This GitHub repository demonstrates the process of using Python to convert multiple weather forecast text files into an interpretable dataset. This repo is designed for the sole purpose of describing and displaying the before and after transformation of files using my developed Python script. Unfortunately, due to proprietary and intellectual property restrictions, I am unable to provide any of my developed code.

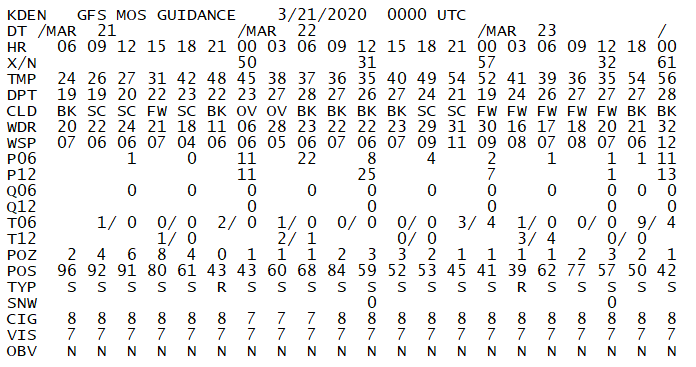
The Datasets we needed to transform from the National Weather Service (NWS) were: **The Short Range Forecast** (gfsmav files) and **The Extended Range Forecasts** (gfsmex files).

**The Short Range Forecast:**

These text files consist of various weather columns such as: temperature, daily high temperature, daily minimum nighttime temperature, etc. However, at first glance at these files it is very confusing to interpret. The data columns are pivoted as row indexes. For the columns full column descriptions you can click on this link: <https://www.nws.noaa.gov/mdl/synop/mavcard.php>

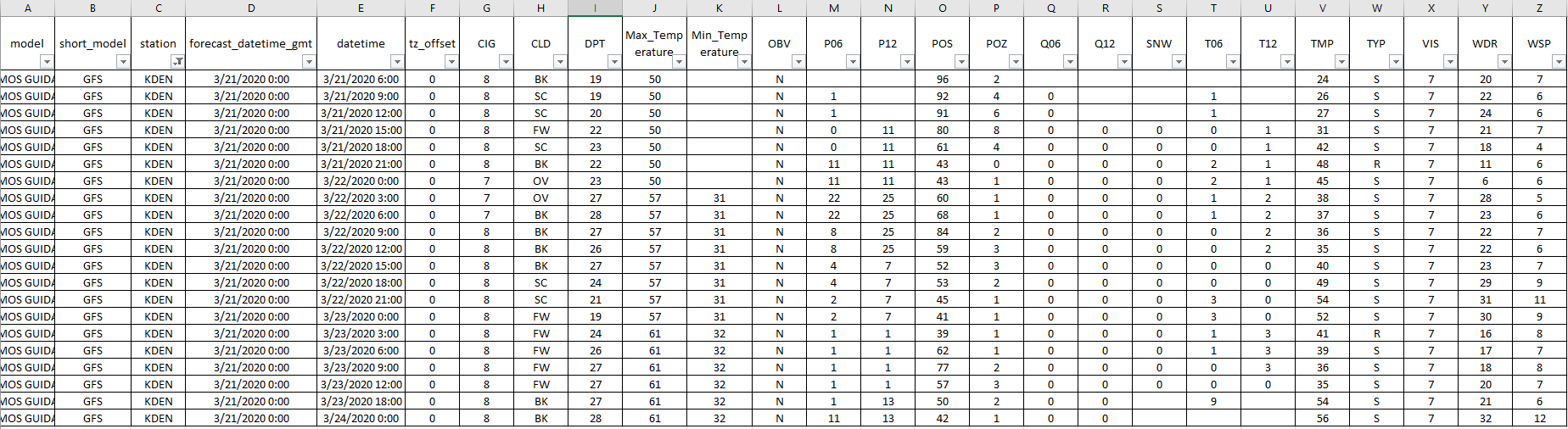
**NWS Short Range Forecast Sample Screenshot**

**mdl\_gfsmav.20200321.t00z\_raw\_20200321012030\_done.txt:**



The goal of my project was to transform these text files to a dataset that ABB Enterprise Software ([Velocity Suite](https://new.abb.com/enterprise-software/energy-portfolio-management/market-intelligence-services/velocity-suite)) could input and display. The short range forecast files are released four times in a day, corresponding with various prediction time intervals. The t##z represent the UTC time the forecast is made. There are various weather stations across the United States and its territories represented within these files.

