Quantifying Disparities in Electricity Reliability

ZIQI WANG,
MARLEY E. RYWELL,
DAKSHA S. MARATHE,
DALIT HENDEL,
UMA MAHESHWARI DASARI

Why Power Outages Matter

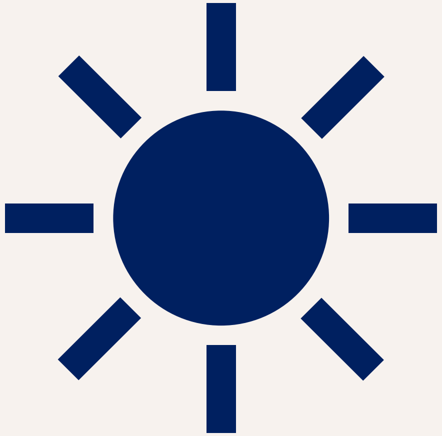
- Climate change is increasing the frequency of extreme weather
  - Extreme weather can cause power outages for thousands at a time
  - Ex. Hurricanes Sandy and Irene made it all the way up the east coast
- Health risks
  - loss of clean water, refrigerated medicine, food storage, medical technologies, safety mechanisms, and indoor air pollution
- Costs
  - Hurricane Sandy cost $1.8 billion in repair and response cost for utilities

Literature Review

- Methodologies previously used: SMOTE, GAMs, Deep Neural Nets, Decision Tree ensembles
- Consensus
  - Weather is a major predictor of power outages
  - There are health risks and economic costs associated with outages
- Conflicting theories
  - No agreed 'best' model to predict future outages
  - No consensus if the SAIDI and SAIFI are correlated with weather
- Gap in research
  - Existing research examines disparities due to extreme weather events
  - Lack of research examining disparities due to non-extreme weather events
    - Lee et al. (2022) "This limitation is partly due to the inability of researchers to access fine-resolution data related to the extent and duration of outages for subpopulations."

Research Question

- Is there a significant difference in the time it takes for power to return for outages which affect different income levels, diversity, and population density by town?



Data Collection

MA Power Outage Data:

- Mass.gov aggregates energy company emergency response (ERP) reports. The dataset includes town/city of outage duration, time, and reason. Biases include human error in outage reports

MA Weather Data:

- The Global Historical Climatology Network - Daily (GHCN-Daily) integrates daily climate observations from ~30 different data sources. Biases include unreported events.

MA Income Data:

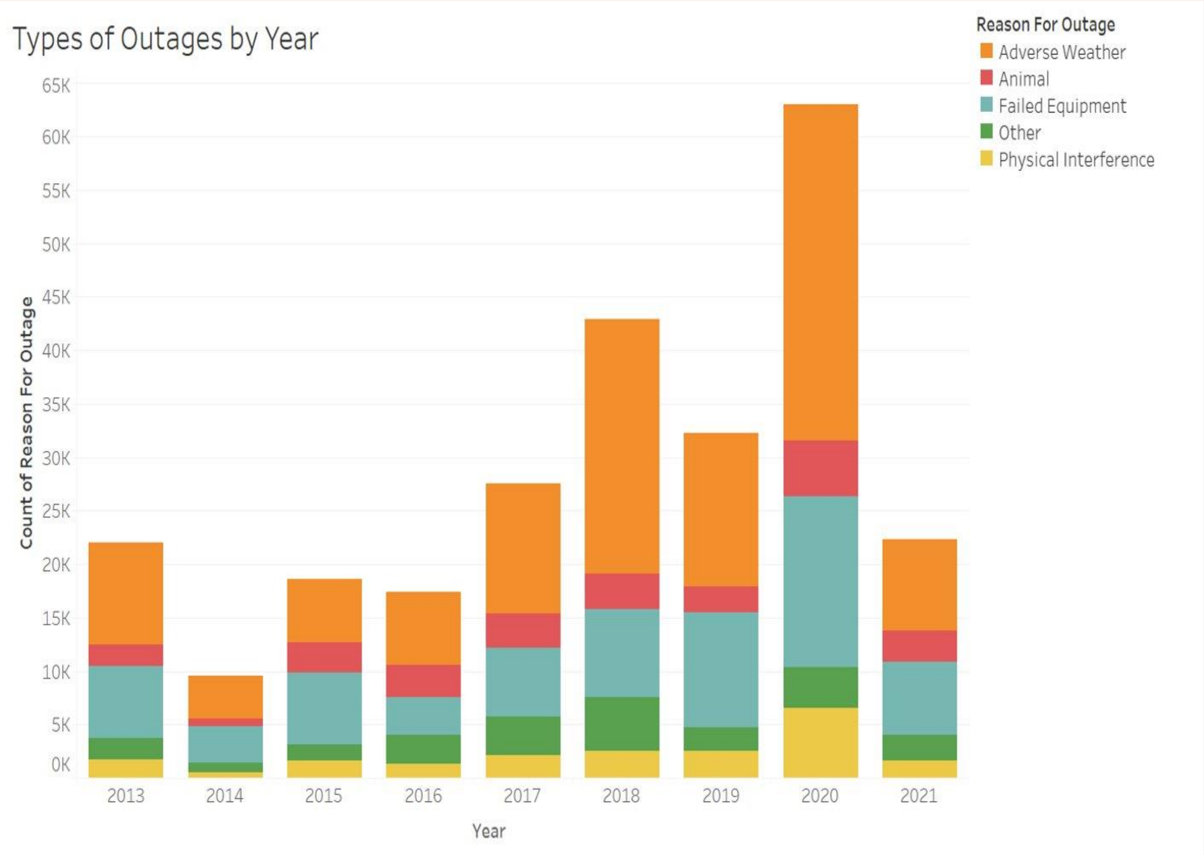
- Income data by town from the MA Department of Revenue

MA 2020 Census Data:

- Collected by the Census Bureau and relies on self-report information.

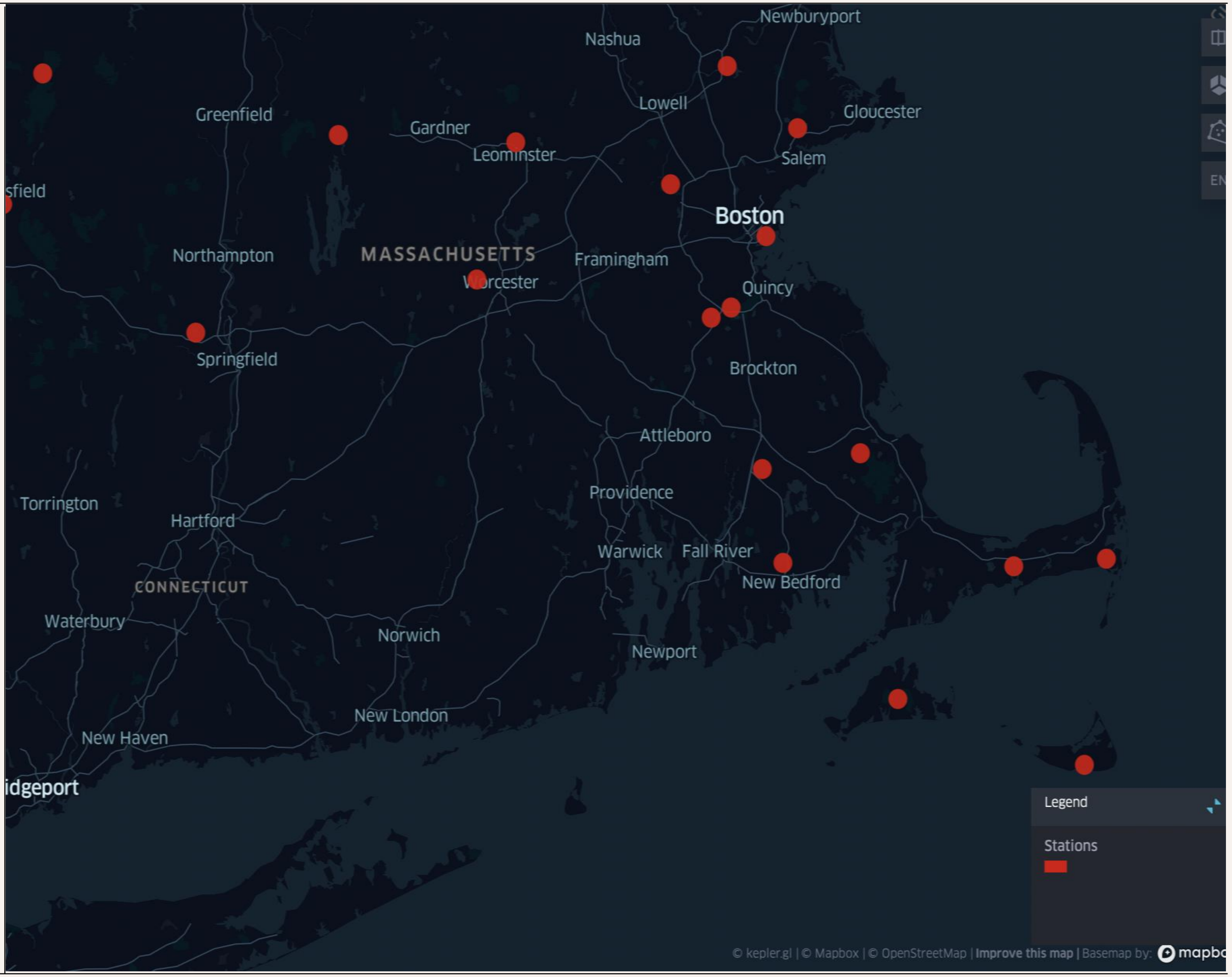
Outage data

- 3 electricity providers in Massachusetts
  - Unitil
  - Massachusetts Electric Company and Nantucket Electric Company
  - Eversource
- Data provided from 2013-2018
- We followed the same schema and collected data from 2019-2021
  - https://www.mass.gov/info-details/power-outages#historic-power-outages-
  - Limitation: Unitil had no outage data for 2020



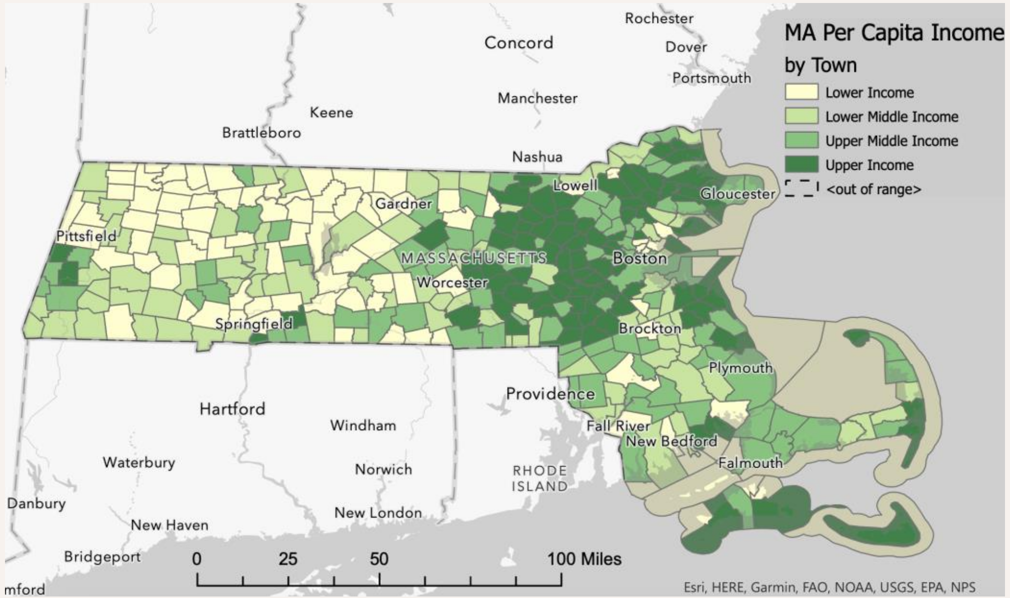
Weather data

- Data collected from National Ocean and Atmospheric Administration (NOAA) https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND
- Only 19 weather stations recorded the data
- Data from 2013-2021
- Daily max or average resolution of variables



