***Final Project Goal Setting:***

Part of Weeks 11 & 12 assignment, I targeted to load data into database table from 2 source system. One source is from website data and another source is from csv file formatted data. Upon importing each source system data into database table, target is to make it as single dataset. Using 2 different data sets or merged dataset goal is to visualize the data in different formats like generic chart, bar chart, linear chart etc…

***Python Code:***

#import os #import the os library (enables operating system dependent functionality)

*import requests*

*import pandas as pd*

*import sqlite3*

*import os*

*import matplotlib.pyplot as plt*

#Step-1 -- Read data from website and load into database table and print the data from database table.

#Read the data from the website and store same into CSV format for easy data access

*url = 'https://apps.who.int/gho/athena/data/GHO/E1\_nat\_tv\_radio,E2\_intl\_tv\_radio,E3\_nat\_print,E4\_intl\_print,E5\_billboards,E6a\_pt\_of\_sale,E7\_internet,E21a\_direct\_ad\_fines,E\_compl\_all\_dir,E9\_free\_distrib,E10\_promo\_discounts,E11\_brand\_stretching,E12\_brand\_sharing,E13\_brand\_placement,E14a\_prod\_tv\_films,E14b\_movies\_anti\_tob\_ads,E15a\_sponsorship,E15b\_sponsor\_contribution,E15c\_sponsor\_publicity,E21b\_indirect\_ad\_fines,E\_compl\_e\_all\_indir,E17a\_csr\_ban,E17b\_csr\_promo\_self,E17c\_csr\_promo\_others,E18\_csr\_anti\_tobacco\_media,E6b\_ban\_display\_pt\_of\_sale,E23\_vending\_machines,E24\_internet\_sales\_ban,E22\_subnational\_authority\_exists,E\_comprehensive\_subnat.html?profile=ztable&filter=COUNTRY:\*;REGION:\*'*

*html = requests.get(url).content*

*df\_list = pd.read\_html(html)*

*df = df\_list[-1]*

# Below can be used to export data into CSV file and read the same

*#df.to\_csv('my\_data.csv')*

*#website\_data = pd.read\_csv('my\_data.csv')*

# create database table to store website data

*query = """*

*CREATE TABLE web\_source\_data*

*(S\_No VARCHAR(20), GHO VARCHAR(100),*

*PUBLISHSTATE VARCHAR(20), YEAR VARCHAR(20),*

*REGION VARCHAR(20), COUNTRY VARCHAR(20), DISPLAY\_VALUE REAL,*

*NUMERIC\_VALUE REAL, LOW\_RANGE VARCHAR(20),*

*HIGH\_RANGE VARCHAR(20), COMMENT VARCHAR(20)*

*);"""*

*con = sqlite3.connect(':memory:')*

*con.execute(query)*

*con.commit()*

# Write the website data into a sqlite table

*df.to\_sql('web\_source\_data', con, if\_exists='replace', index=False)*

# Create a cursor object

*cur = con.cursor()*

# Fetch and display website loaded data into database table, only top 3 rows

*for web\_row in cur.execute('SELECT \* FROM web\_source\_data LIMIT 3'):*

*print(web\_row)*

# Close connection to SQLite database

*con.close()*

#Step-2 -- Read data from CSV file and load into database table and print the data from database table.

*os.chdir('C:/Users/vahin/OneDrive/Documents/GitHub/DSC540Spring2021/DSC540/DSC540/Week11&12') #change directory*

*os.getcwd() #get the current working directory to confirm the directory change*

*csv\_data = pd.read\_csv('xmart.csv')*

# create database table to store csv source data

*query1 = """*

*CREATE TABLE csv\_source\_data*

*(Country VARCHAR(100),*

*YEAR VARCHAR(100),*

*Ban\_ntl\_tel\_radio VARCHAR(100),*

*Ban\_adv\_tel\_radio VARCHAR(100),*

*Ban\_int\_tel\_radio VARCHAR(100),*

*Ban\_local\_mag\_new VARCHAR(100),*

*Ban\_int\_mag\_new VARCHAR(100),*

*Ban\_bill\_out\_mag VARCHAR(100),*

*Ban\_adv\_pos VARCHAR(100),*

*Ban\_adv\_int VARCHAR(100),*

*Law\_vio\_adv\_ban VARCHAR(100),*

*Overall\_compl\_score VARCHAR(100),*

*Ban\_free\_mail VARCHAR(100),*

*Ban\_prom\_dis VARCHAR(100),*

*Ban\_non\_t\_brand\_names VARCHAR(100),*

*Ban\_non\_t\_products VARCHAR(100),*

*Ban\_t\_TV\_Films VARCHAR(100),*

*Ban\_app\_TV\_Films VARCHAR(100),*

*Requiredanti\_timages VARCHAR(100),*

*Ban\_sponsorship\_publicity VARCHAR(100),*

*Ban\_form\_contribution VARCHAR(100),*

*Banning\_publicity\_individuals VARCHAR(100),*

*Law\_requires\_fines VARCHAR(100),*

*Overall\_compliance\_score VARCHAR(100),*

*Ban\_CSR VARCHAR(100),*

*Ban\_tobacco\_CSR VARCHAR(100),*

*Ban\_tobacco\_companies\_CSR VARCHAR(100),*

*Ban\_\_smoking\_prevention\_youth VARCHAR(100),*

*Ban\_product\_POS VARCHAR(100),*

*Ban\_tobacco\_vending\_machines VARCHAR(100),*

*Ban\_internet\_sales VARCHAR(100),*

*Subnational\_bans\_promotion\_sponsorship VARCHAR(100),*

*Subnational\_bans\_comprehensive VARCHAR(100)*

*);"""*

*con = sqlite3.connect(':memory:')*

*con.execute(query1)*

*con.commit()*

# Write the csv data into a sqlite table

*csv\_data.to\_sql('csv\_source\_data', con, if\_exists='replace', index=False)*

# Create a cursor object

*cur = con.cursor()*

# Fetch and display csv loaded data into database table, only top 3 rows

*for csv\_row in cur.execute('SELECT \* FROM csv\_source\_data LIMIT 3'):*

*print (csv\_row)*

# Close connection to SQLite database

*con.close()*

# HAVING CODING ISSUE TO LET JOIN WEBSITE SOURCE AND CSV FILE DATA SOURCE INTO SINGLE DATASET

#Join the datasets together in Python into 1 dataset

*for csv\_web\_row in cur.execute('SELECT \* FROM csv\_source\_data outer join web\_source\_data on web\_source\_data.YEAR = csv\_source\_data.YEAR LIMIT 3'):*

*print (csv\_web\_row)*

# HAVING CODING ISSUE TO VISUALIZE THE DATA

#visualizations the data

#Creating the generic chart

*csv\_web\_row.plot(x="YEAR", y=["Country"])*

*plt.plot(csv\_web\_row['YEAR'], csv\_web\_row['Country'])*

#Show the plot

*plt.show()*

***Pending tasks to complete the project:***

* Had difficulties to import API / JSON file format data into database table, so that it can also be another source form of data.
* Having coding issues to let join website source database table and csv file data source database table into single data source.
* Having coding issues to visualize the data reading from database table.