MGT 6203 Group Project Proposal

Team 37; Team Members:

Carina Grady; cgrady6; I graduated with a B.S. in Mechanical Engineering in 2019 from Iowa State University with a technical focus in energy. I currently work as a senior consultant at Guidehouse for various utility clients across the US. I mostly work in data management and impact energy evaluation, namely for commercial and residential energy efficiency programs. These programs gauge how well clients' customer incentive program is working, which impacts their climate goals. I use data analytics frequently, managing large sets of data in R so the client can make informed decisions.

Amanda Wijntjes; awijntjes3; I earned a B.S. in Biomedical Engineering with a minor in Computer Science, from Georgia Tech in May 2021. After graduating, I joined an early talent engineering rotational program at Merck, a Fortune 500 pharmaceutical company. Through this program, I've gained hands-on technical and operational experience in vaccine manufacturing, including live virus and polysaccharide vaccines. Analytics projects include utilizing SQL and Python to clean and model manufacturing data (i.e. SAP) and Power BI for data visualization.

Ty Underwood; tunderwood30; I graduated with a B.S. in Economics from Texas A&M University in 2017. My current role is working as an engineer/analyst in the Analytics Innovation and Development (Engineering / R&D) group at a mid-sized IT company (Inmar Intelligence). Most of the professional analytics projects I've worked on have been in the area of consumer behavior in the supermarket retailer space, building internal and external analysis applications for our customers and internal business partners (Python/R/YAML/SQL/Google Cloud Platform).

Joshua Feldman; jfeldman31; I love the weather and have pursued a career in meteorology since I was in elementary school. I accepted a position at WeatherOptics, a weather intelligence company, after graduating from Stony Brook University with a dual B.S. in Atmospheric and Oceanic Sciences and in Physics. I primarily work as a forecaster for hospital coalitions and as a forensic and predictive analyst. I commonly work with Python for computational and GIS applications, SQL, and Google Cloud Platform. One of my proudest accomplishments is the development of a model that predicts freight delay times based on weather and environmental conditions. This model has since become WeatherOptics' flagship product, *RightRoute*.

Micah Owens; mowens47; I work as a director of analytics and marketing technology for Sandbox Group, where I lead a centralized team that responds to analytics, testing, and optimization needs from about a dozen educational entertainment and games brands around the US and Europe. This has let me work on and lead analytics projects in consumer behavior, mobile app engagement, user acquisition, paid media, a/b testing, and web optimization (using R, SQL, Python and data testing/visualization tools). I received my undergraduate in 2018 from Belmont University, with a Bachelors in Business Administration. In this program, I received a comprehensive education in financial management and forecasting, and also learned the fundamentals of programming and statistics.

OBJECTIVE/PROBLEM

Project Title:
"Navigating Renewable Energy Markets - A Utility Company's Guide to Effective Forecasting"

Background Information on chosen project topic:
With the increasing global emphasis on sustainable energy, utility companies are under pressure to invest more in renewable energy sources and optimize their current operations. Understanding energy consumption patterns, generation capabilities, and the influence of external factors like weather can help in making informed decisions. Spain has been at the forefront of renewable energy, with a significant portion of its energy consumption coming from renewable sources. With the increasing variability of intermittent renewable energy sources, such as wind and solar, there's a need to understand and predict energy consumption patterns and pricing to optimize operations and investments. A Transmission System Operator (TSO) oversees the high-voltage transmission grid, ensuring electricity transfer from generation sources to consumers. The dataset, sourced from ENTSOE and Spanish TSO Red Electric España, offers hourly electrical consumption and forecast data, complemented by weather insights from the Open Weather API for Spain's major cities.

Problem Statement:
We aim to predict electric demand and pricing by analyzing weather patterns and energy generation options so that utility companies can make informed decisions on renewable energy investments.

Primary Research Question (RQ):
How do energy consumption patterns and pricing correlate with weather patterns, and how can this information be used to predict future energy pricing and demand?

