Rebecca Melo

12/1/2021

**ETL Project**

**Topic**: Daily Travel during the COVID-19 Public Health Emergency vs. Accidents in the US

**Data sources**:

* <https://www.bts.gov/daily-travel>
* <https://www.kaggle.com/sobhanmoosavi/us-accidents>

**ETL:**

Extract – Extract the data from the csv files

Transform – Create tables from file: Based on State, Frequency, and Distance, etc. I will then filter the data and and dropped unwanted columns and null values by using .dropna() and use .fillna(). I will then create an index to act as IDs, where these IDs will be used as primary keys. Once I create the tables, I will use pd.merge() to join the tables with the original data to include the new ids. Excess columns will be removed as well. The IDs in these tables will act as foreign keys in the main data, acting as primary keys in the index tables.

Load – Loaded data into PostgreSQL.