

## **Database: hdb**

The final csv file is in : /Staging/Main\_final.csv

**Or**

The final populated database table: main

### **You may run the file in sequence**

1. HDB01\_Downloader.ipynb
2. HDB02\_DataTransformer.ipynb
3. HDB03\_StatFileProcessor.ipynb
4. HDB04\_DataMerger.ipynb
5. HDB05\_DataLoader.ipynb

### **No need in sequence**

6. HDB06-SQL\_Calls.ipynb
7. HDB07-Data Analysis(sample).ipynb

#### **HDB01\_Downloader.ipynb**

- Read file from website
- Deposit files in <root>/DataLake
- Combine downloaded file into one
- Demo file downloading

#### **HDB02\_DataTransformer.ipynb**

- Using the datasets/Main.csv (From Shazlin)
- Perform transformation and cleaning
- Create new columns
- Deposit files: staging/Main.csv

#### **HDB03\_StatFileProcessor.ipynb**

- Process files:
  - datasets/Births.csv
  - datasets/Inflation.csv
  - datasets/Unemployment.csv
  - datasets/Divorces.csv
  - datasets/GDP.csv
  - datasets/Marriages.csv
- Deposit files:
  - staging/stat\_monthly.csv
  - staging/stat\_yearly.csv

#### **HDB04\_DataMerger.ipynb**

- Merge Files
  - staging/Main.csv

- staging/stat\_monthly.csv'
- staging/stat\_yearly.csv
- Deposit file:
  - staging/Main\_final.csv

#### HDB05\_DataLoader.ipynb

- Populate Databases into Postgress Database: hdb
- Populate following Database tables
  - main
  - stat\_monthly
  - Stat\_yearly
- Populate secondary database tables:
  - (may not use, just demo database operation)
  - births
  - gdp
  - marriages
  - divorces
  - inflation
  - Unemployment
  -

#### HDB06-SQL\_Calls.ipynb

- Demo SQL Call

#### HDB07-Data Analysis(sample).ipynb

- Demo matplotlib, seaborn