**ETL Project Report**

**By Achyut Shrestha, Kiran Babuk, Sam Deschamps, Mark Gotanga**

Extract:

Our original data sources were 7 csv files located within the “Transit Systems of the World” dataset on Kaggle.com.

Transform:

We reviewed the data for any null values within the columns, and then we verified the possibility to drop those null values. We also split columns where necessary, such as the splitting latitude and longitude into two separate columns . Then we removed the duplicate columns that were shared between the individual csvs. Then we loaded the data into postgres tables and queried tables in order to confirm the availability of our data. To verify nothing went wrong while removing duplicates we took a count before we removed them, and then after.

Load:

Our final database in a comprehensive dataset outlining the relevant data regarding Transit Systems around the World. The final dataset reflects our goal of removing redundant and superfluous data and retaining data that may be relevant to the study of global transit systems.