Vedant Raulkar

Dashboard Project Documentation

Project Overview

This project visualizes India-Australia trade data through a web-based dashboard. The data includes export and import trade values, cleaned and presented interactively using Python-based tools.

Libraries Used

- Pandas: Data manipulation and cleaning.
- Dash: Web framework for creating the dashboard interface.
- Plotly: Graph plotting for interactive visualizations.
- Selenium: Web automation for scraping trade data from the source website.
- WebDriver Manager: Managing Selenium WebDriver binaries.
- OpenPyXL: Reading and writing Excel files.

Data Source

The trade data is scraped from the Commerce Ministry Trade Statistics website (https://tradestat.commerce.gov.in/eidb/default.asp).

Important Note: The website currently does not provide data for certain months, leading to gaps in the dataset.

Project Components

- Export_scrap_year.py: Scrapes export trade data for all available years and stores the data in an Excel file (Export_trade_data.xlsx).
- Import_scrape_year.py: Scrapes import trade data for all available years and stores the data in an Excel file (Import_trade_data.xlsx).
- clean.py: Cleans the scraped data by removing invalid rows and formatting numeric columns. Outputs cleaned data to cleaned_Export_trade_data.xlsx and cleaned_Import_trade_data.xlsx.
- Dashboard.py: Creates an interactive dashboard to visualize export and import data. Users can select commodities and years to filter data.

Automation Limitations

The project does not implement the most optimal automation solution. To update the data:

1. Run the Export_scrap_year.py and Import_scrape_year.py scripts to scrape the latest data.
2. Run the clean.py script to clean the exported data.

3. Use the cleaned files (cleaned_Export_trade_data.xlsx and cleaned_Import_trade_data.xlsx) with the dashboard.

Dashboard Features

- Export Visualization: Line and bar charts for export data.
- Import Visualization: Line and bar charts for import data.
- Filters: Dropdown menus to filter by year and commodity.
- Interactive Layout: Allows users to explore trade data visually.

Instructions for Running the Project

1. Install the required Python libraries:

pip install pandas dash plotly selenium webdriver-manager openpyxl

2. Run the scraping scripts to fetch and clean the data:

python Export_scrap_year.py

python Import_scrape_year.py

python clean.py

3. Start the dashboard:

python Dashboard.py

4. Open the dashboard in a web browser (usually at http://127.0.0.1:8050).

Future Improvements

- Automate the scraping and cleaning process into a single script.
- Handle missing months more effectively.
- Optimize the dashboard for larger datasets.