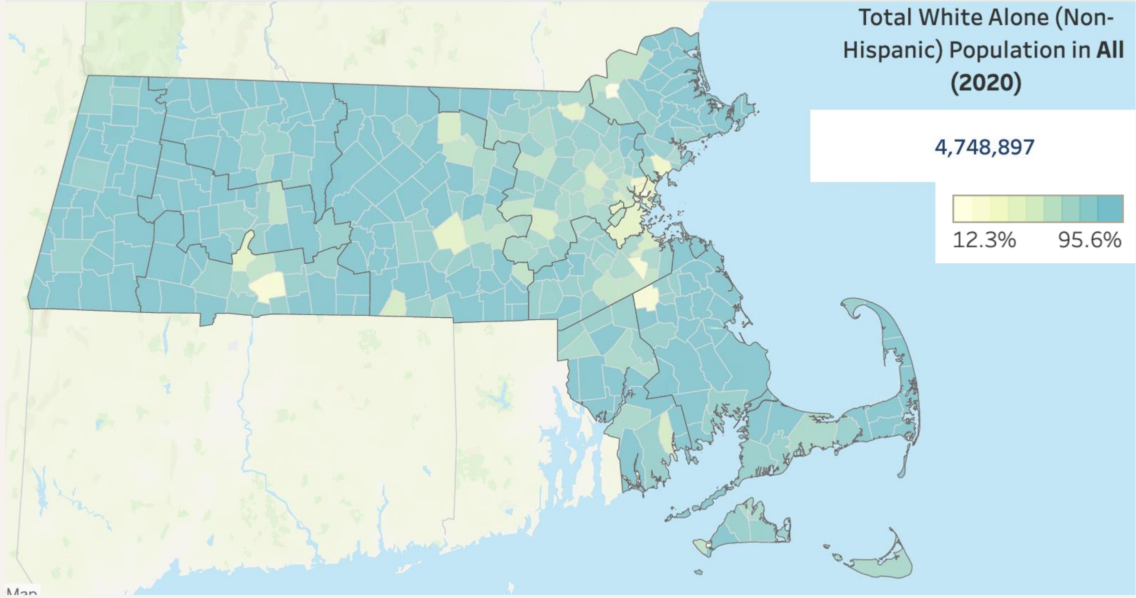
Income Data

- Data from 2013-2021
- Provided by the MA Department of Revenue
- Total income and per capita income by town



Census Data

- Data from the 2020 US census
- Population density by block
- Percent race/ethnicity identification by town



https://www.sec.state.ma.us/census2020/index.html

Merging Data

- Only 19 weather stations but 319 towns in MAs
  - Towns linked to weather events by the closest weather station using ArcGIS Pro
- Outage and weather data merged by **town** and **date**
  - Town names standardized across all datasets (i.e. 'Manchester' to 'Manchester by the sea')
  - Boston neighborhood outages reduced to town level (i.e. 'South Boston' to 'Boston')
- Income data merged by **town** and **year**
- Census data merged by **town**
- **Each row in final dataset was a unique outage event**

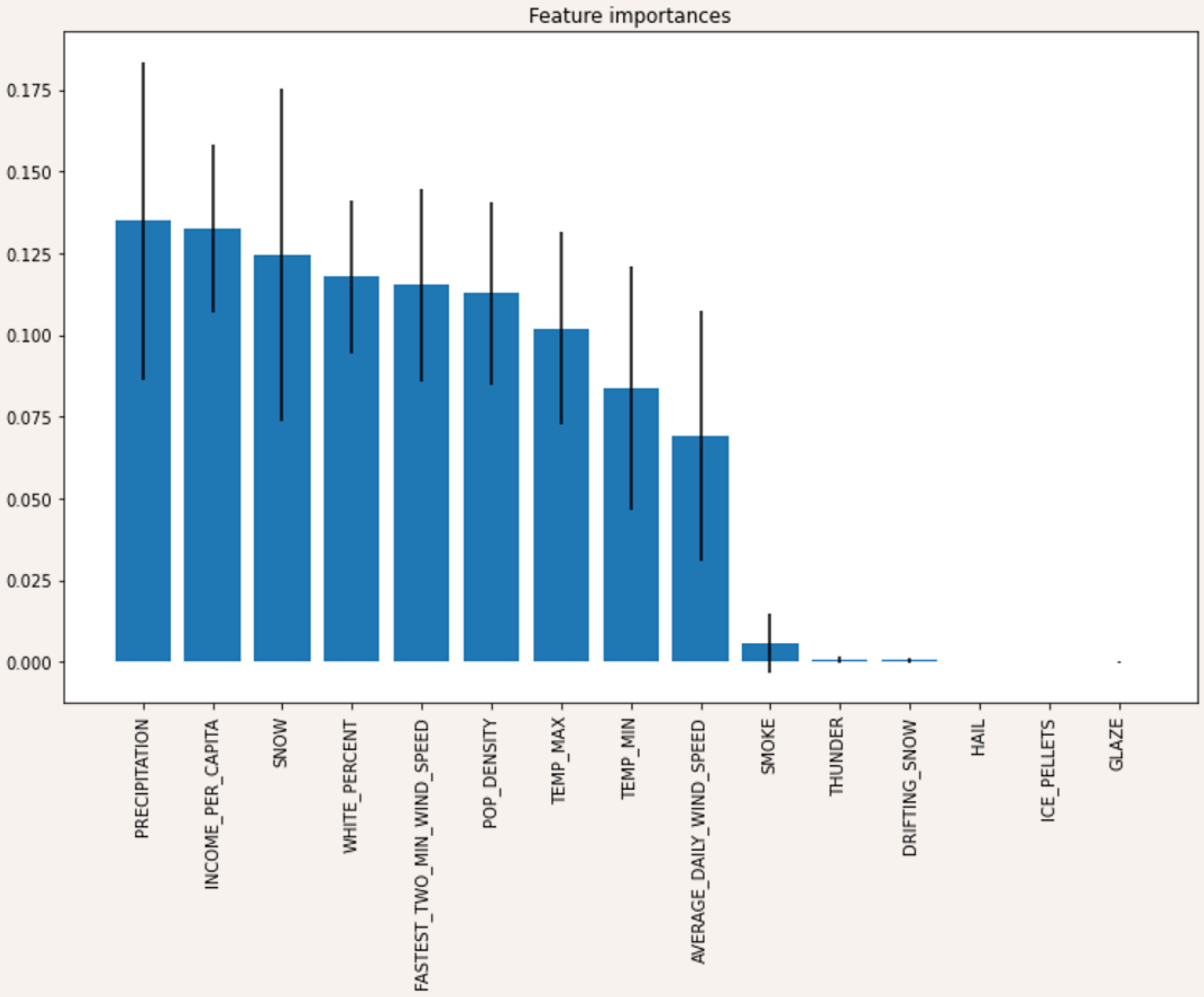
Feature Engineering and Processing

- Dropped: about 3%
  - No weather data available for some outages
- Null Values: Snow (83.7%) and Average Windspeed (0.1%)
  - Filled with average value by month for each year
- Weather "flags" (dummy variable) Nan values replaced with 0
- Features generated
  - Population Density: aggregated up from block to town level
  - Quartiles used to make Income, Diversity, and Population Density levels
  - Customer Outage Hours (response) : (outage duration in hours) x (number of customers affected)
    - Can also be visualized as: $\text{CUSTOMER OUTAGE HOURS} = \sum U_i * N_i$
    - Where **Ui** is outage duration and **Ni** is number of customers affected

Random Forest for Feature Selection

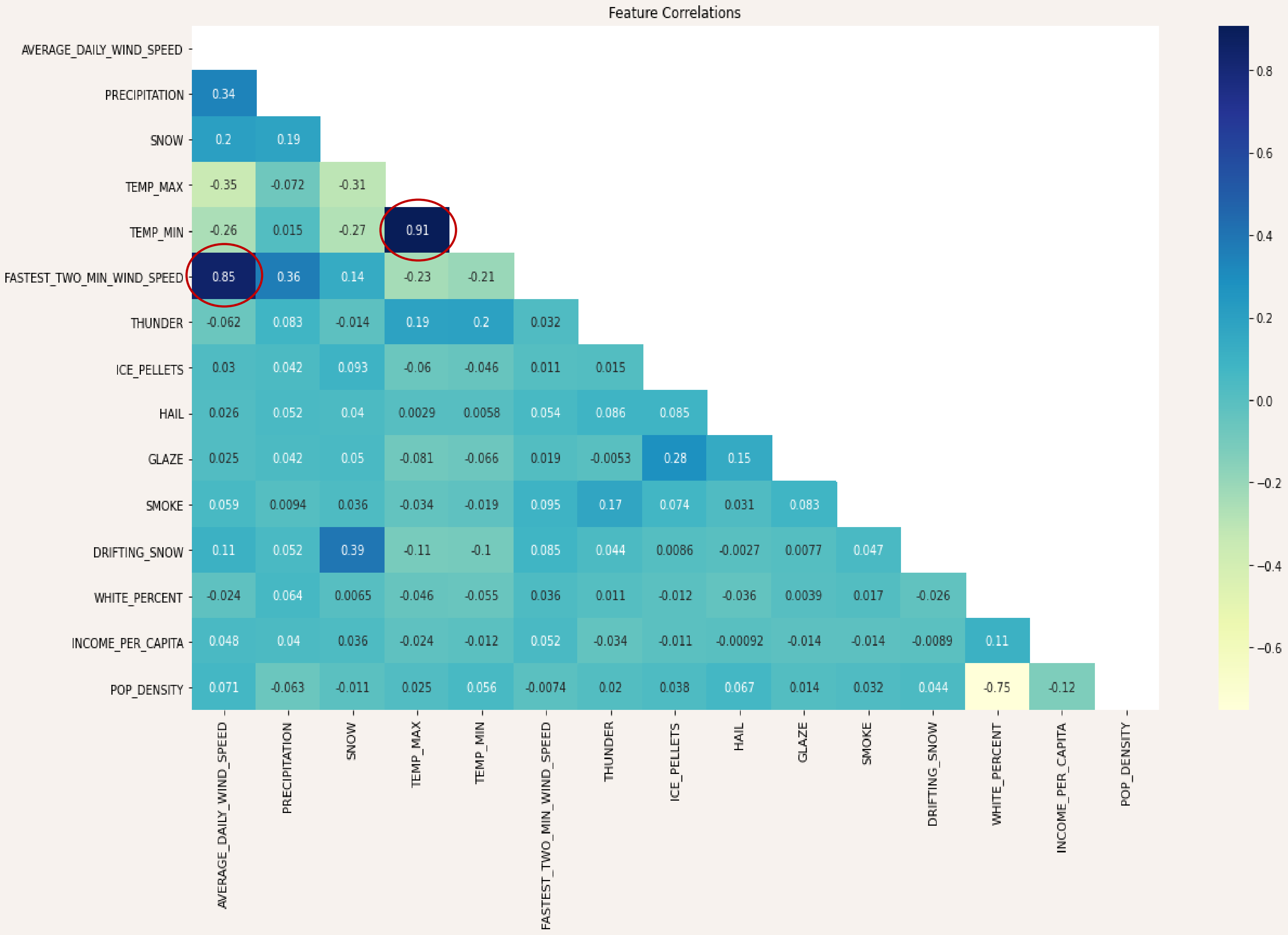
Important Weather Features:

- Precipitation (mm)
- Snow (mm)
- Fastest two Minute Wind Speed (meters per second)
- Temperature Maximum (C)
- Temperature Minimum (C)
- Average Wind Speed (mm)



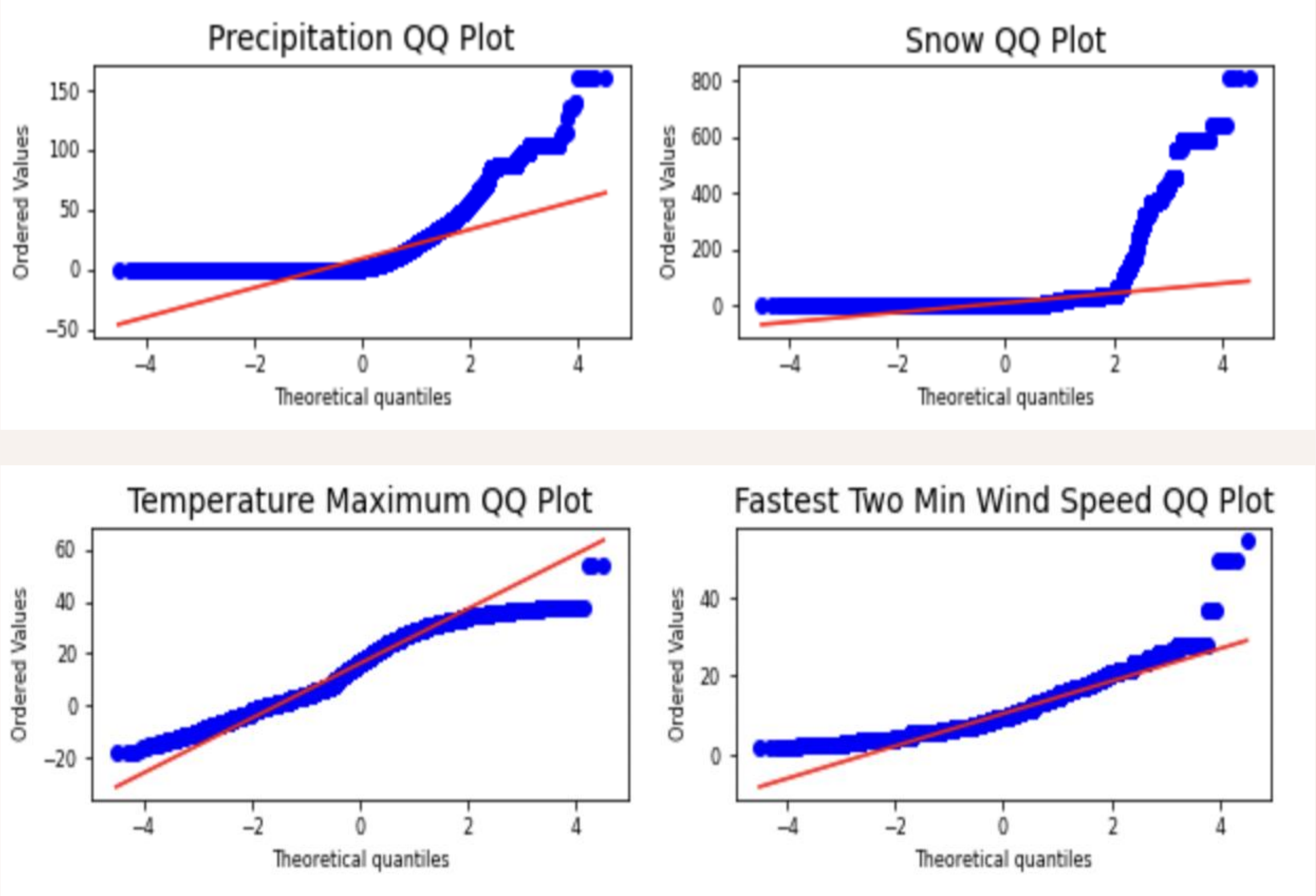
Feature Correlation Analysis

- Temperature Maximum and Minimum are highly correlated
  - Kept Temperature Maximum
- Fastest Two Minute Wind Speed and Average Wind Speed are highly correlated
  - Kept Fastest Two Minute Wind Speed

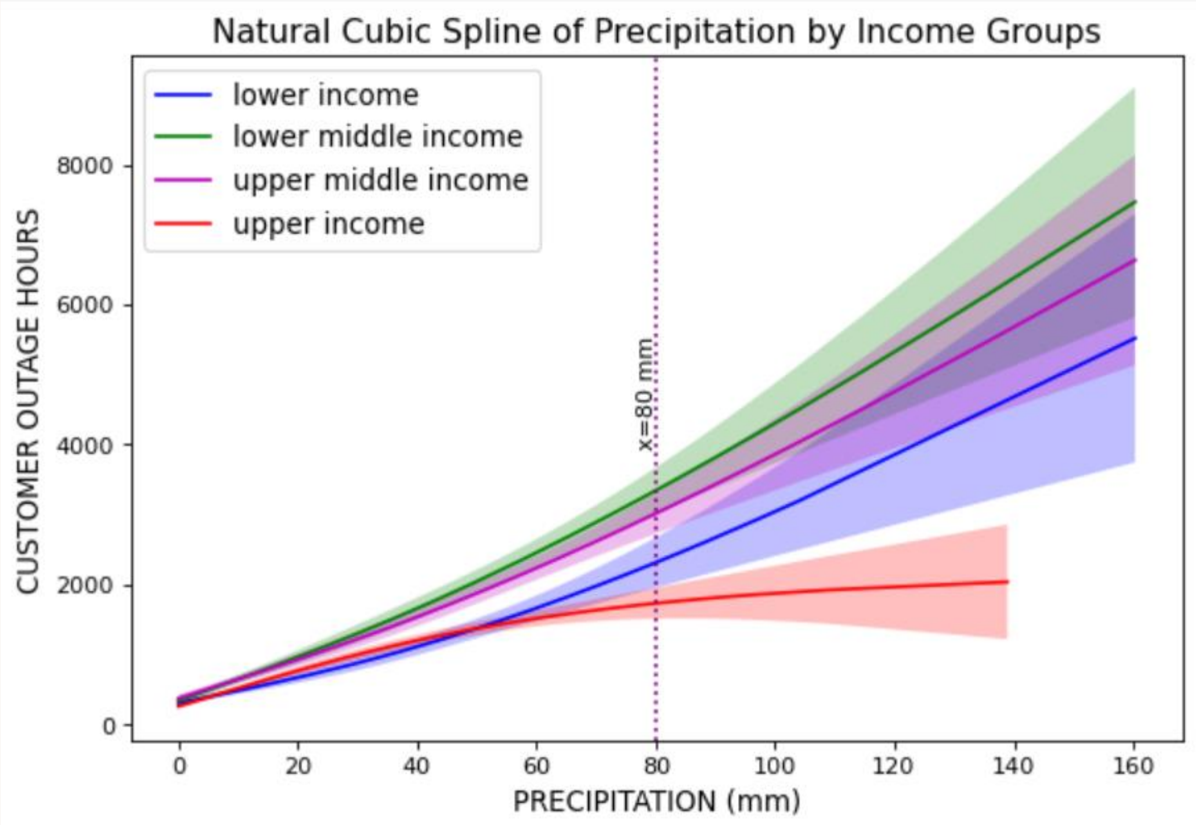


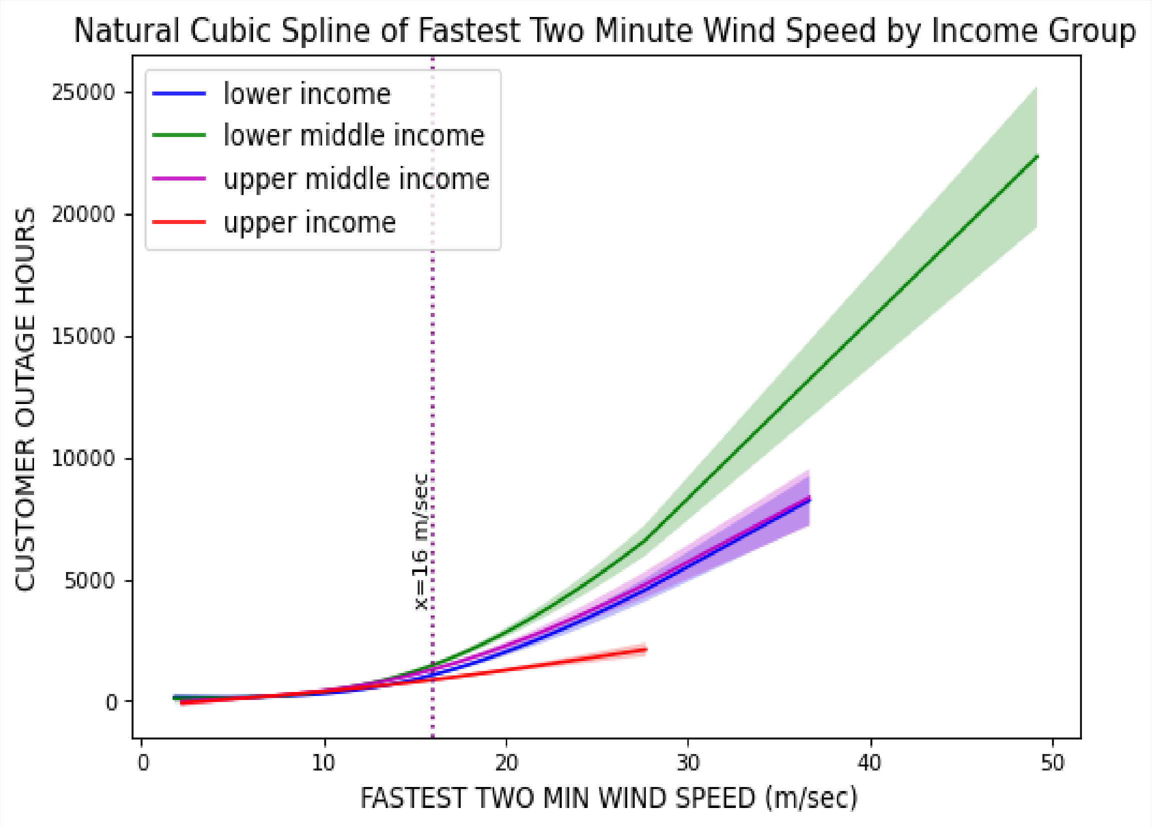
Looking for Normality

- Shapiro-wilk p-values are all 0
  - With an alpha of .05, we reject the null hypothesis that these variables are normally distributed
- Because the variables are not normal, we used a general additive model based on splines




```
Shapiro-wilk p-value for precipitation: 0.0
Shapiro-wilk p-value for snow: 0.0
Shapiro-wilk p-value for temperature max: 0.0
Shapiro-wilk p-value for fastest two minute wind speed: 0.0
```