Supporting Research Questions:

- How well does our predictive model perform in forecasting electrical pricing, and can we demonstrate improved accuracy compared to TSO forecasts, particularly when considering time-of-day variations?
- Which specific weather measurements (e.g., temperature, humidity, wind speed) have the most significant impact on electrical consumption and pricing within the dataset?
- How can we optimize the utility's renewable energy operations based on predicted energy demands?

Business Justification:
Improving our forecasting capabilities for electrical consumption and pricing presents an exciting proposition from a business perspective. Accurate forecasts allow for optimized resource allocation, reducing the need for costly last-minute adjustments and minimizing excess capacity. This operational efficiency directly translates to substantial cost savings. Moreover, aligning our pricing strategies with real-time market dynamics ensures maximum revenue capture while maintaining a competitive edge. From a customer perspective, consistent energy supply, fewer outages, and transparent pricing enhances consumer satisfaction and confidence in the company. On the sustainability front, precise forecasting is pivotal to meeting global climate goals. Predicting and optimizing peak load facilitates the seamless integration of renewable energy sources into our grid. This foresight is invaluable, allowing us to prepare for shifts in energy goals and ensuring our continued growth and relevance in a rapidly evolving market.

By tackling the issue of energy consumption, we would allow our business (and other similar businesses / government entities) to gain insight on the energy utilization trends of Spain and apply prediction methods to forecast pricing and consumer usage among different energy generation methods. Overall, this would be used in most companies / entities' strategic planning departments and can help prepare for future growth and/or decline of energy methods.

DATASET/PLAN FOR DATA

Data Sources (links, attachments, etc.):
[Hourly energy demand generation and weather (kaggle.com)]

Data Description:
Our dataset includes four years of electrical consumption, generation, pricing, and weather data for Spain. The weather dataset contains 17 columns of data: datetime index localized to CET, city, temperature (Kelvin), min temperature (Kelvin), max temperature (Kelvin), pressure (hPa), humidity (%), wind speed (m/s), wind direction (degrees), rain - last hr (mm), rain - last 3 hrs (mm), snow - last 3 hrs (mm), cloud cover (%), weather description code, short current weather description, long current weather description, weather icon code.

Figure 1. Sample data rows from the weather dataset.

1	dt_iso	city_name	temp	temp_min	temp_max	pressure	humidity	wind_speed	wind_deg	rain_1h	rain_3h	snow_3h	clouds_all	weather_id	weather_main	weather_description	weather_icon
2	2015-01-01 00:00:00+01:00	Valencia	270.475	270.475	270.475	1001	77	1	62	0	0	0	0	800	clear	sky is clear	01n
3	2015-01-01 01:00:00+01:00	Valencia	270.475	270.475	270.475	1001	77	1	62	0	0	0	0	800	clear	sky is clear	01n
4	2015-01-01 02:00:00+01:00	Valencia	269.686	269.686	269.686	1002	78	0	23	0	0	0	0	800	clear	sky is clear	01n
5	2015-01-01 03:00:00+01:00	Valencia	269.686	269.686	269.686	1002	78	0	23	0	0	0	0	800	clear	sky is clear	01n
6	2015-01-01 04:00:00+01:00	Valencia	269.686	269.686	269.686	1002	78	0	23	0	0	0	0	800	clear	sky is clear	01n

The energy dataset contains 29 columns of data:

- Datetime index localized to CET
- Energy Demand in unit MW: biomass generation, coal/lignite generation, coal gas generation, gas generation, coal generation, oil generation, shale oil generation, peat

generation, geothermal generation, hydro1 generation, hydro2 generation, hydro3 generation, hydro4 generation, sea generation, nuclear generation,other generation, other renewable generation, solar generation, waste generation, wind offshore generation, wind onshore generation, forecasted solar generation, forecasted offshore wind generation, forecasted onshore wind generation, forecasted electrical demand, actual electrical demand

- Cost per energy in unit EUR/MWh: forecasted price, actual price.

Figure 2. Sample data rows from the energy dataset.

