The Pennsylvania State University

The Graduate School

College of Engineering

ANALYSIS OF HOURLY WEATHER FORECASTS AS AN INDICATOR OF OUTAGE CHARATERISTICS IN AN ELECRIC POWER SERVICE SYSTEM

A Thesis in

Industrial Engineering

by

Vivek Ananthakrishnan

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2016

The thesis of Vivek Ananthakrishnan was reviewed and approved* by the following:

Vittaldas Prabhu
Professor, Industrial and Manufacturing Engineering
Thesis Advisor

Saurabh Bansal
Assistant Professor of Supply Chain Management
Supply Chain and Information Systems

Janis P. Terpenny
Peter and Angela Dal Pezzo Department Head
Harold and Inge Marcus Department of Industrial and Manufacturing Engineering

*Signatures are on file in the Graduate School

ABSTRACT

The world is so dependent on electricity in the 21st century that any interruption or failure in the electric service system could lead to major disruption of daily life and economic losses. Therefore electric utility companies strive to maintain high customer service levels and try to make the best use of past power outage experiences. Increasing capabilities of analytics and improved accuracy of short-term weather forecasts present opportunities for electric utilities to leverage large amount of historic data to forecast outages and proactively make decisions to minimize disruptions. This thesis presents an analytic approach to evaluate the possibility of using hourly weather forecasts to predict outage impact characteristics such as customer minutes interrupted (CMI) and duration of an outage. The key variables in predicting outage durations are identified and their effect on predicting the characteristics is estimated. Furthermore, a cause-effect analysis is done to determine the equipment affected due to various causes of outage, which would help reduce time to repair and enable companies to develop plans to ensure better allocation of resources during an outage and prioritize maintenance schedules in advance of an event. Data from a utility company service location is used as a case study to discuss the methodology proposed and validate the models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would take this opportunity to thank everyone who has helped me in successfully completing this thesis.

Firstly, the support, encouragement and valuable guidance of my advisor Dr. Vittaldas Prabhu was extremely crucial throughout this thesis and I would like to sincerely thank him for the same.

I would also like to thank Dr. Saurabh Bansal for agreeing to be reader for this thesis and providing me with valuable insights.

A big thank you to Mr. Jack Steele for providing me with the data for this thesis and also providing guidance throughout the course of the thesis.

I would like to thank Sharan Srinivas for his guidance and support at every critical stage of the thesis. To all my colleagues, friends and mentors who have helped me in every difficult situation. Lastly, I would like to thank my family for their never ending support and providing me the right platform to achieve my goals.

Chapter 1

INTRODUCTION

In the current world industrial engineers look to optimize every form of the manufacturing system, from the production of the product to the development of an effective supply chain. The priority of every organization is two-fold: Providing customers the best possible product at the lowest cost; maintaining different quality standards. The scenario is not very different in the case of a service industry. In the case of an electric supply service system for an example, every company tries to provide the most reliable service in the form of consistent and uninterrupted service to its customers while reducing cost of supply and maximizing profits. While this would work in an ideal world, often companies have to deal with problems like downtime or outage time due to a failure in the supply system and finding solutions to these problems have been of highest importance in order to provide reliable service.

1.1 Overview

The US Department of Energy believes that a fully functioning smart grid will mean improved reliability, better efficiency and reduce America's dependence on foreign oil. It further adds that since 1998 there has been a growth in peak demand for electricity, exceeding transmission growth by almost 25% every year. This growth has been driven by population growth, bigger houses and more electric/electronic devices per household.

Power companies and electric utilities are increasingly focusing on taking steps to improve reliability of the electric power distribution system.

Past data says that electric power outages lead to heavy economic losses for both the electric utility companies as well as the end users. In 2000, a one-hour outage that hit the Chicago Board of Trade resulted in a loss of $20 trillion due to trades delayed. The US Department of Energy report further adds that the electricity system today is 99.97 percent reliable, but power outages and interruptions still cost Americans at least $150 billion each year, which breaks down to $500 per person [1]. Between the years 2003 and 2012, weather events caused roughly 679 power outages that each affected at least 50,000 customers. That is a significant number considering that the overburdened power grid delivers electricity to more than 144 million end-use customers in the United States [2]. The economic costs of an outage is often passed on to consumers due to the congestion in the system and this drastically impacts the customer satisfaction levels. A one hour outage costs a commercial business a little more than $1000 in addition to impacting public safety and risk, says the American Society of Civil Engineers report [43].

Outage costs are directly proportionate to the customer's dependence on electricity during an outage and with annual electricity use in a typical U.S. home increasing 61 percent since 1970 it is becoming increasingly important to prevent outages or at the least reduce the duration of an outage through better recovery planning and forecasting. Outage costs vary significantly depending on the outage attributes such as timing (season and time-of-day), advance notice, frequency, duration and severity of the outage. For the purpose of the thesis an outage is defined as a complete or total loss of service, typically resulting from a distribution-related cause or transmission failure.

1.2 Significance of the area of research

There are several reasons attributed to an electric power outage that question the reliability of the Electric Power System. The causes could range from equipment failure and overloading of the line to major weather related events. The power systems are most vulnerable to storms and extreme weather events and electric utility companies can improve the recovery from weather related outages by enhancing the overall condition and efficiency of the power delivery system. Seasonal storms combined with wind, snow, rain, ice etc. cause significant outages and the data on weather related outages have been used in the past to estimate the costs of an outage and the impact it has on life and the industries. According to past weather related outage data, 90% of customer outage-minutes are due to events which affect local distribution systems, while the remaining 10% stem from generation and transmission problems, which affect a larger number of end-users as they cause widespread outage [3].

Increasingly companies are looking to tackle outages due to both local distribution systems and larger transmission systems by developing strategies to reduce or prevent outages. 72 hour forecasts of an electric power outage and the cost parameters associated with the outage would help companies optimize their manpower and resource planning thereby helping to reduce costs. These concerns have helped to stimulate research activity in the area of electric power supply systems modelling and analysis during the past decade as discussed in Chapter 2.

1.3 Importance for Utility Companies

Utility companies are responsible for the infrastructure of the power delivery system and improving the resiliency of the system. They stand to lose the maximum due to an outage since they cannot sell electricity and hence it is imperative that they take steps to prevent outages through better prediction via resource planning to improve recovery time to an outage caused by a weather related event. Traditionally, utility companies have planned in advance to tackle any interruptions that could occur due to a severe weather condition and have taken steps to ensure timely recovery form such disasters and minimize the outage time. Generally, the companies look at multiple level Pareto charts to drill down and get to the root cause of the outage and develop a strategy to be followed both in the event of an outage as well as during lean time. They are often most interested in physical damage to the electric power system since this directly affects restoration times and costs. During lean time companies ensure better and timely maintenance of equipment. But, the real concern for companies is to forecast the outage durations or look even further at predicting the customer minutes interrupted.

With the accuracy of weather forecasts improving drastically companies can use this to their advantage to estimate the impact of an outage in terms of the outage duration. Figure 1.1 shows that the forecast error in miles of predicting storms and hurricanes has been improving with every passing year and 96 – 120 hour predictions are nearly as accurate as the actual occurrence. Other recent weather predictions show that forecasts 4-5 days out as nearly as accurate as 1 day forecast which gives enough time for companies to manage their resources and take necessary actions to minimize the impact of an outage.

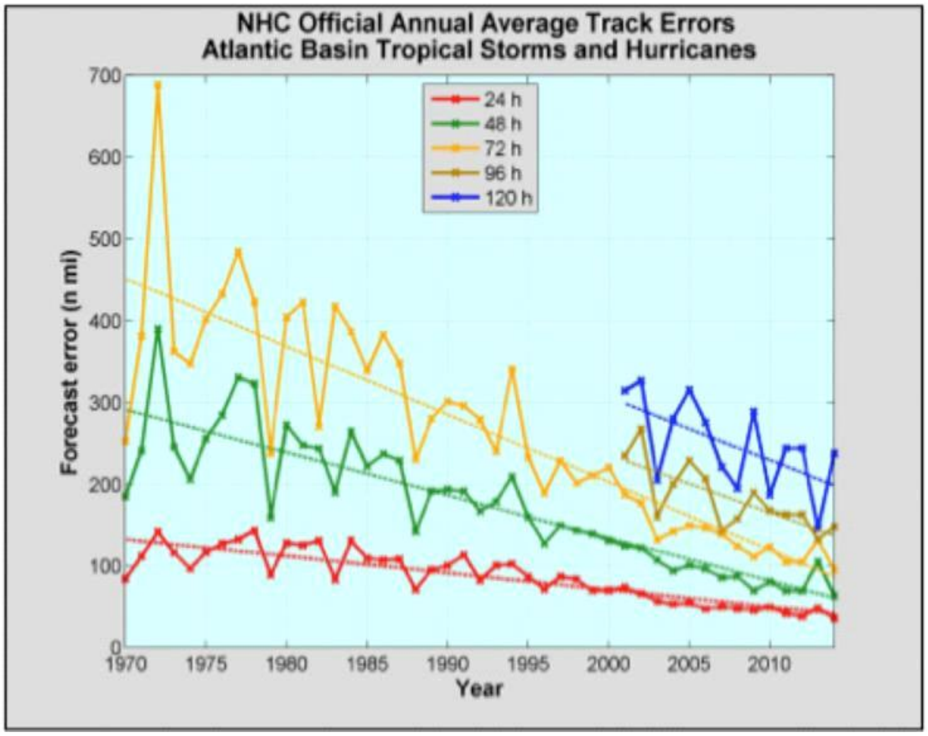


Figure 1.1: Trends in the errors of Atlantic tropical storm / hurricane forecast

Source: (Courtesy NOAA, National Hurricane Center)

Companies must look to use the weather forecast to help answer a part of the variation in the duration of a power outage. With the availability of outage data for model validation, companies must utilize hourly weather forecast in addition to other attributes to help predict the customer units interrupted or the outage duration. Having done this they can then look to prioritize the response to an outage depending on the estimated impact and the equipment that is frequently affected due to the weather event (Cause-Effect), optimizing their man power allotment in expectation of an outage.

The predictive capability of regular weather forecasts to estimate the duration of an outage at the hourly level during the occurrence of a weather related event is analyzed and a method to help decision makers classify the customer minutes interrupted into predefined categories is discussed. Hourly weather data is shown to account for 40% of the variation the customer minutes interrupted. Inclusion of further variables would help decisions makers prioritize their response to an outage prediction. The equipment that is most commonly affected due to each category of the cause of interruption is also identified to give insights into the type of maintenance and work force scheduling that needs to be done in order to ensure prevention or faster recovery from an outage event.

The remainder of this thesis is organized as follows: The past literature on the area of research is reviewed in Chapter 2, explaining the importance of the topic of research, the current issues that electric utility companies face, the impact of outages on the world as well as past research to assess the impact of weather related outages. Chapter 3 presents the process methodology that is followed. Chapter 4 provides a detailed description of the data used for the analysis as well as exploratory data analysis to identify trends in the output variables. A continuous variable prediction method is discussed in Chapter 5 and a classification method to predict the range of the customer minutes interrupted is discussed in Chapter 6. In Chapter 7, association rules to identify interesting trends in the types of equipment affected due to different categories of interruption types are derived. Finally, Chapter 8 concludes the thesis listing the limitations of the model and future scope of research.

Chapter 2

LITERATURE SURVEY

Every product or service that is offered to the end users these days is competitive mainly because of the increasingly efficient supply chain practices being followed. Lummus et.al [4] define a Supply Chain as “all the activities involved in delivering a product from raw material through to the customer including sourcing raw materials and parts, manufacturing and assembly, warehousing and inventory tracking, order entry and order management, distribution across all channels, delivery to the customer, and the information systems necessary to monitor all of these activities”. The authors then go on to add that supply chain management is the process that integrates and coordinates all the activities involved in the movement of raw material from suppliers through the distributor and retailer until the final delivery of the product to the customer. The following Figure 2.1 illustrates the product supply chain depicting both the flow of the product and information through the integrated supply chain.

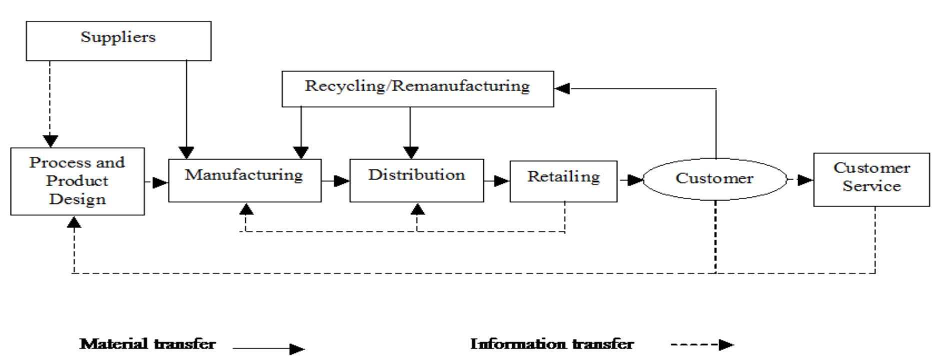


Figure 2.1 Product Supply Chain

2.1 Service System

While it is easier to visualize a supply chain network when the final end product given to the customer is a physical product, it is slightly complex when the customers are linked to the suppliers by the transfer of a service at every level of the supply chain. In a product supply chain the designed product is manufactured using raw material from suppliers, then passed on to the retailers and finally to the end customer through a distribution network. A service supply chain involves a customer-supplier service supply relationships. The end service provided to the customer is a utility service which makes the supply chain design slightly different. The figure 2.2 below shows the flow of the service through the supply chain.

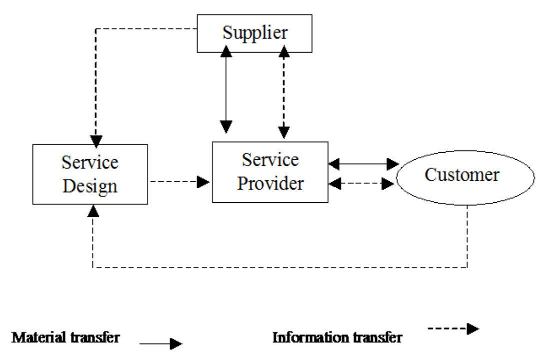


Figure 2.2 Service Supply Chain

2.2 Layout of Electric Supply Distribution network

The power generators own and operate electricity generating facilities or power plants and sell the power produced to the utility service providers. These service providers are the key players in the network as they are responsible for providing a reliable source of electric power while ensuring uninterrupted service at an affordable cost. The electric power supply industry is an integral part of the service system industry and also important for the economy as all businesses rely heavily on electric power to operate. The Energy Information Administration (EIA) predicts that there would be an increase of 29% in the electricity demand in the U.S between 2012 and 2040 [42].

The utility service provider is directly involved in the design of the service and is answerable directly to the customer. Let us consider for illustrative purposes an electric power utility service supply chain network. The products and services provided in such a network are limited to the electric power supplied and transmission services provided. The various decision makers in this system such as the power generators, the power suppliers, the transmitters and the customers operate in a decentralized system. A depiction of the supply chain network for electric power is given in Figure 2.3 [5] [6].

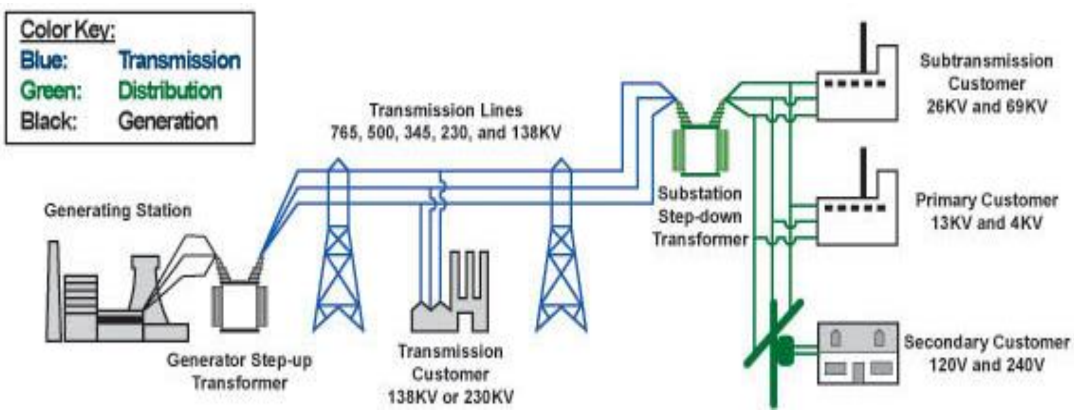


Figure 2.3 Basic Electrical Distribution System

Source: U.S.-Canada Power System Outage Task Force, *Final Report on the August **14, 2003,** Blackout in the United States and Canada: Causes and Recommendations,* April 2004, **p. 5,**

Understanding the outline of the electric power supply system is crucial for identifying the critical points in the service supply network.

- **Power generation -** The electric power generating station could be any power plant such as gas, oil, nuclear or thermal that converts these fuel sources into electricity. The power generated is then stepped up – increase the voltage before being supplied to the transmission lines. The voltage of the power generated is stepped up to as high as 500,000 Volts and passed through transmission lines to the distribution lines.

- **Transmission -** The Transmission stage of the network involves transmitting electricity over long distances at very high voltages since energy transfer efficiency is high at very high voltages. Large players in the transmission market operate thousands of miles of transmission lines across the country.

- **Distribution -** The electricity transmitted via the transmission lines is then stepped down using a step down transformer. The distribution line network which are made up of the feeders and laterals deliver electricity local to commercial, industrial and residential customers. These feeder and laterals handle current with a voltage range of 13kV – 23kV while the service lines handle a voltage range of 120 V – 480 V. The distribution costs for industrial and commercial users are less due to the high voltage power being supplied.

The service utility in a particular geographical area handles the transmission and distribution of electricity to the end-users through a vertically integrated structure. 58% of the final price of electricity is due to power generation costs while transmission and distribution services account for 11% and 31% of the price of electricity provided to the end customers.

The utility sector has not kept pace with other sectors in applying supply chain management techniques but that is changing as leading companies now recognize the importance of efficient supply chains. Companies have shifted focus to include supply assurance and risk management and not merely look to reduce cost and increase efficiency which was the original focus [44]. This thesis looks to address these challenges to help utilities reduce risks and provide electric supply assurance.

2.3 Service System Reliability and Customer Satisfaction

The electric power system forms an essential component of the quality of life in today's world. The reliable operation of the electric power system impacts both daily life and the economy. Implementing the right standards is crucial to prevent major blackouts like the 2003 blackout in north-eastern North America which served as a catalyst for standards implementation, as it encouraged steps to be taken to provide reliable service. Another key aspect of the service utility is service quality which is affected by common sources of interference including switching operations, system faults, operation of nearby telecommunication equipment, and lightning strikes [7].

Malfunction of electric components leads to electric service interruptions and in order to identify the possible causes of failure and locate the failure points, an electric utility company must focus on the equipment after the substation or the step down transformer. The feeders branch off into overhead and underground laterals that are further linked to overhead and underground transformers. It is important to analyze why outages occur and which equipment malfunctions due to the above mentioned reasons. A sample section of the distribution line network is shown in figure 2.3.

Ongoing communication efforts and increased price satisfaction have improved overall customer satisfaction with residential electric utility companies according to a J.D. Power 2015 Electric Utility Residential Customer Satisfaction Study. Ensuring proactive communication during power outages [8] and the availability of a decision making method to predict impact of outages remains a challenge for utility companies. Moreover, the industry lags behind in satisfaction levels when compared with other service industries such as television and telecommunication.

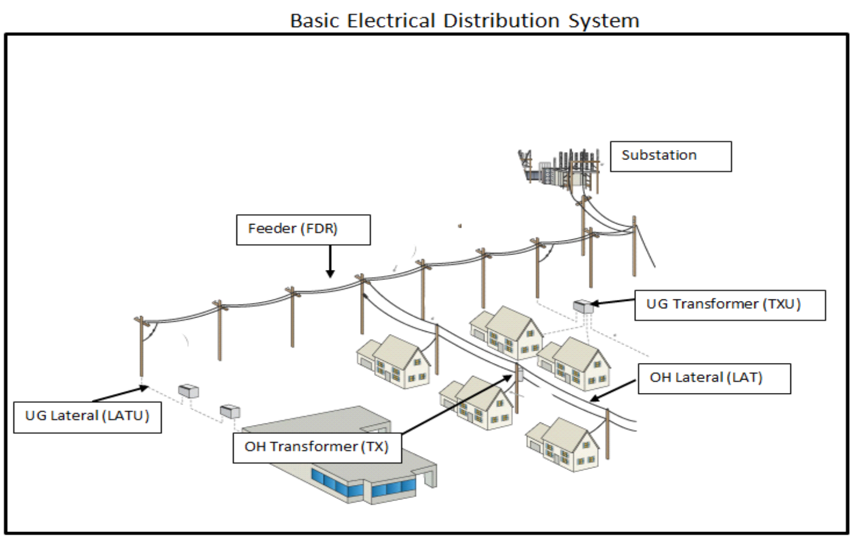


Figure 2.3: Distribution Line Network

In service reliability applications, the contingent valuation approach involves the use of customer surveys in which customers are asked to rate the value of an additional outage in terms of a reduction in prices. The main parameters or characteristics that define an outage in this approach are attributes such as frequency, season and time-of-day of occurrence and duration [9].

Kumar et al. analyze the importance of estimating the value of service reliability for electric utility customers by estimating the electricity outage cost for both residential and commercial customers through the contingent value approach [10]. Kenneth H. Tiedemann estimates residential costs of outages using the contingent valuation method. The outage costs have been explained as a function of outage length, time of day and season by taking into consideration three perspectives; cost/outage, cost/hour and cost per kWh lost [11]. Eto et al. describes three distinct end-use approaches for tracking the reliability needs of both industrial and residential customers and also provides a financial tool for addressing the reliability needs of the customer [12].

2.4 Impact of Power outages

There have been many studies that have looked into increasing reliability and improving infrastructure in the U.S. since the electric power blackout in northeastern U.S. and Canada in 2003. In 2012, massive damages were caused by Hurricane Sandy when it hit the east coast of the United States. The East coast power companies experienced widespread power cuts, resulting in revenue losses of more than $50 billion to the economy, decreased customer satisfaction and impacted the brand image. These events affected both the distribution infrastructure and the supply-chain operations of the companies [13].

The costs of outages can exist in various forms including lost output and wages, spoiled inventory, delayed production, inconvenience and damage to the electric grid. The most common cyber threats to the power system are bypassing controls, integrity violations and authorization violations, since many consumers have become so dependent on electronic systems, a cyber-attack that causes a momentary disruption of power could prove to be costly. For example a 20-min outage at an important manufacturing plant could cost as much as US$30 million [14]. The cost of an outage varies from customer to customer and is defined as a function of the end users' dependence on electricity; the duration/timing and nature of the outage; and finally the economic loss of the activity being disrupted. Costs incurred due to the outage or economic indicators are not the only criteria for measuring the importance of an outage. Sullivan et al. provide a statistical model to estimate the customer expectations for service reliability and power quality, and also estimate cost for customer interruptions for generation, transmission, and distribution outages for different time lengths under various circumstances [15].

The importance of electric outages and electric reliability to the nation needs to be understood and a study to estimate this has been done by assembling and standardizing the outage cost measurements of different utilities into a national database. Furthermore, trends in interruption costs with geographic location as well as differences in the cost by customer type have been studied [16]. Eto et al. developed a bottom-up approach for estimating the cost of power interruptions and power quality (power reliability) issues to US electricity consumers and estimated the base-case estimate of the annual cost for power interruptions to US electricity consumers as $79 billion based on this approach. The range of the costs could be as high as $135 billion or as low as $22 billion based on sensitivity assumptions. The study also concluded that the majority of outage costs are borne by the commercial and industrial sectors, not the residential sector. Surprisingly, a major factor that drives the costs is the frequency of the outages and not the duration of an outage which basically implies momentary interruptions have a larger impact compared to sustained interruptions. This bottom up approach has been later used to assess the benefits of improving reliability of the grid [17].

The Electric Power Research Institute initially estimated the national cost of power interruptions and reliability events (Power quality events) at $26 billion per year which was the first ever attempt at estimating the cost of power outages. This estimate was later revised it to $119 billion per year in subsequent studies after including commercial and residential users as well as more prolonged outages [18]. LaCommare and Eto specify that customers typically report the highest costs from outages during summer weekday afternoons and hence the time of the day must be taken into serious consideration during other reliability studies. They also develop a comprehensive framework for estimating the power interruption costs by taking into consideration the number of customers by region by type, the cost of a reliability event and the frequency and type of these events and finally specify a rough estimate for the cost as $80 billion a year [19]. Reichl et al. tried to estimate the costs of power outages to Austrian households, firms and the public offices and found that the average value of load lost due to the power interruptions of one hour on a weekday morning is 17.1 € per kWh of electricity not supplied. However the paper does not take into account the power outages that last greater than 48 hours [20]. Woo et al. estimate the average residential cost estimate for a 1-h outage as US$45 (HK$350) in Hong Kong [21]. With a shift to a digital society, business activities and industrial sectors have become increasingly sensitive to interruptions in the power supply. The Primen Report estimates outage costs as a function of outage duration and found that longer outages quite naturally create greater costs for businesses, but the relationship between outage length and cost is far from linear. The report specifies the average cost of a one-second outage across sectors surveyed as $1,477, the average cost of a three-minute outage as $2,107 and that of a 1 hour outage as $7,795 which shows that linear models will not be accurate in the prediction of outage costs. The report also states that the Digital Economy sector has a lower impact in terms of costs as compared to the Fabrication and Essential Services (F&SE) and the Continuous process manufacturing sectors [18]

2.5 Causes of Disruption in Electric Power supply

Table 1. Large Blackouts in the United States

Statistics for Outage Cause Categories

	% of events	Mean size in MW	Mean size in customers
Earthquake	0.8	1,408	375,900
Tornado	2.8	367	115,439
Hurricane/Tropical Storm	4.2	1,309	782,695
Ice Storm	5	1,152	343,448
Lightning	11.3	270	70,944
Wind/Rain	14.8	793	185,199
Other cold weather	5.5	542	150,255
Fire	5.2	431	111,244
Intentional attack	1.6	340	24,572
Supply shortage	5.3	341	138,957
Other external cause	4.8	710	246,071
Equipment Failure	29.7	379	57,140
Operator Error	10.1	489	105,322
Voltage reduction	7.7	153	212,900
Volunteer reduction	5.9	190	134,543

Source: *Trends in the History of Large Blackouts in the United States*, http://www.uvm.edu/~phines/publications/2008/Hines_2008_blackouts.pdf.

Notes: Totals are greater than 100% because some events fall into multiple initiating-event categories.

Figure 2.4: List of Large blackouts in the U.S.

2.5.1 Weather-Related Causes of Power Outages

The storm or other natural weather related events can cause prolonged outages that could last a few hours to a few days depending on the type and severity of the event or occurrence.

- Lightning - Lightning strikes can affect the electrical equipment directly or strike trees, which may fall onto power lines and cause outages.
- Ice - Ice storms that create a build-up of ice on power lines and trees can cause tree limbs and entire trees to fall onto power lines, causing an outage.
- Wind, Tornados and Hurricanes - High winds may cause trees, to come in contact with power lines and break the power by engaging the circuit breakers or other protective equipment. Wind also blows tree limbs or entire trees onto the power lines, breaking the lines and poles.
- Rain and Flooding - Heavy rains and Floods can cause damage to both above-ground and underground electrical equipment and these equipment may need to be shut down to prevent damage thereby causing outage.

2.5.2 Other Causes of Power Outages

Outages are also caused due to Equipment Failure such as transformer failure, broken insulators, bad underground cable, etc. Regular maintenance helps prevent equipment failure, but failure still causes a small number of outages.

Additional causes of outages can be listed as man-made outages that such as vehicle and construction accidents with power poles and power lines, maintenance from utilities, and the occasional human error.

Dust can also cause failure of electrical equipment through short circuits especially in areas that are prone to dust storms and sand storms. Preventing unplanned power failures is important by ensuring the equipment is sealed and less vulnerable to failure.

Planned outages such as crew maintenance scheduling as well as unplanned crew request and customer calls lead to power outages that can be seldom controlled or avoided.

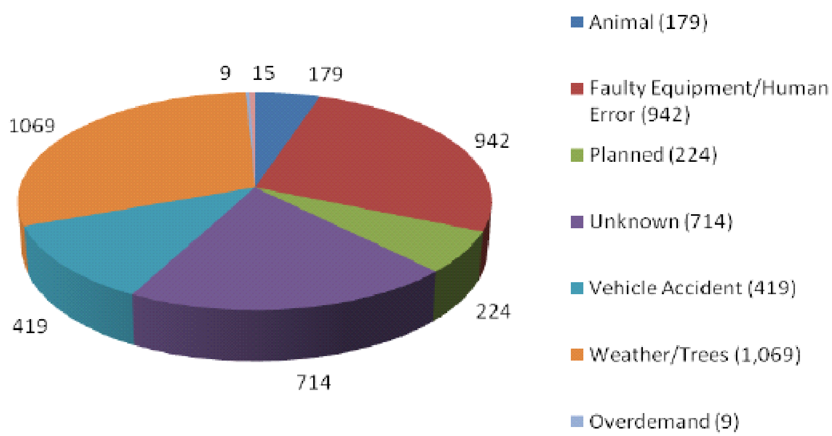


Figure 2.5: Power outages by Interruption Cause [24]

2.6 Impact of Weather related causes

High Winds can be devastating when combined with rain and seasonal storms. These winds can cause heavy damage to electric utility systems network and have been found to be the main reason for service interruptions and outages. The primary effect of this is the falling of trees on distribution lines and poles. The failure of these transmission lines and local distribution lines leads to prolonged outages and maximum customer minutes interrupted. The number of customers affected and the mean time to repair these is high since they carry bulk power over long distances.

The weather related outages are associated with heavy economic loss and this in turn can have real effects, as power outages can impact businesses (through lost orders and damage to perishable goods and inventories), and manufacturers (mainly through downtime and lost production, or equipment damage). In this world of communication both the buyers and the customers will face heavy losses, both monetarily as well as in need.

Some of the solutions suggested to reduce impacts from outages caused due to weather related events include better scheduling of tree trimming activities to keep rights-of-way clear, converting most transmission and distribution lines underground and improving the utility maintenance practices to prevent any errors during repair and improve reliability [23].

The 2013 White house report discusses the economic benefits of increasing the electric grid resilience to weather outages. The report estimates the average annual cost of power outages caused by severe weather between 2003 and 2012 inflation adjusted as $18 billion to $33 billion. The report also says that the annual costs fluctuate significantly and are greatest in the years of major storms such as Hurricane Ike in 2008, with cost estimates ranging from $40 billion to $75 billion, and Superstorm Sandy in 2012, with cost estimates ranging from $27 billion to $52 billion [2].

A U.S. Department of Energy reports suggests that between 2003 and 2012, roughly 679 power outages, each affecting at least 50,000 customers, occurred due to weather events in the U.S. Climate change is causing an increase in the number of outages caused by severe weather as the frequency and intensity of hurricanes, blizzards, floods and other extreme weather events have been increasing at an alarming rate. The United States suffered eleven billion-dollar weather disasters in the year 2012 which was almost equal to damages in any of the recent years.

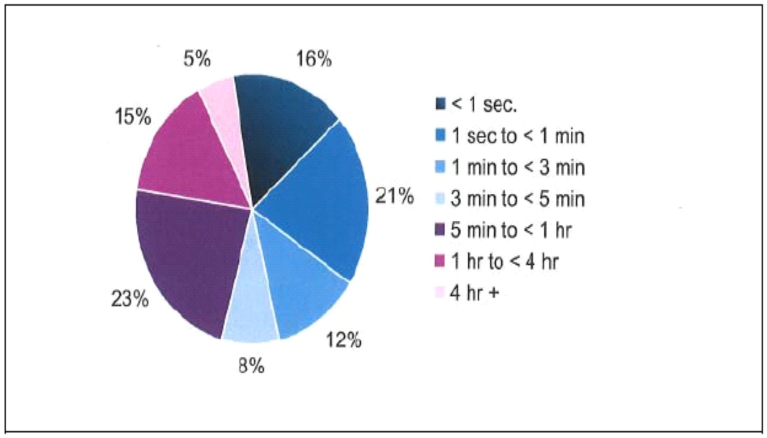
The Eaton Report quotes a Price Waterhouse study on how weather related power outage disturbances disrupt IT systems and states that more than a third of companies take longer than a day to recover from a prolonged outage and 10 percent of companies take more than a week. It also quotes a US Department of Energy study that suggests 33 percent of companies face economic losses of up to a half million dollars per power failure with IT system disruptions [24].

2.7 Research to predict weather related outages

Looking at the heavy costs incurred by users as well as utilities due to the weather related electric power interruptions several attempts have been made to predict the outages caused by weather related instances. Several characteristics of a power outage can be predicted such as i) The number of customers affected in a particular geographic location ii) The Customer Minutes interrupted due to an outage iii) The duration of a particular outage and finally iv) the number of interruptions in a particular location during a period of time. Accurate predictions of the above four parameters would go a long way in assisting the prediction of outage related costs, the planning and distribution of repair/reconstruction works and also man-power planning for an electric utility company. Several attempts have been made in the past to predict the impact of a weather related occurrence or disaster.

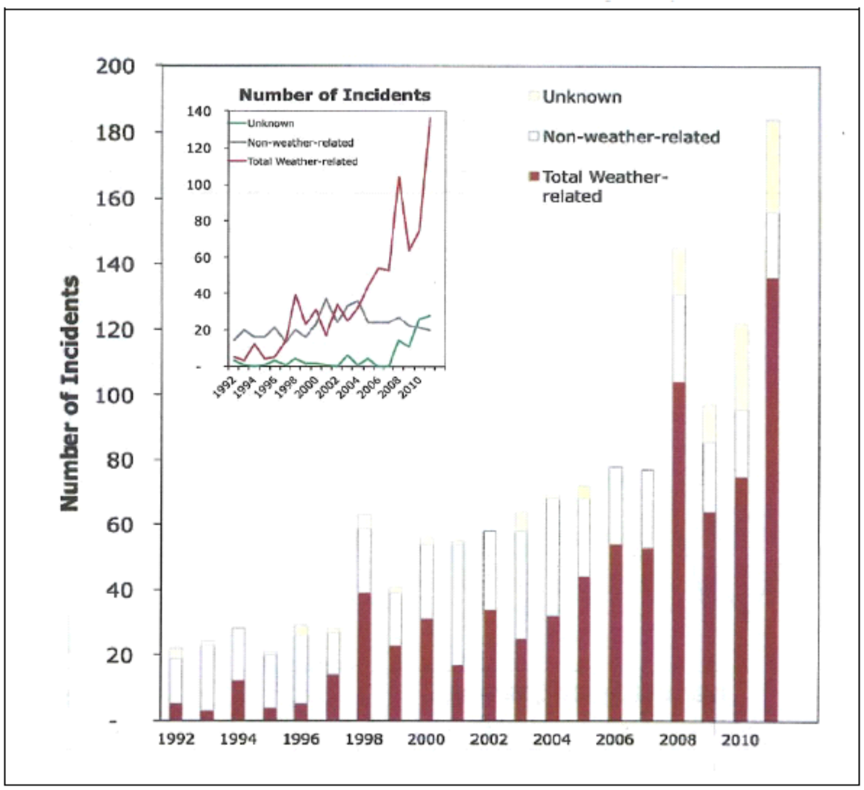
A study by the Lawrence Berkeley National Laboratory and Stanford University states that factoring data related to lightning strikes, precipitation, wind speed and temperatures leads to a better estimation of the impact of an event. The report finds that a mere 5 percent increase in annual average wind speeds produces a 56 percent increase in the Customer Minutes Interrupted for that year and a 10 percent increase in annual precipitation leads to a 10 percent increase in the CMI. Lightning is one of the most significant cause of faults and outages of many electric power systems and is responsible for poor electric system reliability [24]. Detailed modelling of storm characteristics and the response during lighting to assess the system reliability was carried out by Balijepalli et al. using a Monte Carlo simulation to calculate the fault rate of the power system due to lightning [25]. Existing methods exist that predict the interruption frequency and customer minutes affected during an outage, but the methods ignore two important contributing attributes: momentary interruptions and storms. Brown et al. in his paper suggests methods to determine the impact of each of the above discussed phenomena to assess the reliability of a power system [26]. Power outages caused by storm-related or other weather related events can vary in duration but are usually the more prolonged or sustained disruptions. Previous studies note that weather-related events are not always captured in power outage data and that widespread power losses resulting from major weather related events like hurricanes storms and earthquakes are sometimes not considered or recorded for cost estimation as they are not routine losses [27].

In order to minimize losses due to a weather related power outage, utility companies are expected to be aware of an impending storm or weather-related event which may cause outages, and make preparations for restoration of services as efficiently as possible. According to the CRS report for the Congress, preparation and recovery are the two most important aspects of a weather related power outage. Recovery of the system from any such event depends on the intensity of the weather related event and the damage it causes. However, good maintenance, restoration, organization, and communications strategies could hasten recovery and reduce the damage caused. Utilities must be aware of local weather-related events in the past and make use of weather related data to better forecast the impact of a weather related event [23].



Source: The Cost of Power Disturbances to Industrial & Digital Economy Companies. See http://www.onpower.com/pdf/EPRICostOfPowerProblems.pdf.

Figure 2.8: Distribution of Duration of Power Outages



Source: *Electric Grid Disruptions and Extreme Weather*. See http://evanmills.lbl.gov/presentations/Mills-Grid-Disruptions-NCDC-3May2012.pdf.

Figure 2.6: Weather vs Non-Weather related outage incidents

Kristina H. LaCommare et al. in their report to the Department of Energy state that there are statistically significant correlations between the average number of annual power interruptions experienced by a customer and predictor variables such as including wind speed, precipitation, lightning strikes, and the number of customers per line mile, cooling degree-days, the percentage share of underground transmission and distribution lines. The report suggests that an increase in the severity of major events over time is the main reason for the occurrence of a trend in the duration of power interruptions over time [28]. Zhou et al. [16] on the other hand used a Poisson regression model and a Bayesian network model to approximate weather related fault rates. Their results however showed that the largest effect on the reliability of power systems was due to wind, ice, and lightning events. But all these studies are focused on determining how weather related events affect reliability of power system on annual basis. They do not forecast expected outages for specific weather related incidents based on hourly forecasts.

Reed [29] suggests that the gamma distribution gives a good indication of the outage duration and the gust speed squared is found to be the best predictor of the outage durations for an electric power system. Wind-related outage durations during storms and hurricanes show consistent trends in the outage phenomena even though they may exist in different weather conditions. However, the paper further adds that wind speeds cannot be the only parameter that should be used to predict power outages. He further adds that high winds when coupled with rainfall and other hostile weather conditions make the prediction of outage duration more complex. This thesis looks to address this issue.

There have been many widely published papers that discuss power outages due to hurricanes and other large events such as storms. Han et al. used hurricane characteristic variables for estimating the spatial distribution of power outages during hurricanes in the Gulf coast region [30]. Nateghi et al. show that the customer minutes interrupted due to hurricanes can be accurately predicted prior to landfall based several parameters that can be estimated before landfall. They list the most critical parameters as the wind characteristics of the storms and the climatic and geographic characteristics of the service area such as the mean annual precipitation, soil moisture levels, and Standardized Precipitation Index, a measure of local deviations from long-term precipitation. Davidson et al. analyzed power outages and found a statistically significant relationship between the maximum wind gust and the number of outages and furthermore concluded that the wind speed is not sufficient for explaining the pattern of outages and that other explanatory variables are necessary [32].

Liu et al. developed Poisson and Negative Binomial Generalized Linear Models (GLMs) to explain the outage patterns and found that the most important variables for explaining variations in power outages were maximum wind gust, the number of transformers, the power company affected, and the hurricane indicator [33]. Liu et al. found that potentially important tree-related explanatory variables (e.g., number, type, age of trees, and tree trimming frequency) and infrastructure variables (e.g., age and condition of the poles) could lead to better models to predict weather related power outage durations [34].

Guikema et al. used both regression based models and data mining approaches [e.g., classification and regression trees (CART) and Bayesian regression trees (BART)] for predicting physical damage to the power system and outage durations [35]. The data mining approaches were found to outperform the regression-based approaches. Guikema and Ouiring (2012) developed a two-stage model that uses CART and a Poisson regression model to predict the number of power outages. The CART algorithm was used to model if a location will face a power outage and the Poisson Regression Model was used to estimate the number of power outages in a particular location. There has been no work done to estimate the customer minutes interrupted due to an outage, the number of customers affected due to an outage and also the duration of a particular outage in a location [36].

