Civic Data Lab Assignment

1.First I have mined the data from link. Using selenium I have created the driver and then by finding the elements.

    a) Selected the dates.(Calander)
    b) Selected Demand and HOA wise Summary.(Dropdown list)
    c) Selected Rupees (Radio Button).
    d) Clicked on View Data.
    e) Selected tables and rows.

2. Then I wrote the data to a csv file.

3. Read the CSV file using pandas and performed cleaning on the raw dataset.

    a) Filled all NaN values with 01-VIDHANSABHA.
    b) Then dropped the rows containing 'Grand Total' from DmdCd and 'Total' from HOA.
    c) Splitted DmdCd into DemandCode and Demand.
    d) Splitted HOA column to get 10 different columns.
    e) Renamed some of the columns.

4. Created connection with SQLite3 and pushed the dataframe to the database (mydb).

5. Integrated all the steps by creating DAG using Airflow which will run the data pipeline everyday.