Did the U.S. Win the Trade War Against China? A Data-Driven Analysis (2018-2024)

1. Problem Statement

- Did the U.S. win the trade war against China? A data-driven analysis (2018-2024).
- Have they decoupled?

Background

The U.S.'s domestic production is less than its domestic consumption. An economic conflict between China and the United States has been ongoing since January 2018, when U.S. President Donald Trump began setting tariffs and other trade barriers on China with the goal of forcing it to make changes to what the U.S. says are longstanding unfair trade practices and intellectual property theft. For many decades, China and the United States have been locked in such a tight economic embrace that it is challenging to quantify whether, how, or why the embrace may be weakening. Efforts to reduce the deep economic interdependence between the U.S. and China—built over decades—have accelerated due to growing political and ideological tensions. This push for decoupling stems partly from policies introduced during President Donald Trump's administration, notably the trade war, which targeted specific industries with tariffs and export restrictions. Many of these policies have been maintained and expanded under President Joe Biden, indicating bipartisan support for reducing reliance on China.

In addition to trade policies, the need for more diversified and resilient supply chains has emerged as a significant factor driving decoupling efforts, particularly in critical sectors like semiconductors, pharmaceuticals, and rare earth materials. Furthermore, human rights issues, concerns about democracy, and broader geopolitical rivalries have increasingly influenced both U.S. and Chinese policymakers, adding a moral and strategic dimension to the economic separation.

This evolving landscape reflects a broader trend where economic strategy is intertwined with national security and values, leading to selective decoupling rather than a full-scale economic split. But the data also show something else? Let's see.

2. Data Source

- Open-source datasets related to trade, tariffs, and economic indicators.

Sources and Content

Sources - 1

- **Harmonized Tariff Schedule of the United States**: The Home of the U.S. Government's Open Data. Publisher: Office of Tariff Affairs and Trade Agreements
  This dataset is the Harmonized Tariff Schedule plus all revisions for all years of interest. It provides the applicable tariff rates and statistical categories for all merchandise imported into the United States. It is based on the international Harmonized System, the global system of nomenclature that is used to describe most world trade in goods.

Why did I choose this? This contains Special Rate of Duty, Additional Duties for each category of goods imported to the USA along with units in time series. The data is large and should be transformed with time and commodity category.

Structure These yearly published tariff details are in CSV format with all columns consistent every year.

Quality The data has been revised periodically and is systematic. Except for the year 2019, all data are clear and complete.

Access & Use Information

- **Public**: This dataset is intended for public access and use.
- **License**: us-pd (Link)

URLs:

- 2024 Revision 9
- 2023 Revision 11
- 2019 Revision 20
- 2020 Revision 18
- 2021 Revision Basic 12
- 2022 Revision 12
- 2016 Basic Delimited
- 2018 Revision 14

Sources - 2

- **USA Trade® Online**: Provided by the U.S. Census Bureau, USA Trade® Online is a dynamic data tool that gives users access to current and cumulative U.S. export and import data.
  Link

  Why did I choose this? I needed to analyze the consumption of products by USA customers. This provides the raw materials and direct finished goods consumption. Further, I can also analyze the consumption trends if commodities from other countries have been used more.

  Structure These yearly published import details are in CSV format with all columns consistent every year. All the CSVs have many dimensions like countries and their imports.

  Quality The data from Afghanistan imports of sheep in June 2021 has 13. This questions the quality as to how it was counted.

  Access & Use Information

  - **Academic**: This dataset is intended for academic access and use.
    FAQ website question 11 says that we do not need permission to use or publish data. For more information, this toll-free number provided by the website has been contacted: 1-800-549-0595. The census bureau help desk said that the data can be used and published for academic use and citation should be provided.

    This licence will be strictly followed by me by providing citations, un-manipulated and without bias.