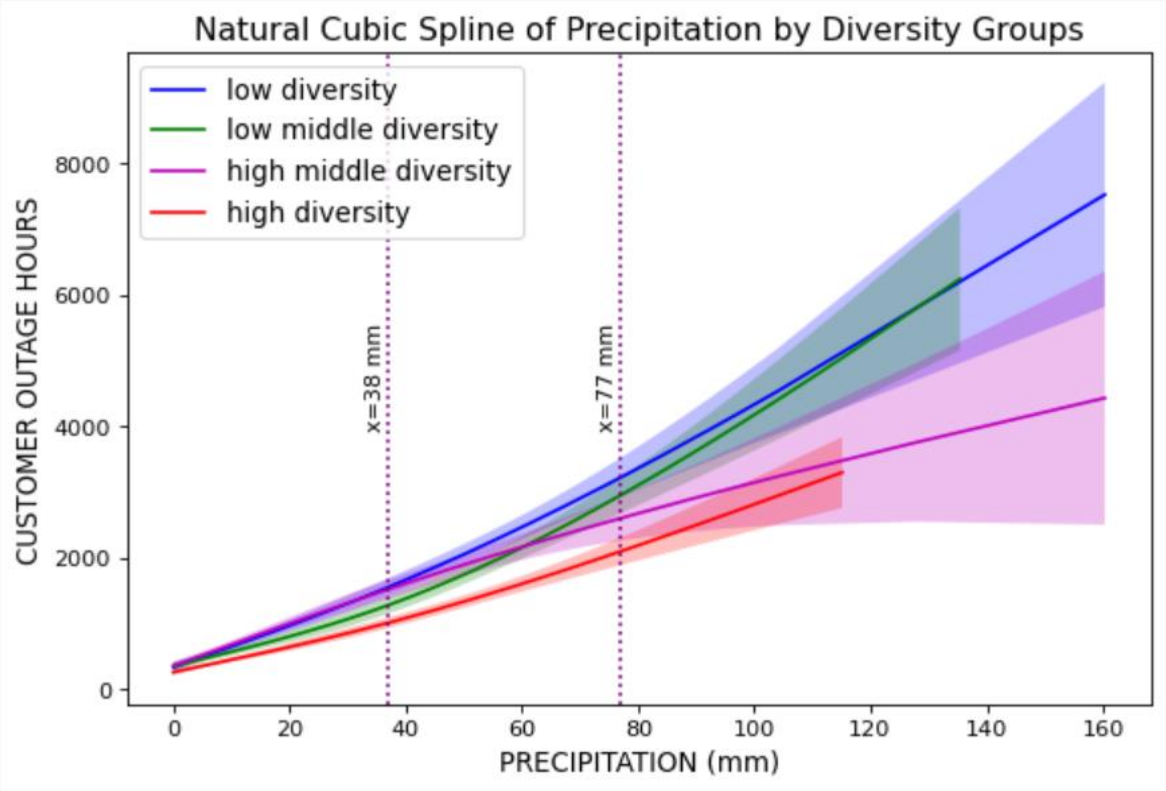


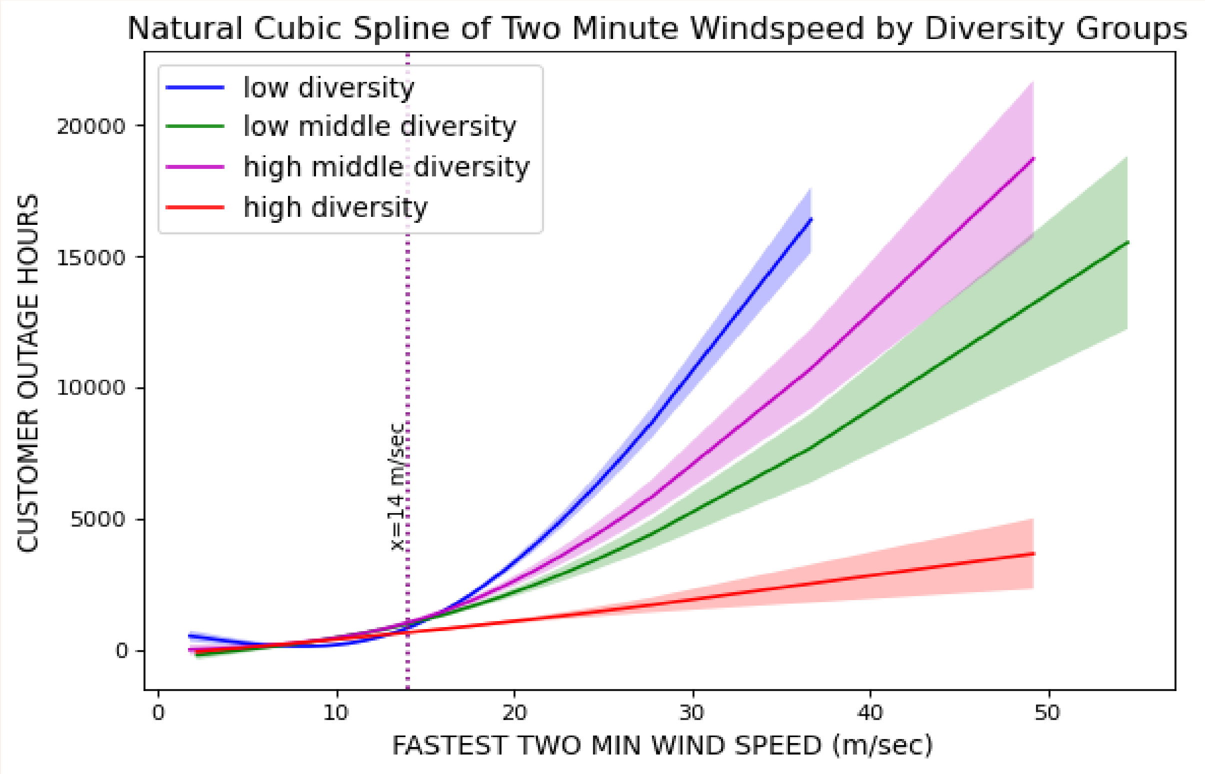


Natural Splines

Income groups

- At a precipitation level of around 80 mm the upper income towns confidence interval stops overlapping with the other groups and has the **lowest customer outage hours**
- At a wind speeds of around 16 m/sec the upper income towns confidence interval stops overlapping with the other groups and has the **lowest customer outage hours**

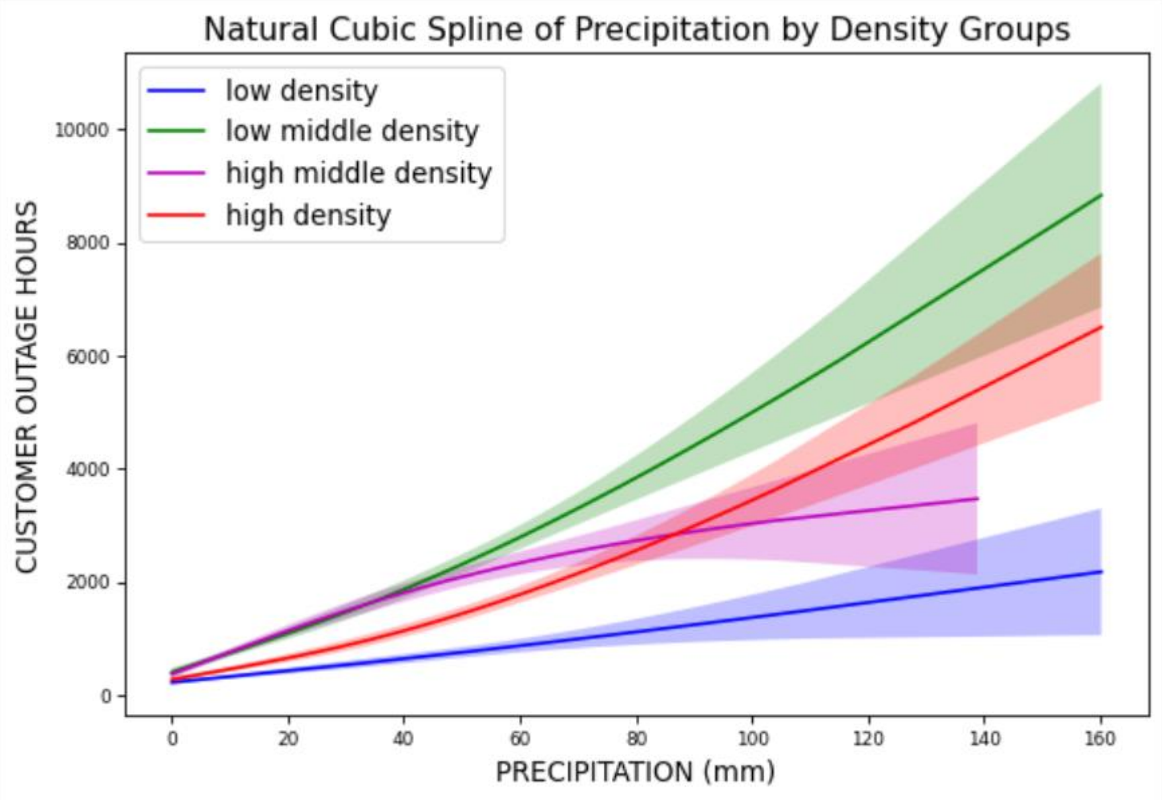


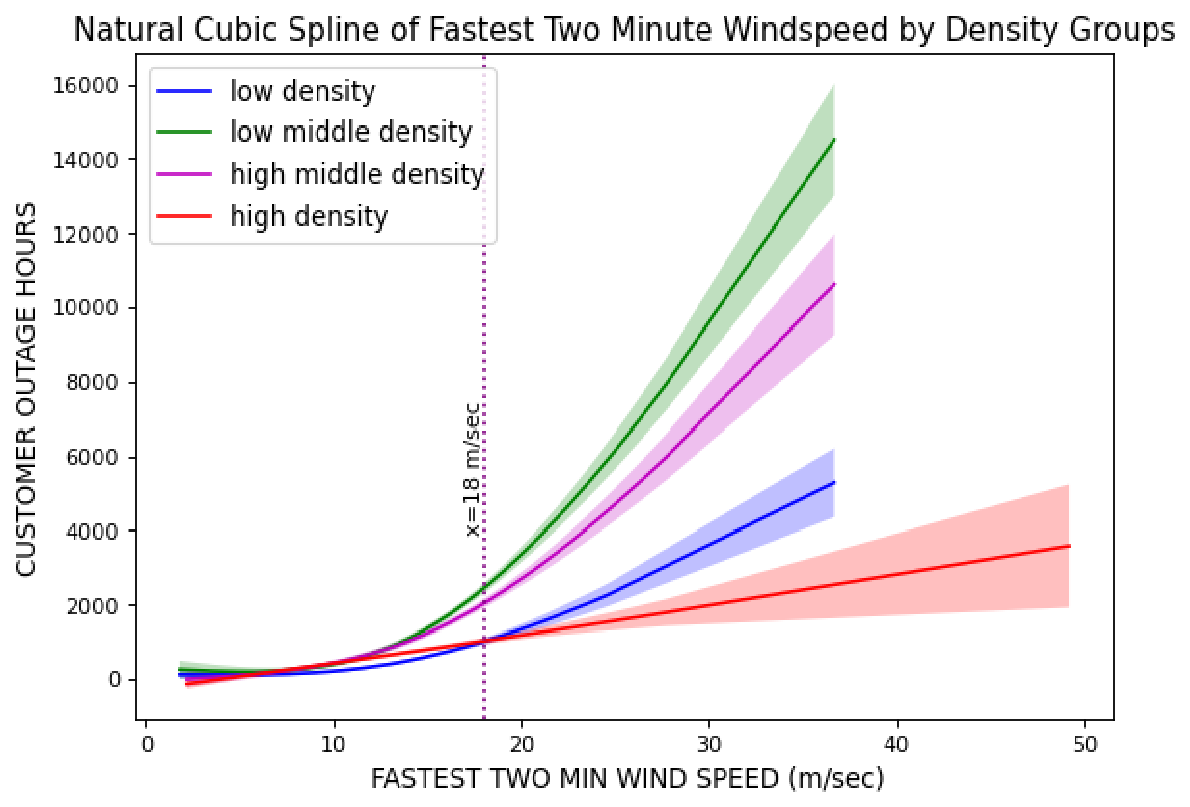


Natural Splines

Race/Ethnicity Diversity groups

- In the range of 38 mm to 77 mm of rain, the highest diversity towns have the lowest customer outage hours
- At wind speeds of above 14 m/sec, the high diversity towns confidence interval stops overlapping and has the **lowest customer outage hours**
- At wind speeds of above 15 m/sec, the low diversity towns confidence interval stops overlapping and has the **highest customer outage hours**



