

MA Outage Data Cleaning

Olukunle Owolabi

Thursday 12th March, 2020

1 Introduction

There are three sources of data for the Massachusetts outage and they include: NATIONAL GRID, EVERSOURCE and UNITIL. These are the three main utilities that provide electricity for the entire state of Massachusetts. Massachusetts provides detailed information on outage incidents for utilities in the state from 2013 - 2018. The electric utility companies file their Emergency Response Plan (ERP) which includes information about historic power outages. More information on the raw data and associated documentations can be found at: <https://www.mass.gov/info-details/power-outages#historic-power-outages-1>. This data has a higher spatial resolution (towns/cities and in some cases streets) relative to the EIA Reliability and the OE417 (Major Outage) data. Similarly to the OE417, the data has a minute-level temporal resolution.

2 Data Description

The cleaned MA outage data has 15 features. A breakdown of these features is given in Table 1. They can be broadly classified as follows: time and date of outage occurrence, geographical location of the outage, nature/cause of the outage, type of equipment and load affected, severity of the outage (customers affected), and Utility/Service Company Information. A total of 138153 outage data samples that represents all outages in MA across the three major utilities in the state is obtained.

3 Assumptions

The following assumptions were made to guide the cleaning process:

1. In cases where no street names were recorded (mostly from data reported by UNITIL), the street name was simply replaced with "nan".
#Replace street name in data with "nan" where unavailable
df10["Street"] = np.nan
2. For years with many missing information e.g. in the 2012 UNITIL outage data where the 'Voltage Levels', 'Circuit Type', 'Failed or Damaged Equipment', 'Weather Condition' and 'Load Type' information which were unreported were set to 'nan'.

Variable/Feature	Description
Date Out	Year, Month and Day of the outage event
Time Out	Time of the outage event (in 24-hour format)
Date In	Year, Month and Day the outage event was curtailed
Time In	Time the outage event was curtailed (in 24-hour format)
Actual Duration	The difference between time of outage and time of restoration
Number of Customers Affected	Number of customers affected by the electric disturbance event
Original Number Customers Affected	Customers affected by power loss due to outage
City/Town	City or Town affected by the outage
Street	Name of street affected by the outage (where applicable)
Company Name	Name of company in charge of electric service
Voltage Level	Voltage level of distribution feeder affected
Circuit Type	Type of circuit affected by outage (e.g. Overhead, Underground)
Reason For Outage	The suspected or known cause of the outage
Failed or Damaged Equipment	Equipment affected by outage
Weather Condition	Weather Condition at the time of outage (Not necessarily the cause of the outage)
Load Type	Nature of load affected (e.g. Residential, Commercial)

Table 1: Feature Description for Cleaned MA Outage Data

4 Methods and Techniques

Some steps were taken in order to clean the data and these include:

1. The data collected per year for each utility was individually inspected before merging. This is to ensure that the columns names are the same for all the dataframes to be merged.

2. In most cases, the column names were renamed in order to ensure uniformity for all dataframes to be merged. An example of a line of code used to do this is given below:

```
#Rename column names  
df1.rename(columns = 'KV Levels':'Voltage Levels', "Reason for Out-  
age": "Reason For Outage", inplace=True)
```

3. After thorough inspection and cleaning of the data for each year, the cleaned yearly data was merged for all the years per utility. A sample code which was used to do this is given below:

```
#Concatenate data frames for one utility  
df_1 = pd.concat([df1, df2, df3, df4, df5, df6, df7, df8, df9], ignore_index  
=True, sort=False).
```

4. To ensure consistency, a quick check was performed to ensure that the column names are unchanged before and after concatenating/merging.

```
#Check  
df1.columns==df_1.columns
```

5. For years with missing information, the appropriate assumption were made in order to clean the data to ensure that it is consistent with other data columns

6. In some cases the city/town affected was separated from a detailed address of isolating device. This was done using the split() function in Python.

7. The date-time column for the outage event occurrence and restoration was split into separate columns of date and time for both outage and restoration events.

8. After all the data were cleaned per year and combined for each utility. The data representing each of the utilities were then merged to obtained a final cleaned dataset.

9. Further checks were conducted per feature to ensure "global" consistency in the data-set. More on that is explained in the next section.

10. A sample inspection where a sample of the cleaned data and was drawn at random was carried out in order check the cleaning accuracy.

