**The Steps**

**YEAR 2002**

* + Filter rows in column date equal (date only)
  + REPLACE Evening to 06:00:00 A.m. date 29/01/2002 in column time event
  + REPLACE Noon November 10 to 10/11/2002 01:33:00 in column Restoration Time
  + REPLACE p.m. TO PM column time event
  + REPLACE a.m.TO AM column time event
  + REPLACE (NA & Unknown) to null in column Restoration Time
  + Merged column time & date event
  + Merged column time & date restoration
  + Changed type to date time
  + Add custom column (Date Time of Restoration - Date Time Event Began)
  + Changed type to duration
  + Trimmed text column (NERC Region- Area- Type of Disturbance)

**YEAR 2003**

* + Filter rows in column date equal (date only)
  + REPLACE p.m. TO PM column time event
  + REPLACE a.m.TO AM column time event
  + REPLACE noon TO PM column time event
  + Changed type with locale1 with type date column date event, restoration date
  + Split column by delimiter (,) in column restoration
  + REPLACE (NA & Unknown) to null in column Restoration Time
  + Merged column time & date event
  + Merged column time & date restoration
  + Changed type to date time
  + REPLACE 18/09/2003 12:00:00 AMTO 19/09/2003 12:00:00 AM column data -TIME restoration
  + REPLACE 30/06/2003 12:00:00 AM TO 01/07/2003 12:00:00 AM column data -TIME restoration
  + Add custom column (Date Time of Restoration - Date Time Event Began)
  + Changed type to duration
  + Trimmed text column (NERC Region- Area- Type of Disturbance)

**YEAR 2004**

* + Filter rows in column date equal (date only)
  + REPLACE p.m. TO PM column time event -restoration
  + REPLACE a.m.TO AM column time event- restoration
  + REPLACE noon TO PM column time event & restoration
  + REPLACE Midnight TO 05:00:00 م column time event
  + REPLACE noon TO PM column time event
  + REPLACE 12 pm TO 12:00 PM column time event
  + REPLACE PM p.m. TO PM column time event
  + Split column by delimiter (,) in column restoration
  + Split column by delimiter space in column restoration12
  + Merged column restoration 1, restoration 2
  + REPLACE (NA & Unknown) to null in column Restoration Time
  + Changed type with locale1 with type date column date event, restoration date
  + Changed type time to date column date event time, restoration date time
  + Merged column time & date event
  + Merged column time & date restoration
  + RENAMED COLUMN Date & Time Event Began, Date Time of Restoration
  + Changed type to date time
  + REPLACE 13/08/2004 12:00:00AM TO 14/08/2004 12:00:00 Am column data -TIME restoration
  + Add custom column (Date Time of Restoration - Date Time Event Began)
  + Changed type to duration
  + Trimmed text column (NERC Region- Area- Type of Disturbance)

**YEAR 2005**

* + Filter rows in column date equal (date only)
  + REPLACE 7/01//05 TO 07/01/2005 column date event
  + REPLACE a.m.TO AM column time event- restoration
  + REPLACE noon TO PM column time event & restoration
  + REPLACE Midnight TO 05:00:00 م column time event
  + REPLACE NA TO 12:00:00 PM column time event
  + REPLACE 5:78 PM TO 06:18:00 PM column time event
  + Split column by delimiter (,) in column restoration
  + Split column by delimiter space in column restoration12
  + REPLACE TO 10/02/05 PM column data restoration1.
  + REPLACE10/02/05 TO null column data restoration2
  + Merged column restoration 1, restoration 2 , restoration3 time
  + REPLACE Ongoing TO null column data restoration2
  + Changed type with locale1 with type date column date event, restoration date
  + Changed type time to date column date event time, restoration date time
  + Merged column time & date event
  + Merged column time & date restoration
  + RENAMED COLUMN Date & Time Event Began, Date Time of Restoration
  + Changed type to date time
  + Trimmed text column (NERC Region- Area- Type of Disturbance)
  + REPLACE 01/02/2005 06:01:00PMTO 02/02/2005 06:01:00 PM column data -TIME restoration
  + Changed type to duration
  + Trimmed text column (NERC Region- Area- Type of Disturbance)

**YEAR 2006**

* + Filter rows in column date equal (date only)
  + Changed type with locale1 with type date column date event, restoration date
  + REPLACE a.m.TO AM column time event- restoration
  + REPLACE noon TO PM column time event & restoration
  + REPLACE Midnight TO 05:00:00 م column time event
  + Split column by delimiter (SPACE) in column restoration
  + REPLACE a.m. TO null column data restoration3
  + Insert column year
  + Fill down column restoration 3
  + REPLACE TO null February data restoration3
  + REPLACE TO null 14 data restoration4
  + Remove another column
  + Merged column restoration 1, restoration 2, restoration3 time
  + REPLACE Ongoing TO null column time restoration
  + Changed type with locale1 with type date column date event, restoration date
  + Changed type time to date column date event time, restoration date time
  + Merged column time & date event
  + Merged column time & date restoration
  + RENAMED COLUMN Date & Time Event Began, Date Time of Restoration
  + Changed type to date time
  + Trimmed text column (NERC Region- Area- Type of Disturbance)
  + REPLACE 01/02/2005 06:01:00PMTO 02/02/2005 06:01:00 PM column data -TIME restoration
  + Changed type to duration
  + Trimmed text column (NERC Region- Area- Type of Disturbance)