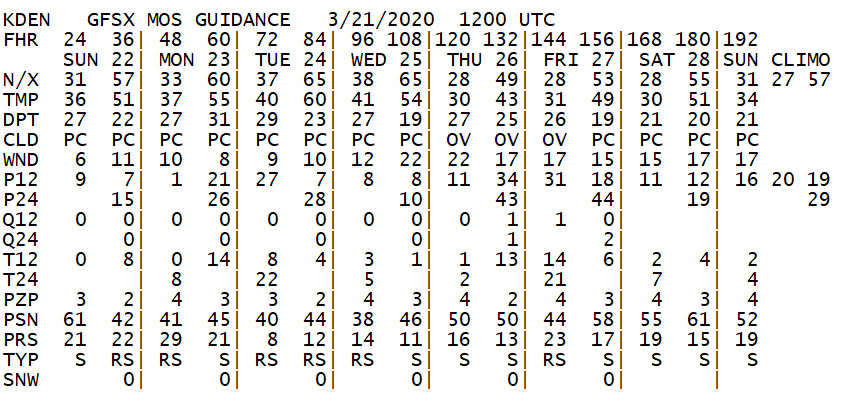
After the Python Script is done executing the files then look like this sample:

**mdl\_gfsmav.20200321.t00z\_cleaned\_20200321012030.csv**

This produced csv makes things much easier to read for the user and allows us to seamlessly import the data using our ETL processes.

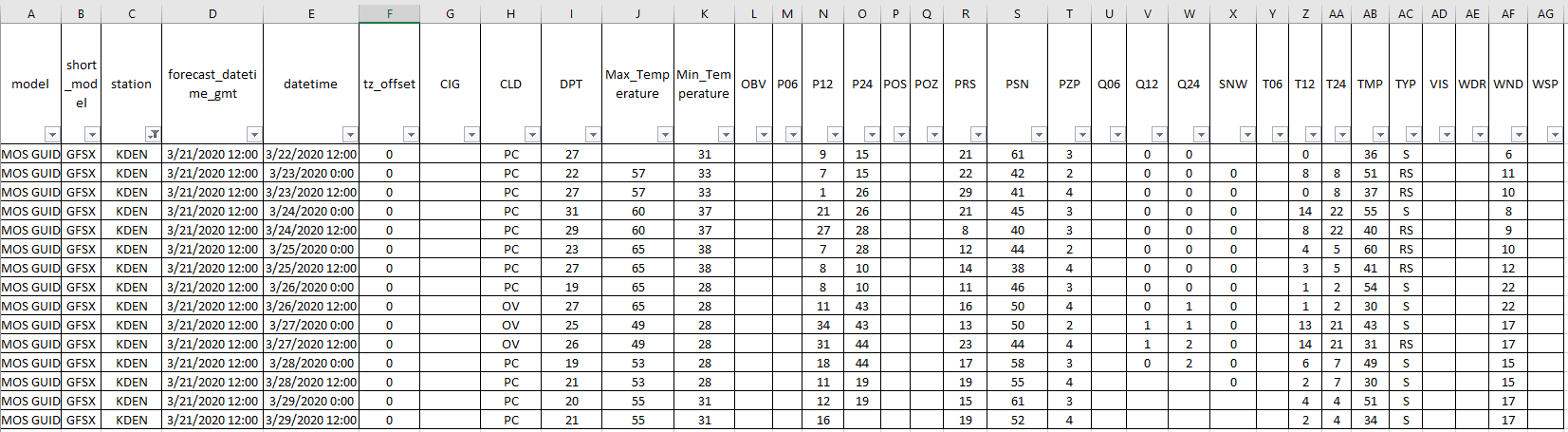
**The Extended Range Forecast:**

The Extended Range Forecast uses a very similar process to the previously mentioned Short Range Forecast. The only major difference is how the data is represented in the raw text file. The Extended Range Forecast uses | as a delimiter. Also these files are released twice per day.

**mdl\_gfsmex.20200321.t12z\_raw\_20200321132113\_done.txt**

Here is the final parsed file screenshot after executing the Python Script:

**mdl\_gfsmex.20200321.t12z\_cleaned\_20200321132113.csv**



Thank you for taking the time to look through this project description. You can contact me directly with any questions you may have via email at: christopher.lach.1@gmail.com