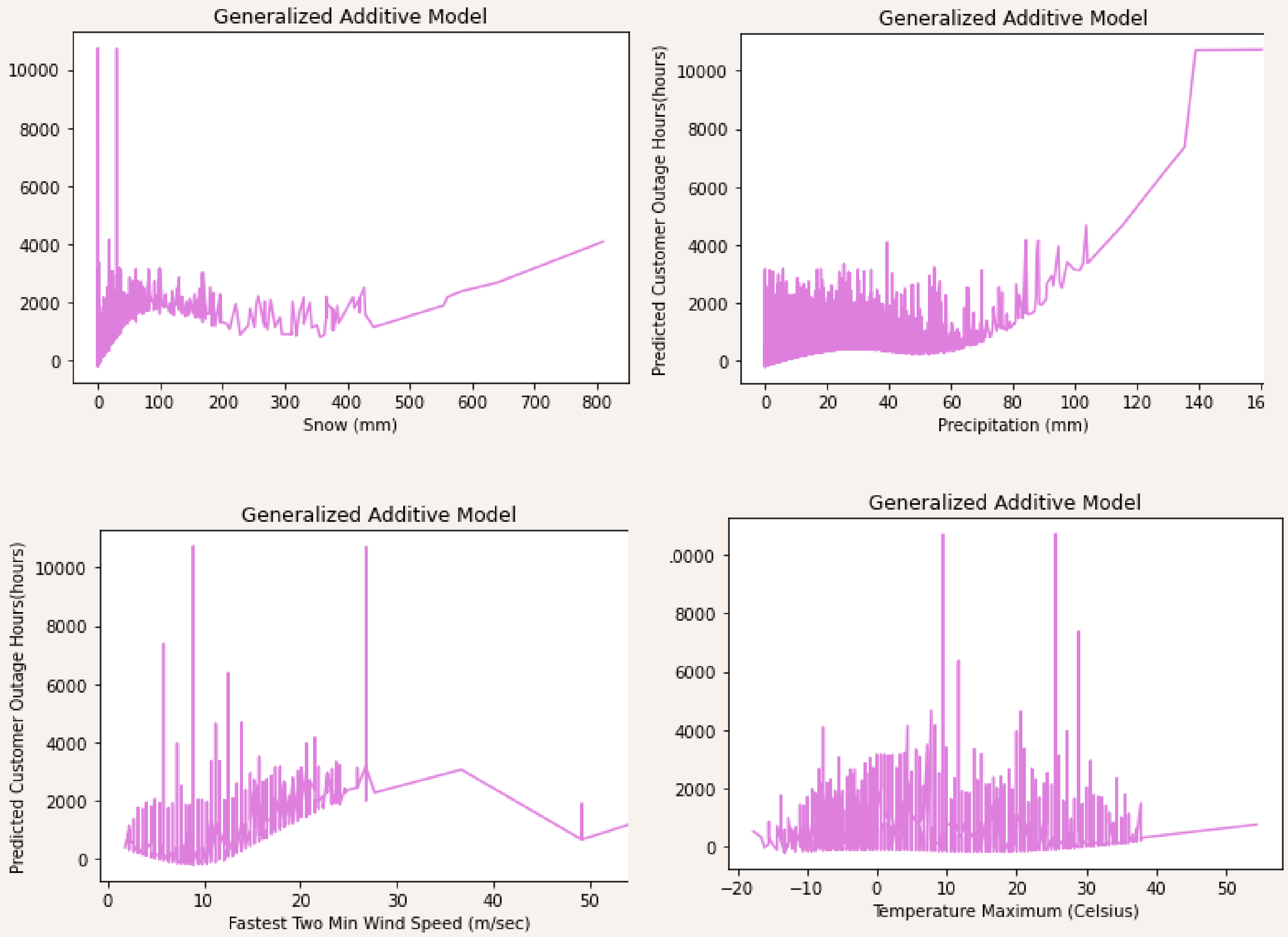


Natural Splines

Population Density groups

- At all precipitation levels the low density towns confidence interval has the **lowest customer outage hours**
- At wind speeds of around 18 m/sec the high density towns confidence interval stops overlapping with the other groups and has the **lowest customer outage hours**

GAMs



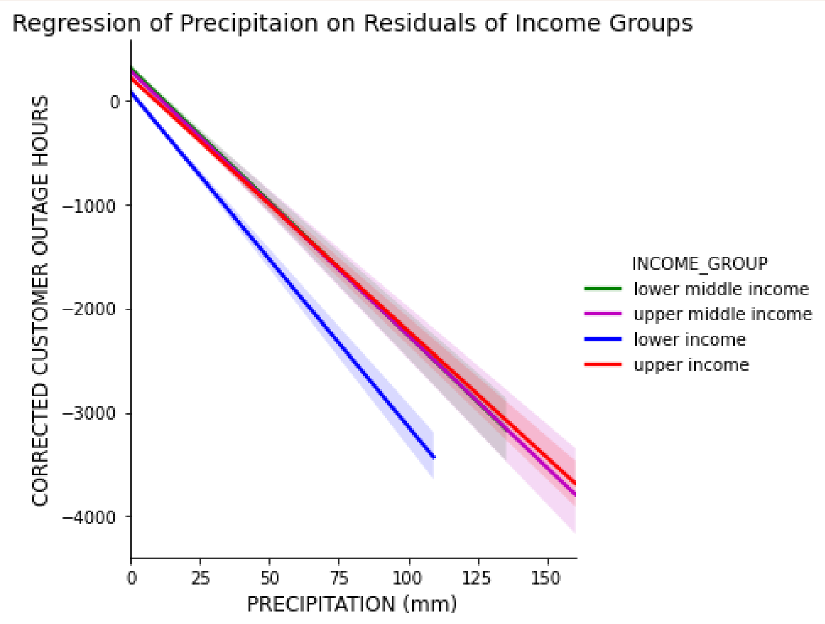
- Goal: correct weather for all variables
- Map from 4-d to 2-d for each dimension

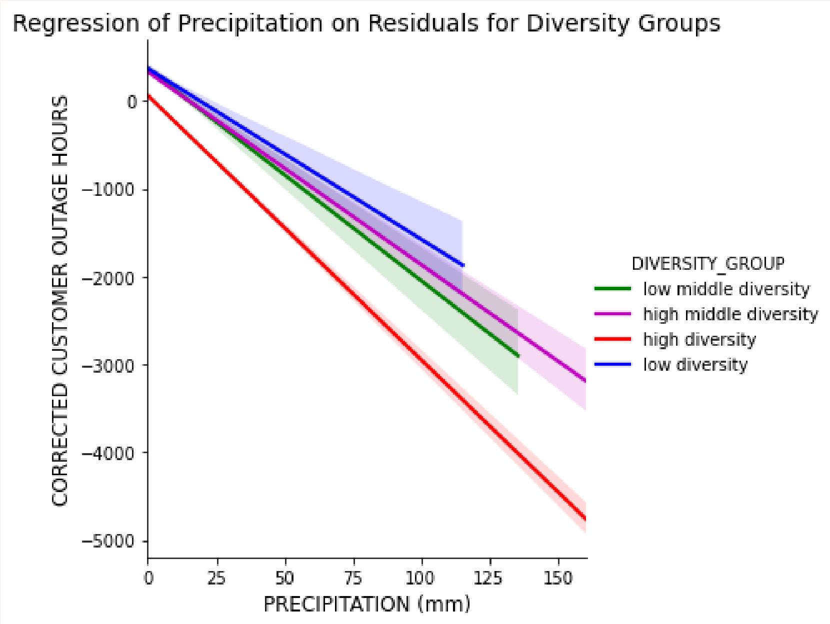
Explaining the variance in GAMs

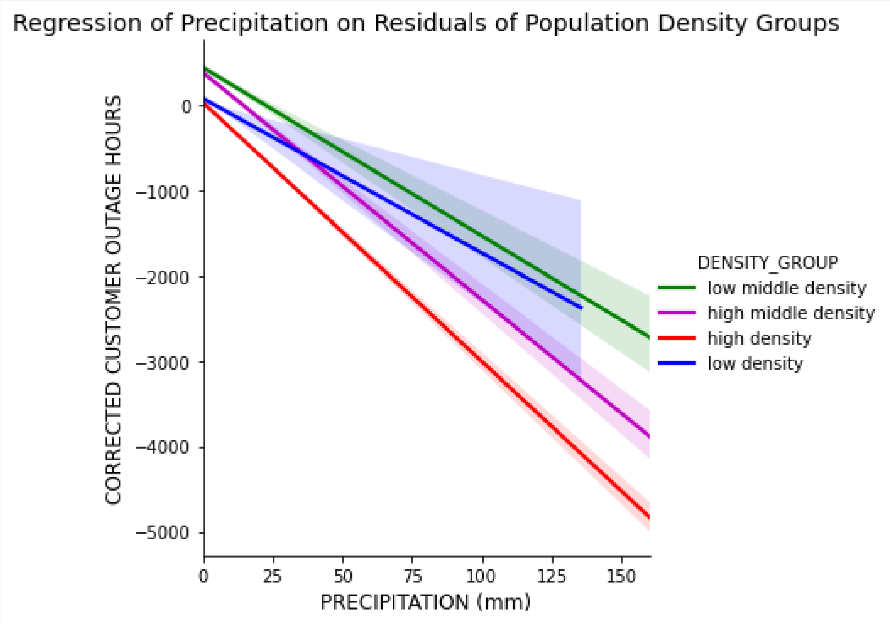
- All models have a random error component, epsilon, which we attribute to factors outside of the model, these are the residuals for each data point
- A natural cubic spline is the piecewise function for each knot

$$S(x) = \begin{cases} S_0(x) = a_0x^3 + b_0x^2 + c_0x + d_0 + \epsilon_0 & t_0 \leq x \leq t_1 \\ \vdots \\ S_{n-1}(x) = a_{n-1}x^3 + b_{n-1}x^2 + c_{n-1}x + d_{n-1} + \epsilon_{n-1} & t_{n-1} \leq x \leq t_n \end{cases}$$