Work done by Zhu et al. [6] who developed outage forecasting models for storms, but focused on short lead-time (one hour and five hour ahead) forecasts is one of the earliest works in predicting modelling of storm related outages. They developed a model to predict outages for approaching storms by taking into consideration variables such as temperature, wind speed, and lightening [37]. Li et al. used a Poisson regression model to predict power outages in advance of 3 days to using forecast of severe weather events such as hurricanes, tornados, and thunderstorms. The predictor variables that were used in the model are the forecasts of wind gusts, gust frequency, rainfall, and temperature, the amount of rainfall in the two weeks preceding the storm event and also considered uncertainties in weather forecast to estimate the damage of a power outage [38].

Finally DeGaetano et al. [39] tried the tackle the issue of ice accretion on power distribution lines by building a forecast model using weather research. The model provided estimates of the outage damage hours before the actual accretion so that preventive measures could be taken by the electric utility to protect the equipment susceptible to damage.

Animal related causes

Electrical Equipment, especially over-ground equipment draw a lot of animals and the Eaton report tracked 179 outages that were animal related in 2015, 150 in 2014, and slightly greater than 200 in both 2012 and 2013. Over-ground equipment comes in contact with the body of animals like squirrels, snakes, birds, raccoons, beavers and even caterpillars creating a short circuit and diverting the path of electricity. This causes a power outage that may be a momentary outage or a slightly prolonged one, but has major economic impacts. The Eaton report also quotes a TE Connectivity study that notes animal related causes cost utilities between $15 and $18 million annually. The Edison Electric Institute study indicates that animals coming into contact with power lines, such as large birds, accounted for 11% of outages in the United States, which is a significant number that needs to be addressed.

To sum, most prior works have looked at predicting outages by looking at very large events (weather related or non-related), though these events constitute for less than 10% of all outages. In the current work, we look at predicting weather related outages (small and large). This thesis has two main objectives. First, it presents a method for predicting the "Outage Minutes" using hourly weather forecast and analyses the predictive capability of regular weather forecasts to estimate the duration of an outage at the hourly level during a weather related event. Furthermore it looks to classify the Customer Minutes Interrupted due to an outage by using a decision making classification method based on weather related variables that would be useful for utility companies. Second, we identify the equipment that is most commonly affected due to each category of Interruption Cause, that would give more insight into the type of maintenance and work force scheduling that needs to be done in order to ensure prevention or faster recovery from an outage event.

Chapter 3

PROBLEM METHODOLOGY

The objective of this thesis as mentioned in the previous chapter is to predict outage characteristics such as Customer Minutes Interrupted and Outage Minutes with the help of an hourly weather forecast and analyze its ability to do so. The model discussed is applied on a recorded outage dataset of a local Electric Utility company. A Cause – Effect Analysis is also done to identify the equipment affected due to a particular interruption cause and list the Interruption type for the same. The steps in the process are described in the Figure

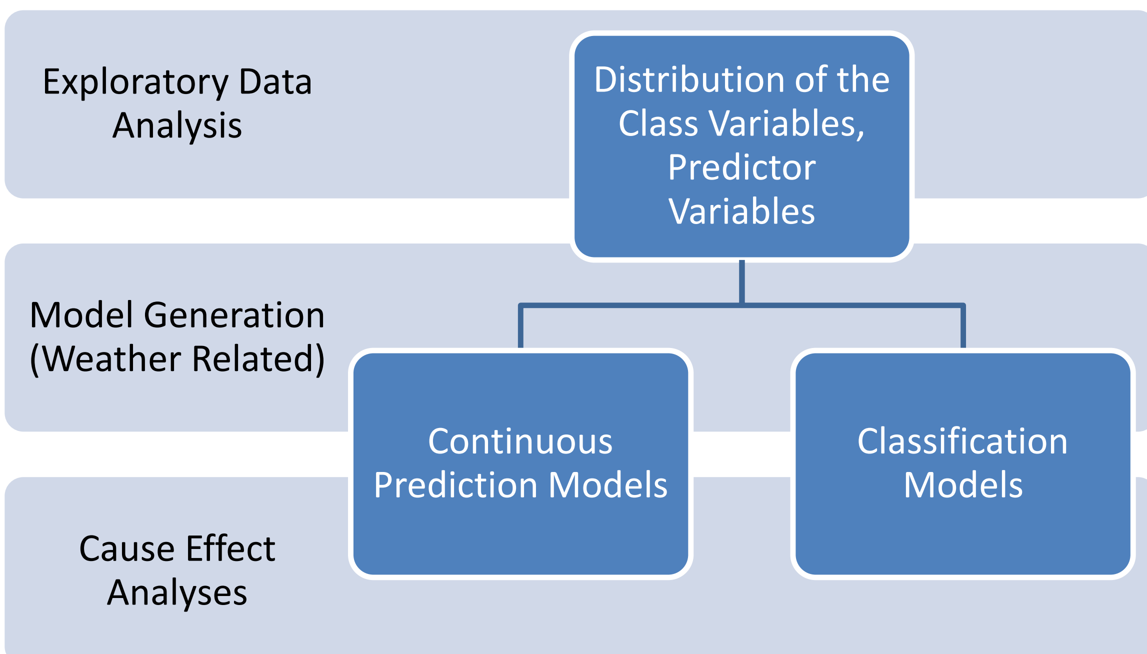


Figure 3.1: Flowchart of the Methodology

STEP 1: EXPLORATORY DATA ANALYSIS

The Initial Exploratory Data Analysis is done on the dataset to identify the parameters or interruption types that could predict outage characteristics. The distributions of the class variables and predictor variables are analyzed to further define the factors for the classification problem. Cleaning of the data is done at this stage and prior knowledge about trends in the Outage characteristics are used to develop more accurate prediction algorithms.

STEP 2: MODEL GENERATION

In this stage the Continuous Prediction Models and Classification Models are defined and evaluated. Regression Analysis is used to predict the continuous form of the outage characteristics. The Decision Tree Methods are used to accurately classify the output parameters and help build a logical, managerial level decision analysis tool.

STEP 3: CAUSE – EFFECT ANALYSIS

The association rule mining technique is used to determine the Equipment affected due to a particular Interruption Cause and also list the Interruption Type specified for each of the causes.

Chapter 4

EXPLORATORY DATA ANALYSIS

4.1 Failure Timeline Data

In this section we determine the trends in the Electric Power Outage patterns faced by an electric company using outage data for a particular geographic location that would help factor the outage characteristics and provide accurate predictions of the Customer Minutes Interrupted thereby enabling better planning and resource allocation.

4.1.1 Data Description

The data set contains over 42,000 outages over the year 2013 with 19 attributes.

- **Town:** This designates the town that is served by ABC Electric
- **Work Centre:** These are the various service center codes within the two towns
- **Outage Ticket #:** These are the numbers assigned to each of the various trouble tickets created to report the outages
- **Ticket Create Time:** Ticket created date and time
- **Ticket Year:** The year that the ticket was initiated
- **Ticket Month:** The month that the ticket was initiated
- **Ticket Calendar Day:** The day that the ticket was initiated
- **Power Off Time:** Power interrupted date and time
- **Power Restore Time:** Power restored date and time

- **Ticket Type:** The ticket type (Types include FDR – feeder outage, LAT Lateral Outage, etc.). There is also a type of ticket for problems other than outages called NLS.
- **Interruption Type:** Interruption report type code
- **Equipment Code:** See Code Table Attachment
- **Interruption Cause Code:** See Code Table Attachment
- **Support Code:** See Code Table Attachment
- **Customers Interrupted:** Number of customers interrupted by the outage
- **Outage Minutes:** Number of minutes that the outage included
- **CMI:** Total customer minutes interrupted (**CMI** = **Customers Interrupted** x **Outage Minutes**)
- **Total Customers Called on the ticket:** The total numbers of customer calls to report the outage on each ticket

The Timeline data is grouped by the hour and rounded to the previous hour, i.e. a timestamp of 12:53 PM is considered as the previous hour 12:00PM.

We define the Customer Minutes Interrupted (CMI) as the Class Variable for the purpose of the exploratory analysis as this is the most important variable for the Utility companies. This is under the assumption that the distribution of the Outage Duration would be similar to that of the Customer Minutes Interrupted.

4.1.1.1 Distribution of CMI by Year/Month/Week and Day.

Figure 4.1 shows that the Customer Minutes Interrupted is found to be the highest for the months June through September (Summer Months) and rather low during the Winter and Spring Months (Oct-Apr) which could indicate that an increase in usage of electrical equipment during the hot summer afternoons leads to more power outages.

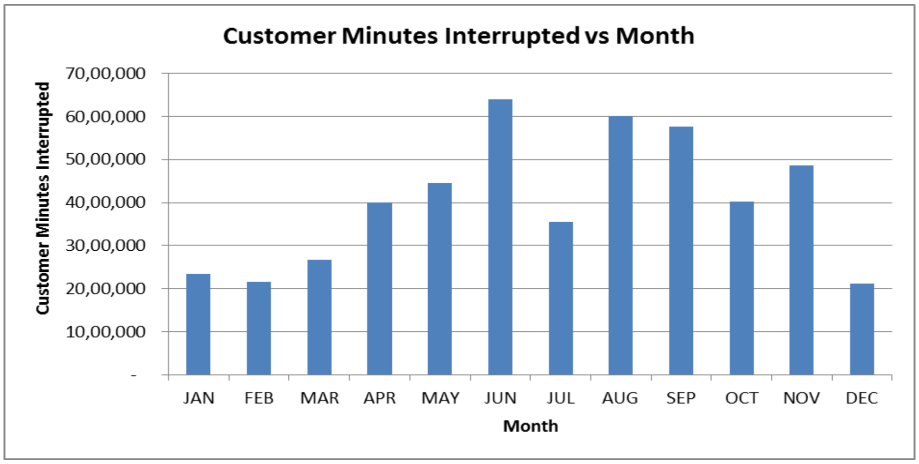


Figure 4.1: Monthly Distribution of CMI

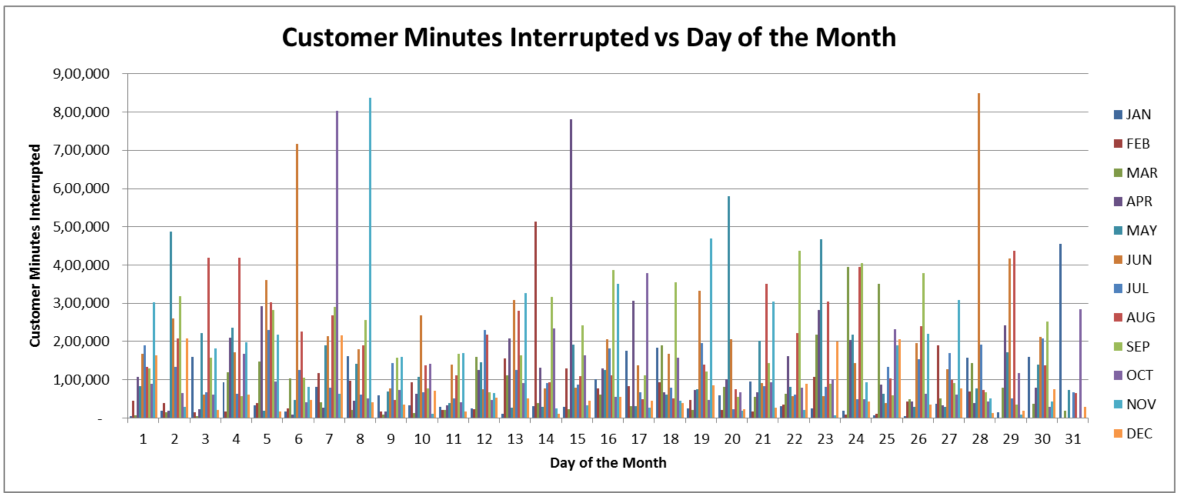


Figure 4.2: Daily Distribution of CMI

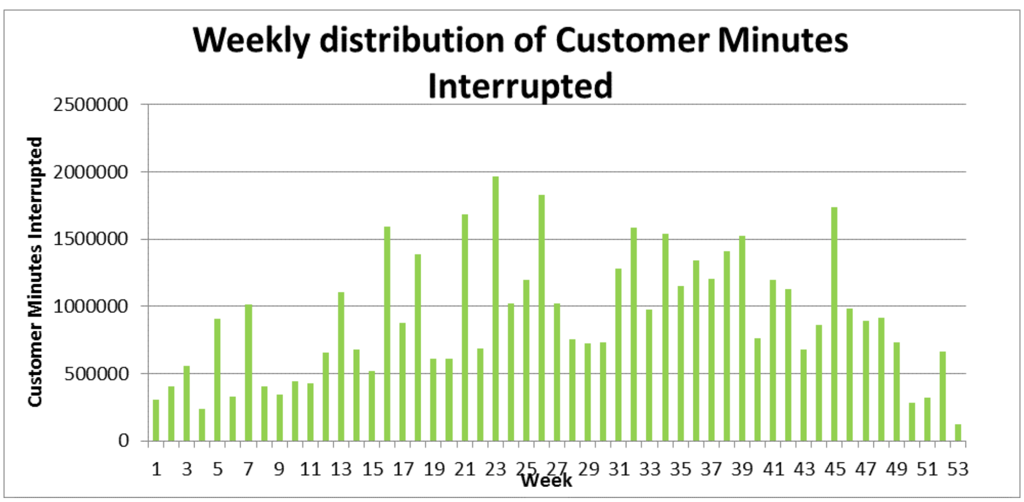


Figure 4.3: Weekly Distribution of CMI

The Figures 4.2 and 4.3 further support the statement to show that the summer months face the highest customer minutes interrupted, but there is no trend in the distribution of CMI across a particular month.

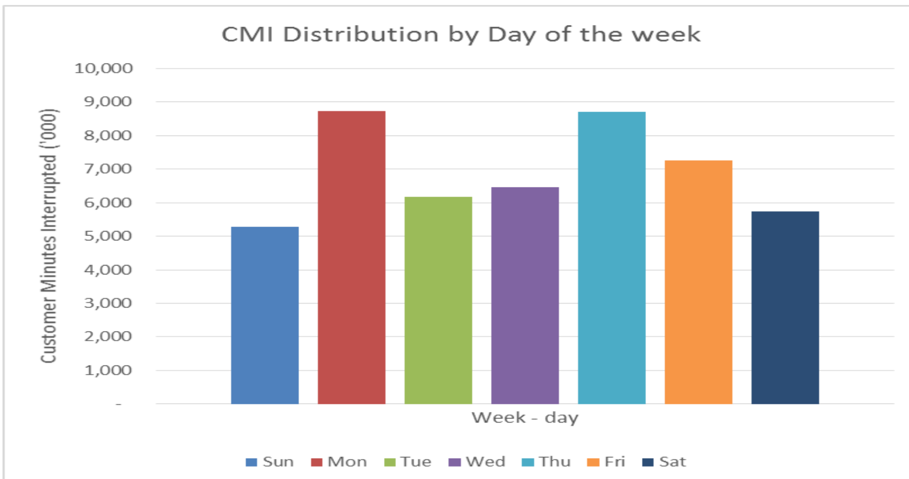


Figure 4.4: Daily Distribution of CMI

4.1.1.2 Distribution of Customer Minutes Interrupted

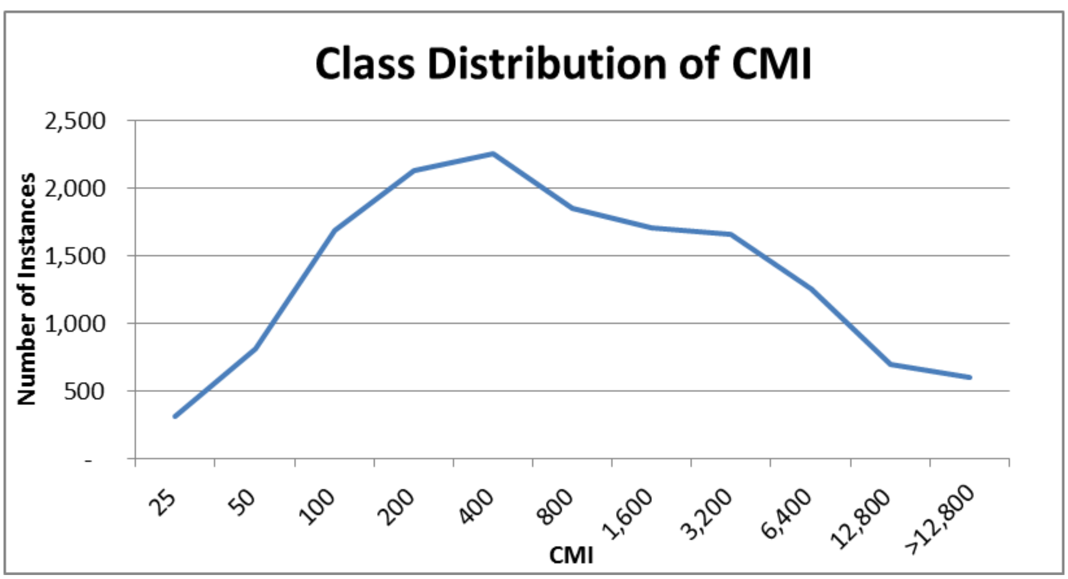


Figure 4.5: Distribution of CMI

4.1.1.3 Distribution of equipment codes/ interruption types

The distribution of the number of instances of each Interruption Cause leading to a power outage and the distribution of the Equipment being affected are shown in Appendix D. Approximate Distribution of Interruption Types in the dataset are:

- Man-Made errors (Negligence or repair crew errors)
  - 9800 - man made errors by the repair crew or service crew, 560 – Vine / Grass
- Failure of Equipment and Planned Equipment related causes
  - 6000+ - Equipment Failures, 39 – Overloading, 2210 - Planned crew request
- Weather and Animal related Causes
  - 1912 - Weather related causes, 1467 - Animal related causes
- Unavoidable Causes
  - 804 - Customer request, 194 – Unplanned Request, 76 – Fires, 312 – Vehicle, 242 – Vandalism
- Unknown Causes or undefined interruption types
  - 19846 - instances of unknown or undefined interruption types

The outages due to the unavoidable causes, man made errors and animal related causes cannot be predicted using deterministic methods and probabilistic models must be developed. Information on equipment failure log/ maintenance log and planned maintenance schedule can be used as variables for prediction of interruptions due to equipment failure and planned work. For the scope of this thesis we analyze weather related data to help predict outage duration during weather related events listed earlier. However, all the above Interruption Types are used in the cause-effect analysis of the outage.

4.2 Weather data

The data set contains over 8500 time-stamped data points of weather during the year 2013. The Weather data is grouped by the hour and rounded to the previous hour, i.e. a timestamp of 12:53 PM is considered as the previous12:00PM.

Data Description

- **Weather Time:** The hourly time stamp of the year 2013.
- **Temperature:** The Hourly Temperature in °F.
- **Heat Index:** The Hourly Heat Index in °F (0 in when there is no Heat Index)
- **Dew Point:** The Hourly Dew Point Temperature of the air in °F.
- **Pressure:** The Hourly Pressure in column inches.
- **Humidity:** The Humidity of the air listed hourly as a percentage.
- **Visibility:** The Air Visibility for the hour (The max visibility is 10 miles)
- **Wind Speed:** The wind Speed in mph (0 when the air is calm)
- **Gust Speed:** The gust Speed in mph (0 when the wind is not defined as gust)
- **Precipitation:** The precipitation in inches during the hour.
- **Customers Interrupted:** Number of customers interrupted by the outage
- **Outage Minutes:** Number of minutes that the outage included
- **CMI:** Customer minutes interrupted (Customers Interrupted x Outage Minutes)
- **Interruptions:** Number of Interruptions during the hour.

Chapter 5

CONTINUOUS PREDICTION MODEL

The first objective of this thesis is to develop a model to predict the Outage Duration. In this chapter we discuss a continuous prediction model that would help predict the hourly Outage Duration based on daily weather forecast attributes.

5.1 Multiple Linear Regression Analysis

A multiple linear regression model is developed to first identify the independent weather related variables that have an effect on the outage duration and then to determine how much these variables are able to answer the variation in the Outage Duration. A Multi-Linear Regression Model of the form is initially defined where the variables are defined as Y -> Dependent Variable (Outage Duration), X_1 -> Temperature, X_2 -> Heat Index, X_3 -> Dew Point, X_4 -> Pressure, X_5 -> Humidity, X_6 -> Visibility, X_7 -> Wind Speed, X_8 -> Gust Speed, X_9 -> Precipitation, X_{10} -> Categorical Weather Event.

The dataset used for the regression analysis consists of only the hourly timestamps during which there is a predicted weather related event.

Normalization of the Variables

The dataset is normalized with center scaling to ensure that the data is distributed between 0 and 1 and remove the variation due to a difference in units.

Transformation of the Variables

The scatterplot of the Dependent variables Vs the Independent variables does not reveal any interesting patterns which can be used as the basis to transform the variables in order to achieve a better regression fit. The only transformations done are:

- In an attempt to reduce the variation in the distribution of the Outage minutes, the square root of the Independent Variable is used for the analysis.

- Past literature as discussed earlier showed that the square of the Gust Speed was found to better predict the impact of an electric power outage. Hence, the square of the Gust Speed is used as a dependent variable in the analysis.

The Model analyzed: $Outage.Minutes\text{^}0.5 = \beta_0 + \beta_1 Temperature + \beta_2 Heat\ Index. + \beta_3 Dew\ Point + \beta_4 Humidity + \beta_5 Pressure + \beta_6 Visibility + \beta_7 Wind\ Speed + \beta_8 (Gust.Speed\text{^}2) + \beta_9 Precipitation + \beta_{10} Events$

Figure 5.1 shows that Precipitation is highly correlated to the Outage Minutes while Gust Speed is a significant variable too. Wind Speed is removed as a variable in the model as it is highly correlated to Gust Speed and hence this might have a negative influence on the prediction. Temperature, Heat Index, Humidity and Dew Point are removed from the model because of the low correlation to Outage Minutes. Moreover Dew point is highly negatively correlated to Pressure. Interestingly Visibility and Pressure have a negative correlation to Outage Minutes and both of these are used to further develop the model.

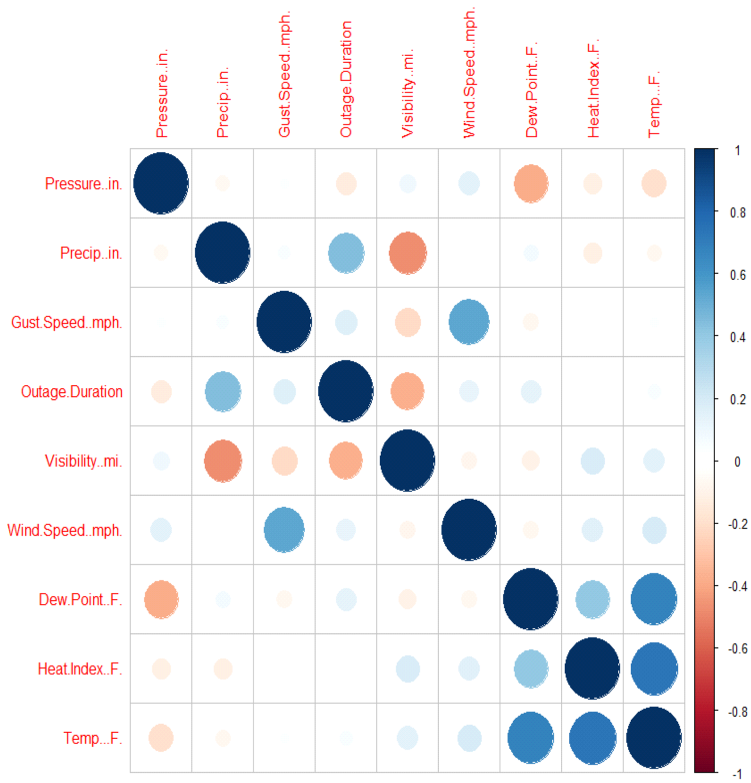


Figure 5.1: Pearson Correlation between attributes

5.2 Analysis of the Results

5.2.1 Initial Complete Model

From the results of the regression analysis, it is shown that there are certain variables that have no impact on the final dependent variable "Outage Duration". The Variables "Temperature", "Heat Index", "Dew Point", and "Humidity" do not have a significant effect in predicting the Outage Duration and can be ignored in the model. The R-Squared value is shown to be 0.3457 which indicates that the above selected variables explain 34.57% of the variation in the Outage Duration. The Diagnostic Plots and the Plot of the Fitted Value Vs the Actual Value is attached in the Appendix A.

5.2.2 Revised Model

The new model that is analyzed can be defined as: $Outage.Minutes^{\wedge}0.5 = \beta_0 + \beta_1 Pressure + \beta_2 Visibility + \beta_3(Gust.Speed^{\wedge}2) + \beta_4 Precipitation + \beta_5 Events$

The R-Squared value is shown to be 0.3318 which indicates that the above selected variables explain 33.18% of the variation in the Outage Duration. From the results of the analysis we can conclude that removing the insignificant independent variables listed above does not increase the error in estimating the Outage Duration significantly and also provides a less complex model. The Diagnostic Plots and the Plot of the Fitted Value Vs the Actual Value is attached in the Appendix A.

- The heavy variation in the outage duration makes it difficult to accurately predict the outage duration with given input weather related characteristics.
- The continuous nature of the independent variables increases the prediction error as the model has to account for inherent varying distribution of the independent variables.

We address both these issues in the next Chapter and try to develop a model that could better predict the Outage Characteristics.

Chapter 6

CLASSIFICATION MODEL

In this Chapter we try to propose a methodology to predict the Customer Minutes Interrupted due to a particular outage. The Customer Minutes Interrupted is an outage characteristic that has been shown to have heavy variation as it includes the variation of both the Customers Interrupted and the Outage Duration.

We divide the Customer Minutes Interrupted into different Classes based on significant time intervals that would be of importance to the electric utility companies. The Class 0 consists of those instances for which there was no recorded outage and the other classes consist of instances where the Customer Minutes Interrupted falls in the time intervals defined as (0-2 hours), (2-5 hours), (5-12 hours), (12-24 hours), (1-2 days), (2-5 days) and (>5 days).

Class Number	CMI (Minutes)	No. Of Instances	% of Distribution in data
Class 0	-	27,781	64.98%
Class 1	120	3,365	7.87%
Class 2	300	2,964	6.93%
Class 3	720	2,442	5.71%
Class 4	1,440	1,716	4.01%
Class 5	2,880	1,687	3.95%
Class 6	7,200	1,662	3.89%
Class 7	>7200	1,138	2.66%

Table 6.1: Distribution of CMI Classes (entire outage dataset)

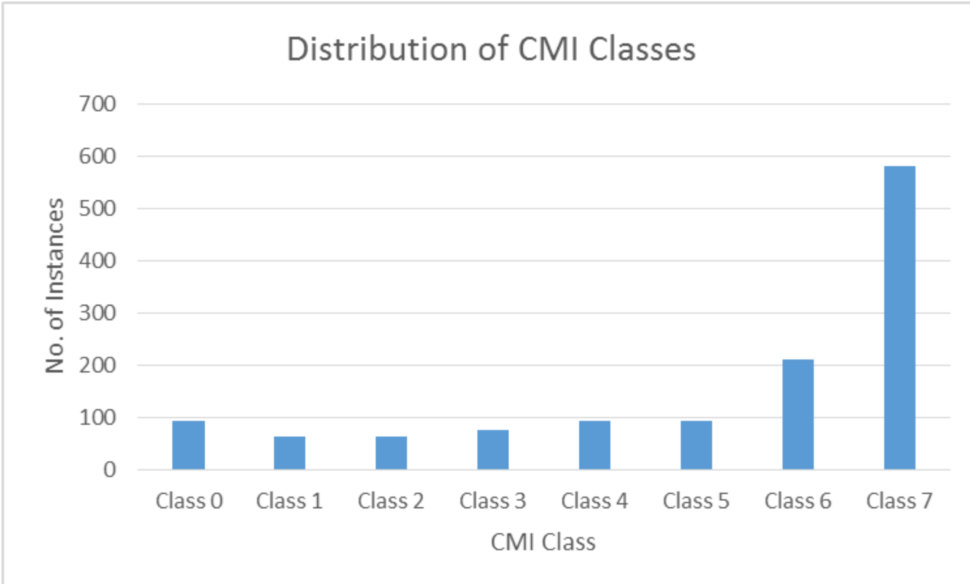


Figure 6.1: Distribution of the CMI Classes for weather events

The Input Variables for the classification approach are the same as that defined in the previous chapter. Only the hourly timestamps during which there is a predicted weather related event is used for the analysis. We use the Decision Tree Classification Method to accurately classify the Customer Minutes Interrupted into the predefined classes.

- ***Decision trees***

A decision tree is a decision support tool that uses a tree-like graph in order to model a decision and their possible consequences. Decision Tree Algorithms make no assumptions about the relation between the outcome and predictors. The algorithm works equally well with linear and non-linear relations. The major advantage of a decision tree is the high interpretability of the tree in terms of a set of rules that can be derived from the tree to be used in managerial decision making.

- ***Binning of Input Variables***

The input variables are binned according to the distribution of the weather attributes as shown in Appendix D with the intention of having an equal Bin size and approximately even distribution of Customer Minutes Interrupted across the bins. The weather variables are grouped into 6 bins for every attribute. Overall, there were 36 bins used in the analysis.

Attribute/Bin No.	1	2	3	4	5	6
Temperature (F)	(,67)	[67,72)	[72,77)	[77,82)	[82,87)	[87,)
Heat Index (F)	0	(0,82)	[82,87)	[87,92)	[92,97)	[97,)
Dew Point (F)	(,66)	[66,68)	[68,70)	[70,72)	[72,74)	[74,)
Humidity (%)	(,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1)
Pressure (in)	(,29.8)	[29.8,29.9)	[29.9,30)	[30,30.1)	[30.1,30.2)	[30.2,)
Visibility (mi)	[0,1)	[1,2)	[2,3)	[3,4)	[4,10)	10
Wind Speed (mph)	[0,5)	[5,10)	[10,15)	[15,20)	[20,25)	[25,)
Gust Speed (mph)	0	(0,21)	[21,26)	[26,31)	[31,36)	[36,)
Precipitation (in)	0	(0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1)	[1,)

Table 6.2: Binning of predictor variables

"Ctree algorithm (Conditional Inference Trees)"

The ctree algorithm recursively performs univariate splits of the dependent variable based on values on a set of covariates. It selects variables that have many possible splits or many missing values and uses a significance test procedure in order to select variables instead of selecting the variable that maximizes an information measure.

The "ctree" algorithms gives a decision tree as shown in figure 6.2. It displays the probability of occurrence of a particular class at the end of each tree node. In our model, we input the same predictor variables as in the revised regression model - Gust Speed (mph), Pressure (in), Precipitation (in), Visibility (mi). We implement the J48 algorithm in R using the caret package.

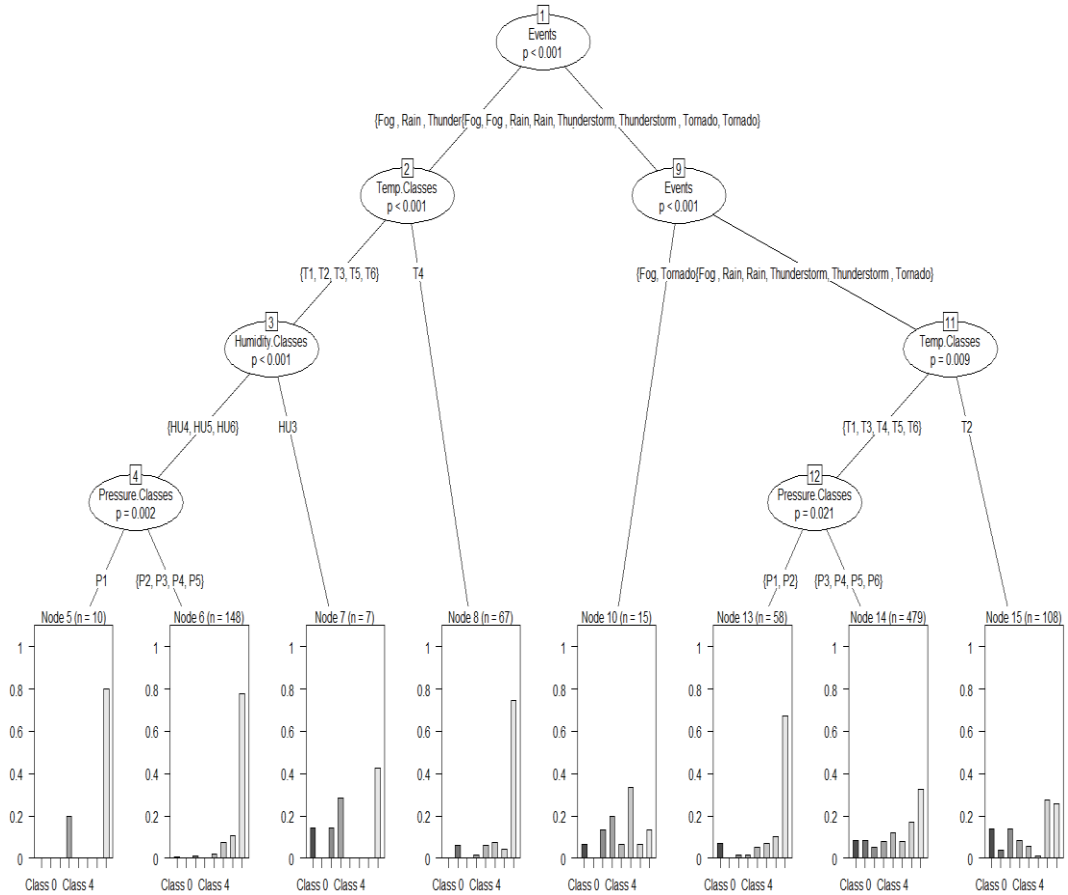


Figure 6.2: Sample section of a decision tree

"J48 Algorithm"

The J48 algorithm used in R implements the 4.5 algorithm for generating a pruned or unpruned C4.5 decision tree. The decision trees generated by J48 can be used for classification of the dependent class variable.

The algorithm for building decision trees using J48 is explained in [41]. For each attribute a, the normalized information gain ratio from splitting on "a" is found. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain. A decision node that splits on the variable with the highest information gain ratio is created. The algorithm then recurs on the smaller sub-lists. We implement the J48 algorithm in R using the caret package. In our model, we input the same predictor variables as in the revised regression model - Gust Speed (mph), Pressure (in), Precipitation (in), Visibility (mi). We implement the J48 algorithm in R using the caret package.

In this method the effect of the unbalanced training set is reduced by over sampling the tuples with lower class instances to match those with the higher class instances. The class instances of classes “Class0”, “Class1”, “Class2”, “Class3”, “Class4” and “Class5” are doubled while the instances of “Class6” and “Class7” are reduced to 100 each to give 846 final tuples.

6.1 Analysis of Results

Ctree algorithm

The overall error rate achieved on implementation of the Conditional Inference Tree is 53% which shows us that 47% of the classes in the tree are classified correctly. The result is satisfactory considering that the distribution of the dataset is highly unbalanced. Figure 6.3 shows a sample section of the decision tree which tells us that the probability of CMI falling a particular class is evenly distributed when the Precipitation is R1. Otherwise, the probability of CMI falling in Class 7 is significantly high with a higher probability if the Visibility is either V1 or V3.

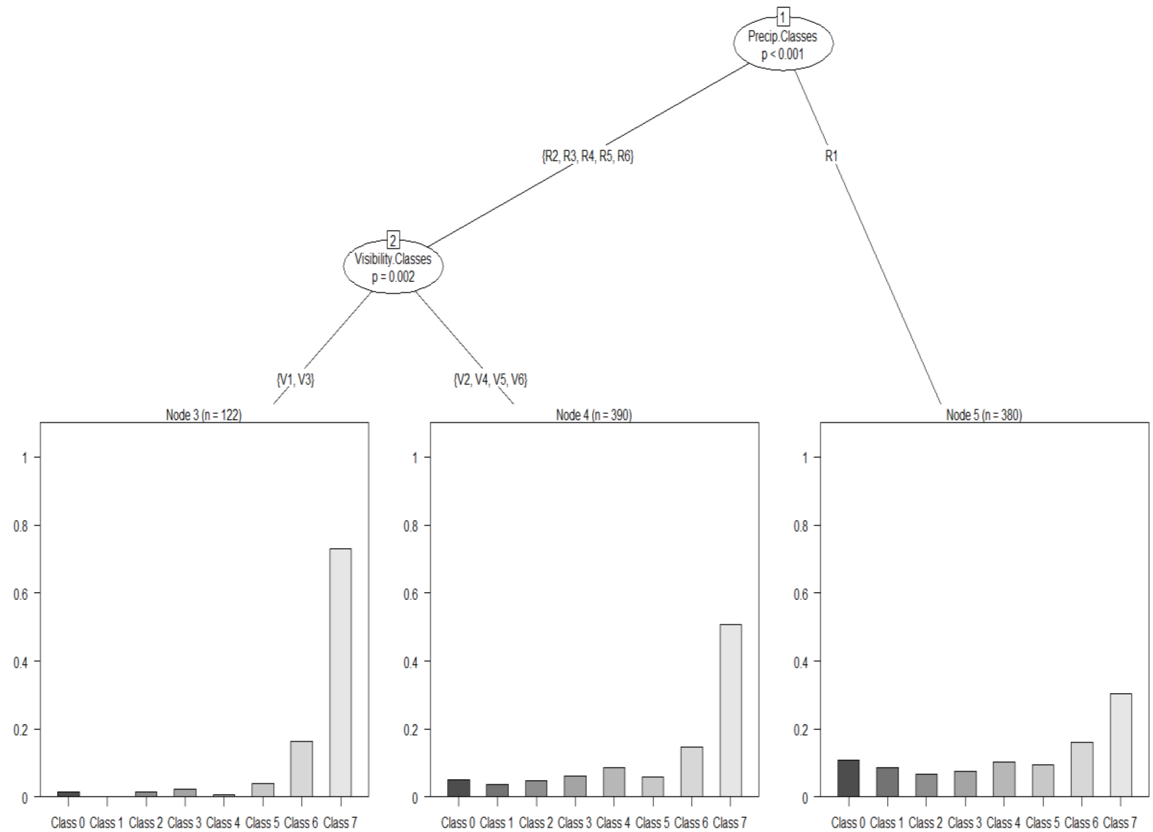


Figure 6.3: Section of the decision tree output

"J 48" – Decision Tree Algorithm

The J48 algorithm classifies the CMI classes with an accuracy of 40%. The results of the algorithm can be found in Appendix B. Table 6.3 shows the confusion matrix of the classification. A majority of the instances of class 7 are found to be classified into the class 7 due to the high imbalance of the test set. As can be seen from the results the balanced accuracies of the classifications of the individual classes are all above 50% which indicates that weather forecasts can be a sufficiently important parameter to predict the outage due to a weather related event. We look at the balanced accuracy to ensure that accuracy is not compromised by unbalanced data and to ensure bias due to an unbalanced dataset is compensated for.

Prediction/Reference	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 0	6	1	4	1	11	3	6	14
Class 1	1	4	0	2	2	5	5	12
Class 2	0	0	0	1	1	1	3	3
Class 3	4	0	5	3	3	4	3	8
Class 4	1	0	0	0	1	0	1	4
Class 5	0	0	1	3	4	4	4	4
Class 6	2	2	2	1	4	0	8	9
Class 7	13	11	10	13	12	12	39	144

Table 6.3: Confusion Matrix

Hourly weather forecast of events is shown to classify the Customer Minutes Interrupted into the defined ranges with an accuracy of 40-50% based on the decision tree methods. The decision tree methods due to their inherent ease in visually communicating a path of choices or set of choices, along with their associated uncertainties and outcomes can be used by decision makers in electric utility companies to estimate the category of customer minutes interrupted using hourly weather data to calculate the CMI due to weather related events. Estimating the intensity of the weather related outage in terms of customer minutes interrupted is merely not enough to warrant some actionable decisions. Decision makers also need to identify the equipment that is affected due to a particular interruption cause. Figure 6.4 shows that a majority of the interruptions are caused by a couple of weather related interruption types across classes. Furthermore, in figures 6.5 and 6.6 we can see that the weather related interruption types that account for 80% of the interruptions vary as we move from the shorter to the longer CMI Classes. Therefore identifying the types of equipment affected by a particular interruption type is essential for decision makers to optimize their man power allocation and ensure pre-event checks in addition to the estimation of the intensity of the forecasted event.

