Ravindra Neralla

DSC540- Data Preparation- Project Milestone 5

This project involves analyzing Chicago crime data for the year 2019 with respect to weather and holidays. One of the major difficulties I found while in this project is on getting holidays data into python from website. I could not get the required the data using web scraping, put some extra effort to learn on how to identify the required data and capture into data frame. Chicago weather data is sourced using weather API.

After sourcing data from all three data sources, worked on transforming the data to standard format to join with other data set, in this process date formatting are changed to make it standard date. Data is finally loaded into three tables in sqlite database, learned on how to connect to database and read data from database.

In the final step, understanding the trends in the crime data and plotting using pyplot and seaborn packages. It is observed from the plots that crime is more during summer months compared to winter months. Also number of crime incidents are more on holidays compared to regular day. As a whole, it was good learning on how to transform data using python. I really enjoyed learning this subject; it is close to my regular ETL work and hope to work on it more in future.