Extract data from CSV

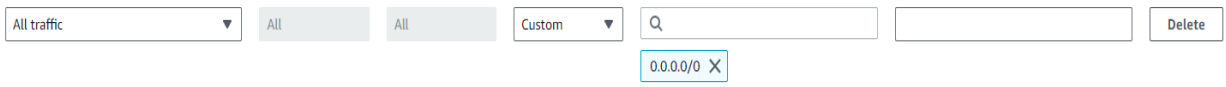
1. The crime data is in csv format. In Python, simply use the Pandas to read in data and store the data as Pandas DataFrame.
2. The FBI crime data is being extracted using API calls from the web. https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/docApi

Data Transform

1. Drop the duplicates in the data.
2. For datetime string, convert it into datetime object and create new columns for year, month, day, weekday and hour respectively.
3. Fill in the null values based on the situation.
4. Merge the crime data and FBI crime data together using Fuzzy Matching.

Load data to AWS Redshift Serverless

1. Create an account for AWS.
2. Instead of using the root, adding a user could be helpful for security
3. Go to the AWS Redshift and create a workgroup, keep every setting in default.
4. Open the Query Editor v2, connect to the workspace using the non-password option.
5. Create a user and password.
6. Go to the workgroup setting. Under network and security, switch on the public accessible to on.
7. Under VPC, add inbound rule:



8. Go to Python or Jupyter notebook, set up the connection using the psycong2 package, where the host and port can be seen from general information of the workgroup setting.