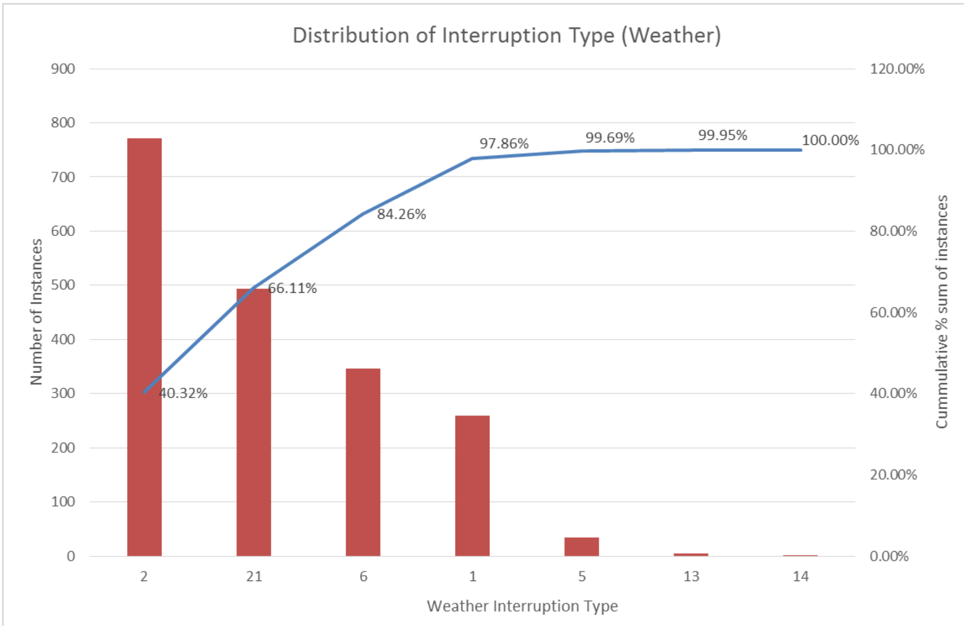


Figure 6.4: Distribution of Interruption Type (Weather)

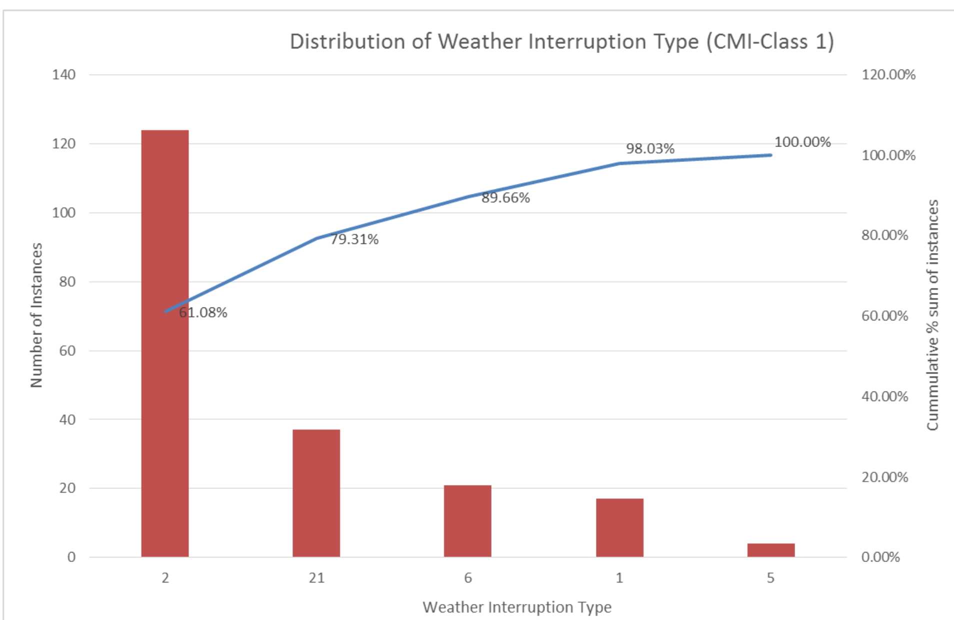


Figure 6.5: Distribution of weather interruption type (Class 1)

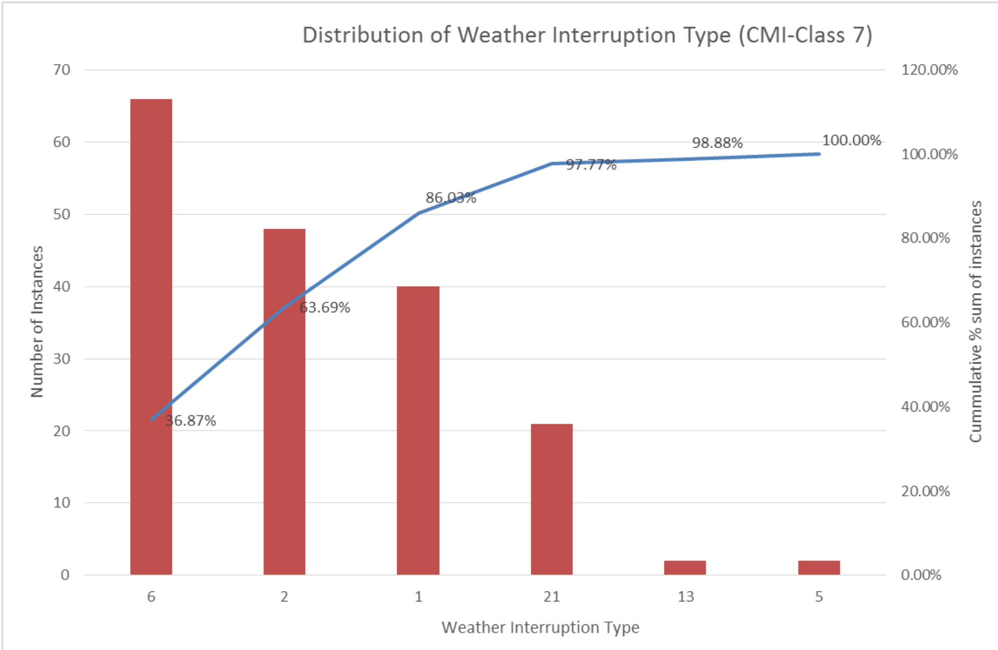


Figure 6.6: Distribution of weather interruption type (Class 7)

This can be found using association mining algorithms discussed in the next chapter. A combination of the impact of an outage in terms of CMI along with insights on the type of equipment that needs to be attended to provides a practical solution to reducing the outage duration due to an outage. Decision makers are left with a simple and logical procedure to help make decisions to take preventive steps and reduce outage minutes before an outage.

Chapter 7

CAUSE – EFFECT ANALYSIS

In this chapter the equipment that could get affected due to a weather related event is identified, which in turn would cause an outage event to occur. Different equipment may be affected due to different Interruption causes and it is imperative that we initially segregate the Interruption Causes into different categories in order to better derive a cause-effect relationship and generalize the result.

7.1 Categorization of Interruption Causes

The different Interruption Causes are divided into the following categories and the number of occurrences of each in the dataset along with the interruption codes used is given below. The Cause Codes used can be found in Appendix D.

Weather related events with the cause codes

001	Lightning with equip damage
002	Storm w/no equip damage
005	Wind (clear day and high winds)
006	Storm/Wind (with equip damage)
013	Tornado
014	Hurricane
015	Ice on Lines
021	Tree/Limb Unpreventable

Animal Related events with the cause codes

007	Squirrel
009	Other Bird
011	Other Animal
016	Monk Parrot Nest
017	Osprey Nest
018	Lizards

Equipment Failure events with the cause codes

172	Overloaded Normal Conditions
173	Overloaded Emergency Conditions
188	Equip Failed-OH
189	Equip Failed-UG
196	Slack Conductors
202	Loose Connection

Unavoidable events with cause codes

003	Accidental Fire (forest fires, etc.)
040	Vehicle
041	Accidental Contact
190	Unknown
191	Vandalism
193	Customer Request
194	Unplanned - Crew Request

Negligence/Avoidable events with cause codes

020	Tree/Limb Preventable
025	Vines/Grass
046	Switching Error
079	Dig-In
170	Wrong size fuse
178	Non-standard Construction
183	Improper Installation
195	Planned - Crew Request

We then used the Association Rule Mining technique to perform the cause-effect analysis.

7.2 Association Rule Mining

Association rule mining is a method for discovering relations between variables in large databases. The rules are identified by measuring certain relationship measures. The measure of the strength of a relationship that we would use for the purpose of this analysis are Support and Confidence.

- **Support**

The support value of X with respect to T is defined as the proportion of transactions in the database which contains the item-set X. Support = Supp(X)

- **Confidence**

The Confidence value of a rule X->Y, with respect to a set of transactions T, is the proportion of the transactions that contains X which also contains Y.

Confidence is defined as: Conf(X->Y) = Supp(XUY)/Supp(X)

To solve the association rule mining problem and identify the interesting rules from the transaction dataset, the Apriori Algorithm is implemented in R.

7.2.1 Apriori Algorithm

Apriori [40] is an algorithm for frequent item set mining and association rule learning for transactional databases. The algorithm first identifies the frequent individual items in the database and then extends the item set based on the interestingness measures satisfying the defined threshold values. The final frequent item sets determined by Apriori are used to determine association rules that highlight general trends in the transactional database.

In the current problem, we adapt the above algorithm to perform the cause effect analysis. The final association rules created would be defined in the form of a cause-effect relationship as the input transaction consists of an Interruption Cause and an Equipment Code. Figure 7.1 explains the Apriori Algorithm.

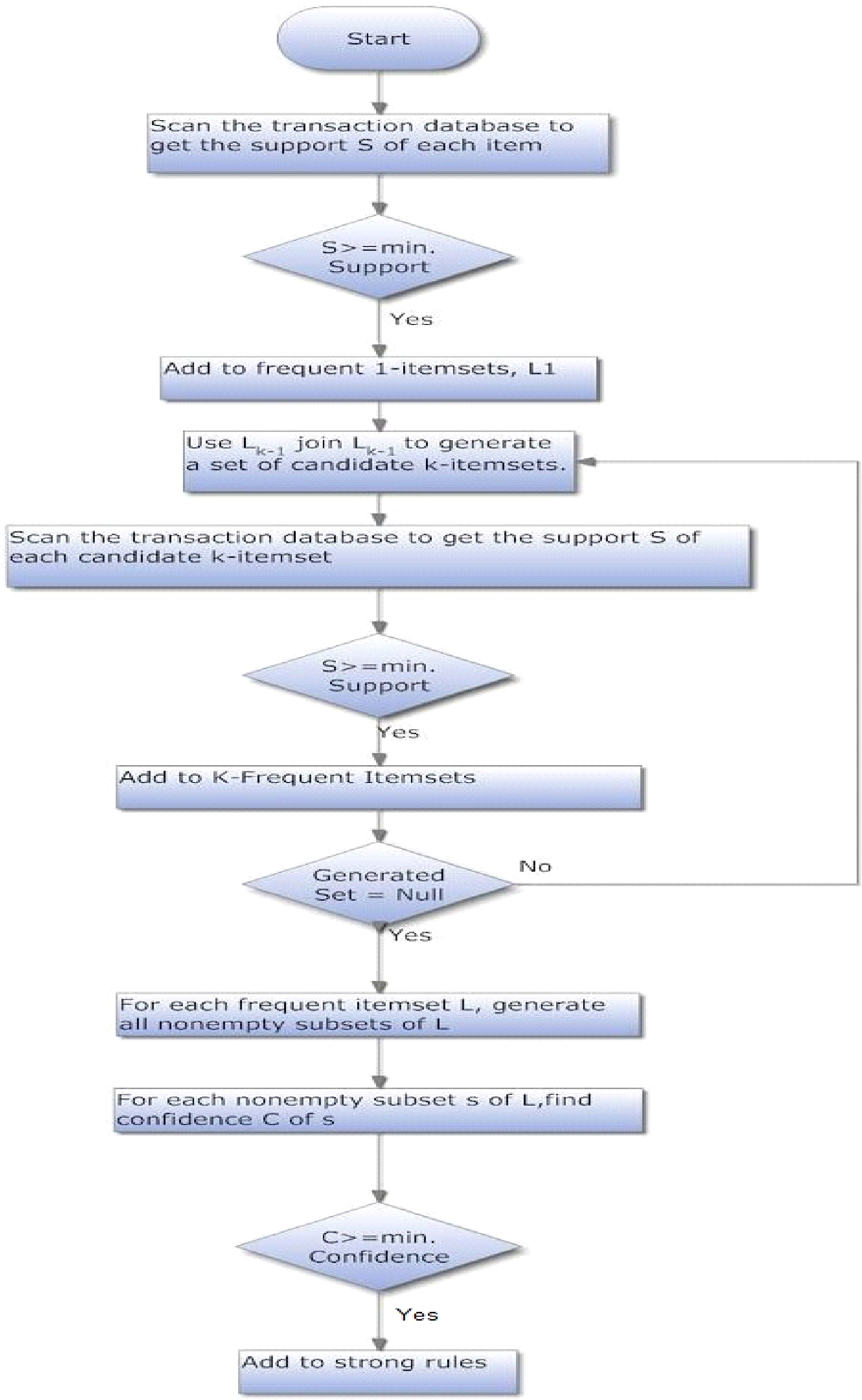


Figure 7.1: Apriori Algorithm

7.3 Analysis of Results

The Apriori Algorithm is applied to the entire dataset and the association rules derived are shown below. The Cause Effect Relationship is listed as Interruption Cause -> Effect. For the analysis of all interruption causes we use a range of support (0.01, 0.1 and 0.3) and confidence (0.5, 0.7 and 0.9).

The results are found to agree with the charts shown in Appendix D.

7.3.1 All Interruption Causes

- **Support – 0.3, Confidence – 0.9**

There are no significant rules with the above support and confidence values which shows that there is no association rule where an interruption cause leads to a failure of a particular equipment in 90% of the occurrences of the interruption cause.

- **Support – 0.3, Confidence – 0.7**

Interruption Cause 195 (Planned – Crew Request) -> *Interruption Type (TX)*

- **Support – 0.3, Confidence – 0.5**

Interruption Cause 195 (Planned – Crew Request) -> *Interruption Type (TX)*

A reduction in the confidence to 0.7 or 0.5 for the same support of 0.3 gives the above rule. There is however no Interruption Cause that causes an equipment failure in more than 30% of the data.

For the same range of confidence the data is then analyzed for a support of 0.1.

- **Support – 0.1, Confidence – 0.9**

Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 94 (Transformer), Interruption Type (TX)

- **Support – 0.1, Confidence – 0.7**

  Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 94 (Transformer), Interruption Type (TX)

  Interruption Cause 20 (Tree/Limb Preventable) -> Interruption Type (SV)

- **Support – 0.1, Confidence – 0.5**

  Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 94 (Transformer), Interruption Type (TX)

  Interruption Cause 20 (Tree/Limb Preventable) -> Interruption Type (SV)

Planned Crew requests affects the transformer in 10% of the instances. A reduction of confidence from 0.7 to 0.5 for a support of 0.1 shows an importance given to the Interruption cause 20 (Tree/Limb Preventable).

For the same range of confidence the data is then analyzed for a support of 0.1.

- **Support – 0.01, Confidence – 0.9**

  Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 94 (Transformer), Interruption Type (TX)

  Interruption Cause 195 (Planned – Crew Request) -> Interruption Type (TX), Equipment Code 81 (Pole)

  Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 111 (Cable), Interruption Type (TX)

  Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 102 (Other

Equipment), Interruption Type (TX)

Interruption Cause 20 (Tree/Limb Preventable) -> Equipment Type 105 (Conductor Damaged), Interruption Type (SV)

Interruption Cause 20 (Tree/Limb Preventable) -> Equipment Type 104 (Conductor Down), Interruption Type (SV)

- **Support – 0.01, Confidence – 0.7**

Interruption Cause 20 (Tree/Limb Preventable) -> Equipment Type 105 (Conductor Damaged), Interruption Type (SV)

Interruption Cause 20 (Tree/Limb Preventable) -> Equipment Type 104 (Conductor Down), Interruption Type (SEC)

Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 94 (Transformer), Interruption Type (TX)

Interruption Cause 195 (Planned – Crew Request) -> Interruption Type (TX), Equipment Code 81 (Pole)

Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 102 (Other Equipment)

- **Support – 0.01, Confidence – 0.5**

Interruption Cause 195 (Planned – Crew Request) -> Interruption Type (TX)

Interruption Cause 189 (Equip Failed -UG) -> Equipment Code 111 (Cable)

Interruption Cause 193 (Customer Request) -> Interruption Type (NLS)

Interruption Cause 189 (Equip Failed -UG) -> Interruption Type (SEC)

Interruption Cause 188 (Equip Failed -OH) -> Equipment Code 93 (Fuse Switch)

Interruption Cause 7 (Squirrel) -> Interruption Type (TX)

Interruption Cause 25 (Vines/Grass) -> Interruption Type (TX)

Interruption Cause 195 (Planned – Crew Request) -> Equipment Code 81 (Pole)

Interruption Cause 188 (Equip Fail-OH) -> Equipment Code 91(Connector),

Interruption Type (SV)

Interruption Cause 11 (Other Animal) -> Interruption Type (TX)

Interruption Cause 202 (Loose Connection) -> Equipment Code 91 (Connector)

The Interruption Causes that affect an equipment in more than 1% of all cases is shown above. A reduction in the support significantly increases the number of interesting association rules for the same range of confidence. Decision makers must decide the priority that must be given to a particular cause-effect relationship and the number of such rules that are required depending on the man power available.

To extract association rules for different interruption cause categories, the support for the analysis is set at a threshold of 0.001 and the confidence for the analysis is set at a threshold of 0.5 which is better than the toss of a coin. This ensures that a significant number of rules are generated.

7.3.2 Weather Related Causes

The results of the Cause-Effect analysis when done on Weather related causes are listed below:

Interruption Cause 2 (Storm w/no equip damage) -> Interruption Type (LAT)

Interruption Cause 1 (Lightning with equip damage) -> Interruption Type (TX)

Interruption Cause 6 (Storm/Wind (with equip damage) -> Equipment Code 85 (Arrester)

Interruption Cause 21 (Tree/Limb Unpreventable) -> Equipment Code 105 (Conductor Damaged)

Interruption Cause 21 (Tree/Limb Unpreventable) -> Interruption Type (SV)

Interruption Cause 21 (Tree/Limb Unpreventable) -> Equipment Code 91 (Connector)

Interruption Cause 6 (Storm/Wind (with equip damage) -> Equipment Code 104 (Conductor Down), Interruption Type (LAT)

Interruption Cause 6 (Storm/Wind (with equip damage) -> Equipment Code 94 (Transformer), Interruption Type (LAT)

From the above nothing can be made of hurricanes and tornados since there are not enough data points for the algorithm to derive meaning full rules from.

7.3.3 Animal Related Causes

The results of the Cause-Effect analysis when done on Animal related causes are listed below:

Interruption Cause 7 (Squirrel) -> Interruption Type (TX)

Interruption Cause 11 (Other Animal) -> Interruption Type (TX)

Interruption Cause 11 (Other Animal) -> Interruption Type (LAT)

Interruption Cause 9 (Other Bird) -> Interruption Type (TX)

Interruption Cause 7 (Squirrel), Equipment Type 93 (Fuse Switch) -> Interruption (TX)

Interruption Cause 18 (Lizards) -> Interruption Type (TX)

Interruption Cause 18 (Lizards) -> Interruption Type (LAT)

Interruption Cause 9 (Other Bird) -> Interruption Type (LAT)

7.3.4 Equipment Failure Related Causes

The results of the Cause-Effect analysis when done on Equipment Failure related causes are listed below:

Interruption Cause 189 (Equip Failed-UG) -> Interruption Type (SV), Equipment Type 111 (Cable)

Interruption Cause 188 (Equip Failed-OH) -> Interruption Type (SV), Equipment Type 91(Connector)

Interruption Cause 189 (Equip Failed-UG) -> Interruption Type (LAT), Equipment Type 111 (Cable)

Interruption Cause 188 (Equip Failed-OH) -> Interruption Type (TX), Equipment Type 94(Transformer)

Interruption Cause 189 (Equip Failed-UG) -> Interruption Type (SEC), Equipment Type 111 (Cable)

Interruption Cause 188 (Equip Failed-OH) -> Interruption Type (SV), Equipment Type 105(Conductor Damaged)

Interruption Cause 202 (Loose Connection) -> Equipment Type 91(Connector)

Interruption Cause 188 (Equip Failed-OH) -> Interruption Type (TX), Equipment Type 93(Fuse Switch)

Interruption Cause 188 (Equip Failed-OH) -> Equipment Type 104(Conductor Down)

Interruption Cause 188 (Equip Failed-OH) -> Interruption Type (LAT), Equipment Type 93(Fuse Switch)

Interruption Cause 188 (Equip Failed-OH) -> Equipment Type 85(Conductor Down)

Interruption Cause 188 (Equip Failed-OH) -> Interruption Type (FDR)

Interruption Cause 188 (Equip Failed-OH) -> Equipment Type 88(Conductor Down)

Interruption Cause 188 (Equip Failed-OH) -> Equipment Type 81(Conductor Down)

Interruption Cause 188 (Equip Failed-OH) -> Equipment Type 85(Conductor Down), Interruption Type (TX)

Interruption Cause 188 (Equip Failed-OH) -> Equipment Type 102(Conductor Down)

7.3.5 Unavoidable Causes

The results of the Cause-Effect analysis when done on Unavoidable causes are listed below:

Interruption Cause 191 (Vandalism) -> Equipment Type 160 (Meter)

Interruption Cause 40 (Vehicle) -> Equipment Type 81 (Pole)

Interruption Cause 191 (Vandalism) -> Equipment Type 160 (Meter), Interruption Type (MTR)

Interruption Cause 193 (Customer Request) -> Equipment Type 91 (Connector)

Interruption Cause 193 (Customer Request) -> Equipment Type 160 (Meter)

7.3.6 Avoidable/Human-Error Related Causes

The results of the Cause-Effect analysis when done on causes that may be due to human-error and can be avoided are listed below:

Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 94 (Transformer), Interruption Type (TX)

Interruption Cause 20 (Tree/Limb Preventable) -> Interruption Type (LAT)

Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 81 (Pole)

Interruption Cause 20 (Tree/Limb Preventable) -> Equipment Type 105 (Conductor

Damaged)

Interruption Cause 20 (Tree/Limb Preventable) -> Interruption Type (SV)

Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 91 (Connector), Interruption Type (TX)

Interruption Cause 20 (Tree/Limb Preventable) -> Equipment Type 104 (Conductor Down)

Interruption Cause 20 (Tree/Limb Preventable) -> Interruption Type (SEC)

Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 111 (Cable), Interruption Type (TX)

Interruption Cause 195 (Planned –Crew Request) -> Equipment Type 102 (Other Equipment), Interruption Type (TX)

Interruption Cause 20 (Tree/Limb Preventable) -> Equipment Type 93 (Fuse Switch)

Chapter 8

CONCLUSION

A method for predicting the most important characteristic of power outage “Outage Minutes” using readily available hourly weather forecast is provided. The predictive capability of regular weather forecasts to estimate the duration of an outage at the hourly level is analyzed. A classification based method to classify the Customer Minutes Interrupted due to an outage into predefined classes thereby aid in decision making at the managerial level is designed using a decision tree based method. The hourly weather forecasts are important for this type of classification. Finally, the equipment that is most commonly affected due to each category of cause of interruption is identified using the association rule mining technique that provides another application to the concept of association rules.

8.1 Limitations of the model

- Decades of research and model development have led to a reduction in daily weather forecast errors, but the above mentioned model still depends on the accuracy of hourly weather forecasts.
- The methods suggested are not dynamic. Weather and event conditions of only the particular hour is used to predict the outage duration.

- Prediction of CMI is majorly dependent on the number of customers affected which depends on the layout of the grid and the design of the electric network.

8.2 Summary of Conclusions

- The hourly weather forecast is able to answer 35% of the variations in the duration of an outage when a weather related event is predicted.
- There are other factors like the maintenance history of the equipment or the number of people assigned to repair a particular fault that would need to be included to improve the prediction of the duration of outage.
- The variables that give more insight into the maintenance schedule and work force scheduling that needs to be done in order to ensure prevention or faster recovery from an outage event.

8.3 Future Works

- More accurate prediction of the Outage Minutes by taking into consideration the maintenance schedules and also more subjective managerial inputs.
- A more dynamic model can be developed that takes into account the weather forecast for the hour along with the average of the weather forecast in the hours or days prior to the prediction hour.

- Prediction of the number of customers affected through the use of Poisson Regression or Zero Inflated Regression Models and better understanding of the network design.

- Build a man power planning model using the prediction of the outage minutes, the availability of man power and also the prediction of the equipment that would fail.

REFERENCES

[1] Litos Strategic Communication, “Exploring the imperative of revitalizing America’s electric infrastructure,” U.S. Department of Energy, 2008.

[2] The President’s Council of Economic Advisers, “Economic Benefits of Increasing Electric Grid Resilience TO Weather Outages,” U.S. Department of Energy, 2013.

[3] Alison Silverstein, Transmission 101, National Association of Regulatory Utility Commissioners, April 20, 2011, p. 30

[4] Lummus, R.R., Vokurka, R.J., "Defining supply chain management: a historical perspective and practical guidelines", Industrial Management and Data Systems, Vol. 99 No. 1, pp. 11-17, 1999

[5] Nagurney A, Matsypura D. A Supply Chain Network Perspective for Electric Power Generation, Supply, Transmission, and Consumption. In: Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 3-27.

[6] Nagurney A, Liu Z, Cojocaru M, Daniele P. Dynamic electric power supply chains and transportation networks: An evolutionary variational inequality formulation. Transportation Research Part E. 2007;43(5):624-46.

[7] Canadian Electricity Association, “The Importance and Benefits of Standards in the Electric Utility Industry”, 2015.

[8] Chase RB. Where does the customer fit in a service operation? Harvard business review. 1978; 56(6):137.

[9] Doane, Michael J. and Hartman, Raymond S. and Woo, chang ki, Households' Perceived Value of Electrical Service Reliability: An Analysis of Contingent Valuation Data (1988). The Energy Journal, 1988.

[10] Jha DK, Sinha SK, Garg A, Vijay A. Estimating electricity supply outage cost for residential and commercial customers. IEEE; 2012.

[11] Tiedemann KH. Estimating the value of reliability for residential customers. IEEE; 2004.

[12] Eto J, Koomey JG, Lehman B, Martin N, Mills E, Webber C, Worrell E. Scoping study on trends in the economic value of electricity reliability to the US economy. Report no. LBNL-47911. Berkeley, California: Lawrence Berkeley National Laboratory; 2001.

[13] National Oceanic and Atmospheric Administration. “Hurricane/Post-Tropical Cyclone Sandy, October 22–29, 2012.” Web. 12 Mar. 2015.

[14] Amin M. North America's electricity infrastructure: are we ready for more perfect storms? IEEE Security & Privacy Magazine. 2003; 99(5):19-25.

[15] M. J. Sullivan, B. N. Suddeth, T. Vardell, and A. Vojdani, “Interruption costs, customer satisfaction and expectations for service reliability,” Power Syst. IEEE Trans. On, vol. 11, no. 2, pp. 989–995, 1996.

[16] United States. Department of Energy.Office of Scientific and Technical Information, United States. Department of Energy, Lawrence Berkeley National Laboratory. Characteristics and trends in a National Study of Consumer Outage Costs. Washington, D.C: United States. Dept. of Energy; 2003.

[17] LaCommare KH, Eto JH. Cost of power interruptions to electricity consumers in the United States (US). Energy. 2006; 31(12):1845-55.

[18] Primen. The cost of power disturbances to industrial and digital economy companies. Report no. TR-1006274 (available through EPRI). Madison, WI: Primen; 2001.

[19] LaCommare KH, Eto JH. Understanding the cost of power interruptions to US electricity customers. Report no. LBNL-55718. Berkeley, California: Lawrence Berkeley National Laboratory; 2004.

[20] Reichl J, Schmidthaler M, Schneider F. The value of supply security: The costs of power outages to Austrian households, firms and the public sector. ENERGY ECONOMICS. 2013;36:256-61.

[21] Woo C, Ho T, Shiu A, Cheng Y, Horowitz I, Wang J. Residential outage cost estimation: Hong Kong. ENERGY POLICY. 2014;72:204-10.

[22] Sanghvi AP. Economic costs of electricity supply interruptions. Energy Economics. 1982;4(3):180-98.

[23] Campbell RJ, Lowry S, Library of Congress. Congressional Research Service. Weather-related power outages and electric system resiliency. Washington, District of Columbia: Congressional Research Service; 2012.

[24] Eaton Corp. “Blackout Tracker United States Annual Report 2015”, 2015 Power Outage Annual Report

[25] Balijepalli N, Venkata SS, Richter CW, Christie RD, Longo VJ. Distribution system reliability assessment due to lightning storms. IEEE Transactions on Power Delivery. 2005;20(3):2153-9.

[26] Brown RE, Gupta S, Christie RD, Venkata SS, Fletcher R. Distribution system reliability assessment: momentary interruptions and storms. IEEE Transactions on Power Delivery. 1997;12(4):1569-75.

[27] Kristina Hamachi LaCommare and Joseph H. Eto, Understanding the Cost of Power Interruptions to U.S. Electricity Consumers, Ernest Orlando Lawrence Berkeley National Laboratory, September 2004

[28] Peter H. Larsen, Kristina H. LaCommare, Joseph H. Eto, James L. Sweeney, Assessing Changes in the Reliability of the U.S. Electric Power System. Report no. LBNL-188741. Berkeley, California: Lawrence Berkeley National Laboratory; 2015.

[29] Reed DA. Electric utility distribution analysis for extreme winds. Journal of Wind Engineering & Industrial Aerodynamics. 2008;96(1):123-40.

[30] Han S, Guikema SD, Quiring SM, Lee K, Rosowsky D, Davidson RA. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. Reliability Engineering and System Safety. 2009;94(2):199-210.

[31] Nateghi R, Guikema SD, Quiring SM. Forecasting hurricane-induced power outage durations. Natural Hazards. 2014;74(3):1795-811.

[32] Davidson, R. A., H. Liu, 1. K. Sarpong, P. R. Sparks, and D. V. Rosowsky, 2003: Electric power distribution system performance in Carolina hurricanes. Nat.Hazards Rev., 4, 36-45.

[33] Liu, H., R. A. Davidson, D. V. Rosowsky, and J. R. Stedinger, 2005: Negative binomial regression of electric power outages in hurricanes. /. Inßrastruct. Syst., 11, 258-267.

[34] Liu H, Davidson RA, Apanasovich T. Statistical Forecasting of Electric Power Restoration Times in Hurricanes and Ice Storms. IEEE Transactions on Power Systems. 2007;22(4):2270-9.

[35] Guikema SD, Quiring SM, Han S. Prestorm Estimation of Hurricane Damage to Electric Power Distribution Systems. Risk Analysis. 2010;30(12):1744-52.

[36] Guikema, S. D., and S. M. Quiring, 2012: Hybrid data mining-regression for infrastructure risk assessment based on zero-inñated data. Reliab. Eng. Syst. Saß, 99,178-182.

[37] D. Zhu, D. Cheng, R. P. Broadwater, and C. Scirbona, ``Storm modelling for prediction of power distribution system outages," Electr. Power Syst. Res., vol. 77, no. 8, pp. 973979, Jun. 2007.

[38] H. Li, L. A. Treinish, and J. R. Hosking, "A statistical model for risk management of electric outage forecasts," IBM J. Res. Dev., vol. 54, no. 3, pp. 8–1, 2010.

[39] A. T. DeGaetano, B. N. Belcher, and P. L. Spier, "Short-term ice accretion forecasts for electric utilities using the weather research and forecasting model and a modified precipitation-type algorithm," Weather Forecast., vol. 23, no. 5, pp. 838–853, 2008.

[40] Rakesh Agrawal and Ramakrishnan Srikant, Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.

[41] S.B. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques, Informatica 31(2007) 249-268, 2007

[42] U.S. Energy Information Administration, Monthly Energy Review and Annual Energy Outlook 2014

[43] The 2013 Report Card for America’s Infrastructure, American Society of Civil Engineers, 2013

[44] Shane Fitzsimmons and Matthew McKenna, Utility Supply Chain Management, Booz & Company, 2008

APPENDIX A

R Code for Regression Analysis

```
install.packages("clusterSim")
library(clusterSim)
library(alr3)
#read the file
#view dataset
regress.featureseve <- read.csv("Event.csv", header=TRUE)
regress.featureseve <- as.data.frame(lapply(regress.featureseve,
as.numeric))
regress.featureseve$Events <- factor(regress.featureseve$Events)
View(regress.featureseve)
attach(regress.featureseve)


regress.featureseve$Heat.Index..F. <- NULL
regress.featureseve$Dew.Point..F. <- NULL
regress.featureseve$Humidity..per. <- NULL
regress.featureseve$Pressure..in. <- NULL
regress.featureseve$Events <- NULL
regress.featureseve$Outage.Minutes <- NULL
regress.featureseve$Cust.Int <- NULL
regress.featureseve$CMI <- NULL
regress.featureseve$CMI.Classes <- NULL
regress.featureseve$New.clusters <- NULL


meanregress.featureseve<-
as.data.frame(cbind(Outage.Minutes,CMI,Cust.Int,scale(regress.featurese
ve[,1:9], center = TRUE, scale = FALSE)))

zregress.featureseve<-
as.data.frame(cbind(Outage.Minutes,CMI,Cust.Int,Events,scale(regress.fe
atureseve[,1:9], center = TRUE, scale = TRUE)))

normregress.featureseve<-
cbind(Outage.Minutes,CMI,Cust.Int,data.Normalization
(regress.featureseve[,1:9],type="n4",normalization="column"))

View(meancluster.featuresregress.featureseve)

View(zregress.featureseve)

View(normregress.featureseve)
```

```
#Power Transform
regress.featureseve<-regress.featureseve+0.000001
#set the seed to any number no meaning
set.seed(123)
library(MASS)
library(caret)
library(VGAM)
lambda=seq(0.3,0.65,length=20)
lambda<-lapply(lambda,as.numeric)
m1<-lapply(m1,as.numeric)
boxcox(m1,lambda=seq(-2,2,length=20))
yeo.johnson(m1, lambda=seq(-2,2,length=20))
par(mfrow=c(2,2))
plot(Temperature,sqrt(Defective),ylab=expression(sqrt(Defective)))
library(alr3)

summary(powerTransform(cbind(Outage.Minutes,Temp...F., Heat.Index..F.,
Dew.Point..F., Humidity..per., Pressure..in., Visibility..mi.,
Wind.Speed..mph., Gust.Speed..mph., Precip..in.) ~ 1,
regress.featureseve))



pairs(regress.featureseve)
plot(regress.featureseve[,1],CMI,xlab="Temp")
plot(regress.featureseve[,2],CMI,xlab="HeatIndex")
plot(regress.featureseve[,3],CMI,xlab="Dewpoint")
plot(regress.featureseve[,4],CMI,xlab="Humidity")
plot(regress.featureseve[,5],CMI,xlab="Pressure")
plot(regress.featureseve[,6],CMI,xlab="Visibility")
plot(regress.featureseve[,7],CMI,xlab="Wind Speed")
plot(regress.featureseve[,8],CMI,xlab="Gust Speed")
plot(regress.featureseve[,9],CMI,xlab="Precipitation")


m1 <- lm(CMI^0.5~Temp...F. + Heat.Index..F. + Dew.Point..F. +
Humidity..per. + Pressure..in. + Visibility..mi. + Wind.Speed..mph.+
I(Gust.Speed..mph.^2)+ Precip..in.+ Events,
data=normregress.featureseve)

m2 <- lm(Outage.Minutes^0.5 ~ Visibility..mi. + Pressure..in. +
I(Gust.Speed..mph.^2)+ Precip..in.+ Events,
data=normregress.featureseve)

m2 <- lm(Outage.Minutes^0.5~Temp...F. + Heat.Index..F. + Dew.Point..F.
+ Humidity..per. + Pressure..in. + Visibility..mi. + Wind.Speed..mph.+
I(Gust.Speed..mph.^2)+ Precip..in.+Events,
data=normregress.featureseve)

m3 <- lm(Cust.Int~Temp...F. + Heat.Index..F. + Dew.Point..F. +
Humidity..per. + Pressure..in. + Visibility..mi. + Wind.Speed..mph.+
Gust.Speed..mph.+ Precip..in., data=normregress.featureseve)



summary(m2)
leverage1 <- hatvalues(m1)
StanRes1 <- rstandard(m1)
plot(Temp...F.,StanRes1,xlab="Temperature", ylab="Standardized
Residuals")
#can print like this for the other predictors too
```