- By regressing the residuals over precipitation (our most significant predictor), we see the error terms, epsilons, grow at different rates depending on socioeconomic groups which demonstrates an injustice







Residuals

CORRECTED CUSTOMER OUTAGE HOURS $= y - y_{pred}$

- Income groups: the lowest income quartile is different from the other 3 quartiles at a statistically significant level
- Diversity groups: the highest diversity quartile is different from the other 3 quartiles at a statistically significant level
- Density groups: the highest density quartile is different from the other 3 quartiles at a statistically significant level

Conclusions

- The random forest demonstrates that socioeconomic factors are predictors of customer outage hours
- Certain groups of people experience higher outage durations given the same weather conditions
  - Higher income municipalities experienced lower outage durations for wind speeds above 16 m/s
    - After accounting for weather predictors, lower income towns have higher variance in their customer outage hours
  - Low diversity towns experienced higher outage durations for wind speeds above 14 m/s
    - After accounting for weather predictors, towns with higher diversity have higher variance in their customer outage hours
  - High density towns experienced lower outage durations for wind speeds greater than 18 m/s
    - After accounting for weather predictors, high density towns have higher variance in their customer outage hours

Limitations

- Every dataset had a different level of data resolution
  - Ex: county, municipality, census block, point data (lat, long) etc.
  - Used resolution of outage data (town level)
- Many missing values in the weather data for Snow (83.7%)
  - Filled in with monthly average for that year
- Removed about 3% of data
  - Missing or impossible values (ex: no weather data or negative customers affected)
- 40 towns are served by municipality power
  - Not included in outage data
  - Statistically significant difference in income medians (tested using Moods-median test)

Works Cited:

[1] P. H. Larsen, K. H. LaCommare, J. H. Eto, and J. L. Sweeney, "Recent trends in power system reliability and implications for evaluating future investments in resiliency," Energy, vol. 117, pp. 29–46, 2016. [Online]. Available: https: //www.sciencedirect.com/science/article/pii/S0360544216314979

[2] M. Angalakudati, J. Calzada, V. Farias, J. Gonynor, M. Monsch, A. Papush, G. Perakis, N. Raad, J. Schein, C. Warren et al., "Improving emergency storm planning using machine learning," in 2014 IEEE PES T&D Conference and Exposition. IEEE, 2014, pp. 1–6.

[3] C. Klinger, O. Landeg, and V. Murray, "Power outages, extreme events and health: a systematic review of the literature from 2011-2012," PLoS currents, vol. 6, p. ecurrents.dis.04eb1dc5e73dd1377e05a10e9edde673, January 2014. [Online]. Available: https://europepmc.org/articles/PMC3879211

[4] G. B. Anderson and M. L. Bell, "Lights out: Impact of the august 2003 power outage on mortality in new york, ny," Epidemiology, p. 189–193, Mar 2012.

[5] U. E. I. Administration, "Reliability metrics of u.s. distribution system," 2020. [Online]. Available: https://www.eia.gov/electricity/annual/html/epa 11 01.html#:~: text=SAIDI%20%3D%20System%20Average%20Interruption%20Duration,year%2C% 20the%20average%20customer%20experienced.

[6] P. L. Watson, A. Spaulding, M. Koukoula, and E. Anagnostou, "Improved quantitative prediction of power outages caused by extreme weather events," Weather and Climate Extremes, vol. 37, p. 100487, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212094722000664

[7] S.-R. Han, S. D. Guikema, and S. M. Quiring, "Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models," Risk Analysis, vol. 29, no. 10, pp. 1443–1453, 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.2009.01280.x

[8] E. Kabir, S. D. Guikema, and S. M. Quiring, "Predicting thunderstorm-induced power outages to support utility restoration," IEEE Transactions on Power Systems, vol. 34, no. 6, pp. 4370–4381, 2019.

[9] A. Arif and Z. Wang, "Distribution network outage data analysis and repair time pre- diction using deep learning," in 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 2018, pp. 1–6.

[10] M. Kezunovic, Z. Obradovic, T. Dokic, B. Zhang, J. Stojanovic, P. Dehghanian, and P.-C. Chen, Predicting Spatiotemporal Impacts of Weather on Power Systems Using Big
4
Data Science. Cham: Springer International Publishing, 2017, pp. 265–299. [Online]. Available: https://doi.org/10.1007/978-3-319-53474-9 12

[11] A. Kenward, U. Raja et al., "Blackout: Extreme weather climate change and power outages," Climate central, vol. 10, pp. 1–23, 2014.

[12] K. H. LaCommare, J. H. Eto, L. N. Dunn, and M. D. Sohn, "Improving the estimated cost of sustained power interruptions to electricity customers," Energy, vol. 153, pp. 1038–1047, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/ pii/S036054421830690X

[13] M. S. Kim, B. S. Lee, H. S. Lee, S. H. Lee, J. Lee, and W. Kim, "Robust estimation of outage costs in south korea using a machine learning technique: Bayesian tobit quantile regression," Applied Energy, vol. 278, p. 115702, 2020.

[14] J. H. Eto, "An examination of temporal trends in electricity reliability based on reports from us electric utilities," 2012.

[15] P. H. Larsen, K. H. LaCommare, J. H. Eto, and J. L. Sweeney, "Assessing changes in the reliability of the us electric power system," 2015.

[16] D. Mitsova, A.-M. Esnard, A. Sapat, and B. S. Lai, "Socioeconomic vulnerability and electric power restoration timelines in florida: The case of hurricane irma," Natural Hazards, vol. 94, no. 2, pp. 689–709, 2018.

[17] J. R. Elliott and J. Pais, "Race, class, and hurricane katrina: Social differences in human responses to disaster," Social Science Research, vol. 35, no. 2, pp. 295–321, 2006, katrina in New Orleans/Special Issue on Contemporary Research on the Family. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0049089X06000135

[18] N. Coleman, A. Esmalian, and A. Mostafavi, "Equitable resilience in infrastructure systems: Empirical assessment of disparities in hardship experiences of vulnerable pop- ulations during service disruptions," Natural Hazards Review, 06 2020.

[19] C.-C. Lee, M. Maron, and A. Mostafavi, "Community-scale big data reveals disparate impacts of the texas winter storm of 2021 and its managed power outage," arXiv preprint arXiv:2108.06046, 2021.

Appendix

Datasets

- Weather Data: weather_data_station.csv
  - Weather conditions recorded in 19 stations across MA by National Oceanic and Atmospheric Administration
- Outage Data: outage_data_final_typescleaned2.csv
  - Power outage data by town from Mass.gov
- Census2020_block: town_blocks.csv
  - Census block data from U.S. Census Bureau 2020
- Census2020: census_race_profile_2020.xlsx
  - Racial identification percentages by town from U.S. Census Bureau 2020
- Income data: 2013_2021_income_data_final.csv
  - Income data by town from the MA Department of Revenue
- GIS merged weather outage data: weather_outage_towns.csv
  - Towns matched to nearest weather station thus, Outage and Weather data merged by town and date

The Variance Inflation Factor (VIF) and Multiple Linear Regression (MLR)

	VIF	Predictive Variable
0	4.140869	AVERAGE_DAILY_WIND_SPEED
1	1.207618	PRECIPITATION
2	1.350076	SNOW
3	1.298113	TEMP_MAX
4	3.877575	FASTEST_TWO_MIN_WIND_SPEED
5	1.113245	THUNDER
6	1.101476	ICE_PELLETS
7	1.042926	HAIL
8	1.114632	GLAZE
9	1.058147	SMOKE
10	1.196295	DRIFTING_SNOW
11	1.022154	INCOME_PER_CAPITA
12	2.380999	POP_DENSITY
13	2.326598	WHITE_PERCENT

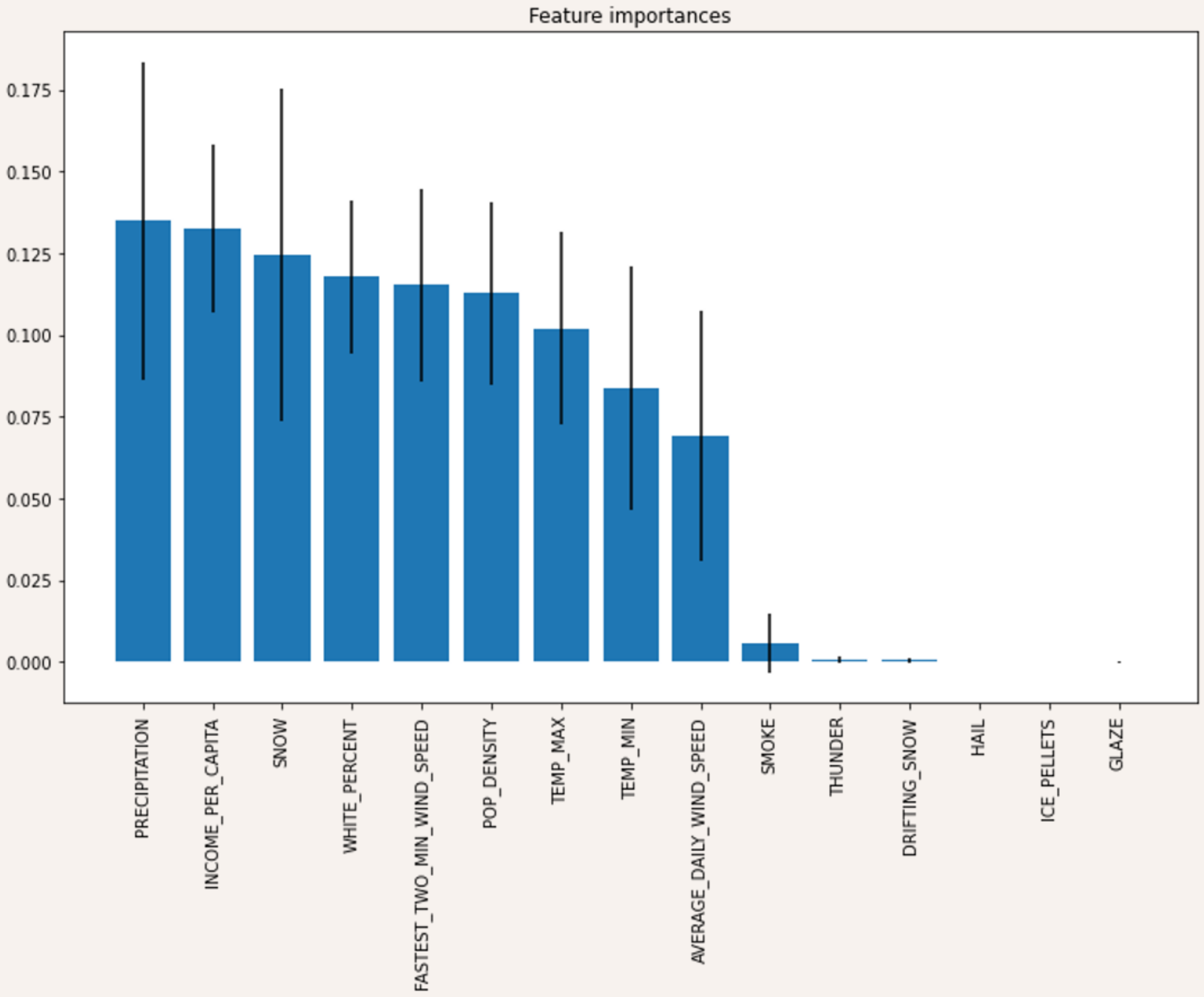
```
                            OLS Regression Results
==============================================================================
Dep. Variable:         FIRST_RESPONSE   R-squared:                       0.034
Model:                            OLS   Adj. R-squared:                  0.034
Method:                 Least Squares   F-statistic:                     516.3
Date:                Thu, 08 Dec 2022   Prob (F-statistic):               0.00
Time:                        14:47:40   Log-Likelihood:            -1.9475e+06
No. Observations:              204548   AIC:                         3.895e+06
Df Residuals:                  204533   BIC:                         3.895e+06
Df Model:                          14
Covariance Type:            nonrobust
================================================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
AVERAGE_DAILY_WIND_SPEED      12.1849      5.773      2.111      0.035       0.870      23.500
PRECIPITATION                 17.7483      0.520     34.161      0.000      16.730      18.767
SNOW                           3.8845      0.240     16.187      0.000       3.414       4.355
TEMP_MAX                      -9.4366      0.777    -12.148      0.000     -10.959      -7.914
FASTEST_TWO_MIN_WIND_SPEED    78.6780      3.382     23.261      0.000      72.048      85.308
THUNDER                     -368.1413     38.547     -9.550      0.000    -443.693    -292.589
ICE_PELLETS                 -722.7639    149.433     -4.837      0.000   -1015.649    -429.879
HAIL                        -444.3928    179.342     -2.478      0.013    -795.898     -92.888
GLAZE                       -751.8692    111.377     -6.751      0.000    -970.165    -533.573
SMOKE                        488.1199     33.989     14.361      0.000     421.503     554.737
DRIFTING_SNOW               -850.7980    124.692     -6.823      0.000   -1095.191    -606.405
INCOME_PER_CAPITA             -0.0025      0.000    -11.733      0.000      -0.003      -0.002
POP_DENSITY                   -0.0040      0.001     -4.081      0.000      -0.006      -0.002
WHITE_PERCENT                169.9709     67.507      2.518      0.012      37.659     302.282
Intercept                   -363.5449     64.980     -5.595      0.000    -490.905    -236.185
==============================================================================
Omnibus:                   456963.597   Durbin-Watson:                   1.506
Prob(Omnibus):                  0.000   Jarque-Bera (JB):      3735615819.893
Skew:                          20.979   Prob(JB):                         0.00
Kurtosis:                     663.717   Cond. No.                     1.40e+06
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.4e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
```

