CSEE5590/490: Big Data Programming

Project Proposal (Increment 1)

Due Date: Friday, Feb 26, 2021

Project Title: Analysis for determination of a relationship between energy demand and weather.

Team Members: Joe Goldsich, Anna Johnson, Jongkook Son, and Bill Yerkes

Goals and Objectives: Utilize the tools and technologies learned from CSEE 5590 to be able to analyze collected data so that it will be possible to determine if there is a relationship between weather and energy consumption and if a relationship exists determine the possibilities of using that relationship to predict future energy needs.

Motivation: The global population continues to increase, and the weather patterns seem to be getting more extreme, from extending periods of both above and below normal temperatures in various parts of the world and in the United States. The demand and consumption of energy increase with the population and with the extreme weather, the need for air conditioning in the summer and for heating in the winter. The recent crisis in Texas has demonstrated what can happen if the energy providers are not able to meet the demands of the consumers. Being able to forecast accurately future demand and plan accordingly can help prevent or mitigate such crises in the future.

Significance: Better planning of resources for Utility Companies can result in reduced cost to the consumers and more reliable service. This also dips into the area of public safety, as loss of power during extreme weather with no warning can be dangerous for vulnerable groups.

Objectives:

Visualize the load and marginal supply curves.

What weather measurements and cities influence most the electrical demand, prices, generation capacity?

Can we forecast 24 hours in advance better than the TSO?

Can we predict electrical prices by time of day better than TSO?

Forecast intraday price or electrical demand hour-by-hour

Potentially identify the abilities of different energy sources/ systems of regulation to keep up with fluctuating demand

What is the next generation source to be activated on the load curve?

Features:

Hive (Map Reduce):

Use Hive and Map Reduce to store and gather metrics on the data to be able to feed the Spark, GraphX and MLLib portions of the project.

Sqoop (MySQL):

Sqoop and MySQL can be used to query our assembled database to identify patterns between extreme weather and fluctuations in energy consumption. MySQL can also be used to identify notable power outages in the past, or times when utility companies were able to keep up with rising demands.

Solr & Lucene (Search Engines):

We have not covered these technologies at this point in time. We need to determine how search engines can be applied to the project.

Cassandra (No SQL):

We will determine if we need to store data in a NoSQL repository, versus storing the data in Hive and MySQL, to be able to perform the required work to make a predictive model. (We have not covered Cassandra as of yet, do not have enough information as to how it can benefit in solving the problem.)

Spark / GraphX / MLLib (Programming and Analytics):

Use Spark, GraphX, and MLLib to analyze the weather and electrical data to be able to create a model which can predict the demand/cost of electricity based upon supplied weather information.

Story Telling:

Life

1. **Who** are the people or communities in need of help?

Consumers of energy utility companies (gas and electricity) are one portion of people who need help, as the recent crisis in Texas has demonstrated. The other people who need help are the producer of the energy being consumed (Utility Companies) and their suppliers. Being able to accurately forecast demand can help them better plan on how much to produce, how much raw materials to keep in stock, and how to better plan to use their resources, including labor, to meet their client's demands. In the case where it is not possible for utility companies to keep up with demand, adequate warning could be given to energy consumers so they can plan for loss of power.

2. **What** problem happened to them?

Recently in Texas with the severe cold weather, the demand for electricity far exceeded the capacity the Utility Companies were able to provide. In addition, the cold weather caused some producers of gas to halt production. If the Utility Companies could have foreseen the spike in demand, it may have been possible for them to take actions to mitigate the negative impacts which were caused by the shortages in energy as compared to the demand that was required to keep people warm.

3. **When** did the problem take place?

This recent issue occurred in February 2021. There were also rolling blackouts in Kansas City Missouri in February 2021 because of the cold. There have been rolling blackouts in California because of the heat over numerous years.

4. **Where** means two things:

   a. The environment and settings that the people or the community is living in, and

Due to a rapidly changing climate, these problems can affect a wide range of environments. Communities that are not accustomed to extreme cold or hot weather are particularly vulnerable to energy shortfalls when hit with unexpected demand. Notably in the case of Texas, homes there were designed to shed heat, so when they were hit with uncharacteristically cold weather, the infrastructure was especially unequipped to deal with power outages.

   b. The place/location where the problem takes place.

The problem of energy shortages can occur anywhere, in any country. The recent cases have been in Kansas City Missouri and in Texas. They have occurred in California. They can occur anywhere.

5. **Why** means the possible causes and/origin of the problem.

The inability of the Utility Companies to accurately predict demand and a lack of excess capacity to handle a reduction in production capacity by outside forces results in a failure to meet the demands of the consumer. IT companies have HA and redundancy built into their server farms to help prevent outages of their services. Utility companies need to have the same HA and redundancy built into their systems, and they need to understand and be able to reasonably predict how much demand there will be from their consumers.

6. **How**: If you would like, you can add a dimension of how. How did it happen? Sometimes, the answer to how can be covered by what, when, and where.

In the case of Texas, lawmakers were warned years in advance of the possibility of this situation happening. They ignored recommendations that they winterize their energy infrastructure, and that contributed greatly to the energy shortages as some systems failed. Analysis of energy consumption vs. weather patterns could provide the necessary information to utility companies and lawmakers in advance of disastrous situations, and could help hold them accountable for their lack of preparedness.

Data

Hourly energy demand generation and weather | Kaggle

This dataset contains 4 years of electrical consumption, generation, pricing, and weather data for Spain. Consumption and generation data were retrieved from ENTSOE a public portal for Transmission Service Operator (TSO) data. Settlement prices were obtained from the Spanish TSO Red Electric España. Weather data was purchased as part of a personal project from the Open Weather API for the 5 largest cities in Spain and made public here.

The scientist:

UMKC Students / CSEE 5590 Big Data Programming. .

Anna Johnson, Joe Goldsich, Jongkook Son, and Bill Yerkes

Users

There are two main users for our application. Consumers of energy utility companies and producers of the energy being consumed.

The Society

Thanks to our application, The overall total utility/energy cost for our society would decrease because each subject would be able to act appropriately according to the prediction of the application. Producers would be able to expand or decrease their production line based on the weather forecast. Consumers of energy would be able to avoid huge amounts of electricity bills because of a more efficient system.

Contribution of Work

Anna Johnson:

Gather data sources. Utilize Hive and Map Reduce to store and gather metrics on the Data. Set up Discord Server. Generate Queries and perform testing. Contribute to project documentation.

Joe Goldsich:

Configure MySQL and possibly Cassandra to store data for analysis. Utilize Sqoop to move data between databases and hive. Generate Queries and perform testing. Contribute to project documentation.

Jongkook Son:

Spark, GraphX and MLLib programming to perform analysis of data to determine trends and relationships contained in the data. Set up google doc for documentation purposes. Contribute to project documentation.

Bill Yerkes:

Spark, GraphX and MLLib programming to perform analysis of data to determine trends and relationships contained in the data. Set up Github repository for storing code and documentation. Contribute to project documentation. Coordinate work efforts.

References:

https://www.kaggle.com/nicholasjhana/energy-consumption-generation-prices-and-weather

Thousands caught off guard by rolling blackouts in Kansas City metro Monday (fox4kc.com)

https://www.texastribune.org/2021/02/17/texas-power-grid-failures/

https://www.forbes.com/sites/arielcohen/2021/02/19/texas-energy-crisis-is-an-epic-resilience-and-leadership-failure/?sh=46d08806eee8

California rolling blackouts during summer heat wave caused by 3 main factors, report says | Fox Business