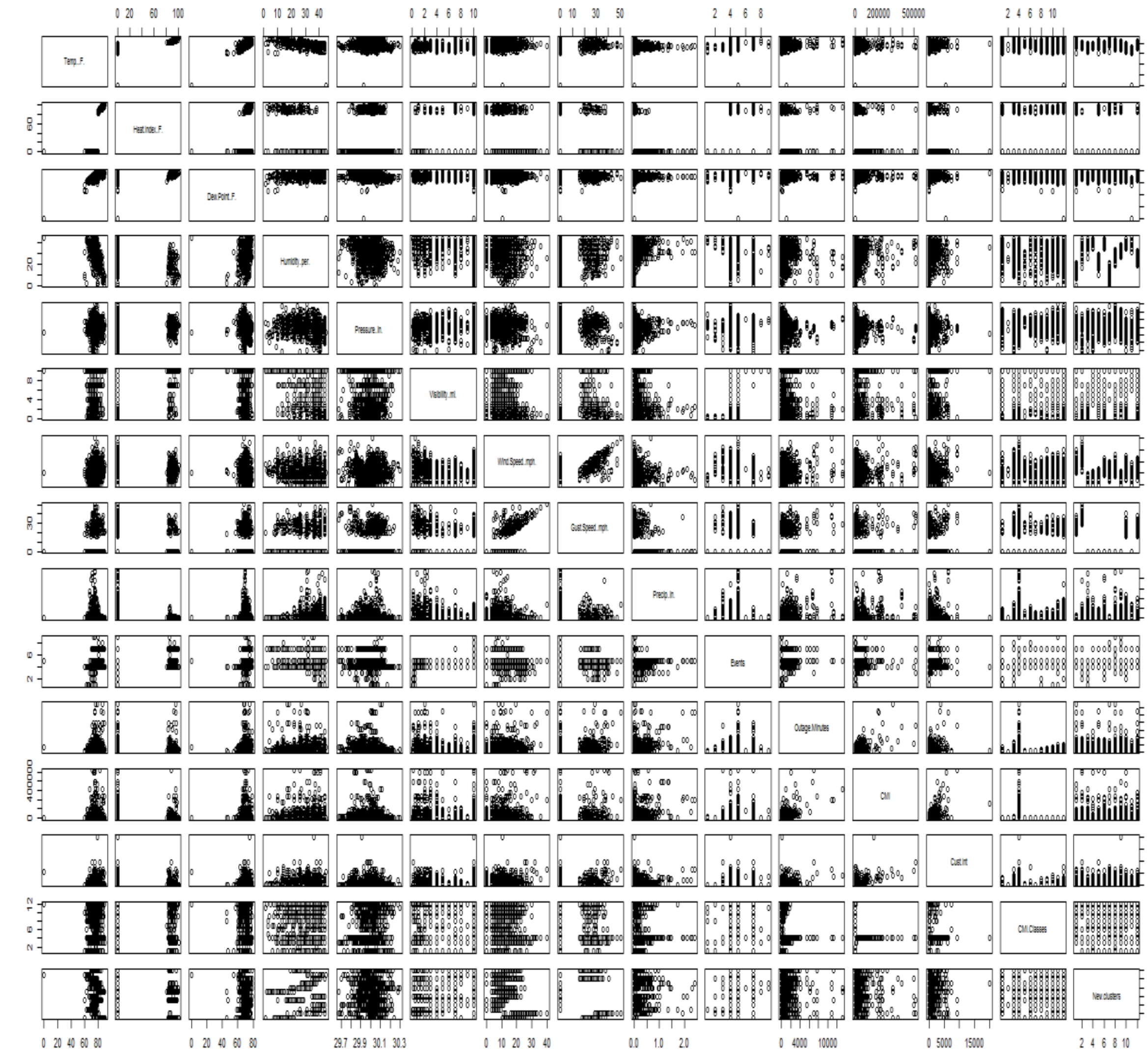
```
par(mfrow=c(2,2))
plot(m2)

#fitted vs actual values for CMI

par(mfrow=c(1,1))

plot(m2$fitted.values,Outage.Minutes,xlab="Fitted Values", ylab="Outage
Minutes")

abline(lsfit(m2$fitted.values,Outage.Minutes))


install.packages('car', repos='http://lib.stat.cmu.edu/R/CRAN/')

install.packages("VGAM")

library(car)

par(mfrow=c(1,1))

avPlots(m1,variable=Temp...F.,ask=FALSE,identify.points=TRUE) #can do
for all variables #added variable plot


library(alr3)

inverseResponsePlot(m1,key=TRUE)


#Figure 6.15 on page 173

library(MASS)

library(caret)

library(VGAM)

lambda=seq(0.3,0.65,length=20)

lambda<-lapply(lambda,as.numeric)

m1<-lapply(m1,as.numeric)

boxcox(m1,lambda=seq(-2,2,length=20))

yeo.johnson(m1, lambda=seq(-2,2,length=20))


#Figure 6.16 on page 173

par(mfrow=c(2,2))
```

```
plot(Temperature,sqrt(Defective),ylab=expression(sqrt(Defective)))


library(alr3)

summary(powerTransform(cbind(Outage.Minutes,Temp...F., Heat.Index..F.,
Dew.Point..F., Humidity..per., Pressure..in., Visibility..mi.,
Wind.Speed..mph., Gust.Speed..mph., Precip..in.) ~ 1,
regress.features))



#to remove all rows with any column=0

row_sub = apply(regress.features, 1, function(row) all(row !=0 ))

##Subset as usual

regress.features1<-regress.features[row_sub,]



m1 <- lm(CMI^0.1~I(Temp...F.^3) + I(Heat.Index..F.^-0.08) +
I(Dew.Point..F.^3) + I(Humidity..per.^2) + I(Pressure..in.^23) +
I(Visibility..mi.^9) + I(Wind.Speed..mph.^0.5)+ I(Gust.Speed..mph.^-
0.5)+ I(Precip..in.^-1.5), data=regress.features)

m1 <- lm(CMI^0.1~I(Temp...F.^3) + I(Heat.Index..F.^-0.08) +
I(Dew.Point..F.^3) + I(Humidity..per.^2) + Pressure..in. +
Visibility..mi. + I(Wind.Speed..mph.^0.5)+ I(Gust.Speed..mph.^-0.5)+
I(Precip..in.^-1.5), data=regress.features)

m1 <- lm(CMI^0.1~I(Temp...F.^3) + I(Heat.Index..F.^-0.08) +
I(Dew.Point..F.^3) + I(Humidity..per.^2) + (Pressure..in.) +
log(Visibility..mi.) + I(Wind.Speed..mph.^0.5)+ I(Gust.Speed..mph.^-
0.5)+ I(Precip..in.^-1.5), data=regress.features)

m1 <- lm(log(CMI)~log(Temp...F.) + log(Heat.Index..F.) +
log(Dew.Point..F.) + log(Humidity..per.) + log(Pressure..in.) +
log(Visibility..mi.) + log(Wind.Speed..mph.)+ log(Gust.Speed..mph.)+
log(Precip..in.), data=regress.features)

m1 <- lm((CMI^0.5)~log(Temp...F.) + log(Pressure..in.) +
log(Visibility..mi.) + log(Wind.Speed..mph.)+ log(Gust.Speed..mph.) +
log(Precip..in.), data=regress.features)

m1 <- lm(CMI^0.5~I(Temp...F.^2) + log(Pressure..in.) +
log(Visibility..mi.) + I(Wind.Speed..mph.^0.5)+ log(Gust.Speed..mph.)+
I(Precip..in.^-0.5), data=regress.features)

m1 <- lm(CMI^0.1~I(Temp...F.^3) + Pressure..in. + Visibility..mi. +
I(Wind.Speed..mph.^0.5) + I(Gust.Speed..mph.^-0.5)+ I(Precip..in.^-
1.5), data=regress.features)

m1 <- lm(CMI^0.1~I(Temp...F.^3) + I(Heat.Index..F.^-0.08) +
I(Dew.Point..F.^3) + I(Humidity..per.^1.5) + I(Pressure..in.^1.2) +
I(Visibility..mi.^8) + I(Wind.Speed..mph.^0.5)+ I(Gust.Speed..mph.^-
0.5)+ I(Precip..in.^-1), data=regress.features)
```

```
m1 <- lm(Outage.Minutes^0.6~Temp...F. + I(Heat.Index..F.^-0.06) +
I(Dew.Point..F.^3) + I(Humidity..per.^2) + Pressure..in. +
Visibility..mi. + I(Wind.Speed..mph.^0.5)+ I(Gust.Speed..mph.^-0.3)+
I(Precip..in.^-1), data=regress.features)

m1 <- lm(Outage.Minutes~Temp...F. + Pressure..in. +
log(Visibility..mi.) + Wind.Speed..mph. + Precip..in.,
data=regress.features)

summary(m1)

plot(m1)



m1 <- lm(Defective ~ Temperature+Density+Rate)

loessfit1 <- loess(CMI ~ Temp...F.,degree=1,span=2/3)

loessfit2 <- loess(m1$fitted.values ~ Temperature,degree=1,span=2/3)

xx <- seq(min(CMI),max(CMI),length=100)

par(mfrow=c(1,2))

plot(Temp...F.,CMI,xlab="Temperature, x1", ylab="CMI, Y")

lines(xx,predict(loessfit1,data.frame(Temp...F.=xx)))

plot(Temp...F.,m1$fitted.values,ylab=expression(hat(Y)),xlab="Temperatu
re, x1")

lines(xx,predict(loessfit2,data.frame(Temp...F.=xx)))


par(mfrow=c(2,2))
mmp(m1,Temp...F.)
mmp(m1,Density,key="topright")
mmp(m1,Rate)
mmp(m1,m1$fitted.values,xlab="Fitted Values")


library(car)

vif(m1)
```

Results of Regression Analysis

- Pairs Plot



- Initial Model Result:

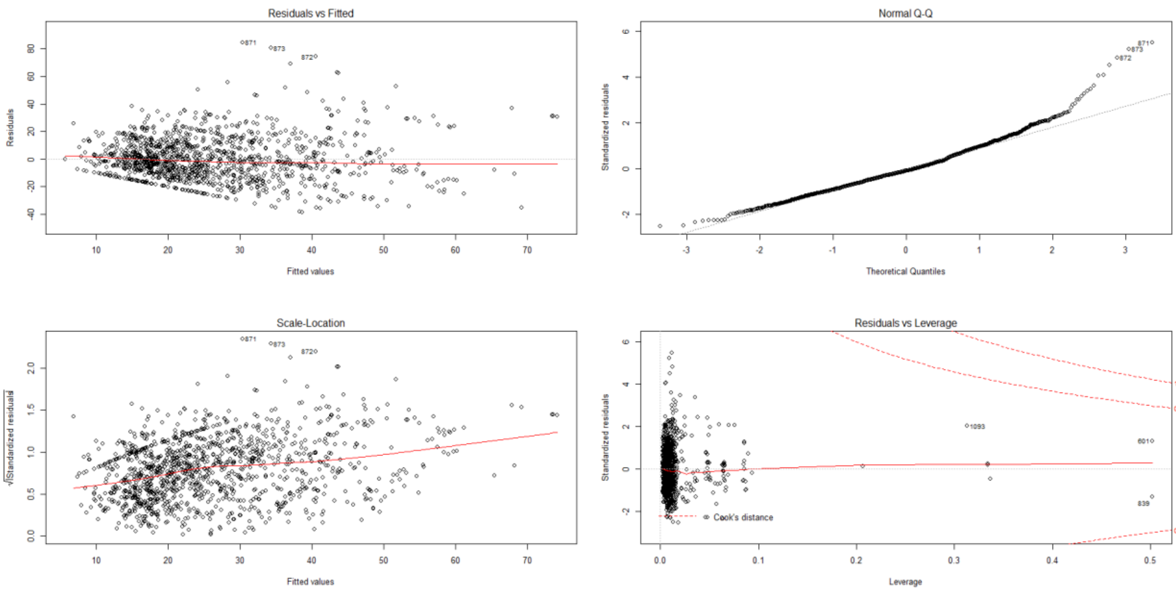
```
Call:
lm(formula = Outage.Minutes^0.5 ~ Temp...F. + Heat.Index..F. +
    Dew.Point..F. + Humidity..per. + Pressure..in. +
Visibility..mi. +
    Wind.Speed..mph. + I(Gust.Speed..mph.^2) + Precip..in. +
    Events, data = normregress.featureseve)

Residuals:
    Min      1Q  Median      3Q     Max
-38.687 -10.127  -1.614   8.957  84.400

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)              11.511     26.559   0.433 0.664784
Temp...F.               -30.503    128.676  -0.237 0.812655
```

```
Heat.Index..F.          -1.455      1.848   -0.788 0.431087    
Dew.Point..F.           53.583    112.786    0.475 0.634808    
Humidity..per.         -14.469     24.801   -0.583 0.559712    
Pressure..in.          -18.943      3.024   -6.265 5.12e-10 ***
Visibility..mi.         -5.686      1.520   -3.740 0.000192 ***
Wind.Speed..mph.        11.120      4.039    2.753 0.005986 ** 
I(Gust.Speed..mph.^2)   11.579      4.046    2.862 0.004280 ** 
Precip..in.             40.908      4.810    8.504  < 2e-16 ***
Events2                 -6.599      6.028   -1.095 0.273832    
Events3                 23.733      5.225    4.542 6.10e-06 ***
Events4                  6.593      4.144    1.591 0.111835    
Events5                 20.303      4.175    4.863 1.30e-06 ***
Events6                 -3.164     16.080   -0.197 0.844041    
Events7                 13.567      4.486    3.024 0.002544 ** 
Events8                 10.334     11.757    0.879 0.379601    
Events9                  2.312      9.870    0.234 0.814815    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.46 on 1257 degrees of freedom
Multiple R-squared:  0.3457,    Adjusted R-squared:  0.3369 
F-statistic: 39.07 on 17 and 1257 DF,  p-value: < 2.2e-16
```

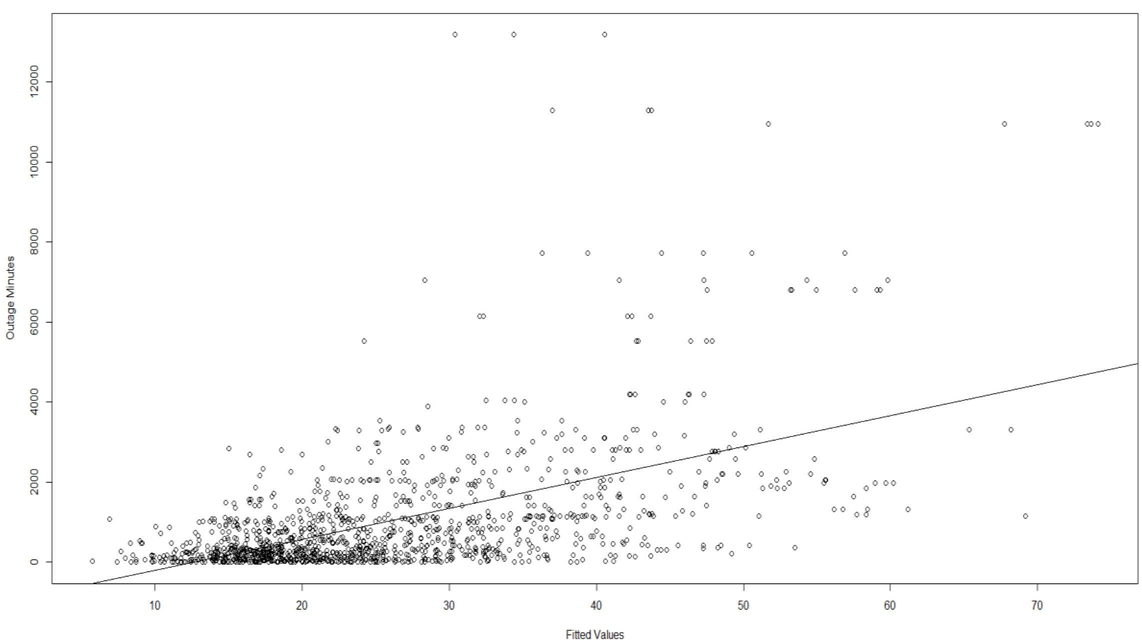
- Revised Model

```
Call:
lm(formula = Outage.Minutes^0.5 ~ Visibility..mi. + Pressure..in. 
+ 
    I(Gust.Speed..mph.^2) + Precip..in. + Events, data = 
normregress.featureseve)

Residuals:
    Min      1Q  Median      3Q     Max 
-36.651  -9.738  -1.957   8.775  86.946 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)             20.737      4.239   4.892 1.13e-06 ***
Visibility..mi.         -5.039      1.493  -3.376 0.000758 ***
Pressure..in.          -16.943      2.965  -5.714 1.37e-08 ***
I(Gust.Speed..mph.^2)   20.950      3.031   6.912 7.58e-12 ***
Precip..in.             38.372      4.810   7.977 3.33e-15 ***
Events2                 -3.208      6.007  -0.534 0.593401    
Events3                 26.233      5.228   5.018 5.97e-07 ***
Events4                  9.630      4.122   2.337 0.019620 *  
Events5                 23.650      4.151   5.697 1.52e-08 ***
Events6                  1.636     16.144   0.101 0.919286    
Events7                 18.657      4.379   4.261 2.19e-05 ***
Events8                 15.110     11.785   1.282 0.200042    
Events9                  5.519      9.928   0.556 0.578376    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.59 on 1262 degrees of freedom
Multiple R-squared:  0.3318,    Adjusted R-squared:  0.3254 
F-statistic: 52.22 on 12 and 1262 DF,  p-value: < 2.2e-16
```

Diagnostic Plots



Fitted Values vs Actual Values Plot



APPENDIX B

R Code for Classification Algorithms.

```
"J48 algorithm"

library(e1071)
library(caret)
#classification tree
treeclassevent <- read.csv("Epredictorclasses.csv", header=TRUE)

#set the seed to any number no meaning
set.seed(123)

treeclassevent$Temp.Classes<-NULL
treeclassevent$Heat.Classes <- NULL
treeclassevent$Dew.Classes <- NULL
treeclassevent$Humidity.Classes <- NULL
#treeclassevent$Events <-NULL
treeclassevent$Outage.Minutes <- NULL
treeclassevent$Cust.Int <- NULL
treeclassevent$CMI <- NULL
treeclassevent$CMI.Classes <- NULL

#sample 2/3's for the training and the remaining for the test set
index <- sample(1:nrow(treeclassevent),size =
(2/3)*nrow(treeclassevent))

trainsetctree <- treeclassevent[index,]
testsetctree <- treeclassevent[-index,]

nrow(trainsetctree)
nrow(testsetctree)
view(trainsetctree)

Confusion Matrix and Statistics


         Reference
Prediction Class 0 Class 1 Class 2 Class 3 Class 4 Class 5 Class 6 Class 7
   Class 0       6       1       4       1      11       3       6      14
   Class 1       1       4       0       2       2       5       5      12
   Class 2       0       0       0       1       1       1       3       3
   Class 3       4       0       5       3       3       4       3       8
   Class 4       1       0       0       0       1       0       1       4
   Class 5       0       0       1       3       4       4       4       4
   Class 6       2       2       2       1       4       0       8       9
   Class 7      13      11      10      13      12      12      39     144

Overall Statistics

               Accuracy : 0.4
                 95% CI : (0.3531, 0.4483)
    No Information Rate : 0.4659
    P-Value [Acc > NIR] : 0.9973

                  Kappa : 0.1319
 Mcnemar's Test P-Value : NA
```

```
Statistics by Class:

                     Class: Class 0 Class: Class 1 Class: Class 2
Class: Class 3 Class: Class 4 Class: Class 5
Sensitivity                 0.22222       0.222222        0.00000
0.125000       0.026316       0.137931
Specificity                 0.89950       0.933661        0.97767
0.932668       0.984496       0.959596
Pos Pred Value              0.13043       0.129032        0.00000
0.100000       0.142857       0.200000
Neg Pred Value              0.94459       0.964467        0.94712
0.946835       0.911483       0.938272
Prevalence                  0.06353       0.042353        0.05176
0.056471       0.089412       0.068235
Detection Rate              0.01412       0.009412        0.00000
0.007059       0.002353       0.009412
Detection Prevalence        0.10824       0.072941        0.02118
0.070588       0.016471       0.047059
Balanced Accuracy           0.56086       0.577942        0.48883
0.528834       0.505406       0.548763
                     Class: Class 6 Class: Class 7
Sensitivity                 0.11594         0.7273
Specificity                 0.94382         0.5154
Pos Pred Value              0.28571         0.5669
Neg Pred Value              0.84635         0.6842
Prevalence                  0.16235         0.4659
Detection Rate              0.01882         0.3388
Detection Prevalence        0.06588         0.5976
Balanced Accuracy           0.52988         0.6213
```

"ctree Algorithm"

```
#plot to show the distribution of the training and test set
library(rpart)
library(ggplot2)
install.packages('partykit')
library(partykit)
library(party)
fit <- rpart(CMI.Classes ~ ., method="class", data=trainsetctree,
control=rpart.control(minsplit=4, minbucket=1, cp=0.001))
fit
fitparty <- as.party(fit)
plot(fitparty)

#Minimal Cp(Complexity parameter) value
fit$cptable[which.min(fit$cptable[,"xerror"]),"CP"]

printcp(fit) # display the results
plotcp(fit) # visualize cross-validation results
summary(fit) # detailed summary of splits

install.packages("rattle")
library(rattle)
rattle()
library(rpart.plot)
library(RColorBrewer)
fancyRpartPlot(fit)
```

```
pfit<- prune(fit, cp=
fit$cptable[which.min(fit$cptable[,"xerror"]),"CP"])
fancyRpartPlot(pfit, uniform=TRUE, main="Pruned Classification Tree")


# plot tree
plot(fit, uniform=TRUE, main="Classification Tree for CMI")
text(fit, use.n=TRUE, all=TRUE, cex=.8)

# create attractive postscript plot of tree
post(fit, title = "Classification Tree for Kyphosis")Summary of the
Conditional Tree model for Classification (built using 'ctree'):
```

Conditional inference tree with 3 terminal nodes

Response: CMI.Classesnew
Inputs: Pressure.Classes, Visibility.Classes, Gust.Classes, Precip.Classes
Number of observations: 892

```
1) Precip.Classes == {R2, R3, R4, R5, R6}; criterion = 1, statistic = 100.927
  2) Visibility.Classes == {V1, V3}; criterion = 0.998, statistic = 69.202
    3)*  weights = 122
  2) Visibility.Classes == {V2, V4, V5, V6}
    4)*  weights = 390
1) Precip.Classes == {R1}
  5)*  weights = 380
```

APPENDIX C

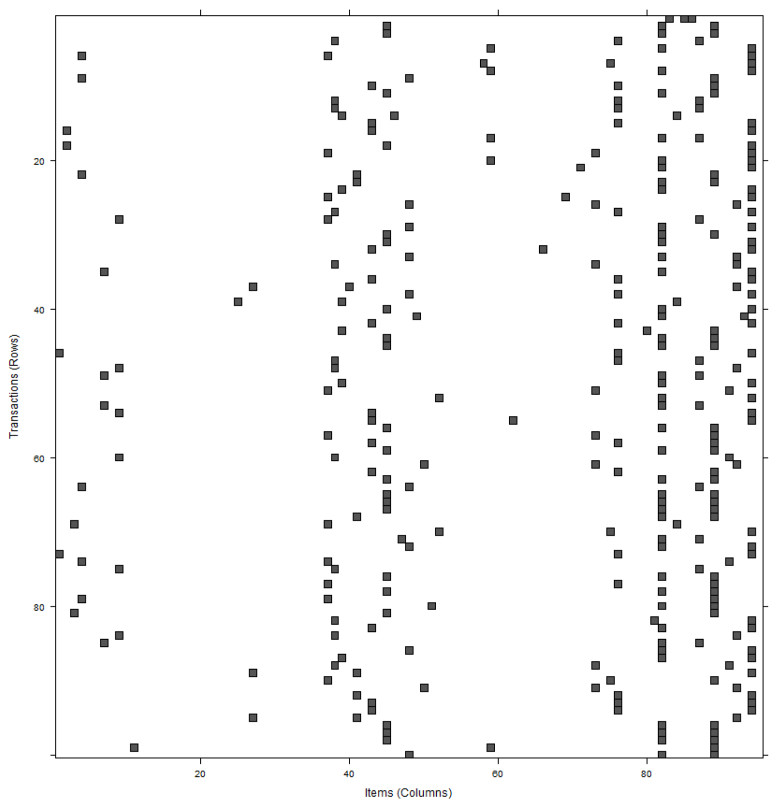
R Code for Association Rule Mining

```
#installing packages
install.packages("arules")
data("Adult")

#load packages
library(xlsx)
library(arules)

#loading the data
mydata <- read.csv("Itemsetminingweather.csv", header=TRUE)
as.data.frame(mydata) #converting to a data frame
#table(mydata)
#table(mydata$Interuption.Type)

#convert the numeric vectors to characters
mydata$Equipment.Code<-as.character(mydata$Equipment.Code)
mydata$Interuption.Cause<-as.character(mydata$Interuption.Cause)
is.character(mydata$Equipment.Code)
is.character(mydata$Interuption.Cause)

#converting the data frame to binary matrix of Interruption Type,
Equipment code
mydatabin <- cbind(mydata, model.matrix( ~ 0 + Interuption.Type,
mydata))
mydatabin1 <- cbind(mydata, model.matrix( ~ 0 + Equipment.Code,
mydata))
mydatabin2 <- cbind(mydata,model.matrix( ~ 0 + Ticket.Type, mydata))
mydatabin3 <- cbind(mydata,model.matrix( ~ 0 + Interuption.Cause,
mydata))
```

```
binarymatdata<-
cbind(mydatabin[,8:17],mydatabin1[,8:53],mydatabin2[,8:18],
mydatabin3[,8:44])


binarymatdata[]<-lapply(binarymatdata,factor)

binarymatdata<-as.factor(binarymatdata[,1:67])

binarymatdata<-as.data.frame(binarymatdata)


#Association Rule Mining Algorithms


#Apriori Algorithm

#defining the transactions

tr<-
read.transactions("Transactionsdataplanned.csv",format="basket",sep=","
) #with duplicates

tr<-
read.transactions("Transactionsdataweather.csv",format="basket",sep=","
,rm.duplicates=TRUE) #without duplicates

inspect(tr)

par(mfrow=c(1,1))

itemFrequencyPlot(tr, support = 0.1)

image(tr[1:100,]) #image([1:10,])

length(tr)


#applying the algorithm

rules <- apriori(tr, parameter= list(supp=0.01, conf=0.5))

summary(rules)

inspect(rules)


subset.matrix <- is.subset(rules,rules)

subset.matrix[lower.tri(subset.matrix,diag=T)] <- NA

redundant <- colSums(subset.matrix,na.rm=T) >= 1

rules <- rules[!redundant]

rules <- sort(rules, decreasing = TRUE, na.last = NA, by = "support")
```

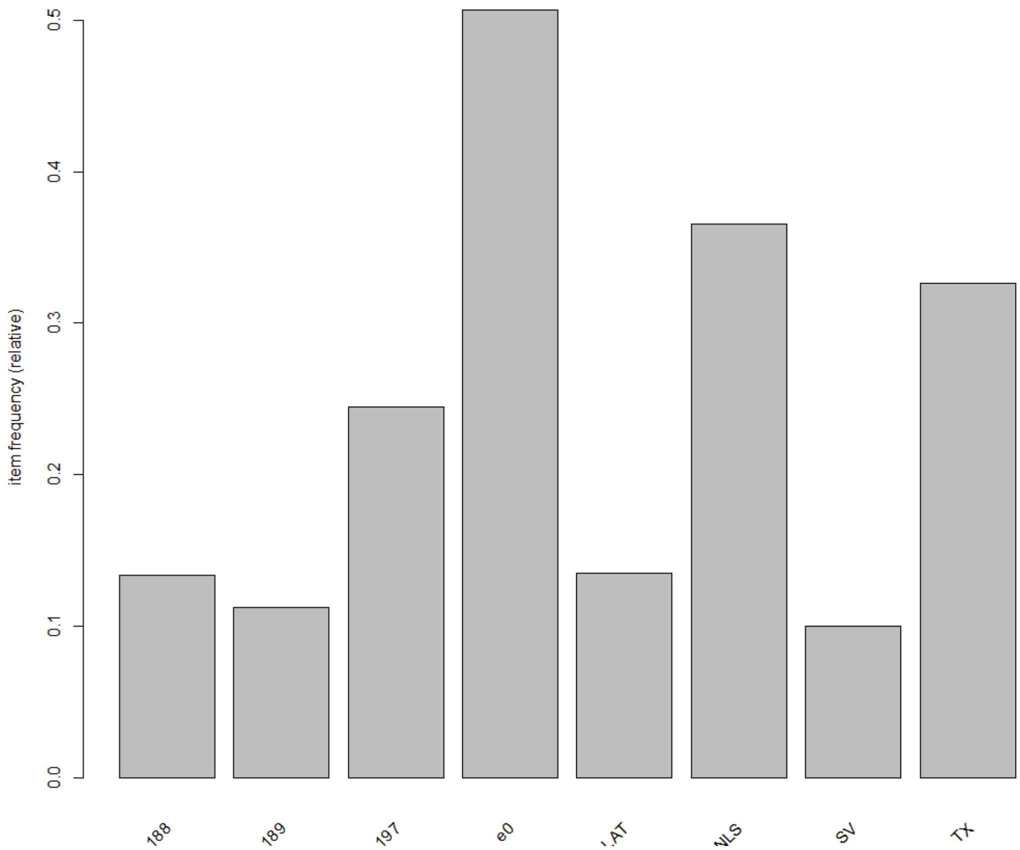
```
inspect(rules)



#wwrite the new data frame to an excel file

write.csv(binarymatdata, "C:/Semester 3/Research/Thesis/Data -
testing/Itemsetminingbinary.csv")

write.xlsx(mydata, "C:/Semester 3/Research/Thesis/Data -
testing/Itemsetminingbinary.xlsx")
```

Item Image Plot



All Causes

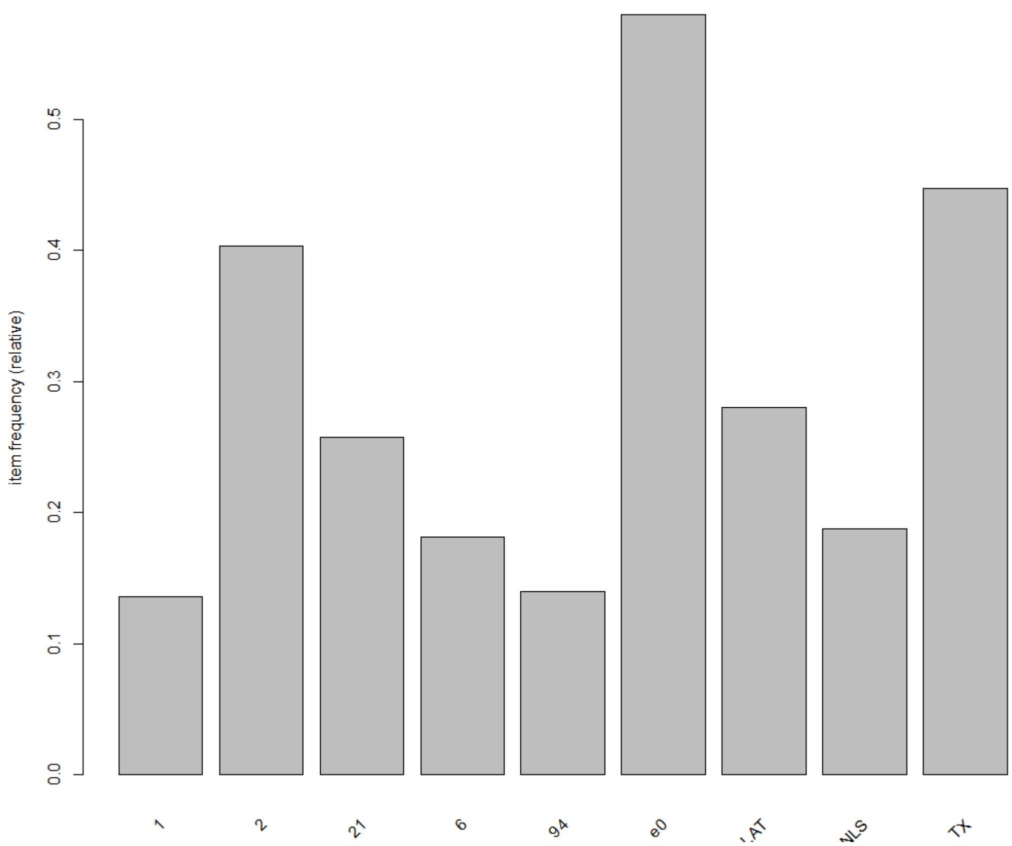
Item Frequency Plot




```
Support - 0.01 (1%) Confidence - 0.5 (50%)

Parameter specification:
 confidence minval smax arem  aval originalSupport support minlen
maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE    0.01      1
10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

apriori - find association rules with the apriori algorithm
version 4.21 (2004.05.09)        (c) 1996-2004   Christian Borgelt
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[95 item(s), 24551 transaction(s)] done [0.02s].
sorting and recoding items ... [34 item(s)] done [0.00s].
creating transaction tree ... done [0.02s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [67 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].


   lhs      rhs      support confidence     lift
```

```
1  {}     => {e0}  0.50645595  0.5064559 1.000000
2  {197} => {NLS} 0.21017474  0.8581407 2.350052
3  {195} => {TX}  0.08060771  0.8954751 2.741590
4  {94}  => {TX}  0.07738992  0.7969799 2.440036
5  {111} => {189} 0.07009898  0.8114097 7.228200
6  {193} => {NLS} 0.02476478  0.7562189 2.070935
7  {105} => {SV}  0.02195430  0.5202703 5.186015
8  {SEC} => {189} 0.02048796  0.5356763 4.771911
9  {93}  => {188} 0.01914382  0.5414747 4.045570
10 {7}   => {TX}  0.01832919  0.6686478 2.047135
11 {25}  => {TX}  0.01824773  0.8000000 2.449283
12 {81}  => {TX}  0.01682213  0.5728155 1.753734
13 {81}  => {195} 0.01649627  0.5617198 6.240174
14 {91,
    SV}  => {188} 0.01596676  0.5178336 3.868938
15 {11}  => {TX}  0.01270824  0.5148515 1.576271
16 {202} => {91}  0.01201580  0.5139373 6.883619
17 {160} => {NLS} 0.01140483  0.5333333 1.460554
```

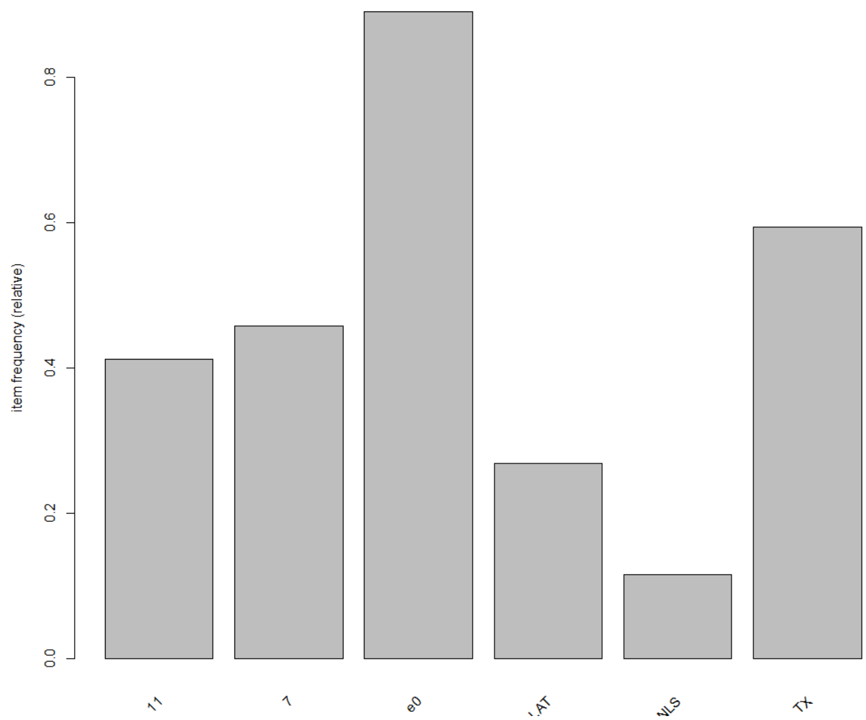
Weather Related Instances




```
set of 33 rules
```

```
rule length distribution (lhs + rhs):sizes
 1  2  3 
 1 14 18 

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.000   2.000   3.000   2.515   3.000   3.000 

summary of quality measures:
    support          confidence          lift       
 Min.   :0.01045   Min.   :0.5149   Min.   :0.9845  
 1st Qu.:0.02457   1st Qu.:0.5708   1st Qu.:1.2251  
 Median :0.05541   Median :0.6455   Median :1.7745  
 Mean   :0.11066   Mean   :0.6860   Mean   :1.9900  
 3rd Qu.:0.14428   3rd Qu.:0.7717   3rd Qu.:2.6054  
 Max.   :0.57972   Max.   :0.9674   Max.   :4.4939  

mining info:
 data ntransactions support confidence
   tr          1913    0.01        0.5



   lhs      rhs     support confidence     lift
1  {}     => {e0} 0.57971772  0.5797177 1.000000
2  {LAT} => {2}  0.14427601  0.5149254 1.277629
3  {94}  => {TX} 0.11604809  0.8314607 1.860332
4  {1}   => {TX} 0.07056979  0.5192308 1.161741
5  {85}  => {6}  0.02822791  0.5294118 2.918630
6  {105} => {21} 0.02456874  0.6714286 2.605361
7  {SV}  => {21} 0.02038683  0.7500000 2.910243
8  {91}  => {21} 0.01254574  0.6153846 2.387892
9  {104,LAT} => {6}  0.01254574  0.5217391 2.876331
10 {94,LAT} => {6}  0.01150026  0.6875000 3.790166
```

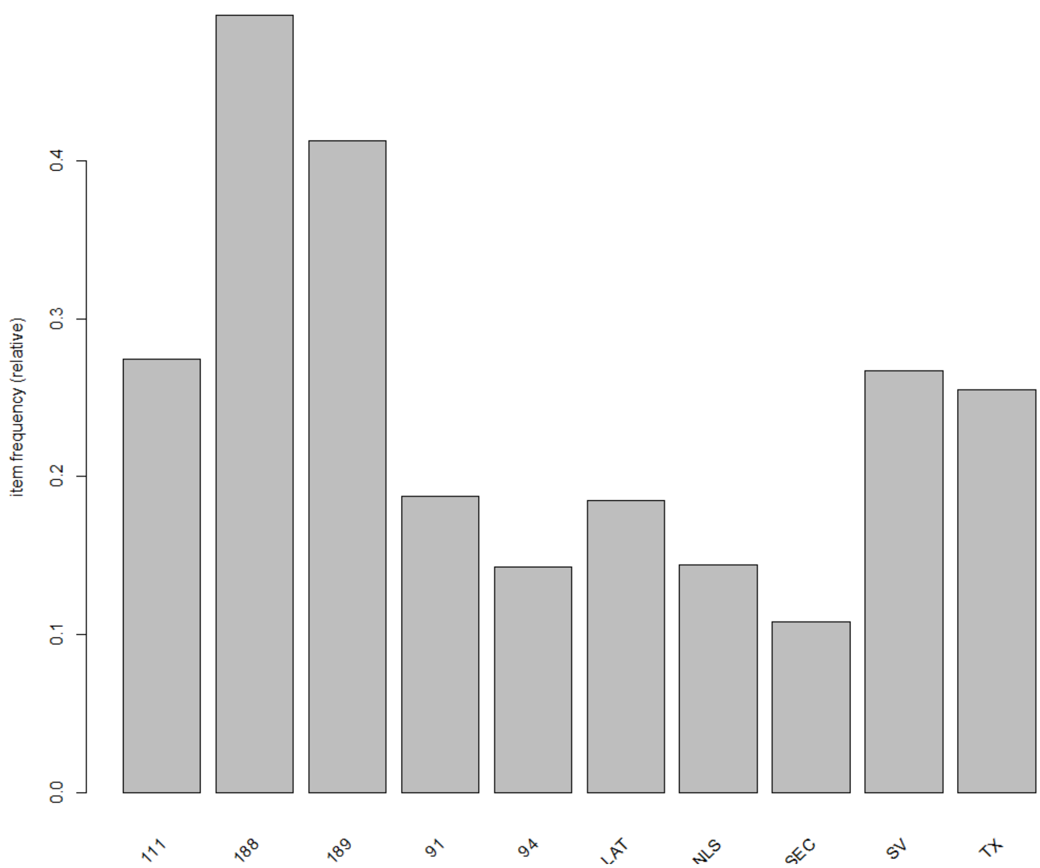
Animal Related Instances




```
Parameter specification:
 confidence minval smax arem  aval originalSupport support minlen
maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE    0.01      1
10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

apriori - find association rules with the apriori algorithm
version 4.21 (2004.05.09)        (c) 1996-2004   Christian Borgelt
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[35 item(s), 1468 transaction(s)] done [0.00s].
sorting and recoding items ... [10 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [33 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].

   lhs      rhs     support confidence      lift
1  {}    => {e0} 0.89032698  0.8903270 1.0000000
2  {}    => {TX} 0.59468665  0.5946866 1.0000000
3  {TX}  => {e0} 0.55381471  0.9312715 1.0459882
4  {e0}  => {TX} 0.55381471  0.6220352 1.0459882
5  {7}   => {e0} 0.41689373  0.9093611 1.0213788
6  {11}  => {e0} 0.36376022  0.8811881 0.9897354
7  {7}   => {TX} 0.30653951  0.6686478 1.1243700
8  {TX}  => {7}  0.30653951  0.5154639 1.1243700
9  {7,TX}  => {e0} 0.28814714  0.9400000 1.0557919
10 {7,e0}  => {TX} 0.28814714  0.6911765 1.1622532
11 {e0,TX}  => {7}  0.28814714  0.5202952 1.1349084
```

```
12 {LAT} => {e0} 0.23024523  0.8578680 0.9635427
13 {11}  => {TX} 0.21253406  0.5148515 0.8657526
14 {11,TX}  => {e0} 0.19550409  0.9198718 1.0331842
15 {11,e0}  => {TX} 0.19550409  0.5374532 0.9037586
16 {LAT} => {11} 0.13555858  0.5050761 1.2235178
17 {11,LAT} => {e0} 0.11648501  0.8592965 0.9651471
18 {e0,LAT} => {11} 0.11648501  0.5059172 1.2255551
19 {NLS} => {e0} 0.09741144  0.8411765 0.9447950
20 {7,LAT} => {e0} 0.09196185  0.8766234 0.9846084
21 {9}   => {e0} 0.07833787  0.9126984 1.0251272
22 {9}   => {TX} 0.05926431  0.6904762 1.1610757
23 {9,TX}  => {e0} 0.05585831  0.9425287 1.0586321
24 {9,e0}  => {TX} 0.05585831  0.7130435 1.1990239
25 {11,NLS} => {e0} 0.04972752  0.8795181 0.9878596
26 {e0,NLS} => {11} 0.04972752  0.5104895 1.2366314
27 {7,NLS} => {e0} 0.03337875  0.8166667 0.9172660
28 {18}  => {e0} 0.02656676  0.7959184 0.8939619
29 {7,93}  => {TX} 0.01362398  0.5405405 0.9089502
30 {93,TX}  => {7}  0.01362398  0.5128205 1.1186040
31 {18,TX}  => {e0} 0.01294278  0.9047619 1.0162131
32 {18,LAT} => {e0} 0.01089918  0.8000000 0.8985463
33 {9,LAT} => {e0} 0.01089918  0.8421053 0.9458382
```

