Extracting:

Links to data source <https://www.kaggle.com/sobhanmoosavi/us-weather-events>

<https://smoosavi.org/datasets/us_accidents>

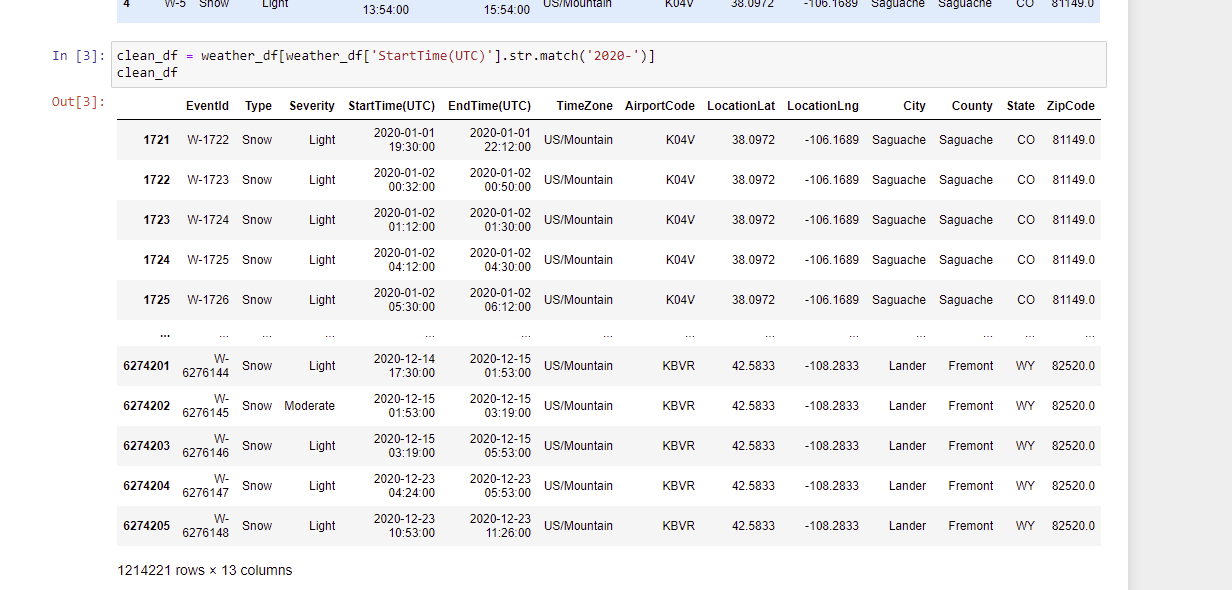
Both data sources were retrieved in csv format.

Accident table Cleaning:

Selected rows with a start time of 2016 with a str.match method

Dropped all columns but (id, severity, start time, end time, description, city , state)

Renamed columns (start time to accident start time, end time to accident end time, severity to accident severity)

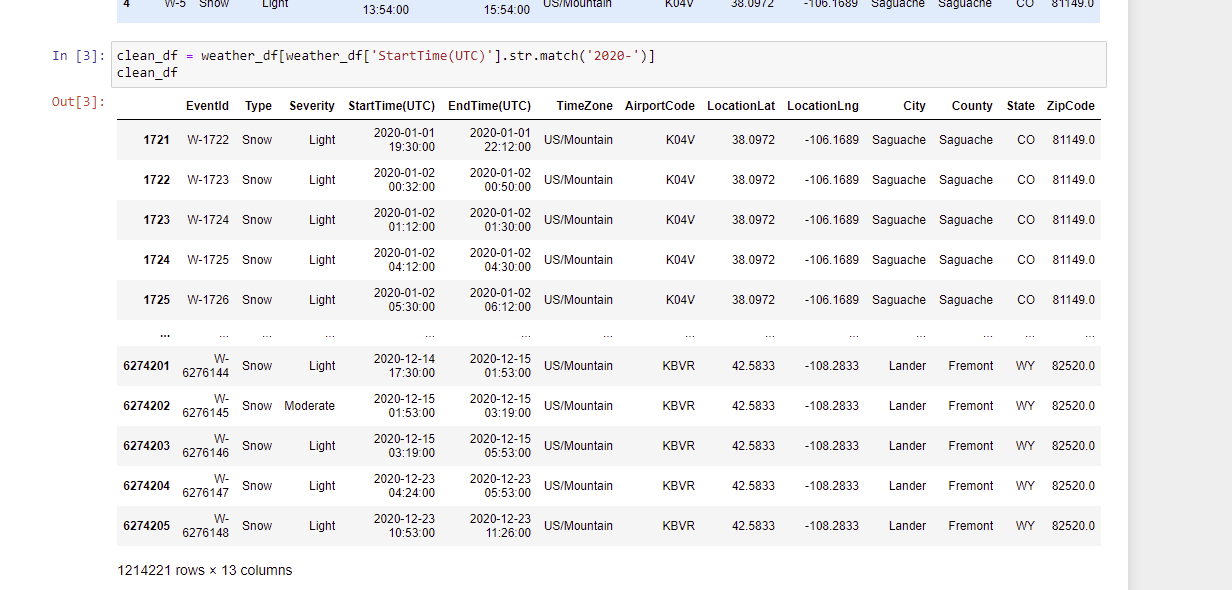


Weather table Cleaning:

Selected rows with a start time of 2016 with a str.match method

Dropped (locationlat, locationlng, county, zipcode)

Renamed columns (start time utc to weather start time utc, end time utc to weather end time utc, severity to weather severity)



While attempting to Join we received the following error

C:\Users\Brian Morton.000\AppData\Local\Microsoft\Windows\INetCache\Content.Word\etl error.png

To resolve we further trimmed both datasets down to a single city.

We split date and time into separate columns and drop time column. We further selected only first occurrence of a date.

Joining:

Tables were combined with an OUTER JOIN in pandas on the City field

Table were loaded into postgres database.

