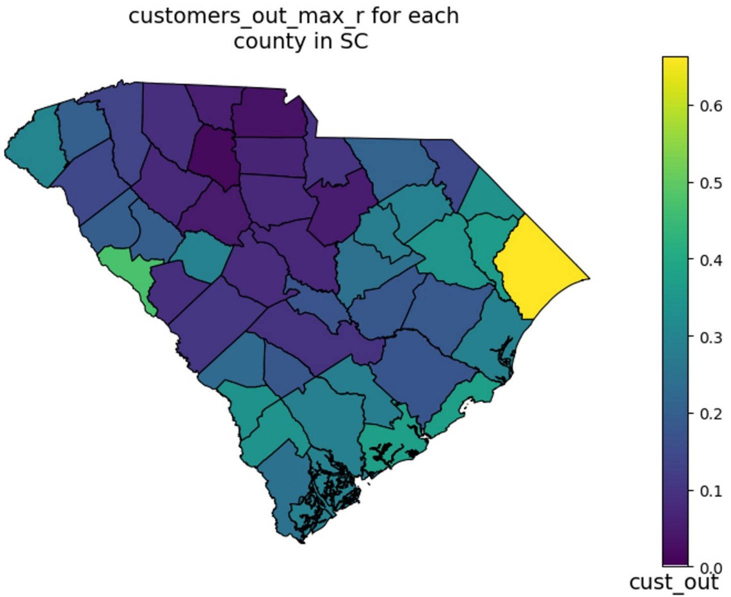
Most of our models make predictions with a considerable lag.





Observation





There is some correlation between the counties with highest RMSE and highest maximum number of customers out per capita. For example, for South Carolina we get a Pearson correlation coefficient of 0.963.

Conclusions

Main product: large aggregated + engineered dataset

datetime	fips_code	customers_out	neighbors	event_count Flood	event_count Storm	event_count Hurricane	event_count Heat	event_count Fire	event_count Wind	wind_speed	sf_12h	tp_24h	t2m	sf	tp
2014-01-11 12:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	5.176102	0.0	0.021837	290.679993	0.0	0.013920
2014-01-11 18:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	4.015108	0.0	0.044741	290.501953	0.0	0.023462
2014-01-12 00:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	4.130178	0.0	0.066655	286.847900	0.0	0.023462
2014-01-12 06:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	2.145273	0.0	0.060844	280.688232	0.0	0.000000
2014-01-12 12:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	1.697687	0.0	0.046925	276.928711	0.0	0.000001

Conclusions

Main product: large aggregated + engineered dataset

Conclusions:

Conclusions

Main product: large aggregated + engineered dataset

Conclusions:

- Our features as used are not very predictive of our target
    - Feature engineering proved ineffective

Conclusions

Main product: large aggregated + engineered dataset

Conclusions:

- Our features as used are not very predictive of our target
- Limited by the sparsity of certain weather events in training data

datetime	fips_code	customers_out	neighbors	event_count Flood	event_count Storm	event_count Hurricane	event_count Heat	event_count Fire	event_count Wind	wind_speed	sf_12h	tp_24h	t2m	sf	tp
2014-01-11 12:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	5.176102	0.0	0.021837	290.679993	0.0	0.013920
2014-01-11 18:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	4.015108	0.0	0.044741	290.501953	0.0	0.023462
2014-01-12 00:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	4.1[illegible]0178	0.0	0.066655	286.847900	0.0	0.023462
2014-01-12 06:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	2.145273	0.0	0.060844	280.688232	0.0	0.000000
2014-01-12 12:00:00	1001	0.0	[1051, 1085, 1101, 1047, 1021]	0.0	0.0	0.0	0.0	0.0	0.0	1.697687	0.0	0.046925	276.928711	0.0	0.000001

Very sparse!

Conclusions

Main product: large aggregated + engineered dataset

Conclusions:

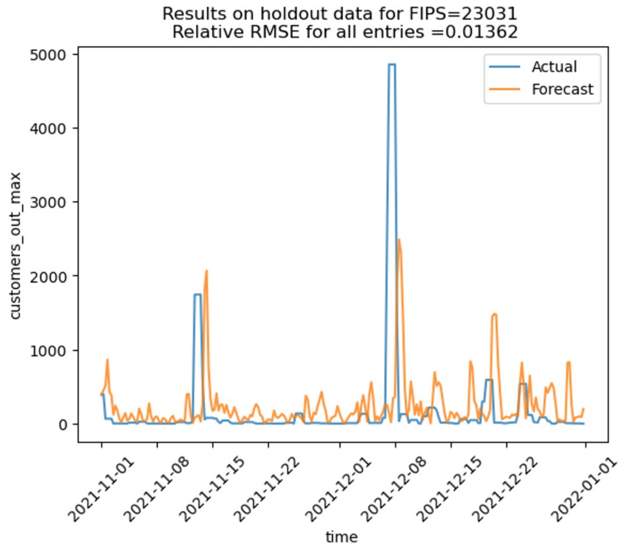
- Our features as used are not very predictive of our target
- Limited by the sparsity of certain weather events in training data
- Perform much better on certain regions

Conclusions

Main product: large aggregated + engineered dataset

Conclusions:

- Our features as used are not very predictive of our target
- Limited by the sparsity of certain weather events in training data
- Perform much better on certain regions
- Predictions tend to lag reality



Conclusions

Main product: large aggregated + engineered dataset

Limitations:

- Predictions are by county, more granular might be more useful
- Models don’t capture long term trends like climate change

Future work:

- Taking geographic relationships into account in a more sophisticated way
- Training a model to predict farther into the future

Works Cited

(1) Ai, Dan and Crago, Christine and Mullins, Jamie, Heat, Power Outages and Mortality in the United States. April 10, 2025. Available at SSRN: https://ssrn.com/abstract=5212367 or http://dx.doi.org/10.2139/ssrn.5212367

(2) Brelsford, C., Tennille, S., Myers, A., Chinthavali, S., Tansakul, V., Denman, M., Coletti, M., Grant, J., Lee, S., Allen, K., Johnson, E., Huihui, J., Hamaker, A., Newby, S., Medlen, K., Maguire, D., Dunivan Stahl, C., Moehl, J., Redmond, D. P., … Bhaduri, B. (2023). The Environment for Analysis of Geo-Located Energy Information's Recorded Electricity Outages 2014-2023 (Version 2). figshare. https://doi.org/10.6084/m9.figshare.24237376.v2

(3) Copernicus Climate Change Service (2023): ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47 (Accessed on 07-MAR-2023)

(4) Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2018): ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47 , (Accessed on 07-MAR-2023)

(5) LaCommare, K.H., Eto, J.H., Dunn, L.N., Sohn, M.D., 2018. Improving the estimated cost of sustained power interruptions to electricity customers. Energy 153, 1038–1047. doi: 10.1016/j.energy.2018.04.082.

(6) National Centers for Environmental Information (NCEI). (2024). Storm Events Database. NOAA National Centers for Environmental Information. Retrieved from https://www.ncdc.noaa.gov/stormevents/

(7) Satterlee, Katie. Feb. 2024. *Predicting Power Outages*. Texas A&M University. https://engineering.tamu.edu/news/2024/02/predicting-power-outages.html

(8) Woods, Alden. May 2, 2023. *These four regions of the US are hardest hit by power outages.* https://deohs.washington.edu/hsm-blog/these-four-regions-us-are-hardest-hit-power-outages

Thank You

Image: NASA