The 4014_parser.py document reads the source csvs from the current working directory, cleanes, transforms them and creates 9 dataframes. Data is split, transformed and cleaned in order to achieve the schema explained in the Documentation file.

After Data extraction, transformation and cleansing 9 csv are produced in the current working directory. The new csvs are named: Characters_master_data, Characters_relationships_data, Episodes_Characters_data, Episodes_LocSubloc_data, Episodes_Master_data, Episodes_Opening_Loc_data, Episodes_Scene_Times_data, Location_Master_Data and Location_Sublocation. As a last step, 4014_parser.py creates a sql file named 4014_data.sql with copy statements for the produced csvs. To do so, it gets the current working directory and writes it in the copy statement.

To automatically load the data and produce the schema explained in the Documentation file navigate to the directory P1 and run 4014_parser.py. Parser_4014.py will read the source csvs and produce the 9 csvs described above. Furthermore, it will produce a new 4014_data.sql file with copy statements. These statements will have updated paths, ready to be loaded in postres. Run 4014_schema.sql file in postgres, this will create the tables needed. Run 4014_data.sql, this sql file will have copy statements to the new csv along with paths pointing at them.