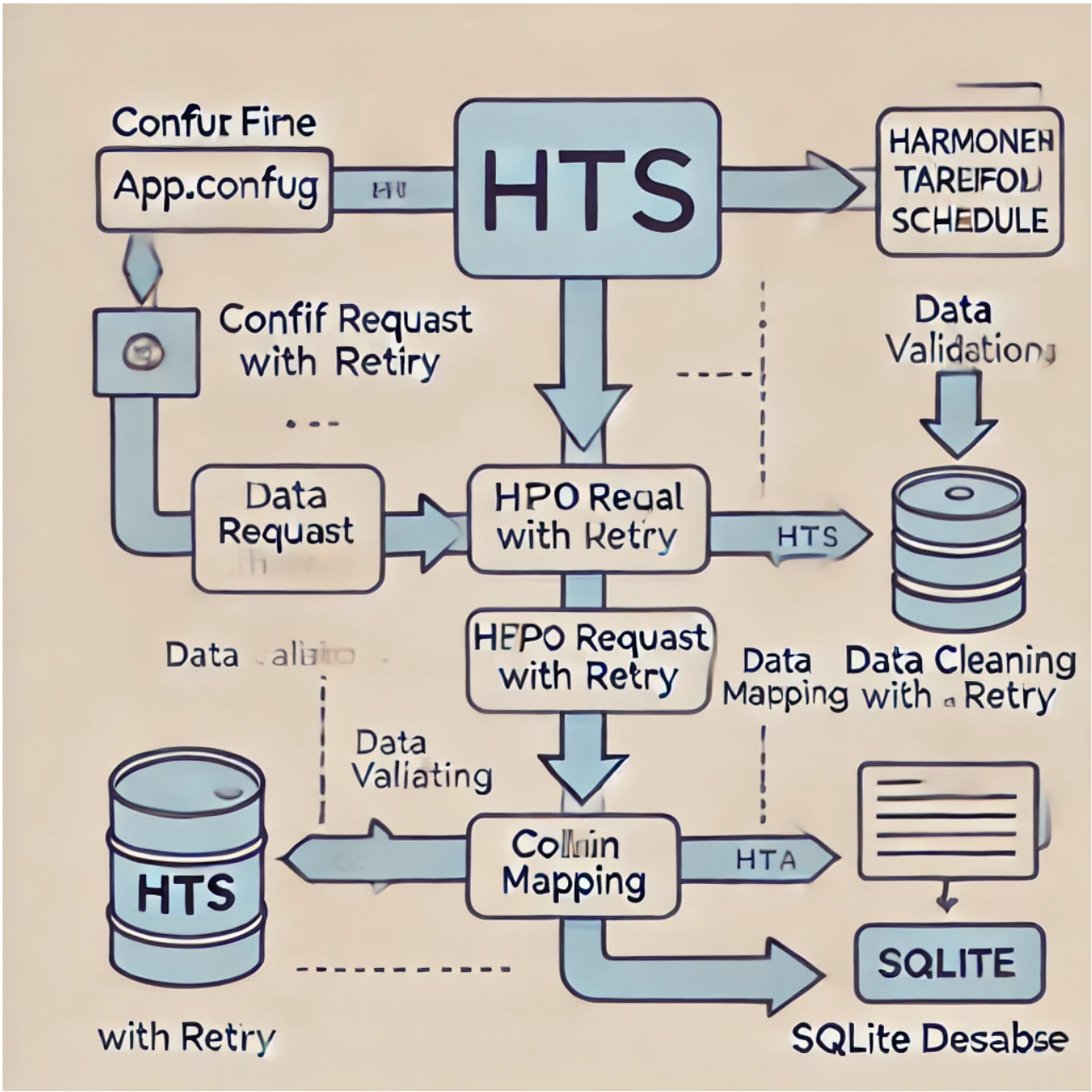
3. Data Pipeline Overview

Technologies Used

- **Python** for data transformation and automation.
- **Pandas** for data wrangling, **SQLite** for storage, and **Selenium** for automation.

