

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