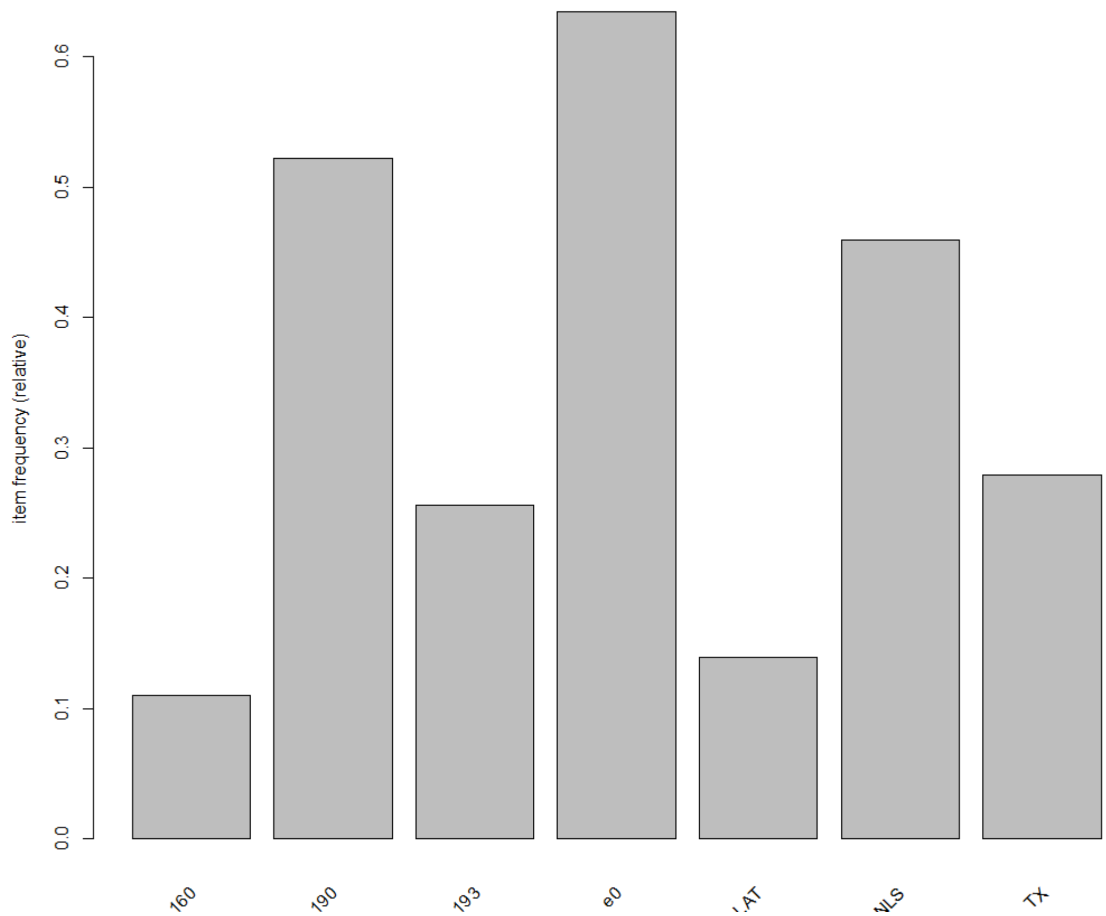
Equipment Failure Instances



```
set of 48 rules

rule length distribution (lhs + rhs):sizes
 2  3
22 26

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   2.000   3.000   2.542   3.000   3.000

summary of quality measures:
    support          confidence          lift
 Min.   :0.01018   Min.   :0.5048   Min.   :1.102
 1st Qu.:0.02328   1st Qu.:0.5839   1st Qu.:1.566
 Median :0.04768   Median :0.7038   Median :1.903
 Mean   :0.05905   Mean   :0.7340   Mean   :2.030
 3rd Qu.:0.07122   3rd Qu.:0.8802   3rd Qu.:2.277
 Max.   :0.25763   Max.   :0.9880   Max.   :4.083

mining info:
 data ntransactions support confidence
   tr          6680    0.01        0.5



1  {111} => {189} 0.25763473  0.9394105 2.276946
2  {189} => {111} 0.25763473  0.6244557 2.276946
3  {TX}  => {188} 0.16107784  0.6310850 1.282912
4  {91}  => {188} 0.10823353  0.5774760 1.173932
5  {LAT} => {189} 0.10404192  0.5622977 1.362899
6  {111,
    SV}  => {189} 0.10389222  0.9612188 2.329805
7  {189,
    SV}  => {111} 0.10389222  0.8685857 3.167114
8  {94}  => {TX}  0.10044910  0.7018828 2.749899
9  {91}  => {SV}  0.09461078  0.5047923 1.888025
10 {NLS} => {188} 0.08997006  0.6247401 1.270013
11 {SEC} => {189} 0.07529940  0.6937931 1.681618
12 {105} => {188} 0.07380240  0.8787879 1.786459
13 {93}  => {188} 0.07035928  0.9090909 1.848061
14 {SEC} => {111} 0.06646707  0.6124138 2.233037
15 {111,
    SEC} => {189} 0.06467066  0.9729730 2.358294
16 {189,
    SEC} => {111} 0.06467066  0.8588469 3.131603
17 {111,
    LAT} => {189} 0.06077844  0.9354839 2.267428
18 {189,
    LAT} => {111} 0.06077844  0.5841727 2.130062
19 {91,
    SV}  => {188} 0.05868263  0.6202532 1.260892
20 {188,
    91}  => {SV}  0.05868263  0.5421853 2.027882
21 {188,
    SV}  => {91}  0.05868263  0.5192053 2.770201
22 {94,
    TX}  => {188} 0.05553892  0.5529061 1.123984
23 {188,
    94}  => {TX}  0.05553892  0.8374718 3.281121
24 {105} => {SV}  0.05074850  0.6042781 2.260122
25 {105,
```

```
    SV}  => {188} 0.04461078  0.8790560 1.787004
26 {105,
    188} => {SV}  0.04461078  0.6044625 2.260811
27 {202} => {91}  0.04416168  0.5139373 2.742093
28 {189,
    94}  => {TX}  0.04146707  0.5831579 2.284748
29 {189,
    TX}  => {94}  0.04146707  0.5843882 4.083382
30 {104} => {188} 0.03502994  0.9360000 1.902763
31 {93,
    TX}  => {188} 0.03488024  0.9357430 1.902241
32 {85}  => {188} 0.02859281  0.7670683 1.559348
33 {FDR} => {188} 0.02500000  0.7660550 1.557288
34 {202,
    91}  => {SV}  0.02440120  0.5525424 2.066620
35 {202,
    SV}  => {91}  0.02440120  0.7056277 3.764851
36 {94,
    LAT} => {189} 0.02365269  0.8777778 2.127560
37 {91,
    TX}  => {188} 0.02215569  0.5421245 1.102067
38 {93,
    LAT} => {188} 0.02110778  0.8703704 1.769347
39 {111,
    NLS} => {189} 0.01511976  0.8145161 1.974226
40 {e0}  => {188} 0.01302395  0.6541353 1.329770
41 {93,
    NLS} => {188} 0.01302395  0.9062500 1.842285
42 {105,
    NLS} => {188} 0.01257485  0.8936170 1.816604
43 {91,
    NLS} => {188} 0.01257485  0.5753425 1.169595
44 {92}  => {188} 0.01227545  0.9879518 2.008374
45 {88}  => {188} 0.01212575  0.7714286 1.568211
46 {85,
    TX}  => {188} 0.01137725  0.8837209 1.796487
47 {81}  => {188} 0.01032934  0.9324324 1.895511
48 {102} => {188} 0.01017964  0.5762712 1.171483
```

Unavoidable Instances




```
set of 39 rules

rule length distribution (lhs + rhs):sizes
 1  2  3
 2 21 16

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   2.000   2.000   2.359   3.000   3.000

summary of quality measures:
    support           confidence          lift
 Min.   :0.01496   Min.   :0.5072   Min.   : 0.8941
 1st Qu.:0.03294   1st Qu.:0.6139   1st Qu.: 1.3279
 Median :0.05697   Median :0.7194   Median : 1.5300
 Mean   :0.13017   Mean   :0.7278   Mean   : 3.2378
 3rd Qu.:0.16932   3rd Qu.:0.8333   3rd Qu.: 3.7460
 Max.   :0.63431   Max.   :0.9600   Max.   :12.3564

mining info:
 data ntransactions support confidence
   tr          3142    0.01        0.5
```

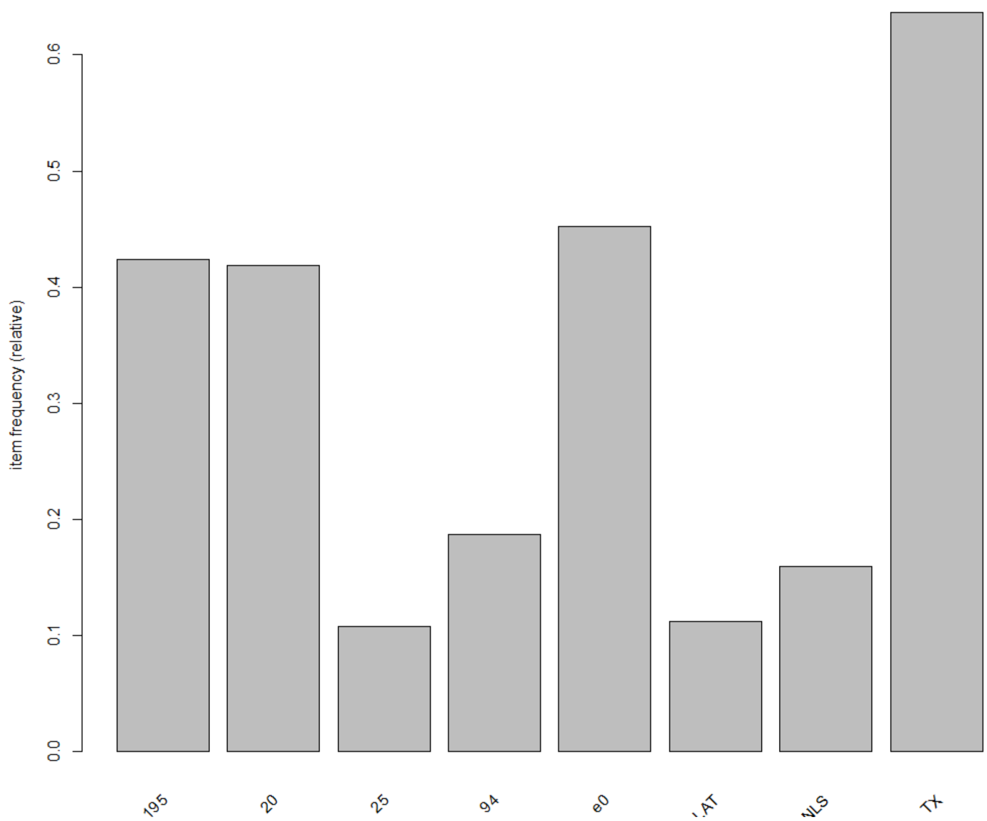
```
   lhs      rhs      support confidence       lift
1  {}    => {e0}  0.63430936  0.6343094  1.0000000
2  {}    => {190} 0.52227880  0.5222788  1.0000000
3  {190} => {e0}  0.45480586  0.8708105  1.3728482
4  {e0}  => {190} 0.45480586  0.7170095  1.3728482
5  {NLS} => {e0}  0.28007638  0.6094183  0.9607588
6  {TX}  => {190} 0.22278803  0.7990868  1.5300004
7  {TX}  => {e0}  0.21451305  0.7694064  1.2129829
8  {190,
    TX}  => {e0}  0.20146404  0.9042857  1.4256225
9  {e0,
    TX}  => {190} 0.20146404  0.9391691  1.7982142
10 {193} => {NLS} 0.19350732  0.7562189  1.6454569
11 {193} => {e0}  0.14513049  0.5671642  0.8941444
12 {193,
    NLS} => {e0}  0.13112667  0.6776316  1.0682983
13 {193,
    e0}  => {NLS} 0.13112667  0.9035088  1.9659450
14 {190,
    NLS} => {e0}  0.12921706  0.8440748  1.3306990
15 {LAT} => {190} 0.11807766  0.8470320  1.6218004
16 {LAT} => {e0}  0.11139402  0.7990868  1.2597745
17 {190,
    LAT} => {e0}  0.10693826  0.9056604  1.4277897
18 {e0,
    LAT} => {190} 0.10693826  0.9600000  1.8380987
19 {40}  => {NLS} 0.06047104  0.6089744  1.3250675
20 {191} => {160} 0.05697008  0.7396694  6.6975254
21 {160} => {191} 0.05697008  0.5158501  6.6975254
22 {160} => {NLS} 0.05601528  0.5072046  1.1036267
23 {81}  => {40}  0.04296626  0.7627119  7.6808996
24 {MTR} => {160} 0.04232973  0.8750000  7.9229107
25 {MTR} => {191} 0.03596435  0.7434211  9.6521857
26 {81}  => {NLS} 0.03405474  0.6045198  1.3153747
27 {191,
    MTR} => {160} 0.03405474  0.9469027  8.5739716
28 {160,
    MTR} => {191} 0.03405474  0.8045113 10.4453489
29 {160,
    191} => {MTR} 0.03405474  0.5977654 12.3564393
30 {160,
    193} => {NLS} 0.03182686  0.7194245  1.5653959
31 {160,
    NLS} => {193} 0.03182686  0.5681818  2.2204319
32 {40,
    81}  => {NLS} 0.02800764  0.6518519  1.4183646
33 {81,
    NLS} => {40}  0.02800764  0.8224299  8.2822909
34 {94}  => {TX}  0.02387015  0.7009346  2.5140827
35 {102} => {NLS} 0.01909612  0.6896552  1.5006209
36 {91}  => {193} 0.01845958  0.6904762  2.6983535
37 {191,
    NLS} => {160} 0.01718651  0.5294118  4.7936938
38 {40,
    e0}  => {NLS} 0.01559516  0.6621622  1.4407988
39 {3}   => {NLS} 0.01495863  0.6184211  1.3456225
```

After Pruning

```
1  {}    => {e0}  0.63430936  0.6343094 1.000000
```

```
2  {}    => {190} 0.52227880  0.5222788 1.000000
3  {193} => {NLS} 0.19350732  0.7562189 1.645457
4  {40}  => {NLS} 0.06047104  0.6089744 1.325067
5  {191} => {160} 0.05697008  0.7396694 6.697525
6  {160} => {NLS} 0.05601528  0.5072046 1.103627
7  {81}  => {40}  0.04296626  0.7627119 7.680900
8  {MTR} => {160} 0.04232973  0.8750000 7.922911
9  {MTR} => {191} 0.03596435  0.7434211 9.652186
10 {81}  => {NLS} 0.03405474  0.6045198 1.315375
11 {94}  => {TX}  0.02387015  0.7009346 2.514083
12 {102} => {NLS} 0.01909612  0.6896552 1.500621
13 {91}  => {193} 0.01845958  0.6904762 2.698353
14 {3}   => {NLS} 0.01495863  0.6184211 1.345623
```

Negligence/Avoidable Instances




```
set of 56 rules

rule length distribution (lhs + rhs):sizes
 1  2  3
 1 28 27

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   2.000   2.000   2.464   3.000   3.000
```

```
summary of quality measures:
    support           confidence          lift
 Min.   :0.01017   Min.   :0.5119   Min.   : 0.9336
 1st Qu.:0.01881   1st Qu.:0.6563   1st Qu.: 1.4071
 Median :0.05654   Median :0.8156   Median : 1.7183
 Mean   :0.08928   Mean   :0.7885   Mean   : 2.1568
 3rd Qu.:0.08639   3rd Qu.:0.9056   3rd Qu.: 2.1395
 Max.   :0.63640   Max.   :0.9947   Max.   :11.5763

mining info:
 data ntransactions support confidence
   tr          5209    0.01        0.5



   lhs       rhs       support confidence       lift
1  {}     => {TX}  0.63639854  0.6363985  1.0000000
2  {195} => {TX}  0.37991937  0.8954751  1.4070980
3  {TX}  => {195} 0.37991937  0.5969834  1.4070980
4  {20}  => {e0}  0.27548474  0.6573523  1.4546083
5  {e0}  => {20}  0.27548474  0.6096007  1.4546083
6  {e0}  => {TX}  0.27222116  0.6023789  0.9465435
7  {94}  => {195} 0.17431369  0.9312821  2.1950444
8  {94}  => {TX}  0.16913035  0.9035897  1.4198489
9  {195,
    94}  => {TX}  0.15972356  0.9162996  1.4398203
10 {94,
    TX}  => {195} 0.15972356  0.9443814  2.2259197
11 {20,
    TX}  => {e0}  0.13476675  0.8162791  1.8062862
12 {NLS} => {20}  0.09771549  0.6125150  1.4615625
13 {25}  => {e0}  0.09541179  0.8875000  1.9638859
14 {NLS} => {e0}  0.08754079  0.5487365  1.2142601
15 {25}  => {TX}  0.08600499  0.8000000  1.2570739
16 {LAT} => {20}  0.08043770  0.7186964  1.7149288
17 {25,
    e0}  => {TX}  0.07794202  0.8169014  1.2836318
18 {25,
    TX}  => {e0}  0.07794202  0.9062500  2.0053765
19 {81}  => {195} 0.07775005  0.9759036  2.3002181
20 {LAT} => {e0}  0.07679017  0.6861063  1.5182362
21 {81}  => {TX}  0.07218276  0.9060241  1.4236741
22 {195,
    81}  => {TX}  0.07160683  0.9209877  1.4471869
23 {81,
    TX}  => {195} 0.07160683  0.9920213  2.3382076
24 {20,
    NLS} => {e0}  0.06623152  0.6777996  1.4998548
25 {e0,
    NLS} => {20}  0.06623152  0.7565789  1.8053228
26 {20,
    LAT} => {e0}  0.06258399  0.7780430  1.7216762
27 {e0,
    LAT} => {20}  0.06258399  0.8150000  1.9447251
28 {105} => {20}  0.05663275  0.9021407  2.1526572
29 {195,
    e0}  => {TX}  0.05644078  0.8855422  1.3914899
30 {SV}  => {20}  0.05279324  0.8435583  2.0128699
31 {91}  => {195} 0.04876176  0.6529563  1.5390269
32 {91}  => {TX}  0.04645805  0.6221080  0.9775446
33 {195,
    91}  => {TX}  0.04415435  0.9055118  1.4228691
34 {91,
```

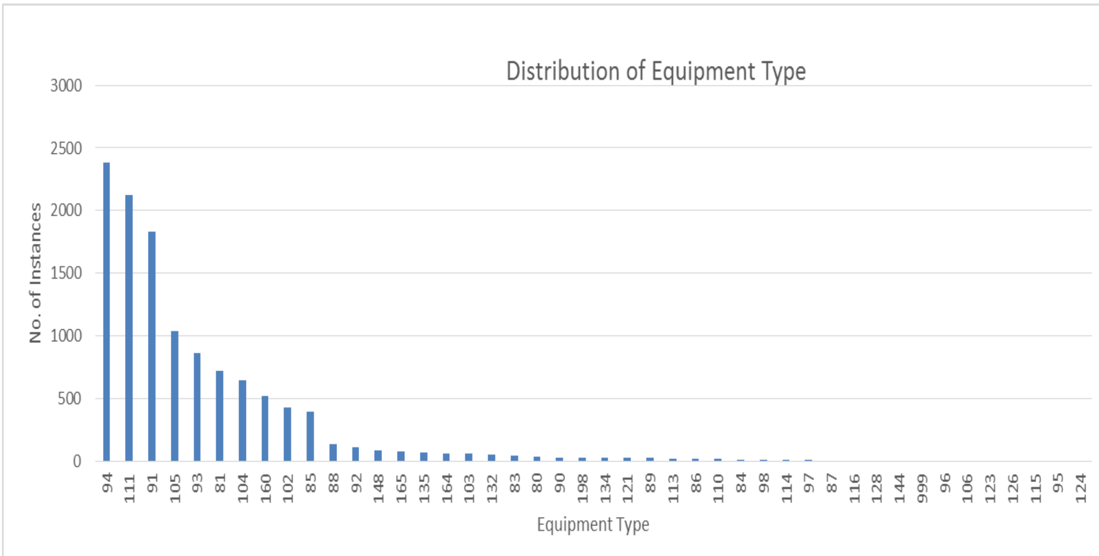
```
    TX}  => {195} 0.04415435  0.9504132  2.2401369
35 {104} => {20}  0.03589940  0.9946809  2.3734735
36 {105,
    SV}  => {20}  0.02898829  0.9679487  2.3096862
37 {20,
    SV}  => {105} 0.02898829  0.5490909  8.7468335
38 {105,
    20}  => {SV}  0.02898829  0.5118644  8.1788396
39 {111} => {195} 0.02169322  0.6647059  1.5667208
40 {SEC} => {20}  0.02130927  0.8604651  2.0532125
41 {111} => {TX}  0.01938952  0.5941176  0.9335622
42 {102} => {195} 0.01881359  0.7313433  1.7237860
43 {111,
    195} => {TX}  0.01881359  0.8672566  1.3627571
44 {111,
    TX}  => {195} 0.01881359  0.9702970  2.2870033
45 {102} => {TX}  0.01631791  0.6343284  0.9967470
46 {93}  => {TX}  0.01516606  0.6320000  0.9930884
47 {102,
    195} => {TX}  0.01478211  0.7857143  1.2346262
48 {102,
    TX}  => {195} 0.01478211  0.9058824  2.1351770
49 {105,
    NLS} => {20}  0.01459013  0.8837209  2.1087047
50 {178} => {NLS} 0.01401421  0.9012346  5.6492550
51 {91,
    SV}  => {20}  0.01363026  0.7977528  1.9035705
52 {20,
    91}  => {SV}  0.01363026  0.7244898 11.5762802
53 {93}  => {20}  0.01247840  0.5200000  1.2408062
54 {94,
    NLS} => {195} 0.01094260  0.7808219  1.8404079
55 {104,
    TX}  => {20}  0.01017470  0.9814815  2.3419776
56 {25,
    LAT} => {e0}  0.01017470  0.8281250  1.8324992
```

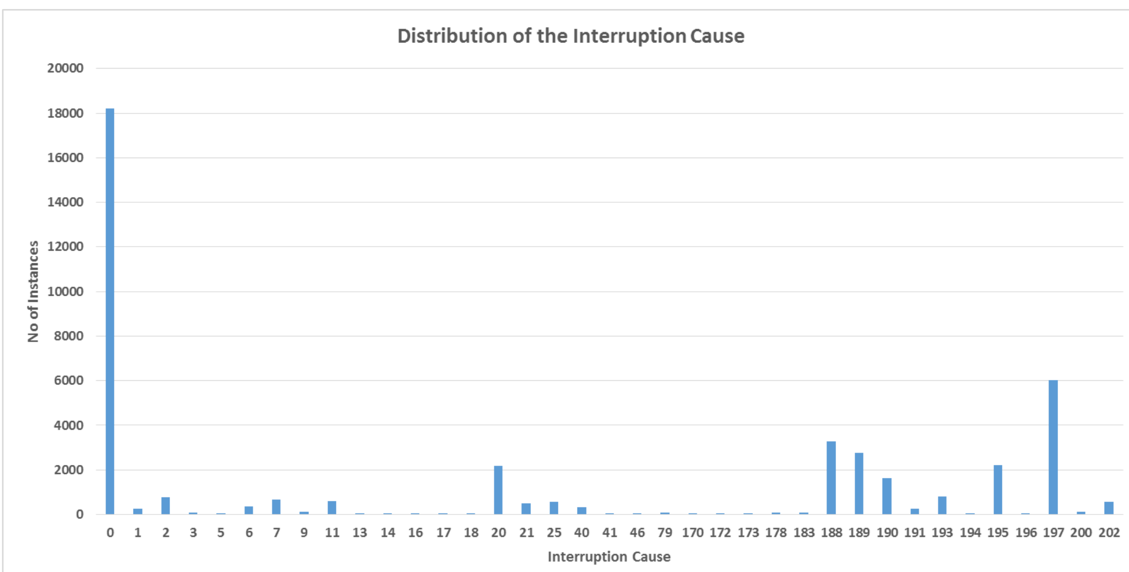
After Pruning

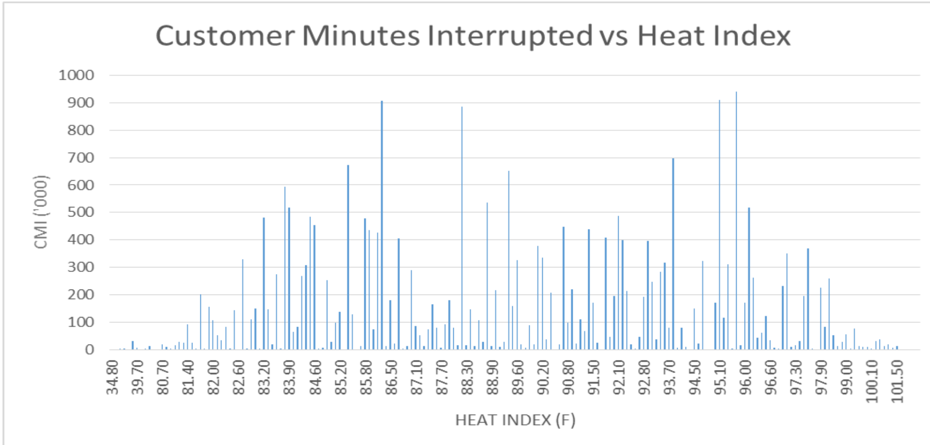
```
   lhs      rhs      support confidence     lift
1  {}    => {TX}  0.63639854  0.6363985 1.000000
2  {20}  => {e0}  0.27548474  0.6573523 1.454608
3  {94}  => {195} 0.17431369  0.9312821 2.195044
4  {NLS} => {20}  0.09771549  0.6125150 1.461562
5  {25}  => {e0}  0.09541179  0.8875000 1.963886
6  {NLS} => {e0}  0.08754079  0.5487365 1.214260
7  {LAT} => {20}  0.08043770  0.7186964 1.714929
8  {81}  => {195} 0.07775005  0.9759036 2.300218
9  {LAT} => {e0}  0.07679017  0.6861063 1.518236
10 {105} => {20}  0.05663275  0.9021407 2.152657
11 {SV}  => {20}  0.05279324  0.8435583 2.012870
12 {91}  => {195} 0.04876176  0.6529563 1.539027
13 {104} => {20}  0.03589940  0.9946809 2.373473
14 {111} => {195} 0.02169322  0.6647059 1.566721
15 {SEC} => {20}  0.02130927  0.8604651 2.053212
16 {102} => {195} 0.01881359  0.7313433 1.723786
17 {178} => {NLS} 0.01401421  0.9012346 5.649255
18 {93}  => {20}  0.01247840  0.5200000 1.240806
```

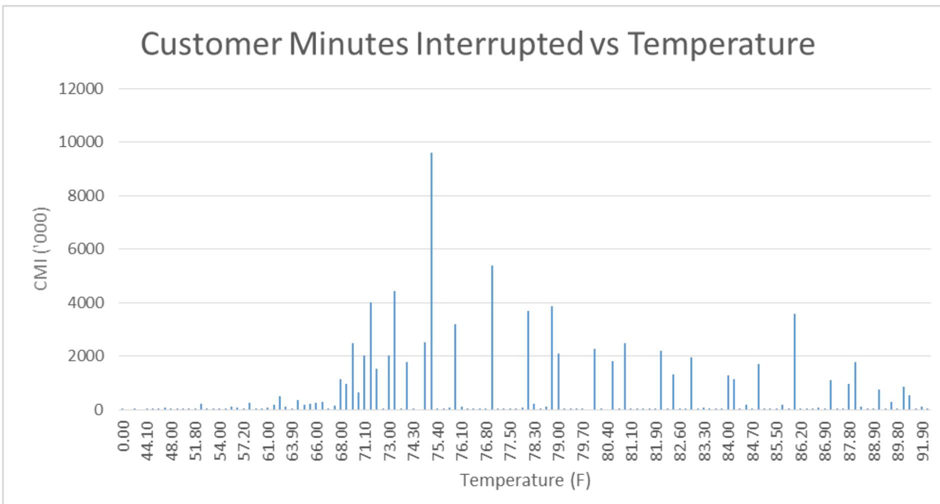
APPENDIX D

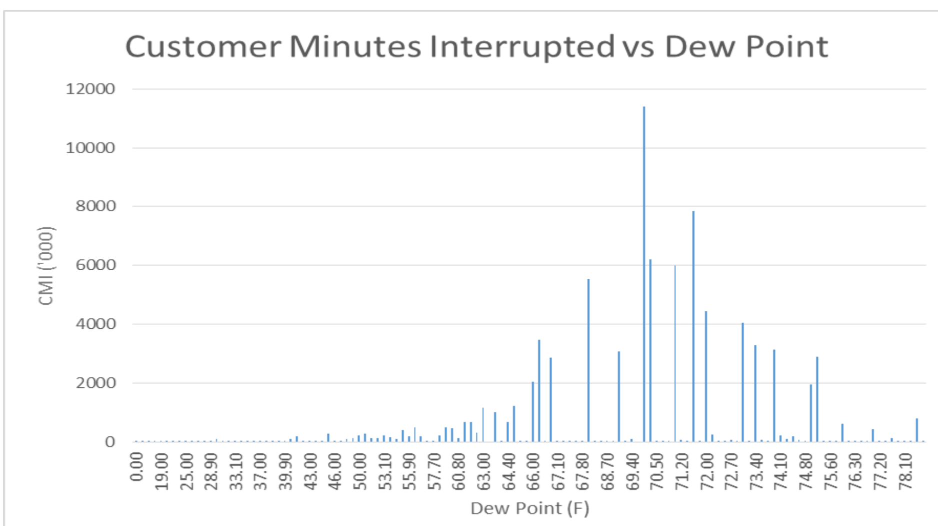
Exploratory Analysis Graphs

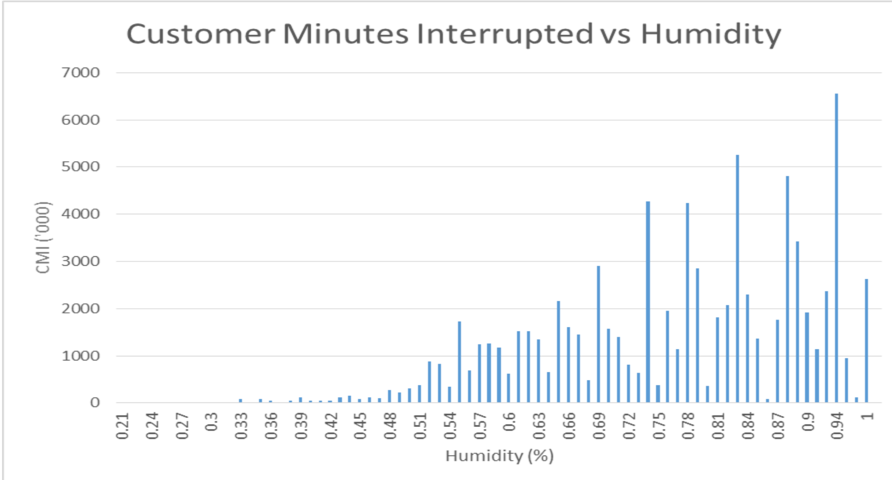


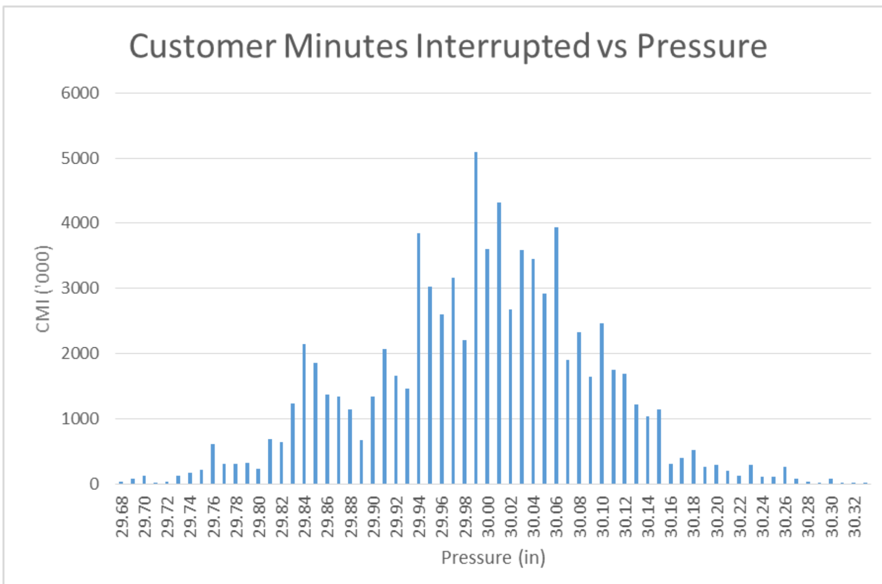


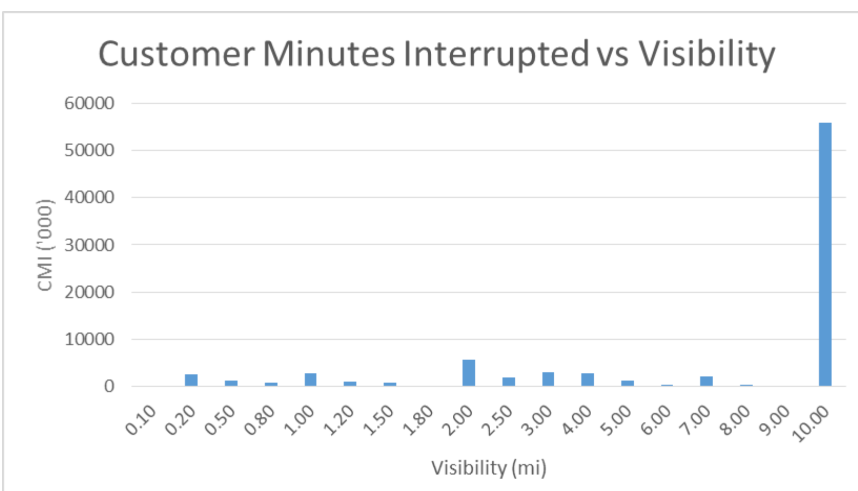

Customer Minutes Interrupted vs Heat Index
CMI ('000)
1000
900
800
700
600
500
400
300
200
100
0
34.80
39.70
80.70
81.40
82.00
82.60
83.20
83.90
84.60
85.20
85.80
86.50
87.10
87.70
88.30
88.90
89.60
90.20
90.80
91.50
92.10
92.80
93.70
94.50
95.10
96.00
96.60
97.30
97.90
99.00
100.10
101.50
HEAT INDEX (F)

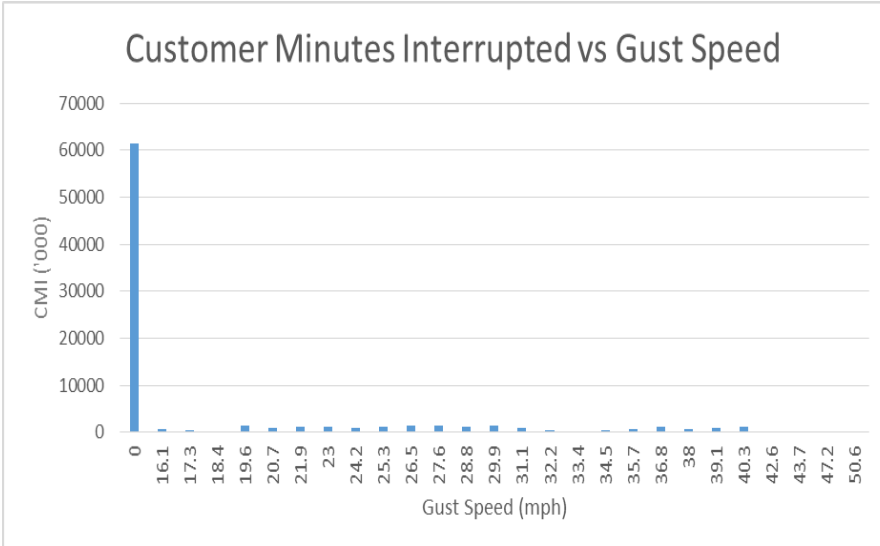

Customer Minutes Interrupted vs Temperature
CMI ('000)
12000
10000
8000
6000
4000
2000
0
0.00
44.10
48.00
51.80
54.00
57.20
61.00
63.90
66.00
68.00
71.10
73.00
74.30
75.40
76.10
76.80
77.50
78.30
79.00
79.70
80.40
81.10
81.90
82.60
83.30
84.00
84.70
85.50
86.20
86.90
87.80
88.90
89.80
91.90
Temperature (F)

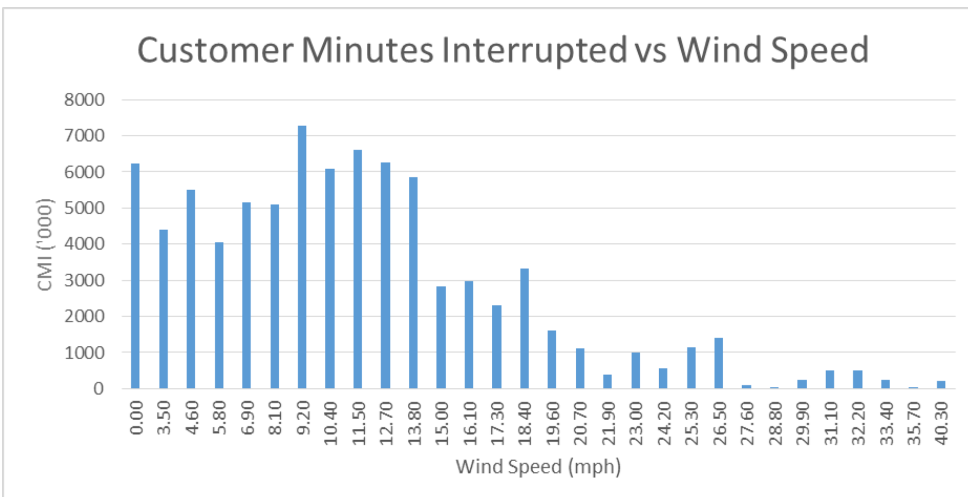

Customer Minutes Interrupted vs Dew Point
CMI ('000)
12000
10000
8000
6000
4000
2000
0
0.00
19.00
25.00
28.90
33.10
37.00
39.90
43.00
46.00
50.00
53.10
55.90
57.70
60.80
63.00
64.40
66.00
67.10
67.80
68.70
69.40
70.50
71.20
72.00
72.70
73.40
74.10
74.80
75.60
76.30
77.20
78.10
Dew Point (F)

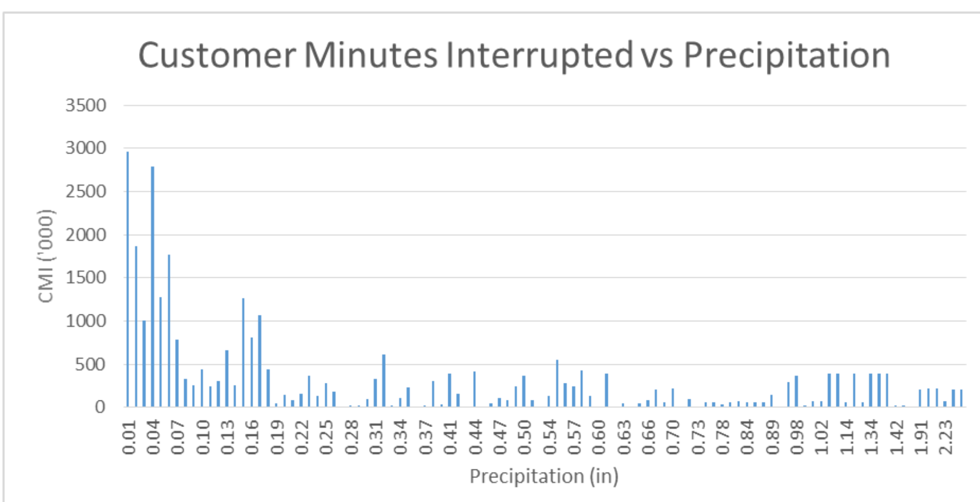

Customer Minutes Interrupted vs Humidity
CMI ('000)
Humidity (%)