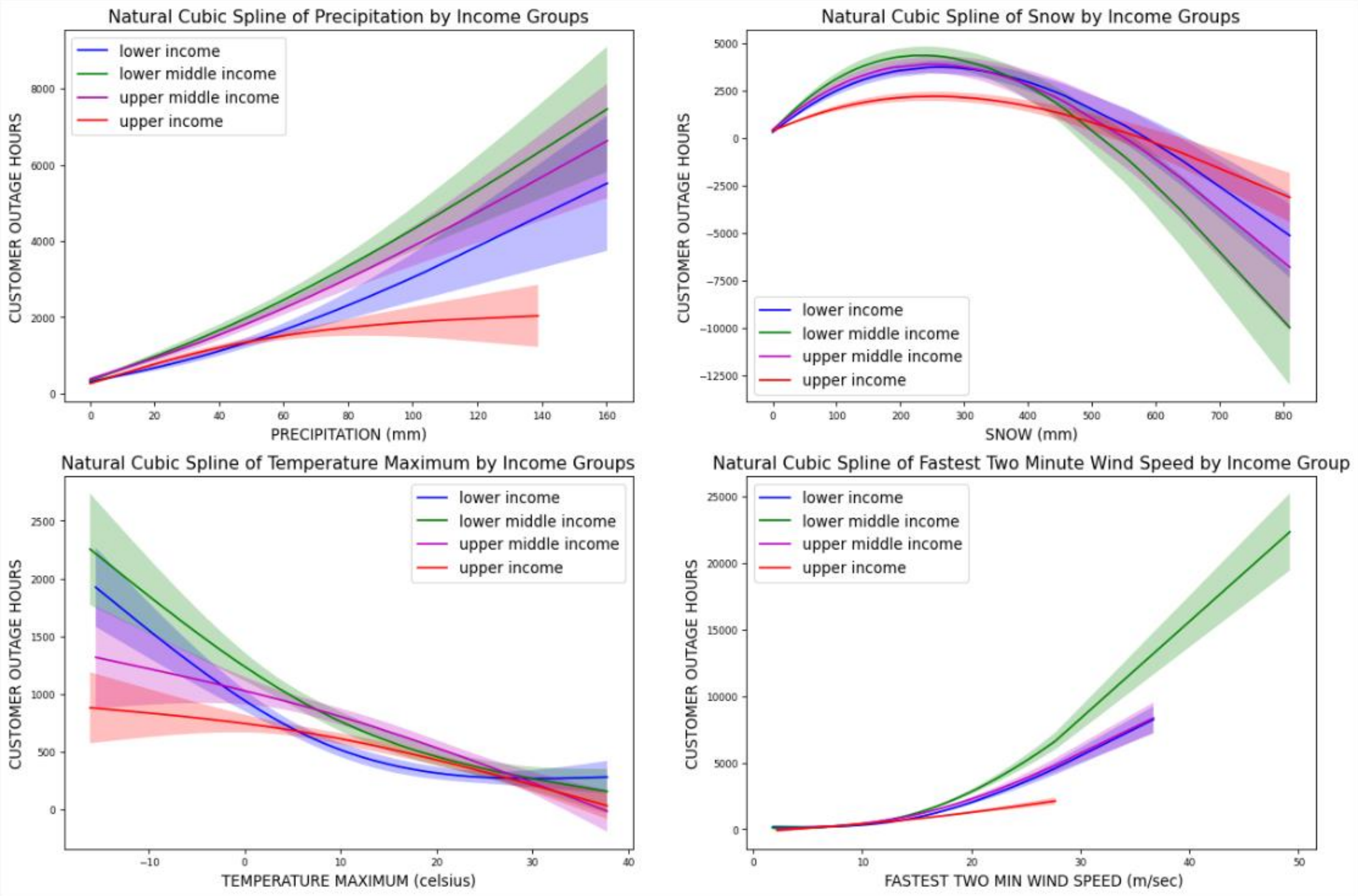
- No features with VIF >5
- Linear regression is not entirely appropriate due to non-linear nature of the data

Random Forest for Feature Selection

Model Tuning:

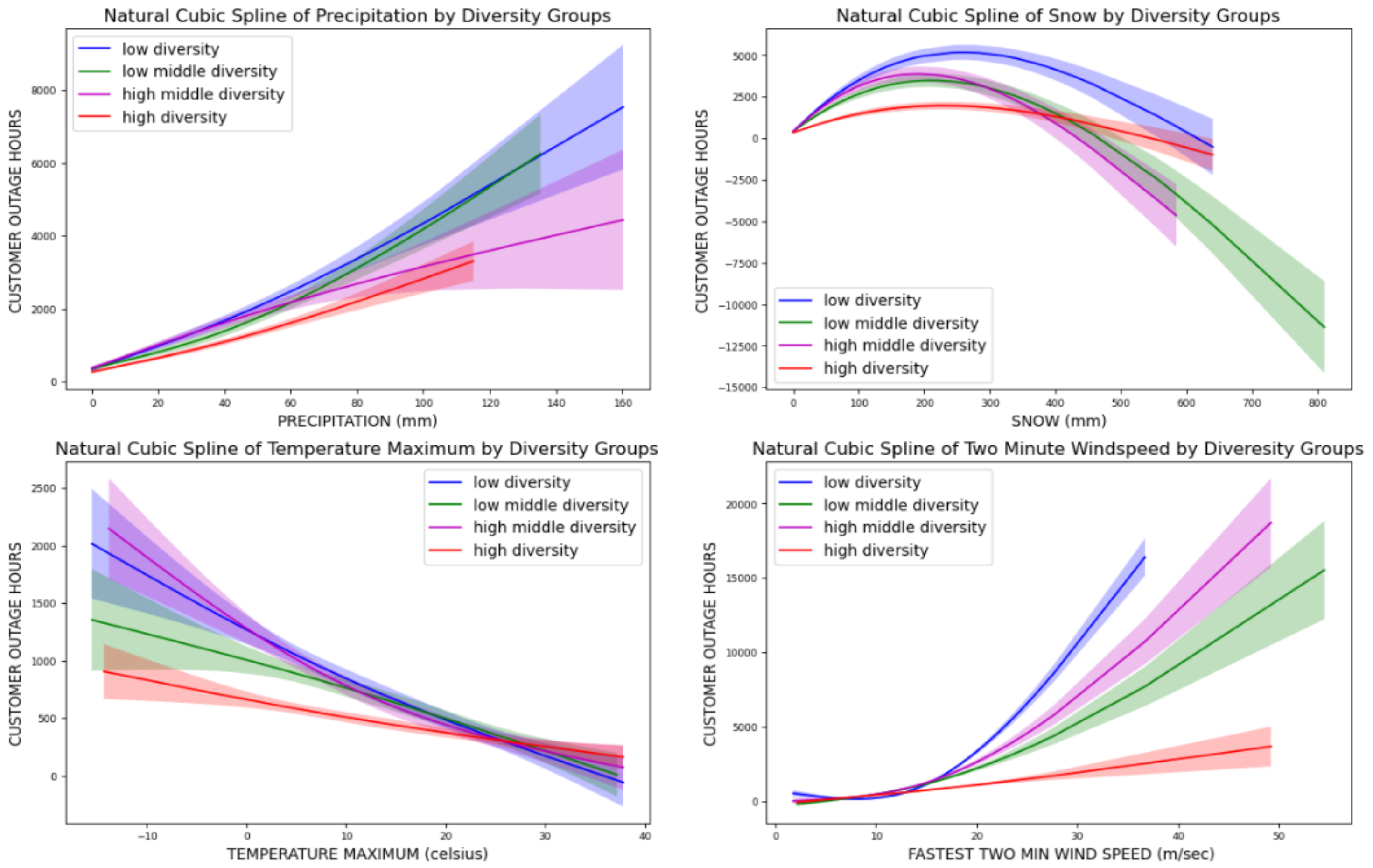
Randomized grid search over 10-100 estimator trees, 3-100 max leaf nodes, 1-10 max depth, 1-30 minimum samples at leaf nodes, 2-20 minimum samples needed to split nodes