**YEAR 2007**

* Filter rows in column date equal (date only)
  + Changed type with locale1 with type date column date event, restoration date
  + REPLACE a.m.TO AM column time event- restoration
  + REPLACE noon TO PM column time event & restoration
  + REPLACE Midnight TO 05:00:00 م column time event
  + Split column by delimiter (SPACE) in column restoration
  + REPLACE a.m. TO null column data restoration3
  + Insert column year
  + Remove another column
  + Merged column restoration 1, restoration 2, time
  + Merged column restoration 1, restoration 2, restoration3 date
  + REPLACE Ongoing TO null column time restoration
  + Changed type with locale1 with type date column date event, restoration date
  + Changed type time to date column date event time, restoration date time
  + Merged column time & date event
  + Merged column time & date restoration
  + RENAMED COLUMN Date & Time Event Began, Date Time of Restoration
  + Changed type to date time
  + Trimmed text column (NERC Region- Area- Type of Disturbance)
  + REPLACE 18/09/2007 12:00:00 AM TO 19/09/2007 12:00:00 AM column data -TIME restoration
  + Changed type to duration
  + Trimmed text column (NERC Region- Area- Type of Disturbance)

**YEAR 2008**

* + Filter rows in column date equal (date only)
  + Changed type with locale1 with type date column date event, restoration date
  + REPLACE a.m.TO AM column time event- restoration
  + REPLACE noon TO PM column time event & restoration
  + REPLACE Midnight TO 05:00:00 م column time event
  + Split column by delimiter (SPACE) in column restoration
  + REPLACE a.m. TO null column data restoration3
  + Insert column year
  + Remove another column
  + Merged column restoration 1, restoration 2, time
  + Merged column restoration 1, restoration 2, restoration3 date
  + REPLACE Ongoing TO null column time restoration
  + Changed type with locale1 with type date column date event, restoration date
  + Changed type time to date column date event time, restoration date time
  + Merged column time & date event
  + Merged column time & date restoration
  + RENAMED COLUMN Date & Time Event Began, Date Time of Restoration
  + Changed type to date time
  + Trimmed text column (NERC Region- Area- Type of Disturbance)
  + REPLACE 01/01/2008 11:30:00 PM TO 01/01/2009 11:30:00 PM column data -TIME restoration
  + REPLACE 08/05/2008 12:56:00 AM TO 09/05/2008 12:56:00 AM column data -TIME restoration
  + REPLACE 02/10/2008 09:50:00 AM 03/10/2008 09:50:00 AM column data -TIME restoration
  + Changed type to duration
  + Trimmed text column (NERC Region- Area- Type of Disturbance)

**YEAR 2009**

* + Filter rows in column date equal (date only)
  + Changed type with locale1 with type date column date event, restoration date
  + REPLACE a.m.TO AM column time event- restoration
  + REPLACE noon TO PM column time event & restoration
  + REPLACE Midnight TO 05:00:00 م column time event
  + Split column by delimiter (SPACE) in column restoration
  + REPLACE a.m. TO null column data restoration3
  + Insert column year
  + Remove another column
  + Merged column restoration 1, restoration 2, time
  + Merged column restoration 1, restoration 2, restoration3 date
  + REPLACE Ongoing TO null column time restoration
  + Changed type with locale1 with type date column date event, restoration date
  + Changed type time to date column date event time, restoration date time
  + Merged column time & date event
  + Merged column time & date restoration
  + RENAMED COLUMN Date & Time Event Began, Date Time of Restoration
  + Changed type to date time
  + Trimmed text column (NERC Region- Area- Type of Disturbance)
  + REPLACE 12/08/2009 10:00:00 AM TO 13/08/2009 10:00:00 AM column data -TIME restoration
  + Changed type to duration
  + Trimmed text column (NERC Region- Area- Type of Disturbance)