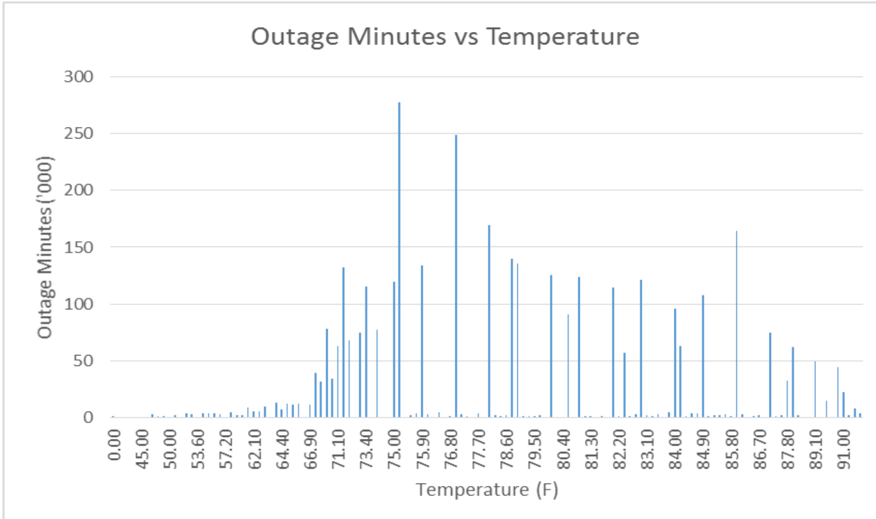

Customer Minutes Interrupted vs Pressure
CMI ('000)
Pressure (in)

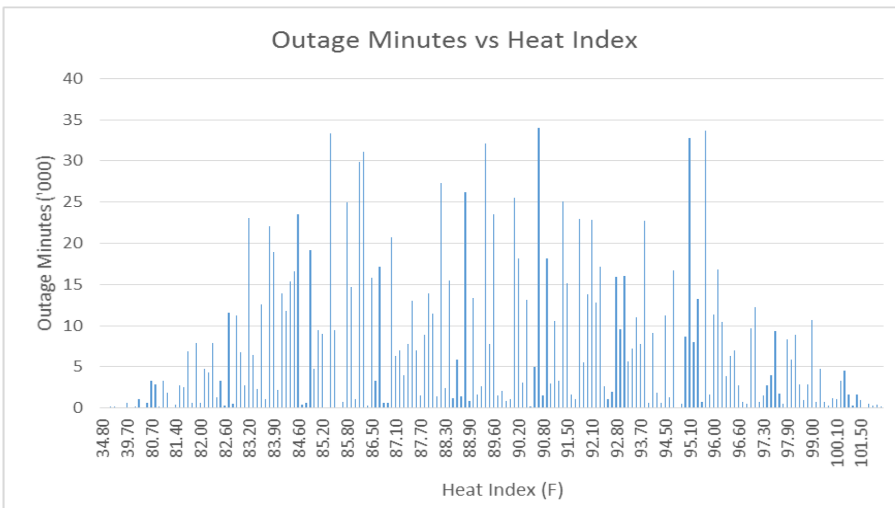

Customer Minutes Interrupted vs Visibility
CMI ('000)
Visibility (mi)

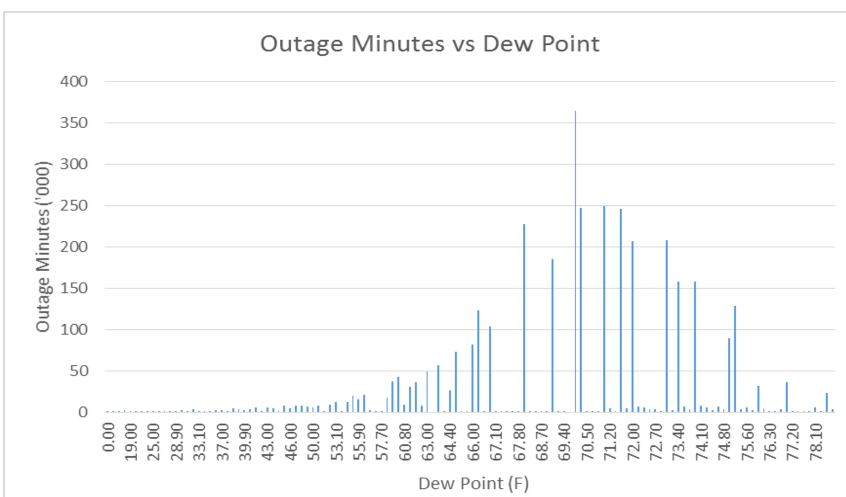

Customer Minutes Interrupted vs Gust Speed
CMI ('000)
70000
60000
50000
40000
30000
20000
10000
0
0 16.1 17.3 18.4 19.6 20.7 21.9 23 24.2 25.3 26.5 27.6 28.8 29.9 31.1 32.2 33.4 34.5 35.7 36.8 38 39.1 40.3 42.6 43.7 47.2 50.6
Gust Speed (mph)

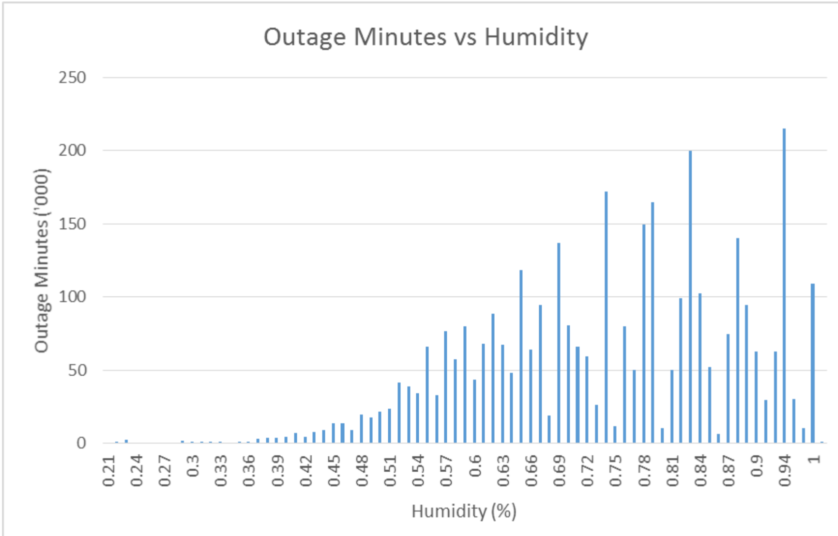

Customer Minutes Interrupted vs Wind Speed
CMI ('000)
8000
7000
6000
5000
4000
3000
2000
1000
0
0.00 3.50 4.60 5.80 6.90 8.10 9.20 10.40 11.50 12.70 13.80 15.00 16.10 17.30 18.40 19.60 20.70 21.90 23.00 24.20 25.30 26.50 27.60 28.80 29.90 31.10 32.20 33.40 35.70 40.30
Wind Speed (mph)

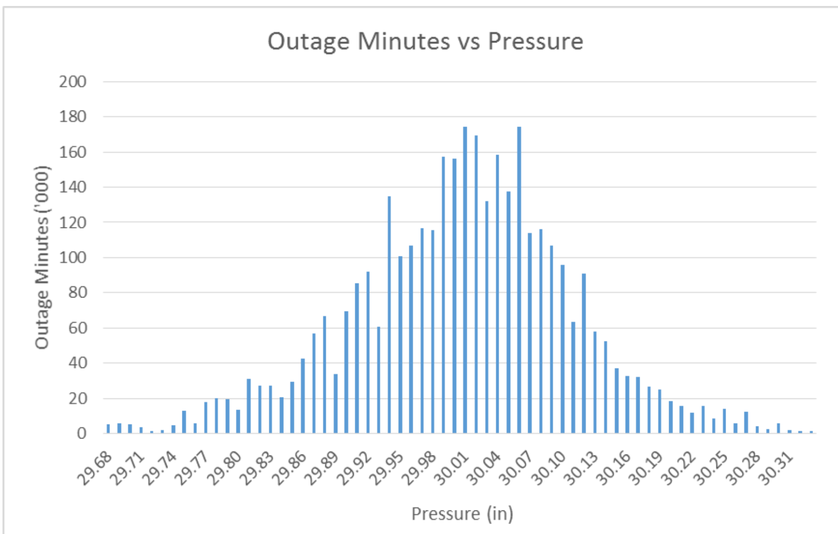

Customer Minutes Interrupted vs Precipitation
CMI ('000)
3500
3000
2500
2000
1500
1000
500
0
0.01 0.04 0.07 0.10 0.13 0.16 0.19 0.22 0.25 0.28 0.31 0.34 0.37 0.41 0.44 0.47 0.50 0.54 0.57 0.60 0.63 0.66 0.70 0.73 0.78 0.84 0.89 0.98 1.02 1.14 1.34 1.42 1.91 2.23
Precipitation (in)

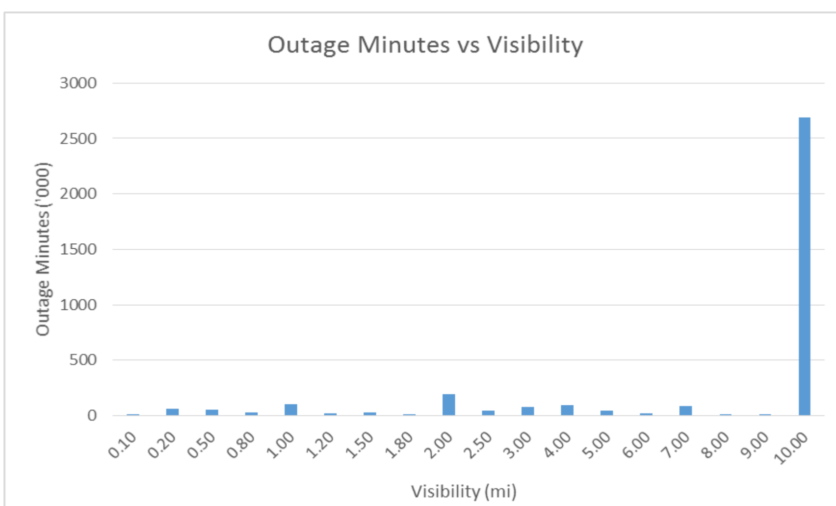

Outage Minutes vs Temperature
Outage Minutes ('000)
Temperature (F)

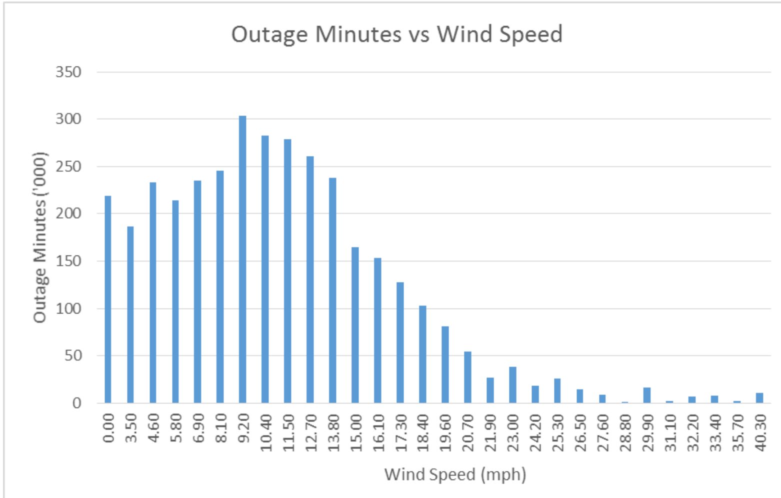

Outage Minutes vs Heat Index
Outage Minutes ('000)
Heat Index (F)

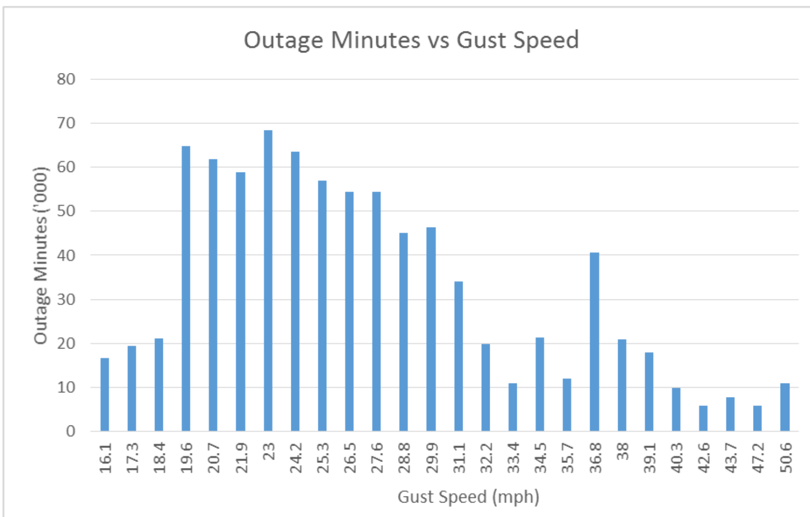

Outage Minutes vs Dew Point
Outage Minutes ('000)
Dew Point (F)

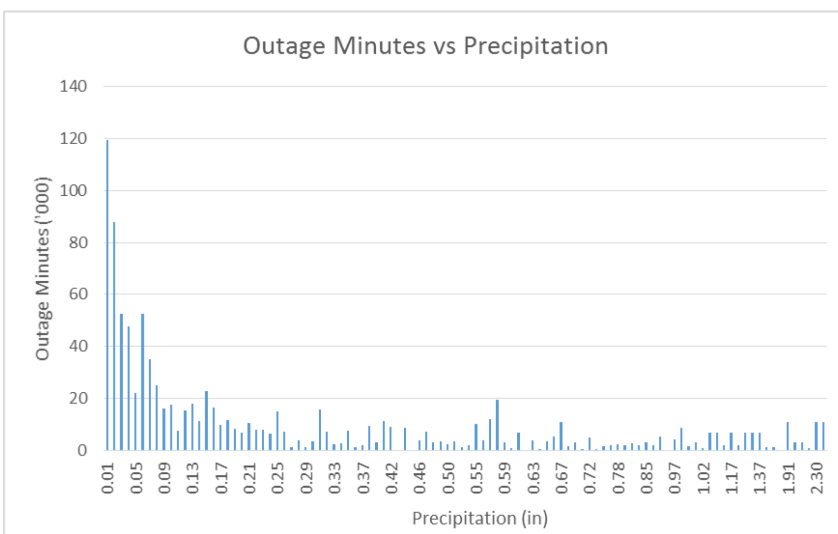

Outage Minutes vs Humidity
Outage Minutes ('000)
250
200
150
100
50
0
0.21
0.24
0.27
0.3
0.33
0.36
0.39
0.42
0.45
0.48
0.51
0.54
0.57
0.6
0.63
0.66
0.69
0.72
0.75
0.78
0.81
0.84
0.87
0.9
0.94
1
Humidity (%)

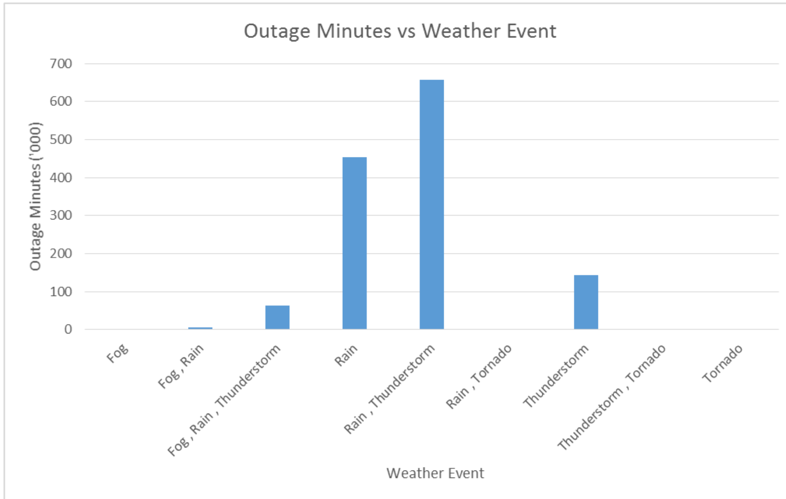

Outage Minutes vs Pressure
Outage Minutes ('000)
200
180
160
140
120
100
80
60
40
20
0
29.68
29.71
29.74
29.77
29.80
29.83
29.86
29.89
29.92
29.95
29.98
30.01
30.04
30.07
30.10
30.13
30.16
30.19
30.22
30.25
30.28
30.31
Pressure (in)

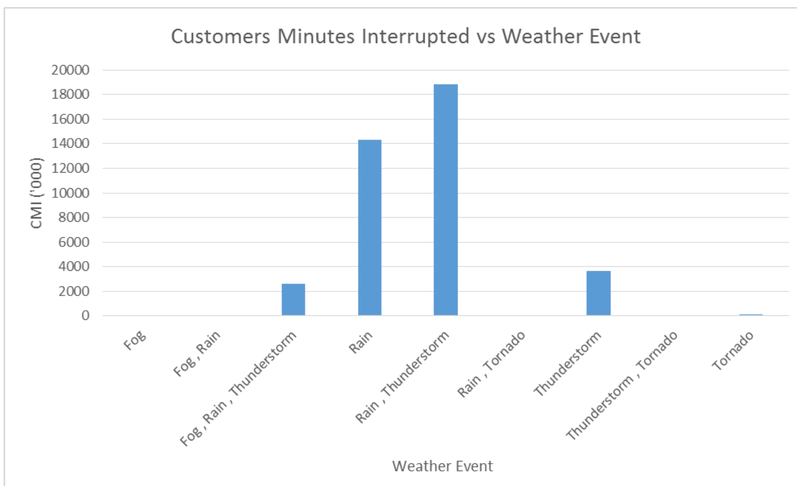

Outage Minutes vs Visibility
Outage Minutes ('000)
3000
2500
2000
1500
1000
500
0
0.10
0.20
0.50
0.80
1.00
1.20
1.50
1.80
2.00
2.50
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
Visibility (mi)

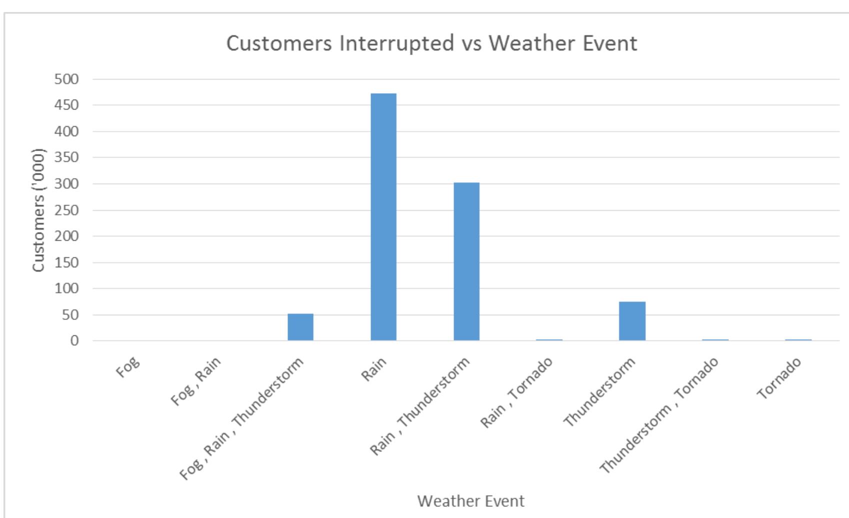

Outage Minutes vs Wind Speed
Outage Minutes ('000)
350
300
250
200
150
100
50
0
0.00 3.50 4.60 5.80 6.90 8.10 9.20 10.40 11.50 12.70 13.80 15.00 16.10 17.30 18.40 19.60 20.70 21.90 23.00 24.20 25.30 26.50 27.60 28.80 29.90 31.10 32.20 33.40 35.70 40.30
Wind Speed (mph)

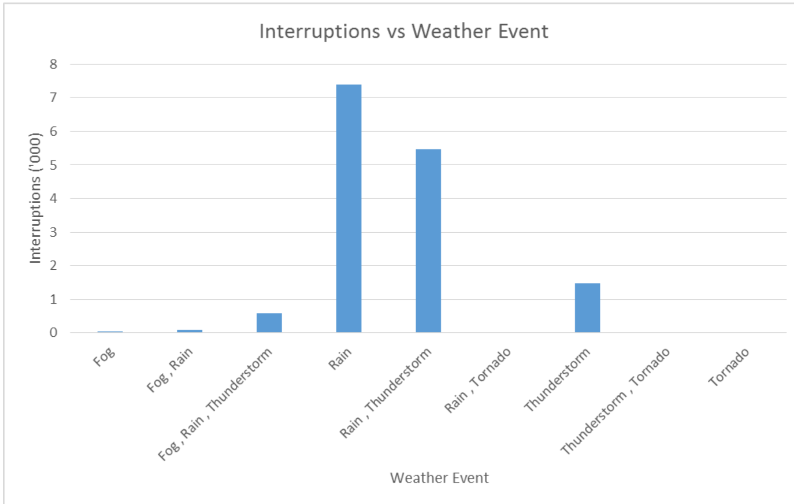

Outage Minutes vs Gust Speed
Outage Minutes ('000)
80
70
60
50
40
30
20
10
0
16.1 17.3 18.4 19.6 20.7 21.9 23 24.2 25.3 26.5 27.6 28.8 29.9 31.1 32.2 33.4 34.5 35.7 36.8 38 39.1 40.3 42.6 43.7 47.2 50.6
Gust Speed (mph)

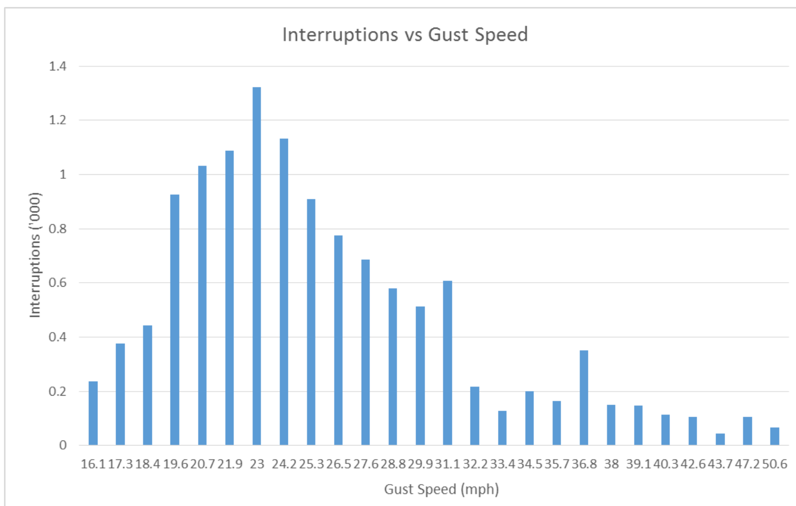

Outage Minutes vs Precipitation
Outage Minutes ('000)
140
120
100
80
60
40
20
0
0.01 0.05 0.09 0.13 0.17 0.21 0.25 0.29 0.33 0.37 0.42 0.46 0.50 0.55 0.59 0.63 0.67 0.72 0.78 0.85 0.97 1.02 1.17 1.37 1.91 2.30
Precipitation (in)

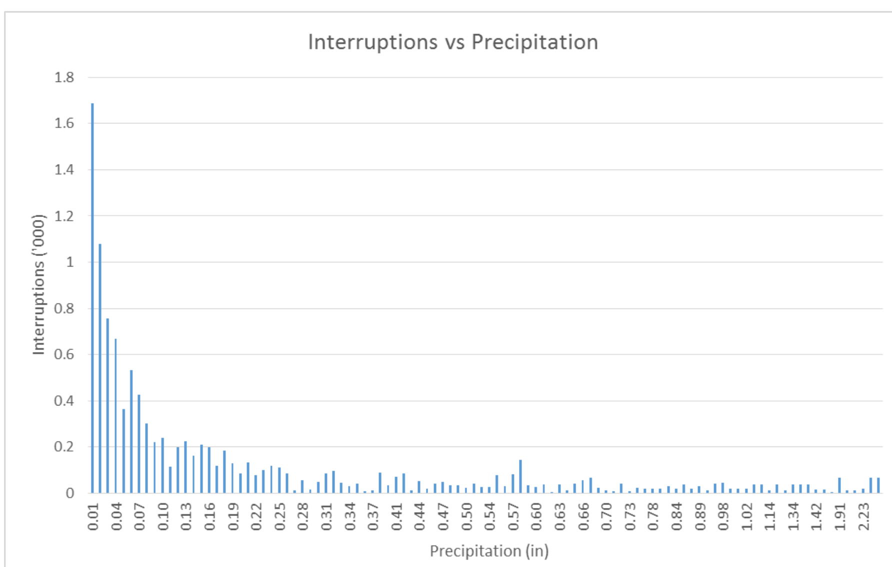

Outage Minutes vs Weather Event
Outage Minutes ('000)
700
600
500
400
300
200
100
0
Fog
Fog , Rain
Fog , Rain , Thunderstorm
Rain
Rain , Thunderstorm
Rain , Tornado
Thunderstorm
Thunderstorm , Tornado
Tornado
Weather Event

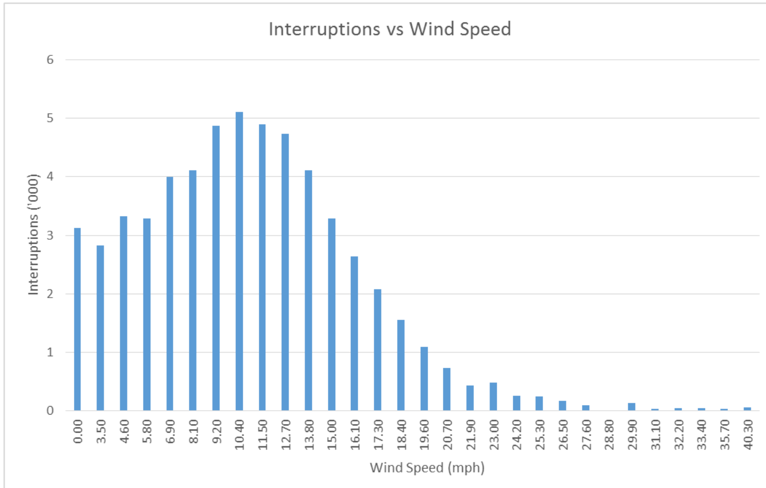

Customers Minutes Interrupted vs Weather Event
CMI ('000)
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Fog
Fog , Rain
Fog , Rain , Thunderstorm
Rain
Rain , Thunderstorm
Rain , Tornado
Thunderstorm
Thunderstorm , Tornado
Tornado
Weather Event

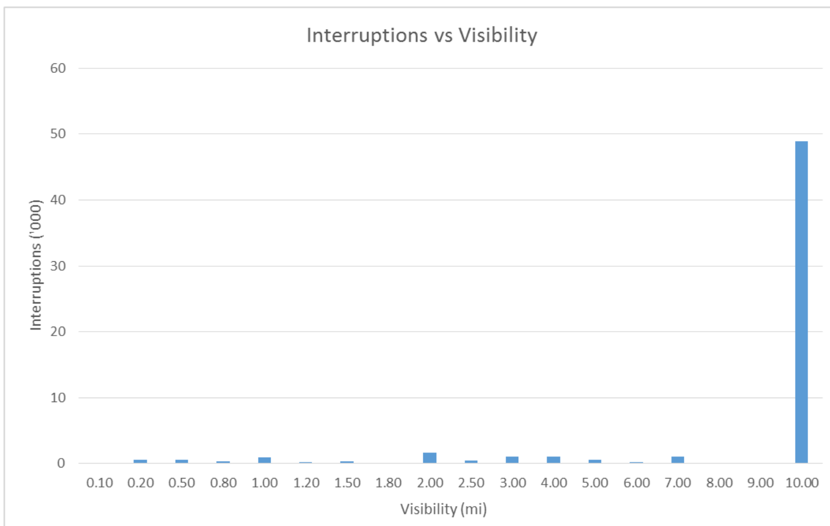

Customers Interrupted vs Weather Event
Customers ('000)
500
450
400
350
300
250
200
150
100
50
0
Fog
Fog , Rain
Fog , Rain , Thunderstorm
Rain
Rain , Thunderstorm
Rain , Tornado
Thunderstorm
Thunderstorm , Tornado
Tornado
Weather Event

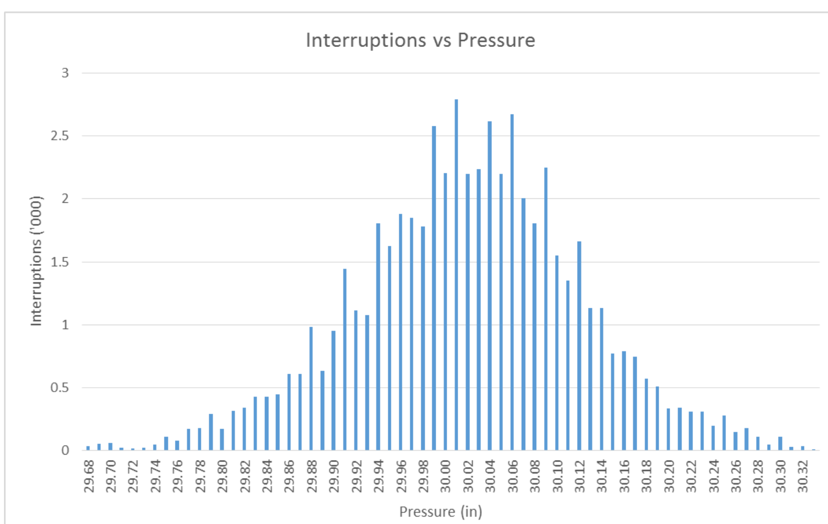

Interruptions vs Weather Event
Interruptions ('000)
8
7
6
5
4
3
2
1
0
Fog
Fog , Rain
Fog , Rain , Thunderstorm
Rain
Rain , Thunderstorm
Rain , Tornado
Thunderstorm
Thunderstorm , Tornado
Tornado
Weather Event

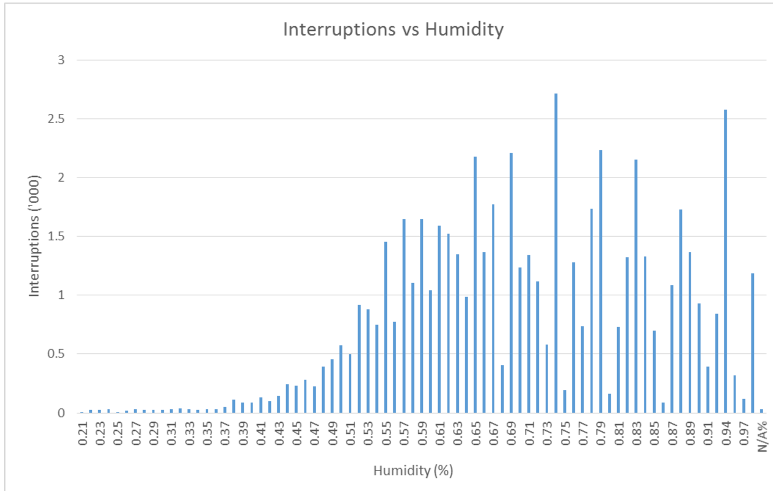

Interruptions vs Gust Speed
Interruptions ('000)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
16.1 17.3 18.4 19.6 20.7 21.9 23 24.2 25.3 26.5 27.6 28.8 29.9 31.1 32.2 33.4 34.5 35.7 36.8 38 39.1 40.3 42.6 43.7 47.2 50.6
Gust Speed (mph)

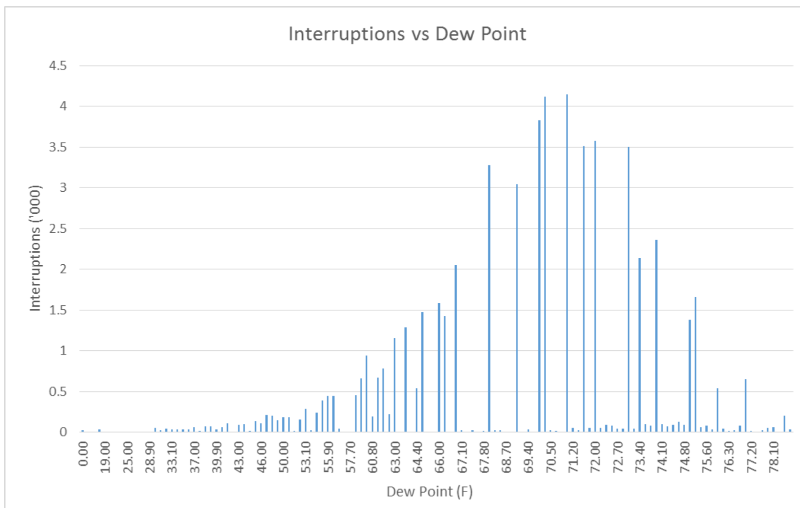

Interruptions vs Precipitation
Interruptions ('000)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.01 0.04 0.07 0.10 0.13 0.16 0.19 0.22 0.25 0.28 0.31 0.34 0.37 0.41 0.44 0.47 0.50 0.54 0.57 0.60 0.63 0.66 0.70 0.73 0.78 0.84 0.89 0.98 1.02 1.14 1.34 1.42 1.91 2.23
Precipitation (in)

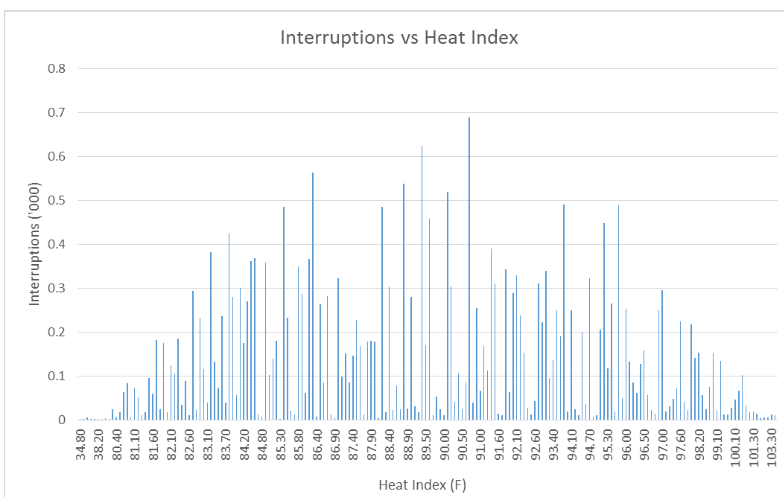

Interruptions vs Wind Speed
Interruptions ('000)
0
1
2
3
4
5
6
0.00
3.50
4.60
5.80
6.90
8.10
9.20
10.40
11.50
12.70
13.80
15.00
16.10
17.30
18.40
19.60
20.70
21.90
23.00
24.20
25.30
26.50
27.60
28.80
29.90
31.10
32.20
33.40
35.70
40.30
Wind Speed (mph)

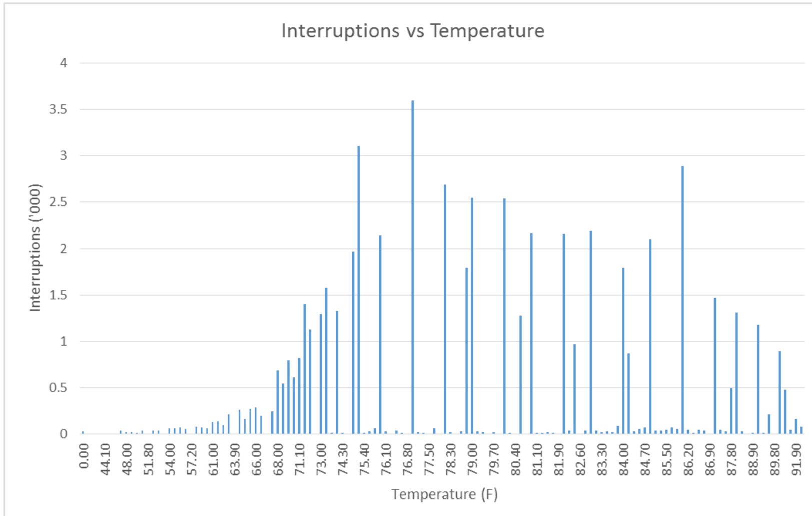

Interruptions vs Visibility
Interruptions ('000)
0
10
20
30
40
50
60
0.10
0.20
0.50
0.80
1.00
1.20
1.50
1.80
2.00
2.50
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
Visibility (mi)

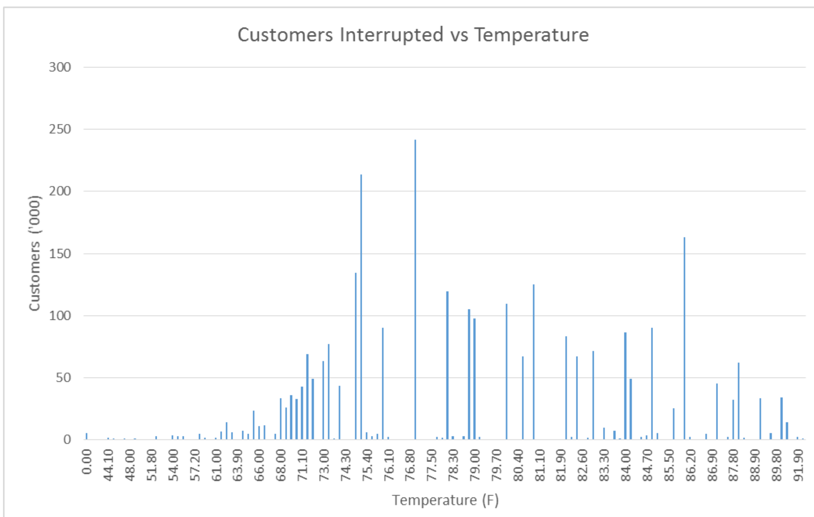

Interruptions vs Pressure
Interruptions ('000)
0
0.5
1
1.5
2
2.5
3
29.68
29.70
29.72
29.74
29.76
29.78
29.80
29.82
29.84
29.86
29.88
29.90
29.92
29.94
29.96
29.98
30.00
30.02
30.04
30.06
30.08
30.10
30.12
30.14
30.16
30.18
30.20
30.22
30.24
30.26
30.28
30.30
30.32
Pressure (in)

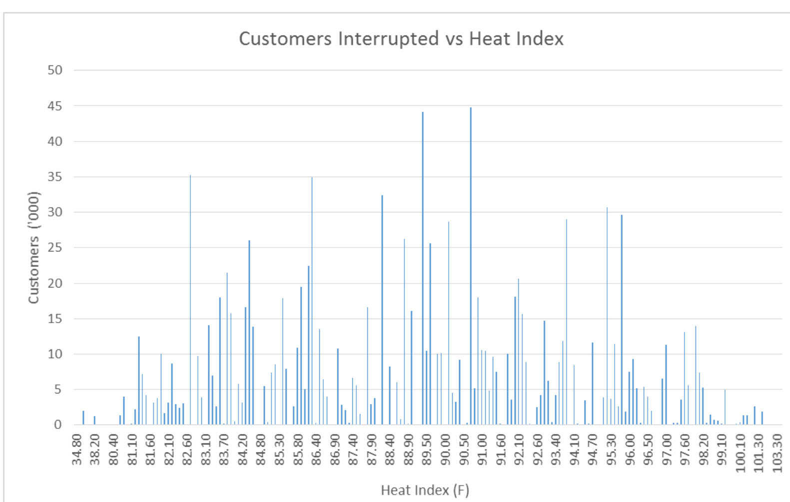

Interruptions vs Humidity
Interruptions ('000)
Humidity (%)

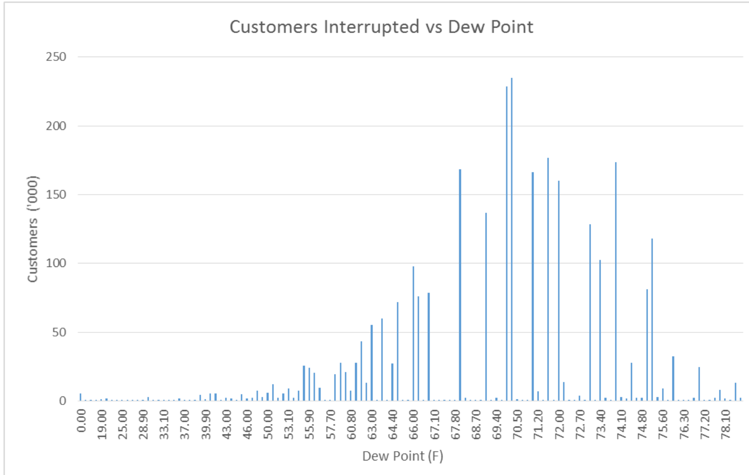

Interruptions vs Dew Point
Interruptions ('000)
Dew Point (F)