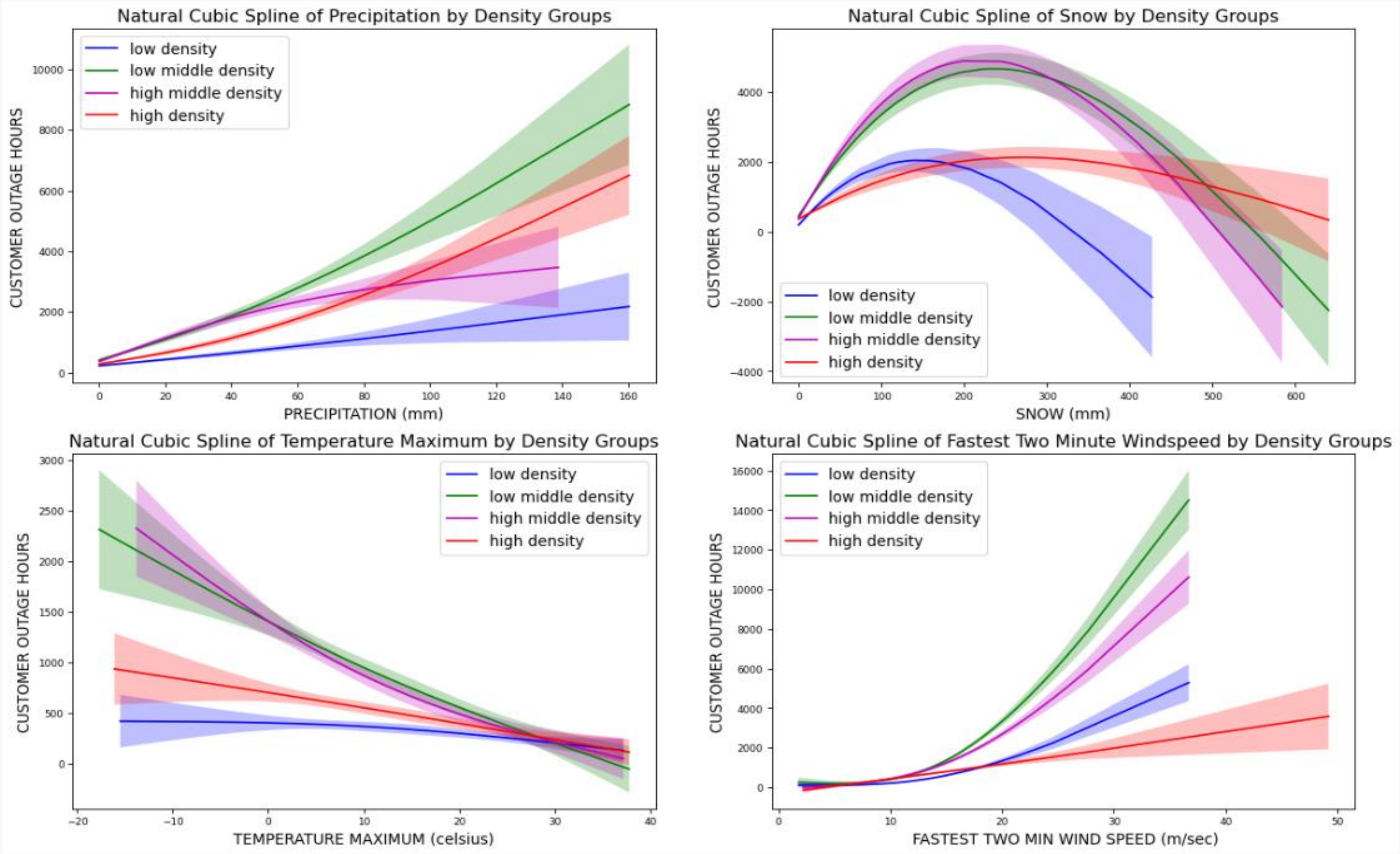

```
LOWER INCOME MSE PRECIPITATION is:  9725784.50
LOWER MIDDLE INCOME MSE PRECIPITATION is:  17380614.52
UPPER MIDDLE INCOME MSE PRECIPITATION is:  11618896.27
UPPER INCOME MSE PRECIPITATION is:  6370298.43

LOWER INCOME MSE SNOW is:  9673299.85
LOWER MIDDLE INCOME MSE SNOW is:  17397606.94
UPPER MIDDLE INCOME MSE SNOW is:  11724358.37
UPPER INCOME MSE SNOW is:  6442654.15

LOWER INCOME MSE TEMP_MAX is:  9770031.86
LOWER MIDDLE INCOME MSE TEMP_MAX is:  17529461.44
UPPER MIDDLE INCOME MSE TEMP_MAX is:  11805575.64
UPPER INCOME MSE TEMP_MAX is:  6451798.23

LOWER INCOME MSE FASTEST_TWO_MIN_WIND_SPEED is:  9591239.21
LOWER MIDDLE INCOME MSE FASTEST_TWO_MIN_WIND_SPEED is:  17140938.23
UPPER MIDDLE INCOME MSE FASTEST_TWO_MIN_WIND_SPEED is:  11511595.84
UPPER INCOME MSE FASTEST_TWO_MIN_WIND_SPEED is:  6367808.28
```

- Splines trained using five-fold cross validation and one standard error rule on optimal degrees of freedom




```
LOWER DIVERSITY MSE PRECIPITATION is:  15968231.67
LOWER MIDDLE DIVERSITY MSE PRECIPITATION is:  13798459.94
UPPER DIVERSITY MSE PRECIPITATION is:  15914845.68
UPPER MIDDLE DIVERSITY MSE PRECIPITATION is:  4598342.42

LOWER DIVERSITY MSE SNOW is:  15982293.79
LOWER MIDDLE DIVERSITY MSE SNOW is:  13883164.39
UPPER MIDDLE DIVERSITYMSE SNOW is:  15914630.34
UPPER DIVERSITY MSE SNOW is:  4654937.14

LOWER DIVERSITY MSE TEMP_MAX is:  16176116.25
LOWER MIDDLE DIVERSITY MSE TEMP_MAX is:  13982188.23
UPPER MIDDLE DIVERSITY MSE TEMP_MAX is:  16001096.32
UPPER MSE TEMP_MAX is:  4679671.67

LOWER DIVERSITY MSE FASTEST_TWO_MIN_WIND_SPEED is:  15421005.83
LOWER MIDDLE DIVERSITY MSE FASTEST_TWO_MIN_WIND_SPEED is:  13697676.95
UPPER MIDDLE DIVERSITY MSE FASTEST_TWO_MIN_WIND_SPEED is:  15701788.62
UPPER DIVERSITY MSE FASTEST_TWO_MIN_WIND_SPEED is:  4619300.91
```

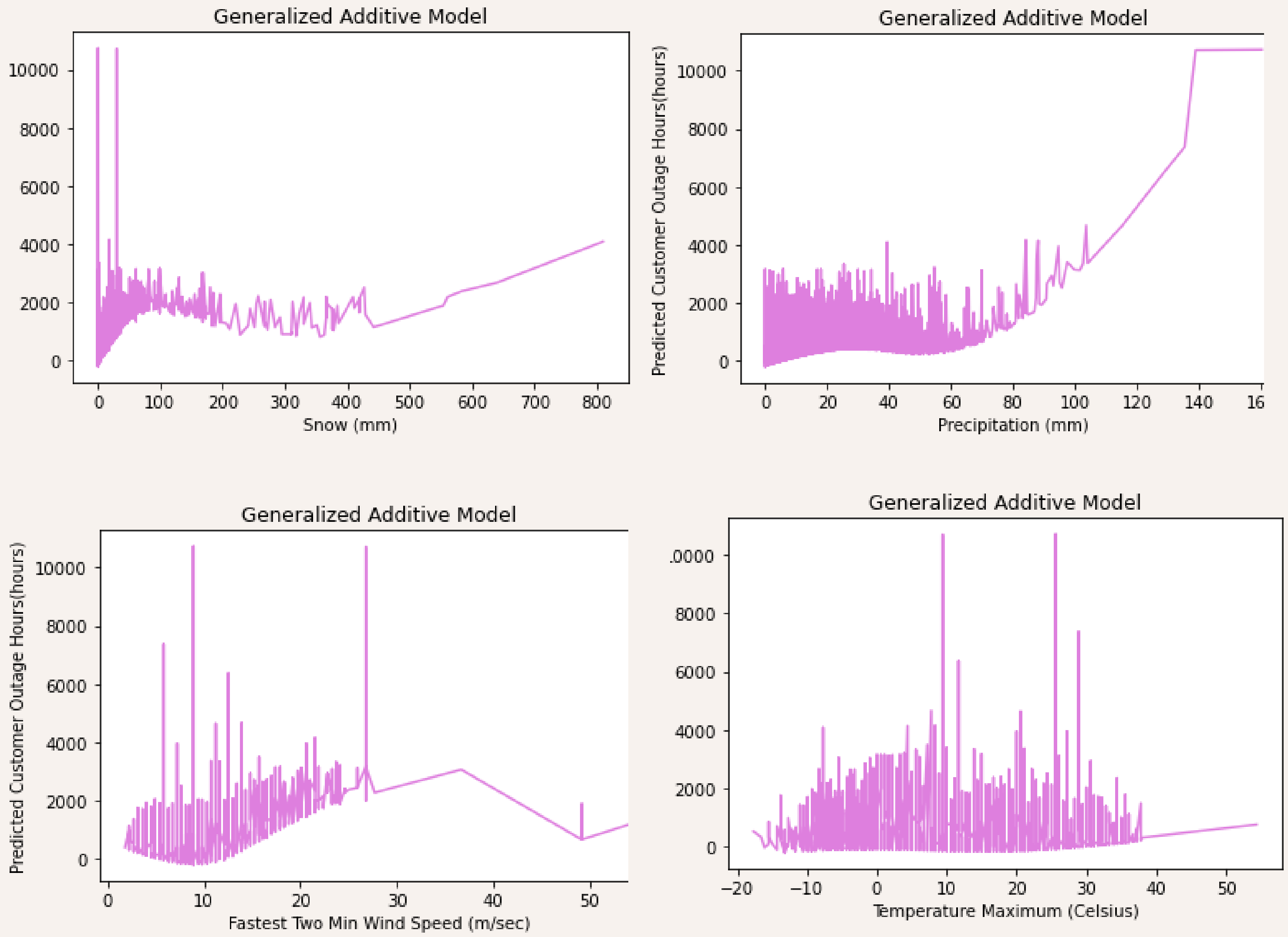
- Splines trained using five-fold cross validation and one standard error rule on optimal degrees of freedom




```
LOWER DENSITY MSE PRECIPITATION is:  4910824.27
LOWER MIDDLE DENSITY MSE PRECIPITATION is:  17857405.32
UPPER MIDDLE DENSITY MSE PRECIPITATION is:  15937911.55
UPPER DENSITY MSE PRECIPITATION is:  5252088.97

LOWER DENSITY MSE SNOW is:  4887869.03
LOWER MIDDLE DENSITY MSE SNOW is:  17885758.61
UPPER MIDDLE DENSITY MSE SNOW is:  15979967.24
UPPER DENSITY MSE SNOW is:  5331482.00

LOWER DENSITY MSE TEMP_MAX is:  4930993.30
LOWER MIDDLE DENSITY MSE TEMP_MAX is:  17997000.89
UPPER MIDDLE DENSITY MSE TEMP_MAX is:  16085817.39
UPPER MSE TEMP_MAX is:  5352891.14

LOWER MSE FASTEST_TWO_MIN_WIND_SPEED is:  4855199.24
LOWER MIDDLE MSE FASTEST_TWO_MIN_WIND_SPEED is:  17537329.30
UPPER MIDDLE MSE FASTEST_TWO_MIN_WIND_SPEED is:  15695909.00
UPPER MSE FASTEST_TWO_MIN_WIND_SPEED is:  5286036.48
```

- Splines trained using five-fold cross validation and one standard error rule on optimal degrees of freedom

GAMs



- Model Tuning: Built using natural cubic splines; Splines trained using five-fold cross validation and one standard error rule on optimal degrees of freedom