5 Feature Inspection

After aggregating the cleaned data set, individual features were examined to ensure elemental consistency within the feature space. The feature columns examined include include:

1. **Company Name:** In cases where a company name was represented with different names, the names were simply unified to ensure consistency. Example is 'FG&E', 'FITCHBURG', 'FITCHBURG GAS ELECTRIC' which all refer to the same company. In this case, "FG&E" was chosen as the unique name representation for this company.
2. **Voltage Levels:** Feature inspection revealed some inconsistencies in the 'Voltage Level' column. In order to obtain a consistent representation for this feature, decimal-type voltage levels were approximated to the nearest standard voltage level. For example, the UNITIL data which contains voltage levels denoted as "13.8kv" and "4.16kv" were approximated to 15 and 5 respectively. Note that the 'kv' added to the voltage level element was also eliminated if any element within this feature contains such. At the end of the inspection, this feature was reduced to 8 unique elements: 5, 15, 16, 25, 35, 69, TR, SR which represents standard voltage levels for power outage lines in MA. A sample line of code used to effect this change is given below:


```
#correction
df_2["Voltage Levels"] = df_2["Voltage Levels"].replace( 'Secondary': 'SY',
'15':15, '15kV': 15, '5kV': 5, '13.8kV': 15, '4.16kV':5, '69kV':69)
```
3. **Circuit Type:** Spelling errors were corrected in this feature. The feature inspection process ensured that the cleaned feature contains three unique elements which represents the main circuit types affected by the outage: "UG", "OH", "CO".
4. **Reason for outage:** Minor spelling errors were corrected in this feature. The feature inspection process ensured that this feature contains eleven unique elements which represents the main reasons for outage in Massachusetts:


```
['Adverse Weather', 'Physical Interference / Other', 'Animal', 'Tree Contact', 'Unknown', 'Failed Equipment', 'Adverse Environment', 'Physical Interference / Company', 'Intentional', 'Loss of Supply', 'Lightning'].
```

 A detailed "find and replace" was carried out to ensure that the "reason for outage" feature as reported by the UNITIL data conformed to the elements in the list above. The Table 2 below gives some details on the transformation during the post-inspection/correction phase.
5. **Failed or Damaged Equipment:** Also a comprehensive search and replace was carried out on this feature in order to ensure consistency. The feature inspection process ensured that this feature contains the following list that represent the nature of failed or damaged equipment during outage in MA:


```
['Other - specify', 'Pole - wood', 'No failure or Unknown', 'Cutout - open', 'Overhead Device Other', 'Cable - EPR', 'Fuse Unit / Fuse Link', 'Circuit breaker', 'Connector', 'Conductors', 'Cable - XLPE - PE', 'Splice', 'Cutout door/fuse holder', 'Crossarm', 'Cable - paper - VC', 'Transformer O/H conv.', 'Transformer Ratio', 'Insulator - pin or post', 'Anchor and Guy', 'Transformer padmount', 'Substation device other', 'Switch - Solid Blades', 'Tap', 'Recloser', 'Switch OFC(Oil Fused Cutout)', 'Transformer Other', 'Lightning arrester', 'Insulator - disc', 'Cable - rubber', 'Switch
```

- Vacuum Fault Interrupter (VFI)', 'Switch Loadbreak', 'Riser/Siphon', 'Switch - Airbreak', 'Cable - Other', 'Cutout - enclosed', 'Elbow', 'Termination', 'Switch - Vacuum Loadbreak', 'U/G Other Device', 'Switchgear', 'Animal Guard', 'Relay', 'Capacitor', 'Bushing', 'Bus', 'Transformer O/H CSP', 'Conductor (broken)', 'Insulator', 'Cable', 'Transformer - Polemount', 'Pole', 'Cutout', 'Connector, Transformer - Polemount', 'Arrester', 'Recloser / Breaker', 'Terminator', 'S/S Breaker', 'Clamp/Auto D.E.', 'Sectionalizer', 'Regulator', 'Network Protector', 'Customer - Meter', 'Customer - Other', 'Control Cable', 'Spacer', 'Switch - Submersible', 'Recloser', 'Fault Indicator', 'Generator', 'Regulator / Autobooster', 'Metering', 'Customer - Weatherhead', 'Tower', 'Switch', 'Transformer - Station']

6. **Weather Condition:** The weather condition feature also contained several elemental inconsistencies and spelling errors which were corrected using a search and replace function (similar to what was carried out for the 'Failed or Damaged Equipment' feature). The feature inspection process ensured that this feature contains the following list that represent all unique weather conditions during outage in MA:
['NORM', 'SNOW', 'OTHR', 'WIND', 'RAIN', 'ICE', 'THUN', 'HURR', 'HEAT', 'BLIZ', 'TORN', 'FLOD']
7. **Street and City/Town:** Capitalized names of streets and city/town were changed to have first letter of each word in capital letters (as per proper nouns).

