Group 9

Spr. 2018 95-797 A3

ETL Procedure Summary

First we cleaned our data available on from the NYC Transportation Department website using Python and R. Next, we used Microsoft SSIS to import data into dimension tables using identity functionality to create surrogate keys. To simulate a Type II dimension change, we changed the source data using R and reloaded the Dim\_Location and Dim\_ Issuer using SCDII Wizard. We then imported data into temp staging table which includes all the natural keys from different dimension tables and used the lookup function in SSIS to map them with the primary keys in fact table to build up the warehouse.