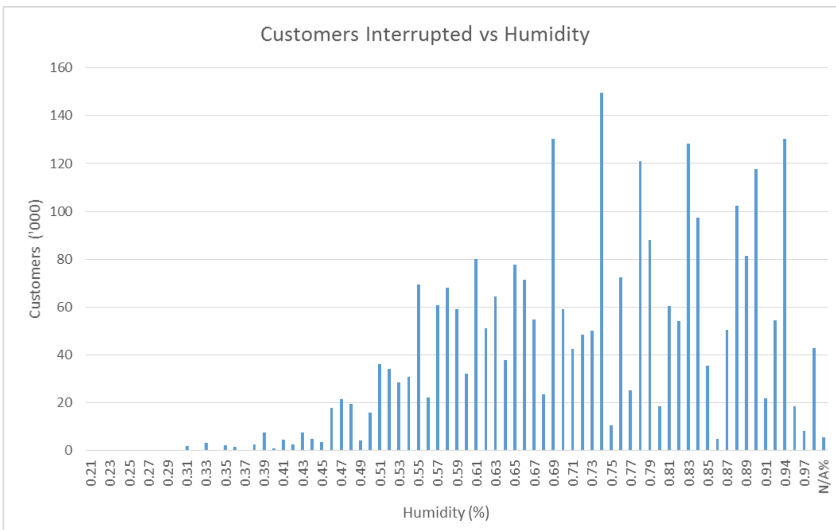

Interruptions vs Heat Index
Interruptions ('000)
Heat Index (F)

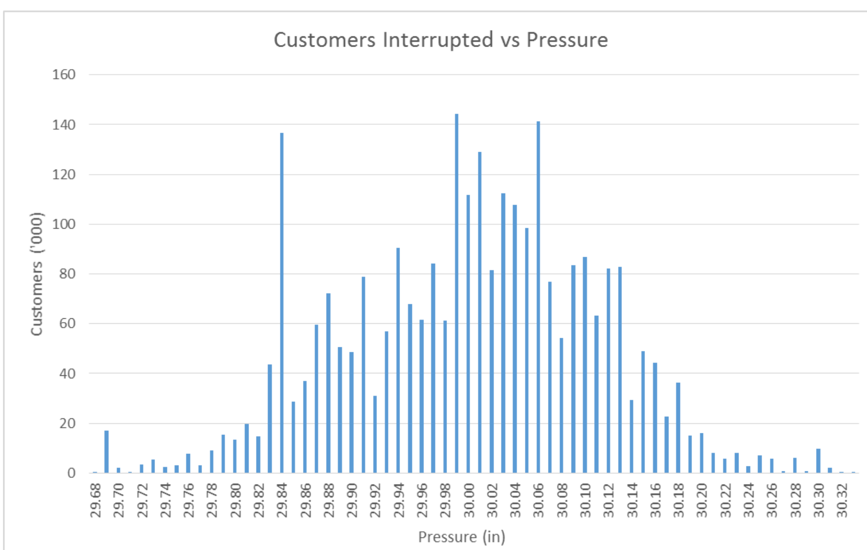

Interruptions vs Temperature
Interruptions ('000)
Temperature (F)

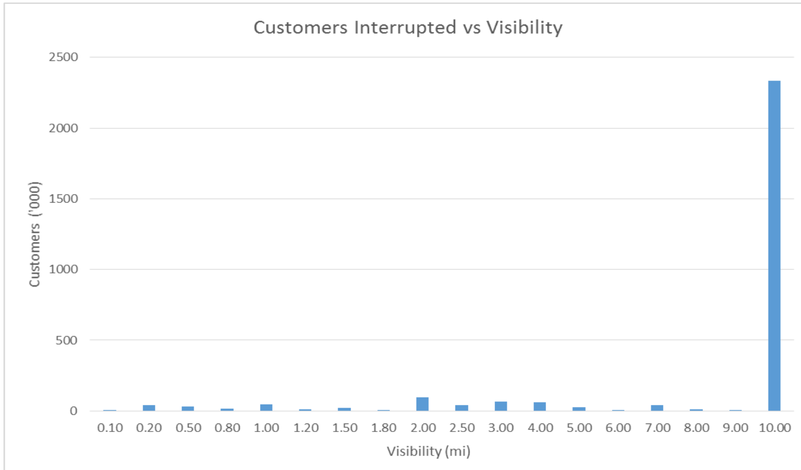

Customers Interrupted vs Temperature
Customers ('000)
Temperature (F)

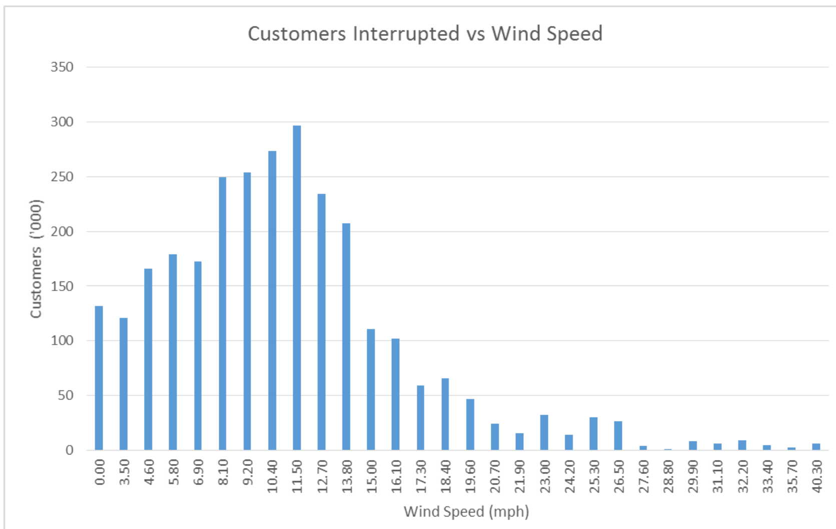

Customers Interrupted vs Heat Index
Customers ('000)
Heat Index (F)

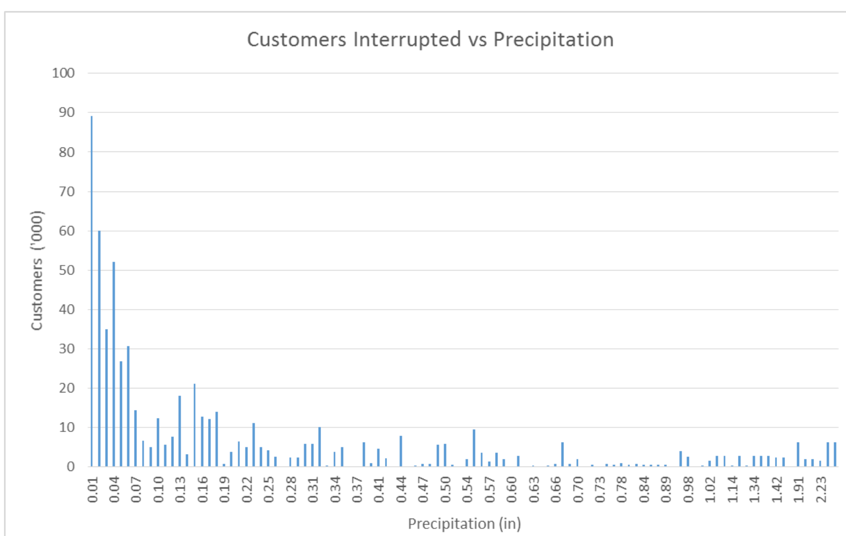

Customers Interrupted vs Dew Point
Customers ('000)
Dew Point (F)

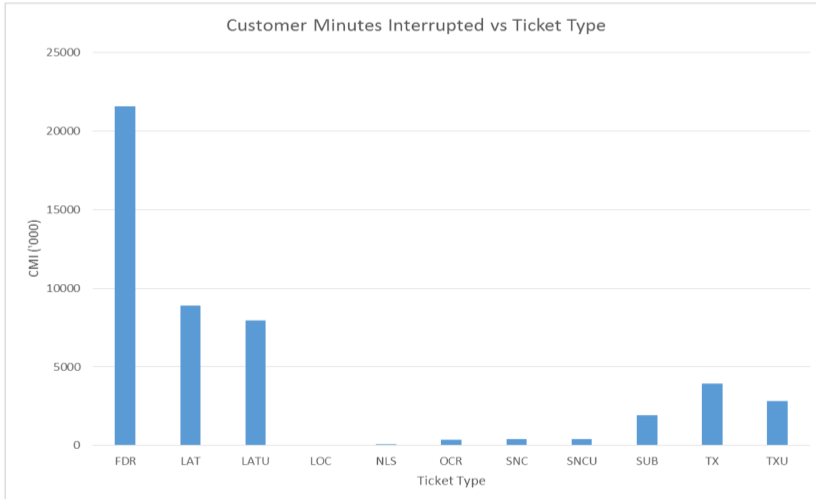

Customers Interrupted vs Humidity
Customers ('000)
Humidity (%)

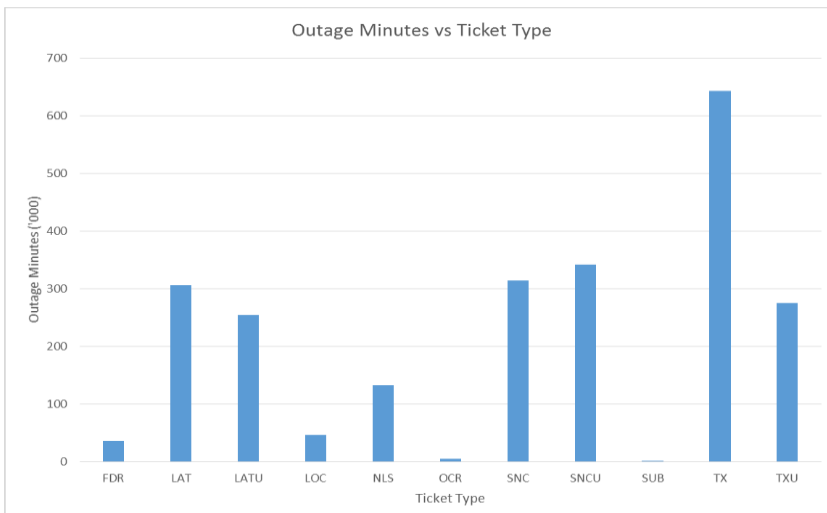

Customers Interrupted vs Pressure
Customers ('000)
Pressure (in)

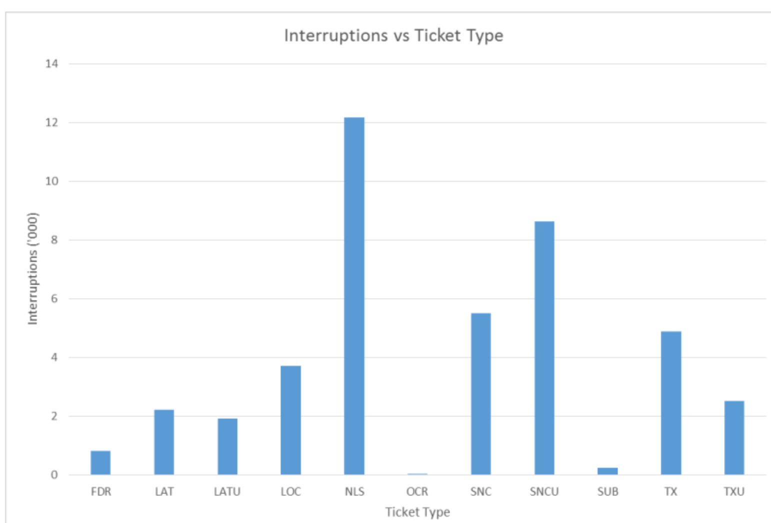

Customers Interrupted vs Visibility
Customers ('000)
2500
2000
1500
1000
500
0
0.10 0.20 0.50 0.80 1.00 1.20 1.50 1.80 2.00 2.50 3.00 4.00 5.00 6.00 7.00 8.00 9.00 10.00
Visibility (mi)

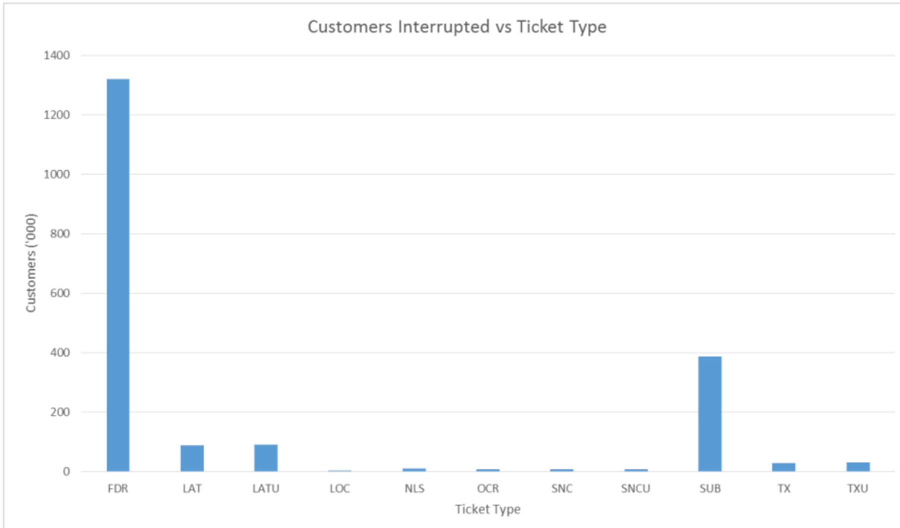

Customers Interrupted vs Wind Speed
Customers ('000)
350
300
250
200
150
100
50
0
0.00 3.50 4.60 5.80 6.90 8.10 9.20 10.40 11.50 12.70 13.80 15.00 16.10 17.30 18.40 19.60 20.70 21.90 23.00 24.20 25.30 26.50 27.60 28.80 29.90 31.10 32.20 33.40 35.70 40.30
Wind Speed (mph)

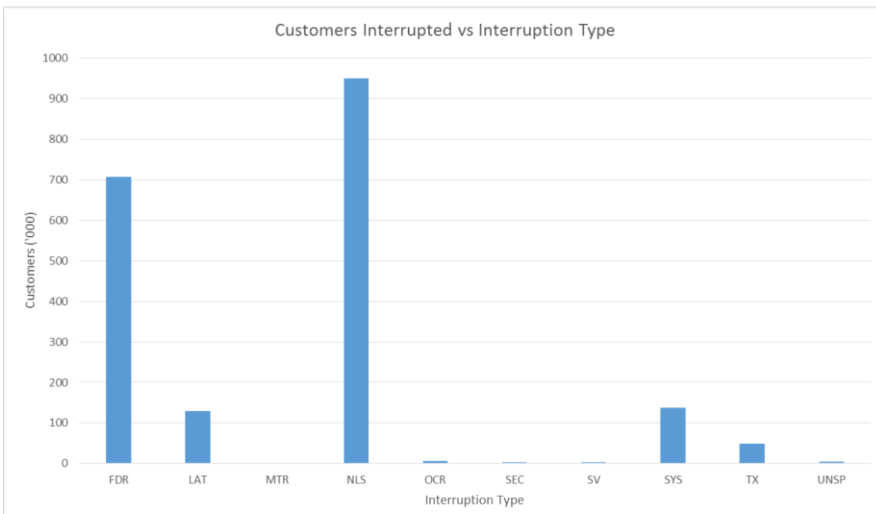

Customers Interrupted vs Precipitation
Customers ('000)
100
90
80
70
60
50
40
30
20
10
0
0.01 0.04 0.07 0.10 0.13 0.16 0.19 0.22 0.25 0.28 0.31 0.34 0.37 0.41 0.44 0.47 0.50 0.54 0.57 0.60 0.63 0.66 0.70 0.73 0.78 0.84 0.89 0.98 1.02 1.14 1.34 1.42 1.91 2.23
Precipitation (in)

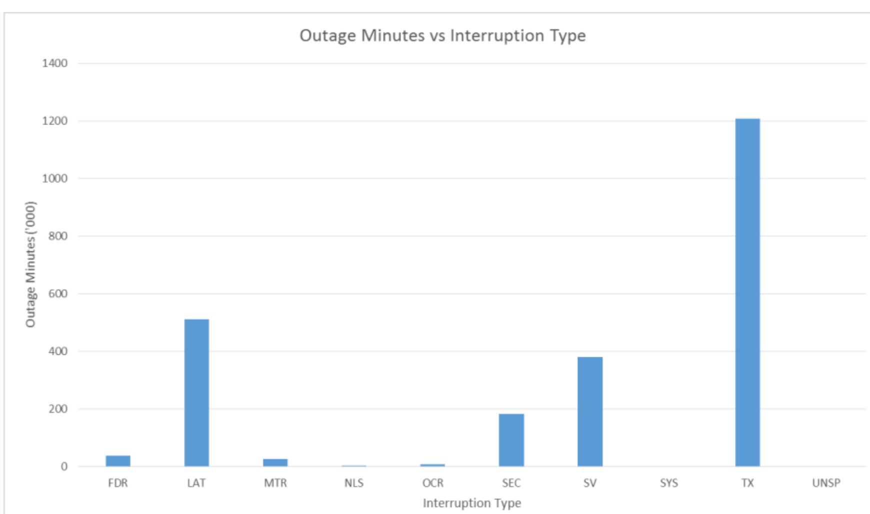

Customer Minutes Interrupted vs Ticket Type
CMI ('000)
25000
20000
15000
10000
5000
0
FDR
LAT
LATU
LOC
NLS
OCR
SNC
SNCU
SUB
TX
TXU
Ticket Type

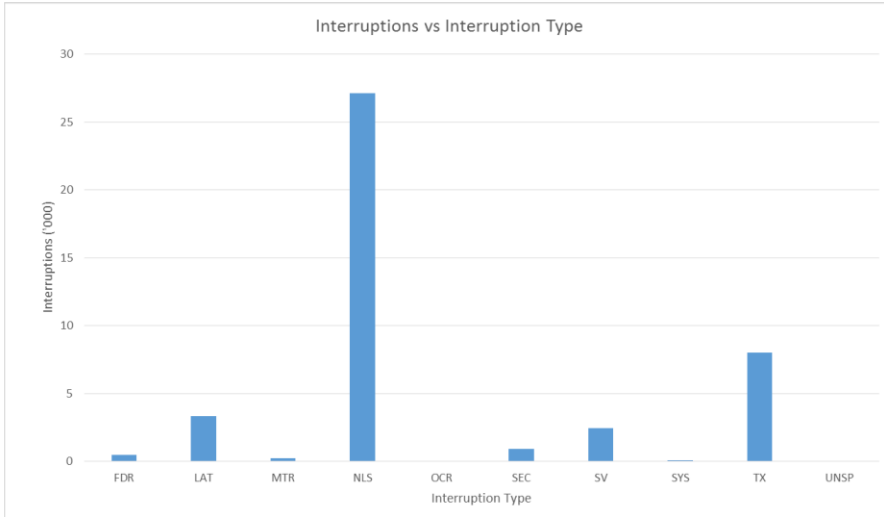

Outage Minutes vs Ticket Type
Outage Minutes ('000)
700
600
500
400
300
200
100
0
FDR
LAT
LATU
LOC
NLS
OCR
SNC
SNCU
SUB
TX
TXU
Ticket Type

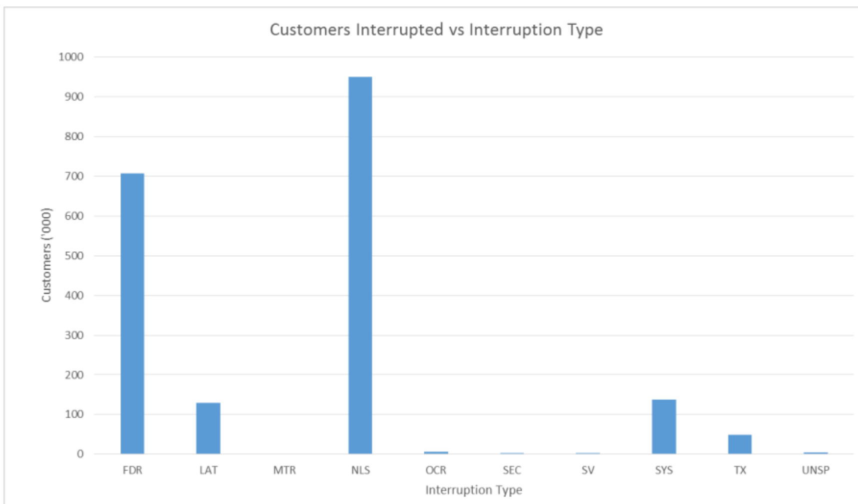

Interruptions vs Ticket Type
Interruptions ('000)
14
12
10
8
6
4
2
0
FDR
LAT
LATU
LOC
NLS
OCR
SNC
SNCU
SUB
TX
TXU
Ticket Type

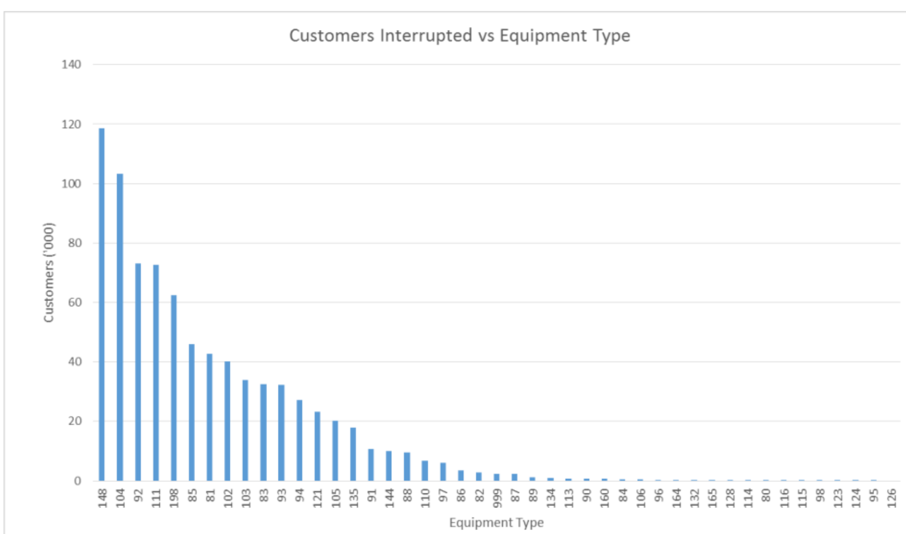

Customers Interrupted vs Ticket Type
Customers ('000)
1400
1200
1000
800
600
400
200
0
FDR
LAT
LATU
LOC
NLS
OCR
SNC
SNCU
SUB
TX
TXU
Ticket Type

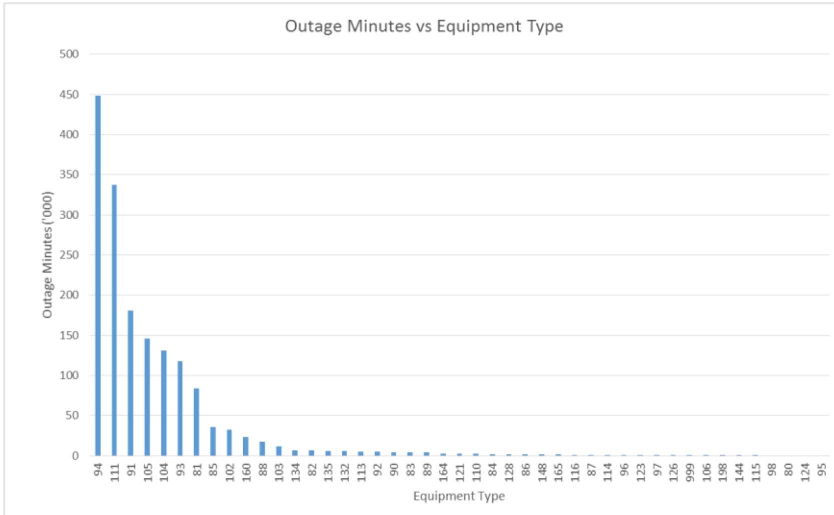

Customers Interrupted vs Interruption Type
Customers ('000)
1000
900
800
700
600
500
400
300
200
100
0
FDR
LAT
MTR
NLS
OCR
SEC
SV
SYS
TX
UNSP
Interruption Type

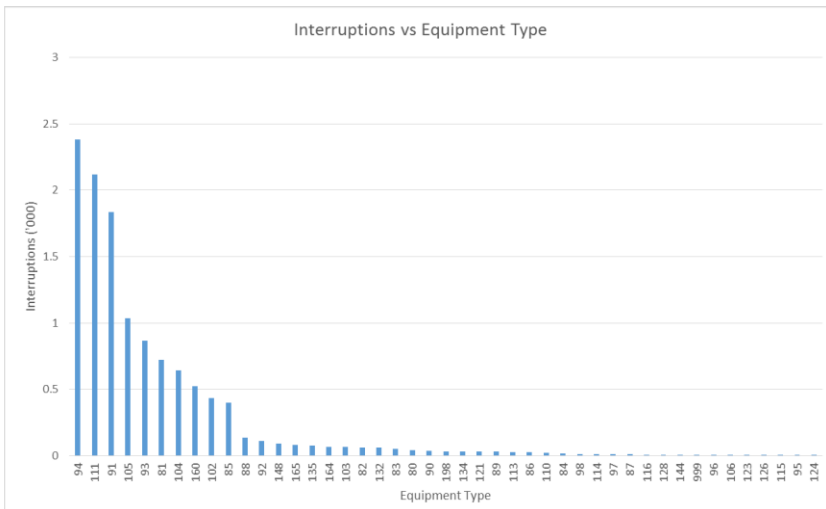

Outage Minutes vs Interruption Type
Outage Minutes ('000)
1400
1200
1000
800
600
400
200
0
FDR
LAT
MTR
NLS
OCR
SEC
SV
SYS
TX
UNSP
Interruption Type

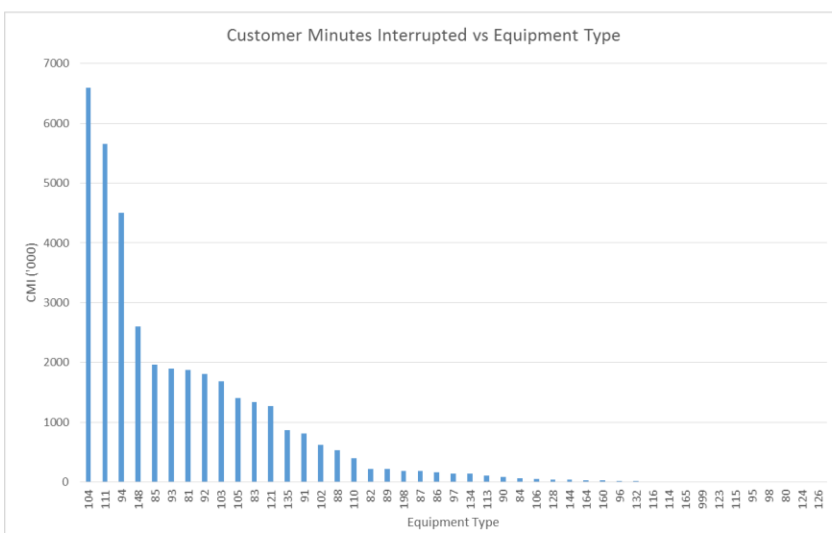

Interruptions vs Interruption Type
Interruptions ('000)
30
25
20
15
10
5
0
FDR
LAT
MTR
NLS
OCR
SEC
SV
SYS
TX
UNSP
Interruption Type

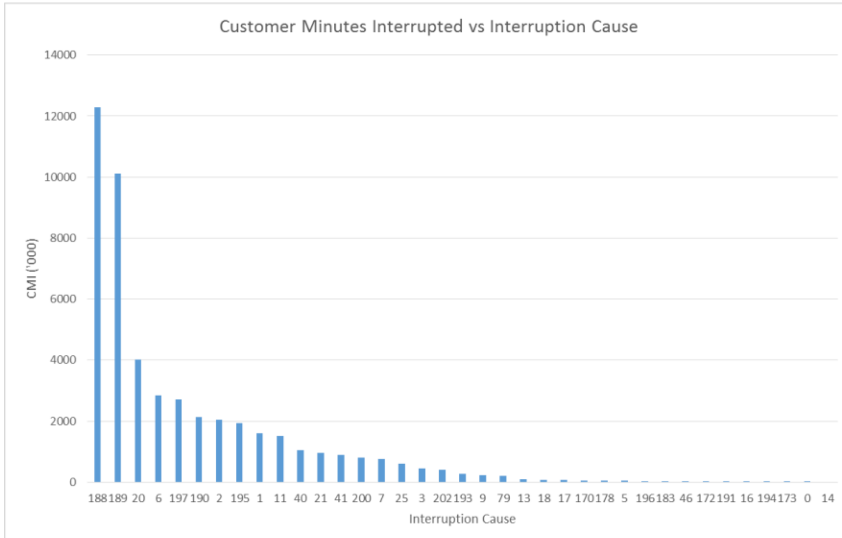

Customers Interrupted vs Interruption Type
Customers ('000)
1000
900
800
700
600
500
400
300
200
100
0
FDR
LAT
MTR
NLS
OCR
SEC
SV
SYS
TX
UNSP
Interruption Type

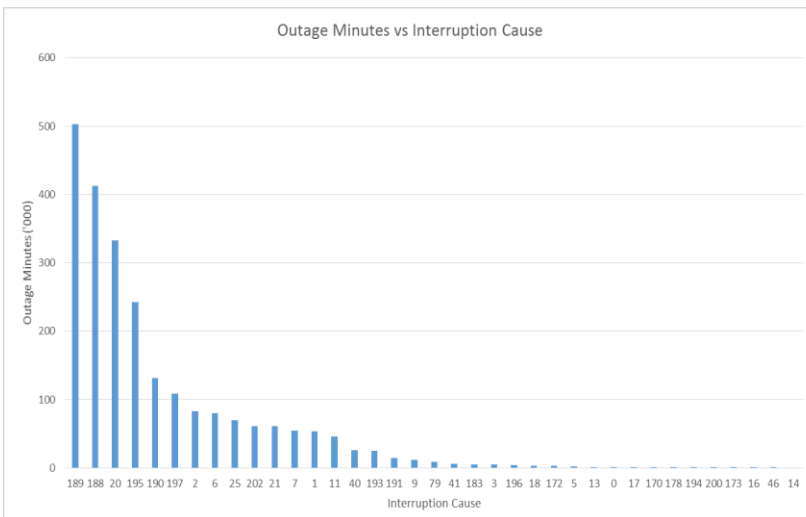

Customers Interrupted vs Equipment Type
Customers ('000)
140
120
100
80
60
40
20
0
148
104
92
111
198
85
81
102
103
83
93
94
121
105
135
91
144
88
110
97
86
82
999
87
89
134
113
90
160
84
106
96
164
132
165
128
114
80
116
115
98
123
124
95
126
Equipment Type

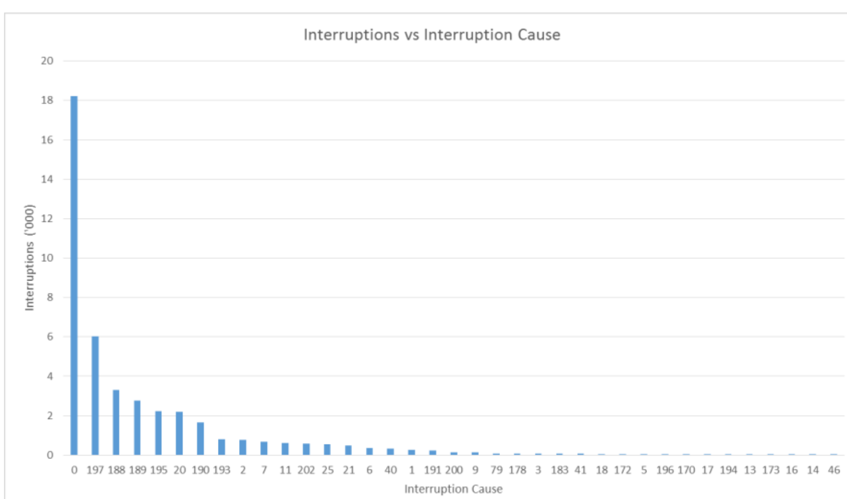

Outage Minutes vs Equipment Type
Outage Minutes ('000)
Equipment Type

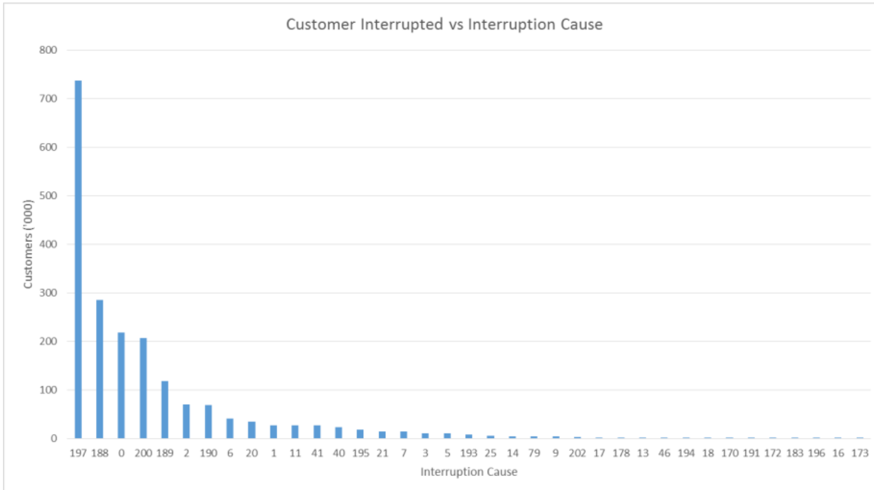

Interruptions vs Equipment Type
Interruptions ('000)
Equipment Type


Customer Minutes Interrupted vs Equipment Type
CMI ('000)
Equipment Type

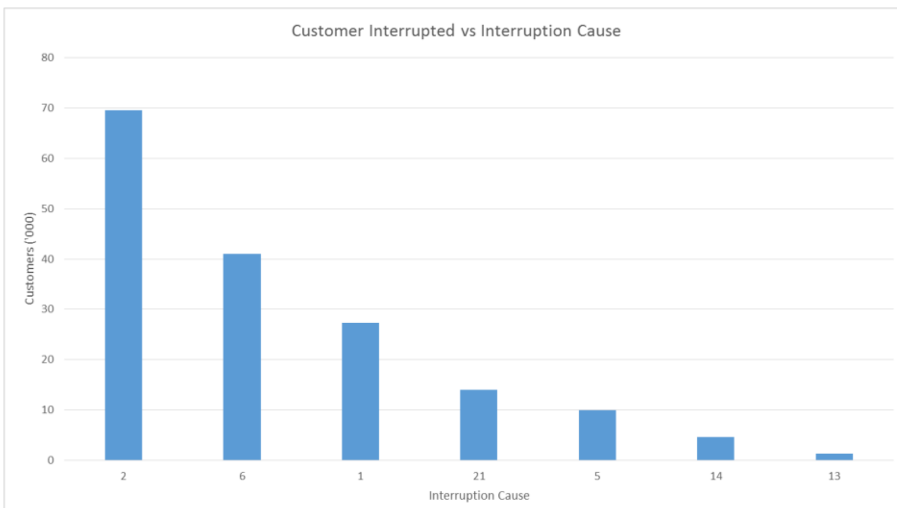

Customer Minutes Interrupted vs Interruption Cause
14000
12000
10000
8000
6000
4000
2000
0
CMI ('000)
188 189 20 6 197 190 2 195 1 11 40 21 41 200 7 25 3 202 193 9 79 13 18 17 170 178 5 196 183 46 172 191 16 194 173 0 14
Interruption Cause

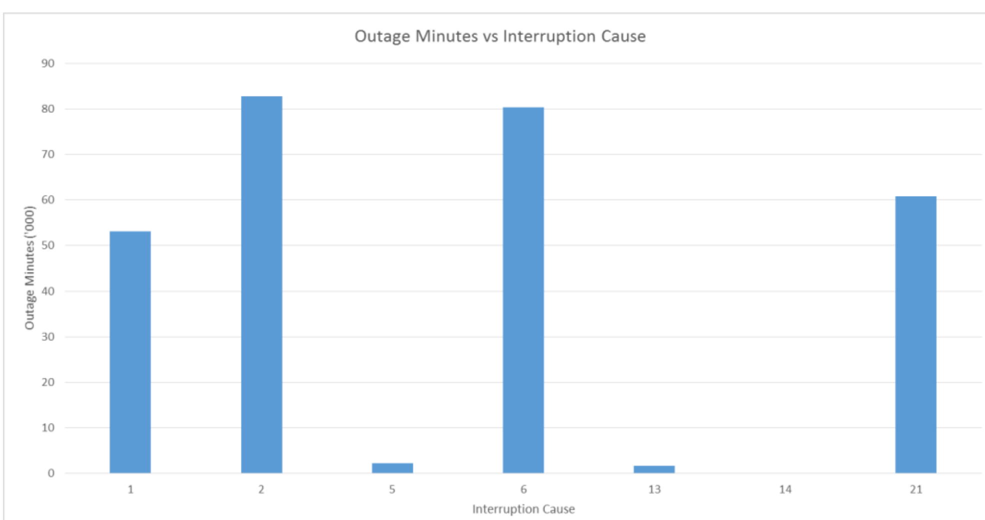

Outage Minutes vs Interruption Cause
600
500
400
300
200
100
0
Outage Minutes ('000)
189 188 20 195 190 197 2 6 25 202 21 7 1 11 40 193 191 9 79 41 183 3 196 18 172 5 13 0 17 170 178 194 200 173 16 46 14
Interruption Cause

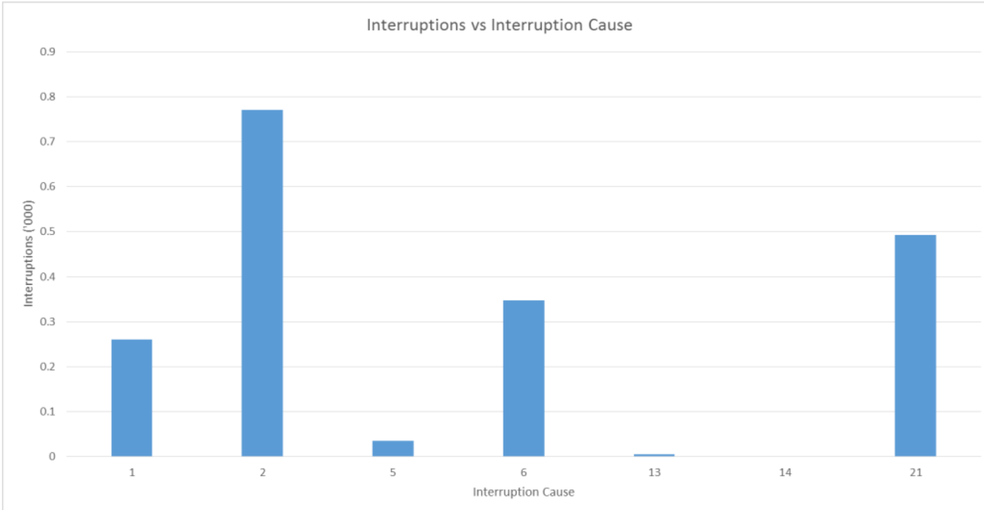

Interruptions vs Interruption Cause
20
18
16
14
12
10
8
6
4
2
0
Interruptions ('000)
0 197 188 189 195 20 190 193 2 7 11 202 25 21 6 40 1 191 200 9 79 178 3 183 41 18 172 5 196 170 17 194 13 173 16 14 46
Interruption Cause

