CS315 - Database
Assignment 1
Date: 14/02/2020

Name: Sarthak Singhal Roll No.: 170635

1. I created 6 tables in the database. First table stores the State code, Rural or Urban, Age group, Bilingual people and Trilingual People. Second table is similar to first table with instead of "Age group" it has "Education level". As I stored just the "State code" in these tables I created one more table to store the "state code" corresponding to the "district code" and the "area name". In fourth table I stored the state code, Rural or Urban, Age group, Total population, literate population and illiterate population. In another table I store the population of different categories of literate people with the state code, rural or urban and the age group.
   In some places there is redundancy of data for e.g. "Total" is stored together with "Male" and "Female". Another example can be data of India is stored although it could be calculated by data of the states. But most of the queries were using this data so I decided to store the redundant data because solving queries from raw data will be less efficient. In this trade off between space and time I chose time over space.
   Also there were differences in the age groups in multilingual-age.csv and age-education.csv for e.g. one had age group 5-9 while other had 0-6,7,8,etc.. In some queries both these tables were required so to handle this difference between age groups I created another table which stores the mapping from age group present in one csv file to age groups present in other csv file.

2. As in some tables I didn't store district code and area name(as I just stored state code), I used "awk" to take the desired columns from the given csv files. As some lines in the csv file included headers I used "tail" command to ignore those lines as they were in the starting. I created the csv file storing mapping between age groups using "echo" command. Then I created the tables by "create table" command. Then I imported the data from the csv files to the tables.
   I have included the script for creating and inserting in the database named **create.sh**. I have also included the csv files that were given with the assignment. They should be present in the same directory in which the script is present.
   Execution Instructions:
   - chmod 700 create.sh
   - ./create.sh
3. I have included a file named **query.sh** to process the given queries.
   Execution Instructions:
   Query number is given as a command line argument to the script. For e.g. if you want the answer of query 2 write "./query.sh 2". The database created using above script should be in the same directory in which query script is present. To run the script run these commands:
   - chmod 700 query.sh
   - ./query.sh *query number*