**YEAR 2010**

* + Filter rows in column date equal (date only)
  + Changed type with locale1 with type date column date event, restoration date
  + REPLACE a.m.TO AM column time event- restoration
  + REPLACE noon TO PM column time event & restoration
  + REPLACE Midnight TO 05:00:00 م column time event
  + Split column by delimiter (SPACE) in column restoration
  + REPLACE a.m. TO null column data restoration3
  + Insert column year
  + Remove another column
  + Merged column restoration 1, restoration 2, time
  + Merged column restoration 1, restoration 2, restoration3 date
  + REPLACE Ongoing TO null column time restoration
  + Changed type with locale1 with type date column date event, restoration date
  + Changed type time to date column date event time, restoration date time
  + Merged column time & date event
  + Merged column time & date restoration
  + RENAMED COLUMN Date & Time Event Began, Date Time of Restoration
  + Changed type to date time
  + Trimmed text column (NERC Region- Area- Type of Disturbance)
  + REPLACE 12/08/2009 10:00:00 AM TO 13/08/2009 10:00:00 AM column data -TIME restoration
  + Changed type to duration
  + Trimmed text column (NERC Region- Area- Type of Disturbance)

**Year 2011-2014**

* Replace unknown values and ongoing in dates with null
* In date of restoration replace 18/3/2001 to 18/3/2011
* And 29/4/2014 to 30/4/2014
* And 27/8/2077 to 27/8/2011

**Year 2015-2022**

Replace unknown values in dates and times with null

**Year 2023**

* Remove the first row to get to the actual data.
* Promote Headers: Set the first row as the column headers for the table.
* Change Data Types: Adjust the "Date Event Began" column to date format.
* Remove Unused Columns: Eliminate the "Alert Criteria" column from the table.
* Adjust Data Types Again: Change various columns to appropriate types (date, time, integer).
* Merge Date and Time: Combine "Date Event Began" and "Time Event Began" into a single datetime column.
* Create a new column for "Date of Restoration", setting it to null if it's "Unknown".
* Remove the original "Date of Restoration" column.
* Rename the newly created column to "Date of Restoration"
* Create a new column for "Time of Restoration", similar to the date handling.
* Merge Restoration Date and Time.
* Insert Month Name: Add a column for the month name based on "Date&Time Event Began".
* Rename Month Name Column: Rename the column to "Event Month".
* Calculate Downtime Duration: Add a column to calculate the difference between "Date&Time of Restoration" and "Date&Time Event Began".
* Set Downtime Column Type: Change the Downtime Duration to a duration type

**Demand loss column**: -

* Replace (N\A, all, UNK, Unknown, none) to null
* Remove words like (approx., Est, less than, peak)
* Using Peak value for demand loss
* Getting average demand loss when there is range of values like (8000-10000) so demand loss will estimate 9000

**People Effected column**: -

* Replace (N\A, all, UNK, Unknown, none, PG&E) to null
* Remove words like (approx., Est, less than, greater than, over, peak)
* Assume utilities and industrial with zero
* Assume any missing data with null
* Getting average of people affected when there is range of values like (8000-10000) so number of people affected will estimate 9000

**We took the recommended analysis into consideration**

1. Are there any trends in power outages over time?

2. What types of events are most important?

3. Are there any special events that skew the data?

4. What is the most common downtime due to power outages?

**Project Team Assumptions**

As part of our analysis of the power outage data, our project team made several key assumptions that guided our conclusions. These assumptions were based on the available data and the nature of the events being analyzed:

1. **Comprehensive event coverage**: We assumed that the dataset provided a complete and accurate record of all significant power outage events from 2002 to 2023. This assumption was critical to ensure that our analysis covered the full scope of potential disruptions without missing key incidents.
2. **Broad "Other" category**: Given the high number of events categorized as "Other" (64 events), we assumed that this grouping includes a wide range of causes that were either too varied or lacked sufficient detail for classification. We relied on this assumption to simplify the categorization process, acknowledging that a more granular breakdown may have been possible with further investigation.
3. **Temporal trend consistency**: Our team assumed that the data collection and reporting processes were consistent across the entire time period. This was important for us to accurately track trends in both the number of outages and their durations over time, allowing us to draw reliable conclusions from year-over-year comparisons.
4. **Most affected region (WECC)**: We assumed that the WECC (Western Electricity Coordinating Council) was genuinely the most affected NERC region, based on standardized data collection across all regions. This assumption was key to ensuring that regional comparisons were fair and that the conclusions drawn from the data accurately reflected the impact across different geographic areas.
5. **Consistent event categorization**: We assumed that the classification of outage causes (e.g., "Vandalism or Theft," "Physical Attack") remained consistent over time. This allowed us to trust that changes in the number of events attributed to each category were genuine reflections of external factors, rather than changes in reporting or categorization standards
6. **Outage frequency and duration correlation**: In analyzing the relationship between the frequency of outages and their average duration, we assumed that there is a meaningful correlation between these variables. This assumption led us to explore the possibility that higher outage frequency may contribute to longer downtime, whether due to resource strain or other systemic factors.

By basing our analysis on these assumptions, we aimed to ensure that our findings were grounded in the most accurate representation of the data available. However, we recognize that certain assumptions, particularly regarding data completeness and categorization, could be revisited with further detailed data or refinement