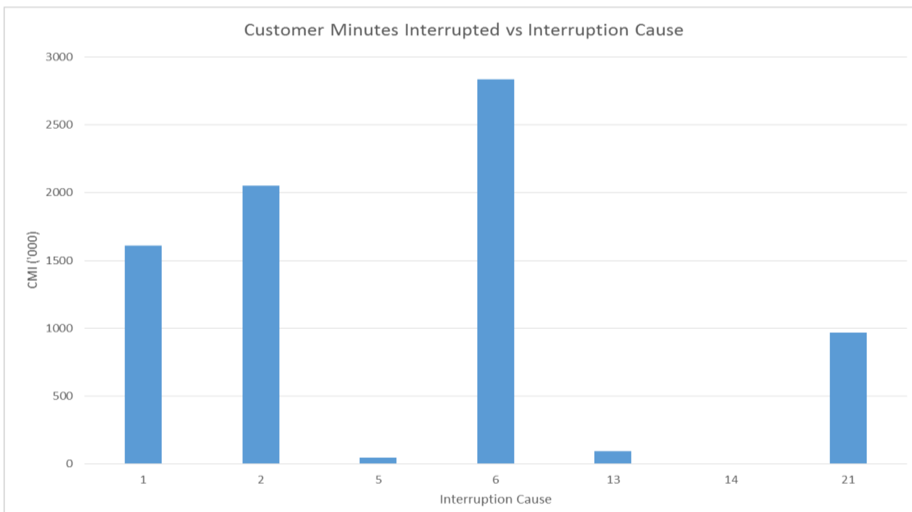


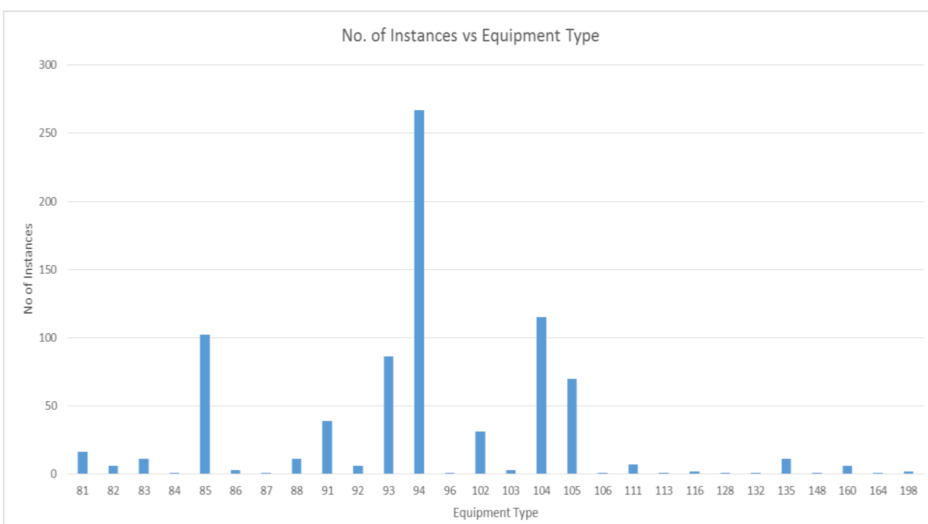
Only Weather Graphs

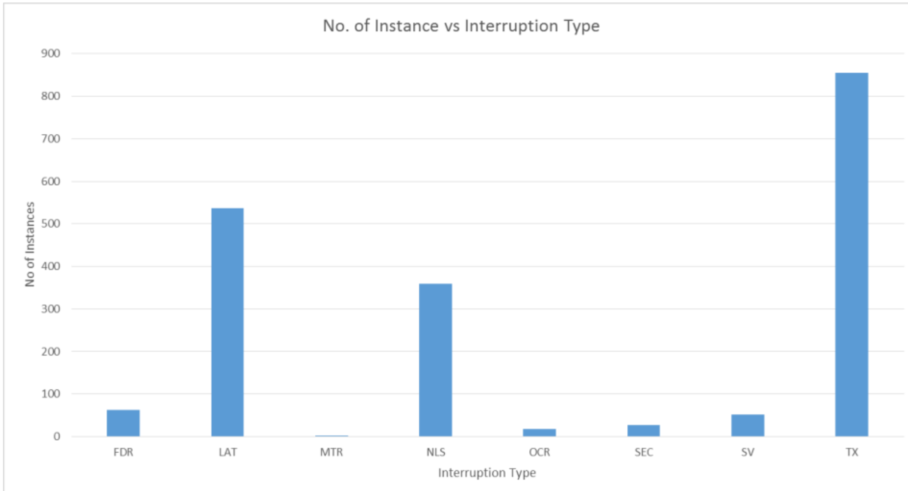


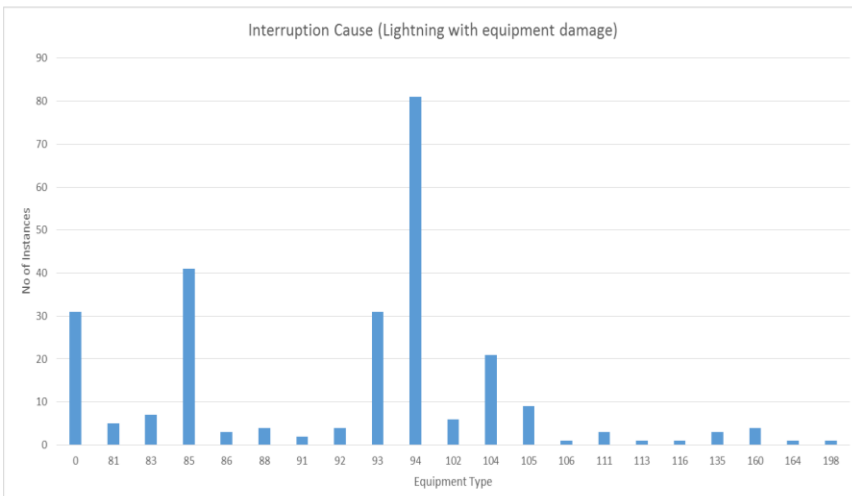


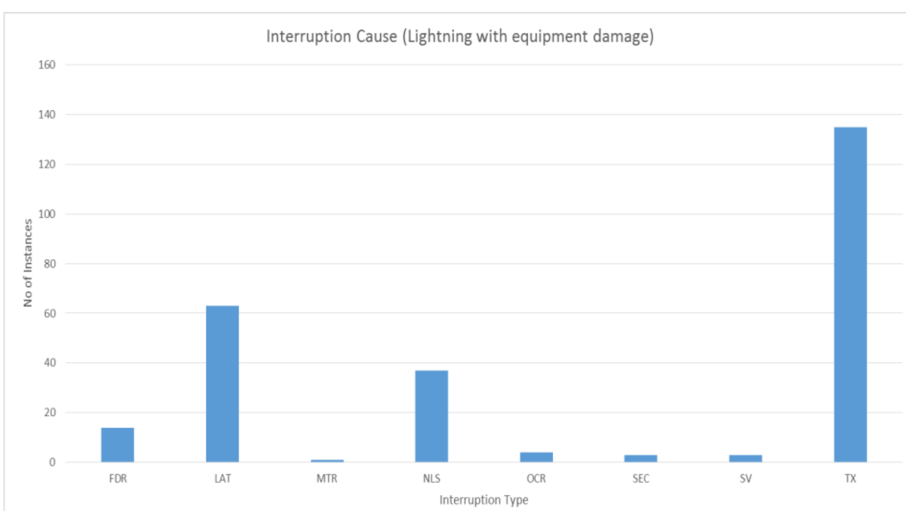

Interruptions vs Interruption Cause
Interruptions ('000)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
5
6
13
14
21
Interruption Cause

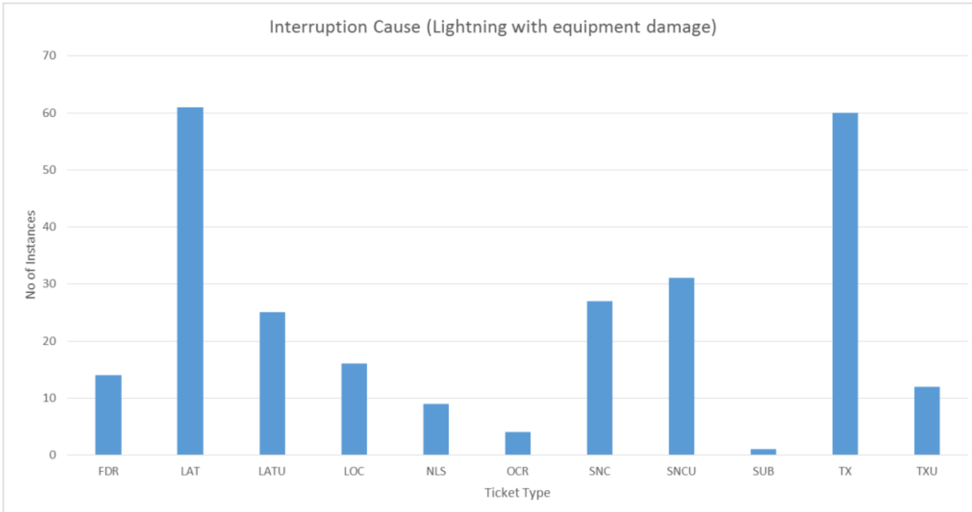

Customer Minutes Interrupted vs Interruption Cause
CMI ('000)
3000
2500
2000
1500
1000
500
0
1
2
5
6
13
14
21
Interruption Cause

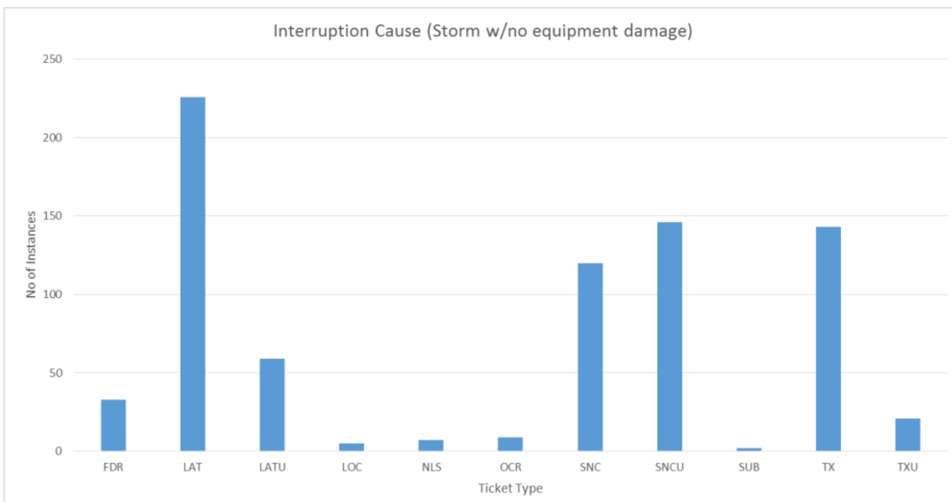

No. of Instances vs Equipment Type
No of Instances
300
250
200
150
100
50
0
81 82 83 84 85 86 87 88 91 92 93 94 96 102 103 104 105 106 111 113 116 128 132 135 148 160 164 198
Equipment Type

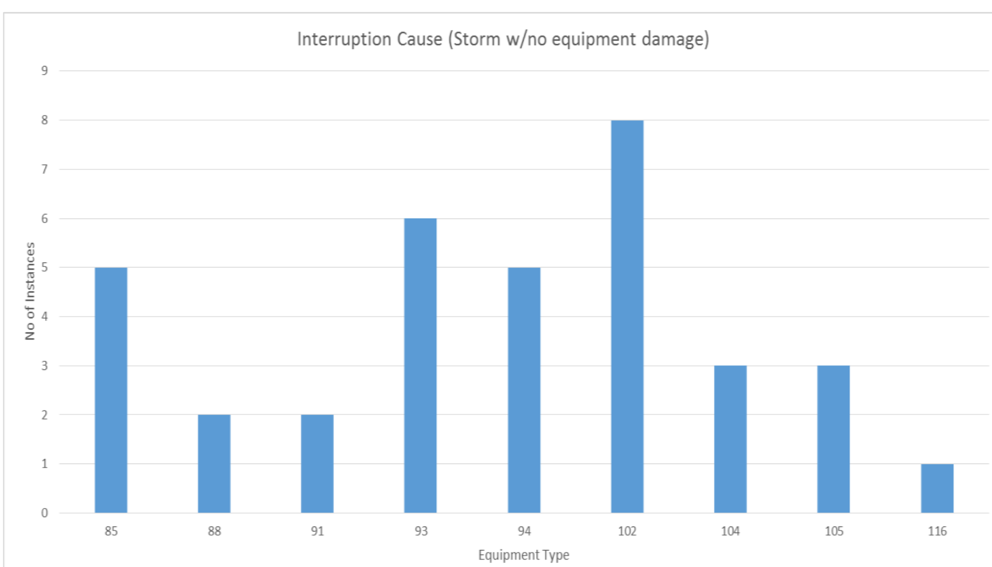

No. of Instance vs Interruption Type
900
800
700
600
500
400
300
200
100
0
No of Instances
FDR
LAT
MTR
NLS
OCR
SEC
SV
TX
Interruption Type

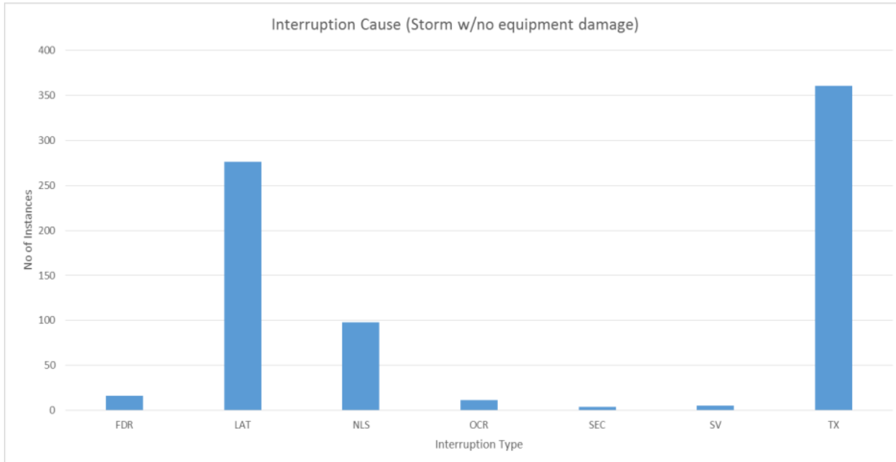

Interruption Cause (Lightning with equipment damage)
90
80
70
60
50
40
30
20
10
0
No of Instances
0
81
83
85
86
88
91
92
93
94
102
104
105
106
111
113
116
135
160
164
198
Equipment Type

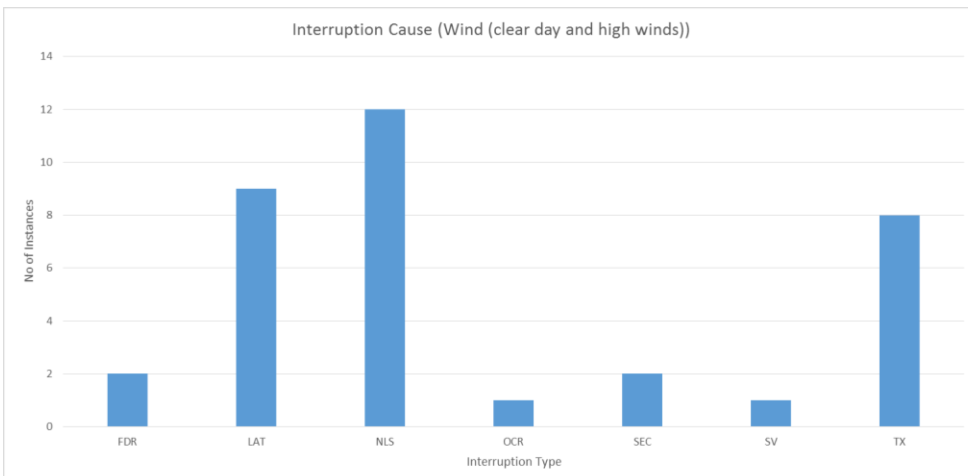

Interruption Cause (Lightning with equipment damage)
160
140
120
100
80
60
40
20
0
No of Instances
FDR
LAT
MTR
NLS
OCR
SEC
SV
TX
Interruption Type

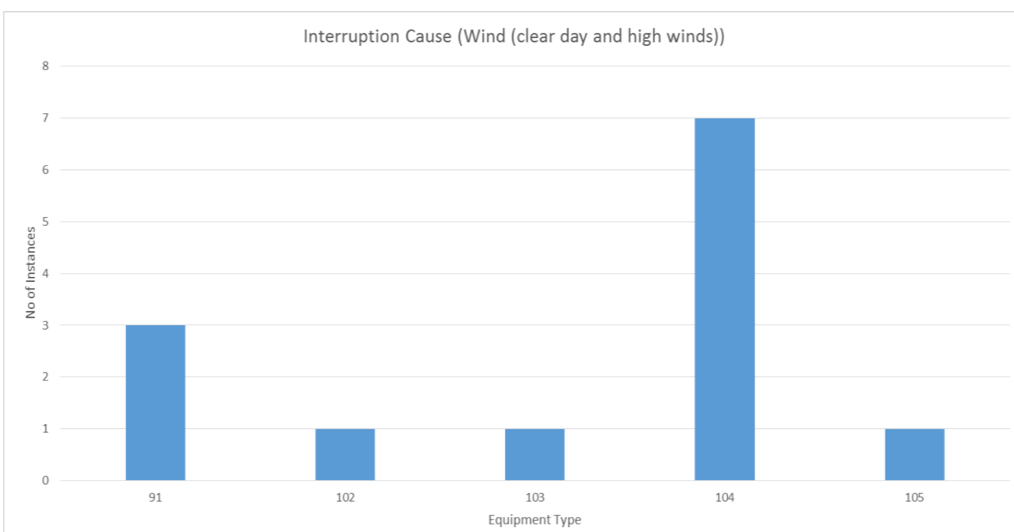

Interruption Cause (Lightning with equipment damage)
No of Instances
70
60
50
40
30
20
10
0
FDR
LAT
LATU
LOC
NLS
OCR
SNC
SNCU
SUB
TX
TXU
Ticket Type

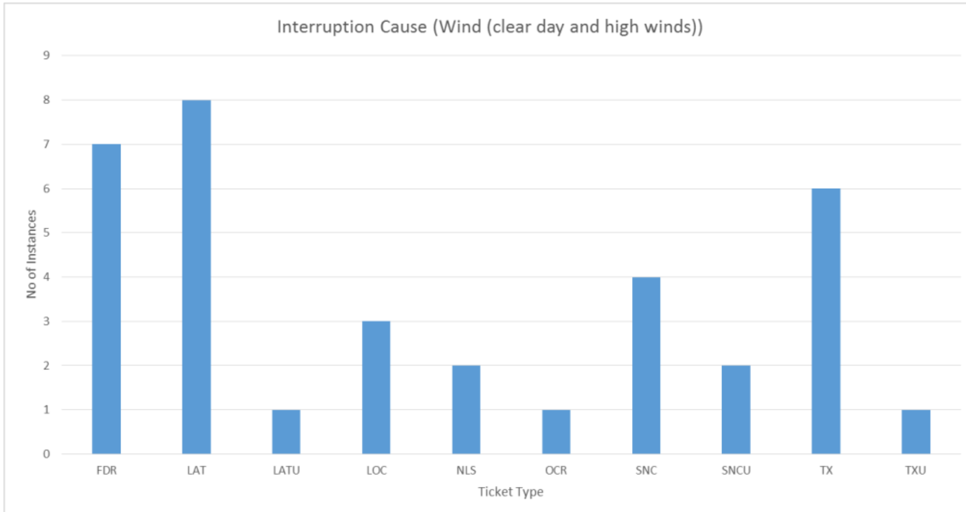

Interruption Cause (Storm w/no equipment damage)
No of Instances
250
200
150
100
50
0
FDR
LAT
LATU
LOC
NLS
OCR
SNC
SNCU
SUB
TX
TXU
Ticket Type

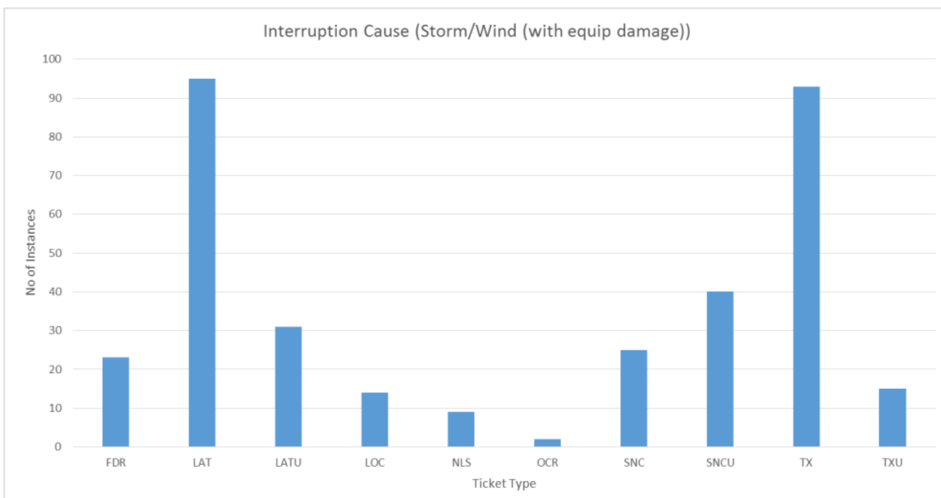

Interruption Cause (Storm w/no equipment damage)
No of Instances
9
8
7
6
5
4
3
2
1
0
85
88
91
93
94
102
104
105
116
Equipment Type

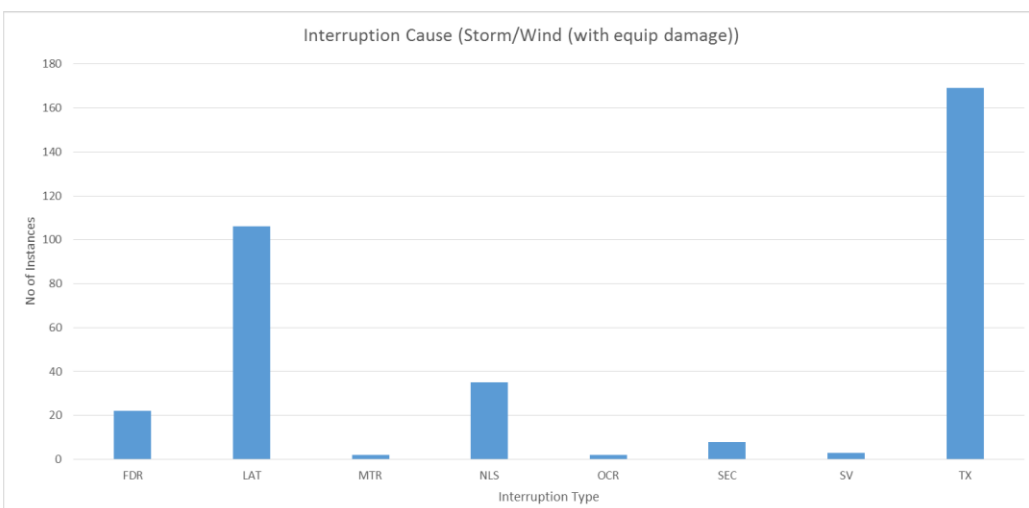

Interruption Cause (Storm w/no equipment damage)
400
350
300
250
200
150
100
50
0
No of Instances
FDR
LAT
NLS
OCR
SEC
SV
TX
Interruption Type

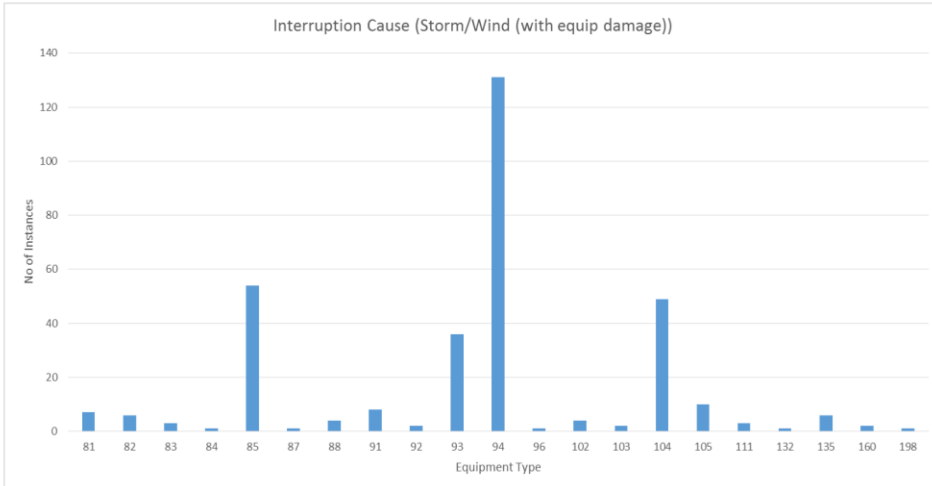

Interruption Cause (Wind (clear day and high winds))
14
12
10
8
6
4
2
0
No of Instances
FDR
LAT
NLS
OCR
SEC
SV
TX
Interruption Type

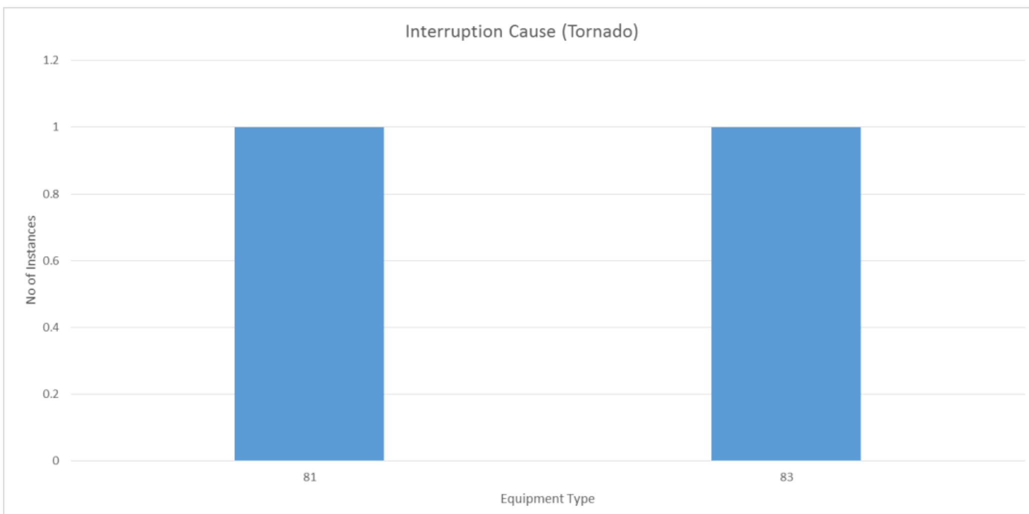

Interruption Cause (Wind (clear day and high winds))
8
7
6
5
4
3
2
1
0
No of Instances
91
102
103
104
105
Equipment Type

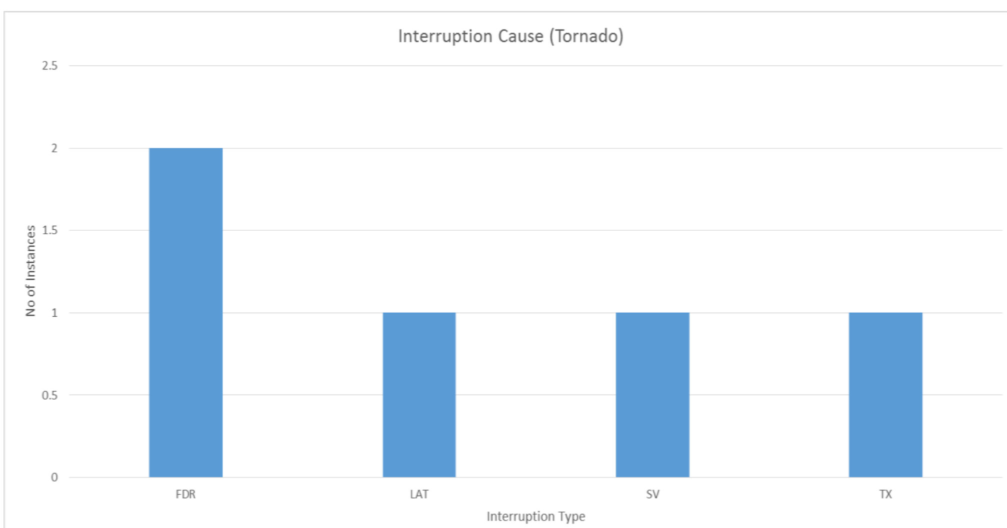

Interruption Cause (Wind (clear day and high winds))
No of Instances
0
1
2
3
4
5
6
7
8
9
FDR
LAT
LATU
LOC
NLS
OCR
SNC
SNCU
TX
TXU
Ticket Type

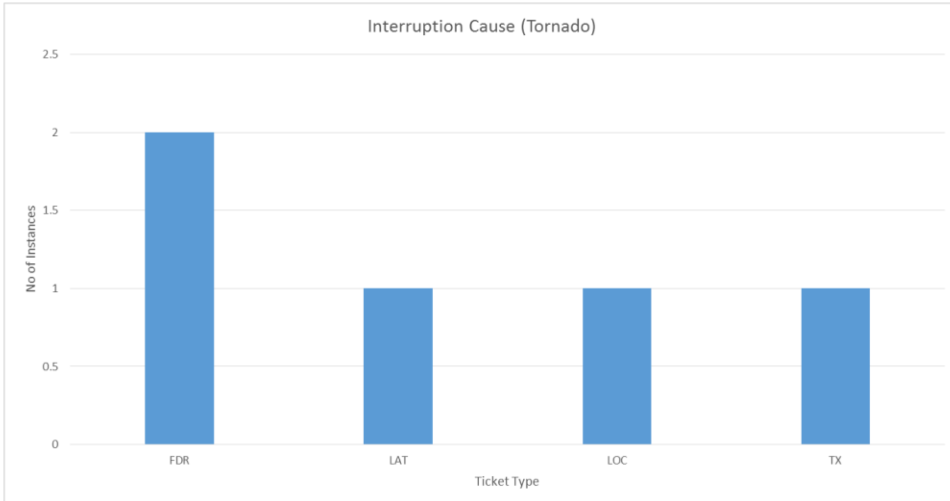

Interruption Cause (Storm/Wind (with equip damage))
No of Instances
0
10
20
30
40
50
60
70
80
90
100
FDR
LAT
LATU
LOC
NLS
OCR
SNC
SNCU
TX
TXU
Ticket Type

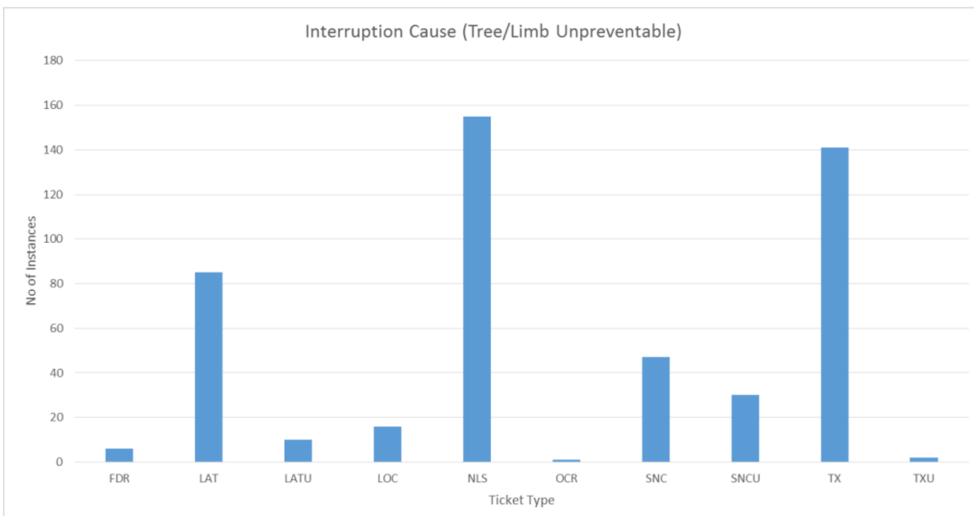

Interruption Cause (Storm/Wind (with equip damage))
No of Instances
0
20
40
60
80
100
120
140
160
180
FDR
LAT
MTR
NLS
OCR
SEC
SV
TX
Interruption Type

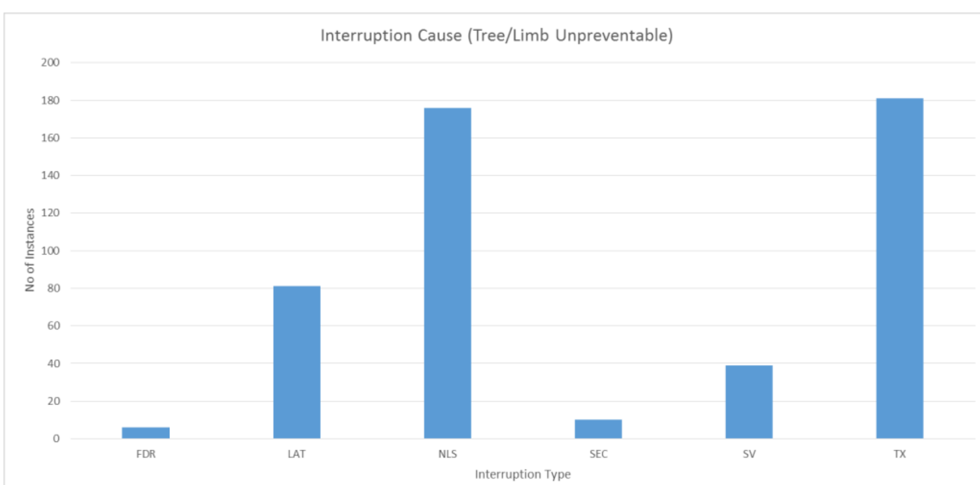

Interruption Cause (Storm/Wind (with equip damage))
No of Instances
0
20
40
60
80
100
120
140
81 82 83 84 85 87 88 91 92 93 94 96 102 103 104 105 111 132 135 160 198
Equipment Type

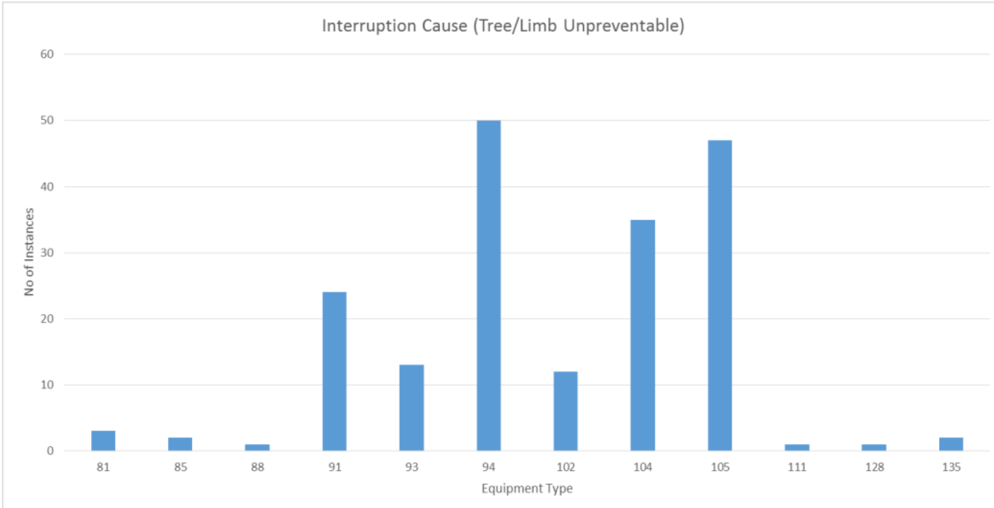

Interruption Cause (Tornado)
No of Instances
0
0.2
0.4
0.6
0.8
1
1.2
81
83
Equipment Type


Interruption Cause (Tornado)
No of Instances
0
0.5
1
1.5
2
2.5
FDR
LAT
SV
TX
Interruption Type


Interruption Cause (Tornado)
No of Instances
2.5
2
1.5
1
0.5
0
FDR
LAT
LOC
TX
Ticket Type


Interruption Cause (Tree/Limb Unpreventable)
No of Instances
180
160
140
120
100
80
60
40
20
0
FDR
LAT
LATU
LOC
NLS
OCR
SNC
SNCU
TX
TXU
Ticket Type


Interruption Cause (Tree/Limb Unpreventable)
No of Instances
200
180
160
140
120
100
80
60
40
20
0
FDR
LAT
NLS
SEC
SV
TX
Interruption Type


Interruption Cause (Tree/Limb Unpreventable)
No of Instances
0
10
20
30
40
50
60
81
85
88
91
93
94
102
104
105
111
128
135
Equipment Type

CODE TABLE

CAUSE CODES

001	E	Lightning with equip damage
002		Storm w/no equip damage
003	E	Accidental Fire (forest fires, etc.)
005		Wind (clear day and high winds)
006	E	Storm/Wind (with equip damage)
007		Squirrel
009		Other Bird
011		Other Animal
013		Tornado
014		Hurricane
015		Ice on Lines
016		Monk Parrot Nest
017		Osprey Nest
018		Lizards
020		Tree/Limb Preventable
021		Tree/Limb Unpreventable
025		Vines/Grass
040		Vehicle
041		Accidental Contact
046		Switching Error
079		Dig-In
170		Wrong size fuse
172		Overloaded Normal Conditions
173		Overloaded Emergency Conditions
178		Non-standard Construction
183		Improper Installation
188	E	Equip Failed-OH
189	E	Equip Failed-UG
190		Unknown
191	E	Vandalism
193	E	Customer Request
194		Unplanned - Crew Request
195	E	Planned - Crew Request
196		Slack Conductors
197		Other (explain)
202	E	Loose Connection

LEGEND:

CAUSE CODES
(TO BE USED FOR ALL INTERRUPTIONS)
Natural Causes
Equipment Related Causes
Other Causes
Accidental Causes

EQUIPMENT CODES
Overhead
Underground
Overhead or Underground

Meter
Substation

SUPPORT CODES
(USE AS SUPPORT OR FOLLOW-UP ONLY)

EQUIPMENT CODES

080	Down Guy or Anchor
081	Pole
082	Cross Arm
083	Insulator
084	Pole Top Pin
085	Arrester
086	Polymer Arrester
087	Tie Wire
088	Jumper
089	Stirrup
090	Hot Line Clamp
091	Connector
092	Disconnect Switch
093	Fuse Switch
094	Transformer
095	Step Down Transformer
096	Line Recloser
097	Line Capacitor
098	Line Regulator
102	Other Equipment
103	Splice
104	Conductor Down
105	Conductor Damaged
106	Automated Switch (DA)
110	Terminator
111	Cable
113	Elbow
114	Tx Fuse Switch
115	Tx Blade Switch
116	Bayonet Switch
121	Padmount Switch
123	RA Switch
124	Mech for Throwover Switch
125	PT Fuse
126	Conduct CKT Fuse
127	Control Cable
128	Vault Transformer
132	Handhole
134	Bushing
135	Pothead
136	Cable - Live End
144	RC-Off (Transmission Only)
148	Other Substation Equipment
160	Meter
161	Meter Blocks - Repairable
162	CT's
163	PT's
164	Other Meter Equipment
165	Meter Blocks - Not Repairable
198	RC-Off (Distribution Only)
200	Transmission Related
211	Injected Cable

SUPPORT CODES

004	Salt Spray Corrosion
012	No Animal Guard
022	Palm Tree
023	Decay/Deterioration
024	Corrosion (Non Salt Spray)
026	Contamination (Non S/S)
050	Foreign Crew or Customer
066	FPL Crew
067	FPL Distribution Contractor
068	FPL Line Clearing Contractor
069	Transmission Contractor
100	Inadequate/No Ground
101	Pole Fire -insulator contamination
153	Oil Circuit Recloser 3 PH
154	Electronic Recloser 3 PH
155	Oil Circuit Recloser 1 PH
156	TripSaver
192	Unplanned - Restoration
199	Defective Material - UPR
222	Power Temp Used
240	Injection Elbow (Installed)
241	Injection Elbow (Not Installed)
242	Flow (Positive)
243	Flow (None)
244	Injection Complete
245	Injection Job Pending
250	Cable Replace - Job Pending
251	Cable Replace - Job Complete
260	Sect Used
261	Sect Not Available
262	Sect Not - Trained
263	Sect Not - Failed
264	Sect Not - 3 Phase
265	Sect Not - FI Used
266	Sect Not - Other
270	Unplanned - Police/Fire
271	Cleared by Phone
274	Repl Cable - Existing PVC
275	Repl Cable - New PVC
293	Cbl EPR/Poly 1st Run
294	Cbl Pilc 1st run
295	Cbl Dir Buried 1st run
296	Cbl Ductbank 1st run
297	Cbl Conduit 1st run
298	Cbl Submarine 1st run
299	Data Corrected
300	Cable - PILC
301	Cable - Direct Buried
302	Cable - In Duct
303	Cable - In Conduit
304	Cable - Submarine
305	In Hanhdole
306	Cable - EPR/Poly
310	Wire - Small (#2 and smaller)
311	Wire - Large (1\0 and greater)
312	Live Front-UG
313	Dead Front-UG
316	1 Phase Dry Vault
317	1 Phase Vault
318	3 Phase Vault
320	AMI Upgrade
401	Planned - Const/Maint
402	Planned - Restoration
403	Planned - Another Reason
410	In Line Dead End
411	In Line Triangular
412	In Line X Arm
413	In Line Vertical
414	Base Horiz Bracket
415	Base Dbl Wide X Arm
416	Base Metal X Arm
417	Base Vert Bracket
999	Named Storm Exclusions