1	time	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	generatio	forecast s	forecast w	forecast w	total load	total load	price day	price actual
2	2015-01-0	447	329	0	4844	4821	162	0	0	0		863	1051	1899	0	7096	43	73	49	196	0	6378	17		6436	26118	25385	50.1	65.41
3	2015-01-0	449	328	0	5196	4755	158	0	0	0		920	1009	1658	0	7096	43	71	50	195	0	5890	16		5856	24934	24382	48.1	64.92
4	2015-01-0	448	323	0	4857	4581	157	0	0	0		1164	973	1371	0	7099	43	73	50	196	0	5461	8		5454	23515	22734	47.33	64.48
5	2015-01-0	438	254	0	4314	4131	160	0	0	0		1503	949	779	0	7098	43	75	50	191	0	5238	2		5151	22642	21286	42.27	59.32
6	2015-01-0	428	187	0	4130	3840	156	0	0	0		1826	953	720	0	7097	43	74	42	189	0	4935	9		4861	21785	20264	38.41	56.04

Key Variables:

Independent variables: weather measurements, cities, energy generation/types, time
Dependent variables: electrical demand/consumption, prices/cost

We hypothesize weather measurements and time will be the most important variables for predicting electrical demand and prices/cost.

APPROACH/METHODOLOGY

Planned Approach

Load and cost are two critical factors we wish to examine in this study. We feel that a random forest would be a good choice for load prediction due to non-linear relationships between features like time of day and load. The data are auto-correlated so randomly splitting the data would not be ideal for this model. Since the dataset spans Jan 1, 2015 through Feb 26, 2016, we have one complete year of data for our study. The 2016 dates can be left for testing. Out-of-bag error will be used for verification of our model throughout each permutation of the training phase. Feature selection will be decided after building a model with all of our independent variables. Only features responsible for large variance reduction will be selected for the remainder of model training. The number of estimators, tree depth, and minimum samples required to split a leaf can then be “optimized” as the value at which out-of-bag error levels off.

While understanding load patterns is crucial for operations efficiency and reliability, it’s important for us to understand electricity cost for business growth and renewable investment. We’re specifically interested in solar and wind. We can examine the relationship between total solar/wind generation and cost using a simple linear regression model. Assuming we find these sources significant, R^2 can be used to interpret how much of the price can be explained by wind and solar. A second model can include other sources of electricity to compare the price sensitivity of each generation method. There may be interaction with features like time of day, temperature, and precipitation, as peak demand times and inclement weather may naturally lead to increased demand. A third regression model with interaction features can inform us if solar and wind generation during those peak demand times help reduce cost.

Anticipated Conclusions/Hypothesis:
We anticipate clear patterns will emerge in the data to indicate one or more of the analyzed predictors have a significant influence on electrical demand, prices, and generation capacity. We expect that weather measurements will have a significant impact on electrical demand and pricing, which can inform the energy company's decisions and budget for green energy investment. It is expected that our decision tree approach and error verification using out-of-bag sampling will provide a more accurate and reliable predictive model for electrical consumption compared to the Transmission System Operator forecasts.

What business decisions will be impacted by the results of your analysis? What could be some benefits?
We expect benefits for companies in the distribution of energy generation resources. For example a government energy provider could use this analysis to inform its level of investment in solar and wind generation plants. The analysis will also inform decisions on how a utility should generate, purchase, or store electricity, leading to cost savings. Additionally, we anticipate that our model's precise pricing forecasts will improve decisions about how we price electricity. By adjusting rates based on anticipated usage, the company may achieve significant savings. The outcomes of this analysis will enhance operational efficiency by minimizing the need for last-minute adjustments and costly emergency measures. This will translate into a reduction in power outages and supply shortages, ultimately boosting customer satisfaction and retention. Finally, the improved utilization and integration of renewable energy sources may align with companies' environmental, social, and governance (ESG) mandates.

PROJECT TIMELINE/PLANNING

Project Timeline/Mention key dates you hope to achieve certain milestones by:
10/06/2023: Finalize project plan
10/15/2023: Complete data cleaning and joining in preparation for analysis
10/31/2023: Build decision tree model and forecast, and begin analysis
11/04/2023: Compile and submit progress report
11/15/2023: Finish model analysis and regression testing
11/27/2023: Finalize and compile findings into final report
12/03/2023: Submit final report