Reflection

The project on creating an ETL data processor gave insight to how large, useful data is digested and manipulated to find resourceful information regarding the digital world. In our project, we focused on car accidents and inclement weather, using data from Open-Meteo Weather API source to detect the effects of harsher environments on the likelihood of motor vehicle collisions. Information about public accidents were established in a CSV file from a Kaggle data set and was merged with the API source information into an SQL database.

Since this assignment works with large amounts of data, cleaning and detailing the information needed was challenging as it strained google colab and personal devices. Following that similar idea, finding the right API that was accessible and efficient for the project was difficult. While there are many public open API sources, some did not have sufficient rate limits for testing. This led to a shift in our original ideas, but we were able to find a public weather source API for our project. Additionally, debugging the program as the code program is longer than our other assignments, which led to a lot of time in fixing any mechanical and syntax errors.

Along with these challenges, we also had aspects of this project that were easier to handle. After our class discussions and prior assignments on API, data extraction and retrieval from an open source was easier than expected. The API we utilized does not require a key and allowed 10,000 API calls a day, which was more than enough for testing. This was extremely helpful in collecting our data as we could look at temperature, humidity, cloudiness, precipitation, etc. Additionally, finding the right libraries to allow for set functions to be performed is vital. From class, it was known that a wide range of libraries are available to set functions within our code. Datetime library was used to process the chronological factors, such as start times for all the car crash incidents in the CSV file, and map it to the temperature and weather conditions at that time at a given location. Pandas was also used to create data frames and simplify our data cleaning.

However, there were unexpected difficulties with finding accessible APIs for our project. We originally planned on using Yelp restaurant reviews to assess against health code violations of the top restaurants in the area. However, this API that was essential and efficient for this idea was not accessible to us as students, which led to some more brainstorming. Although this process is natural in the brainstorming process, it was more difficult than originally planned.

This project's output and data can be useful for future references as it provides useful information on the connection between car accidents and weather. The data correlation found between various temperatures and precipitation can be used to articulate a resource weather app that warns users about possible accidents and collisions on the road. This information can also be used to enhance car safety and manufacturing based on the weather, creating safer vehicles for consumers. By doing so, the car industry will be affected as weather changes can manipulate consumerism and the stock market. Buyers may choose certain cars over the others based on weather compliances, saturating and manipulating the market to favor specific car brands or models.

Our project works to utilize ETL as a tool to analyze information on the connection between the weather and car accidents, creating a resource to help drivers understand road safety and the effects of weather on their driving ability in a given car.