6 Data Visualization

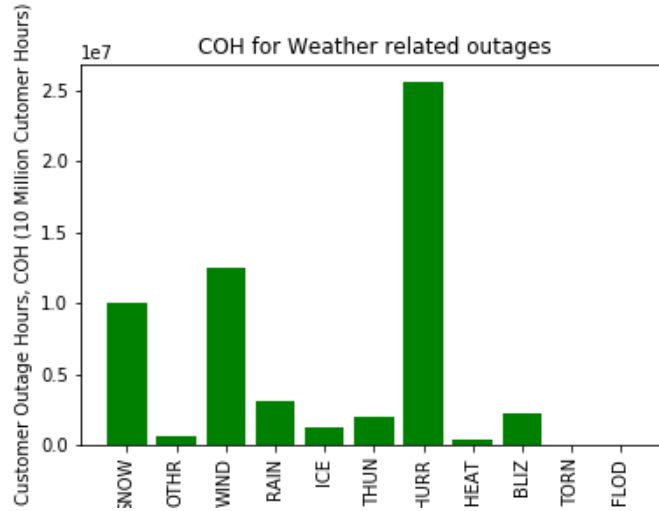


Figure 1: Statistics of Weather related Outages

Before	After
'INTENTIONAL'	'Intentional'
'TREE CONTACT'	'Tree Contact'
'VEHICLE DAMAGE TO POLE/EQUIP'	'Physical Interference / Other'
'ANIMAL'	'Animal'
'FAILED EQUIPMENT'	'Failed Equipment'
'UNKNOWN'	'Unknown'
'FIRE'	'Adverse Environment'
'WIND - NON-VEGETATION RELATED'	'Adverse Weather'
'PHYSICAL INTERFERENCE/OTHER'	'Physical Interference / Other'
'LIGHTNING'	'Lightning'
'OPERATING/TESTING'	'Physical Interference / Company'
'PHYSICAL INTERFERENCE/COMPANY'	'Physical Interference / Company'
'Squirrel'	'Animal'
'Equipment Failure Company'	'Failed Equipment'
'Tree/Limb Contact - Growth into Line'	'Tree Contact'
'Patrolled, Nothing Found'	'Unknown'
'Improper Installation'	'Loss of Supply'
'Scheduled,Planned Work'	'Physical Interference / Company'
'Tree/Limb Contact - Broken Limb'	'Tree Contact'
'Tree/Limb Contact - Broken Trunk'	'Tree Contact'
'Loose/Failed Connection'	'Loss of Supply'
'Vehicle Accident'	'Physical Interference / Other'
'Operator Error/System Malfunction'	'Loss of Supply'
'Broken Tree/Limb'	'Tree Contact'
'Action by Others'	'Physical Interference / Other'
'Lightning Strike'	'Adverse Weather'
'Tree/Limb Contact - Vines'	'Tree Contact'
'Corrosion/Contamination/Decay'	'Adverse Environment'
'Tree/Limb Contact - Uprooted Tree'	'Tree Contact'
'Bird'	'Animal'
'Equipment Failure Customer'	'Failed Equipment'
'Animal - Other'	'Animal'
'Other'	'Physical Interference / Other'
'Civil Emergency (fire,flood, etc.)'	'Adverse Environment'
'Overload'	'Loss of Supply'
'Unknown - Under Investigation'	'Unknown'

Table 2: Elements of 'Reason for outage' feature for UNITIL data before and after feature inspection & correction

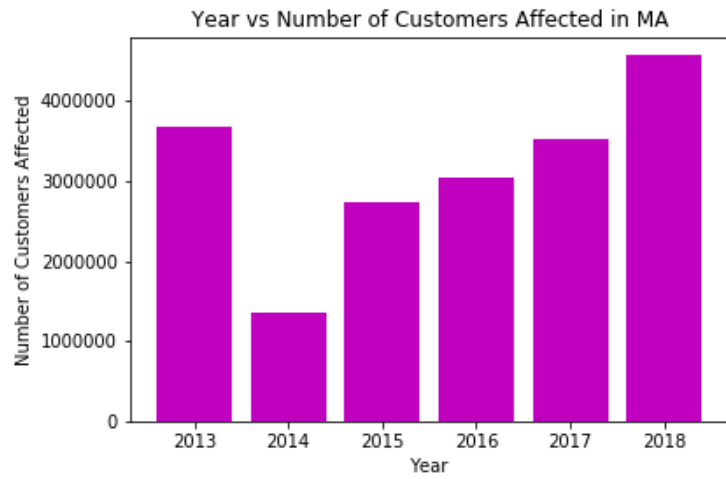


Figure 2: Outage Duration over time

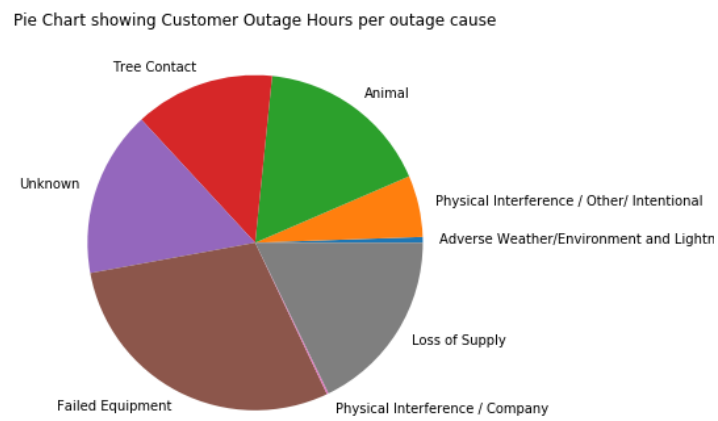


Figure 3: Customer Outage Hours and Causes of Outages

References

- [1] “Historic power outages in Massachusetts.”