**Chicago Crime and Weather project**

**Notes:**

Data dictionary:

* PRCP = Precipitation (tenths of mm)
* SNOW = Snowfall (mm)
* SNWD = Snow depth (mm)
* TMAX = Maximum temperature (tenths of degrees C)
* TMIN = Minimum temperature (tenths of degrees C)

Database GHCND Stations

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Variable** | **Columns** | **Type** | **Example** | | --- | --- | --- | --- | | ID | 1-11 | Character | EI000003980 | | LATITUDE | 13-20 | Real | 55.3717 | | LONGITUDE | 22-30 | Real | -7.3400 | | ELEVATION | 32-37 | Real | 21.0 | | STATE | 39-40 | Character |  | | NAME | 42-71 | Character | MALIN HEAD | | GSN FLAG | 73-75 | Character | GSN | | HCN/CRN FLAG | 77-79 | Character |  | | WMO ID | 81-85 | Character | 03980 | |

Firstly, I located the appropriate data which had been pre-processed for this project. BigQuery hosts various of free datasets ranging in sizes for a variety of domains. For this project I used the GHCND dataset which collects global weather data from thousands of distributed weather stations and sensors and stores in a tabular format.

Data Prep –

SQL in BigQuery

The data was not initially presented as useful immediately. I needed to

1. Work out which stations are in Chicago using LAT/LON values from GMaps and then filter this in the WHERE statement with a nested query.
2. Find a matching column (I decided to use dates and re-cast the table with dates in UTF timestamp format).
3. INNER JOIN to match each record from crime dataset with records from stations dataset matching on dates using the previous sub-select query to search within the correct set of Chicago stations before exporting the results of the query to Tableau.

**Findings**

* Higher temperatures are associated with much higher incidences of theft, battery and criminal damage.
* The more extreme the weather value the higher the frequency of serious property and violent crimes

Now the data has been imported from BQ into Tableau Public I performed some initial EDA to observe and clear relationships. By plotting a horizontal bar graph, we can see the weather types (TMAX, TMIN, SNOW etc) and which crimes they are associated with and to what extent. I decided to drop the row values that has weak correlation and focus on the weather types which have a strong relationship with both the frequency and the type of crime in the city.

In this stage we also need to explore the crimes that have the highest frequency *within* the weather group type. E.g., we can see that TMAX is associated with much higher levels of violent incidence but some types such non-motor vehicle theft and battery have a much higher correlation with higher temperatures than say criminal trespassing.

Initial findings indicate that the crime associated with the highest values in weather types is theft. High incidences are associated with higher weather values i.e. the data shows the higher/ lower the temperature the more arrest reports of theft. Interestingly, in the top 3 weather categories associated with high numbers of crimes, arrests for battery are also exceptionally high. This could be due to the level of violence required for a successful theft. In all three categories, criminal damage also ranks highly.