Read Me

(Group 20)

(Anamitra Maji (MT19112), Pragya Dara (MT19126), Swati Verma (MT19073))

1. Use the .sql file to create the relational database and tables in it. This database would be our data warehouse.
2. To insert the data into tables use the python code. Tu run the code following libraries are required:
    - My-sql connector
    - Pandas
    - Numpy
    - Datetime
    - time
3. For each table a function has been created to insert the data into that table. The input of the function is a tuple containing the values to be inserted in the sequence of the insert query written in the function.
4. To insert the data insert dimensions first and then facts. If data is tried to be inserted in facts without inserting data in dimensions then a message would be printed and data won't be inserted.
5. Data can also be inserted using the csv files.
6. Following preprocessing is done in the transformation steps.
    - Data type conversion.
    - Handling null values by replacing with default values.
    - Removing duplicates.
    - Validating the key data for facts.

Flow of the code is as follows:

Base Dimensions -> Base Facts , Aggregate Dimensions->Aggregate Facts