Steven Nodalo

6/28/2021

Open-Ended Capstone Step 4

Exploratory Data Analysis

# Archived Data

NEXRAD Archived dataset is organized in nested, zipped folders by location, date, and time. The data contained within the zipped folders is for visual analysis purposes only. No human-readable data is contained in the datasets.

# Live Data

NEXRAD Live dataset

# Twitter Data

1. Is the data homogenous in each column?

Both NEXRAD datasets are homogenous in each column, as the data will serve as images and visuals. The Twitter data is not homogenous in each column as some columns will have strings (such as the free\_text column depicting user tweets) while others will have integers (such as the date/time columns).

1. How do you anticipate this data will be used by data analysts and scientists downstream?

This data will be used to analyze the severity of the inclement weather, as well as the impacts of lives of people who are affected by the inclement weather. Additionally, this data can serve in predicting which areas are in need of updating evacuation guidelines by analyzing user Tweets of where they are evacuating to, and the ease of ability to evacuate.

1. Does your answer to the last question give you an indication of how you can store the data for optimal querying speed and storage file compression?
2. What cleaning steps do you need to perform to make your dataset ready for consumption?
3. What wrangling steps do you need to perform to enrich your dataset with additional information?