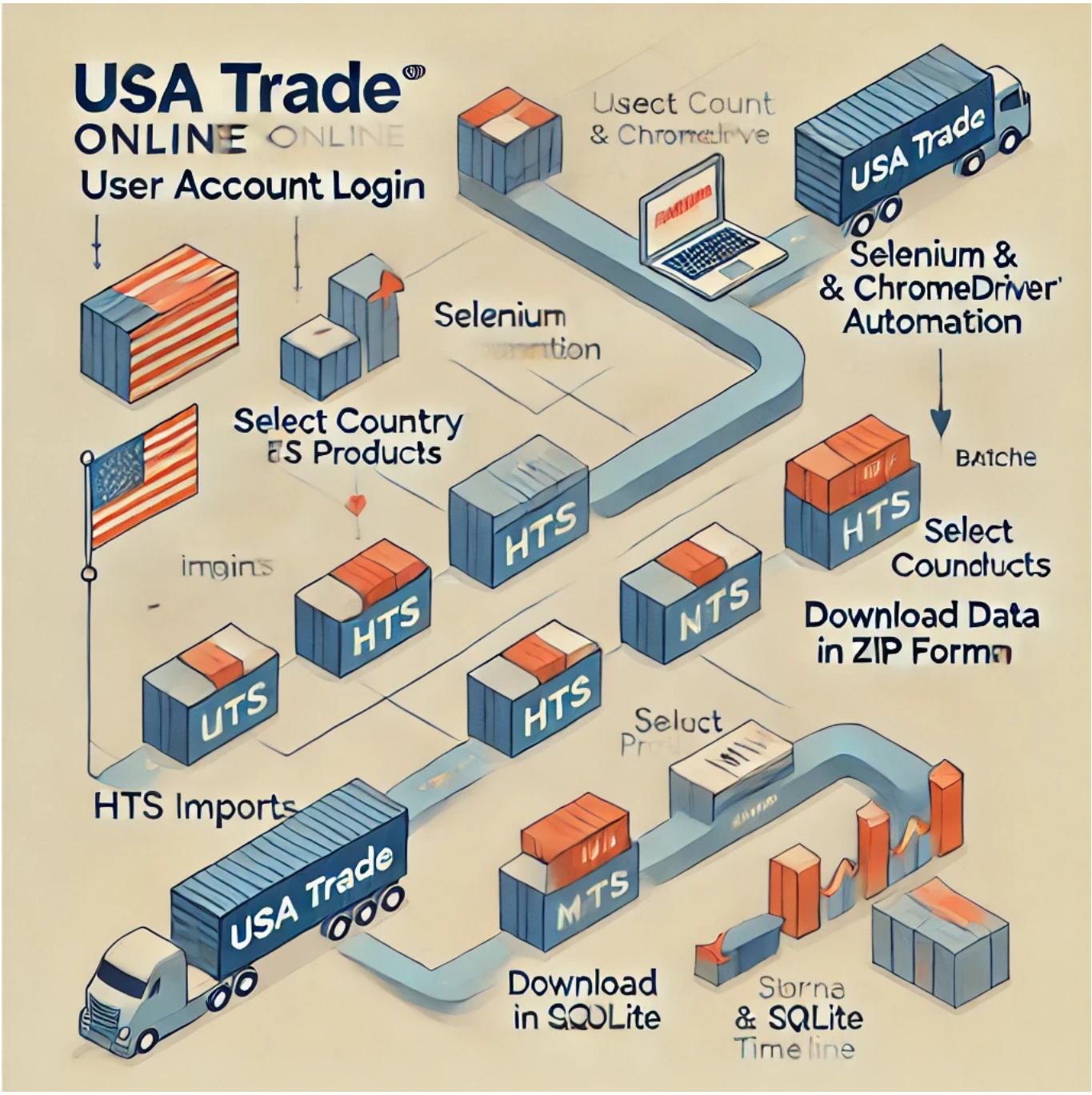
Sources - 1

- **Harmonized Tariff Schedule of the United States**: The pipeline involves importing data from URLs with multiple retries if the database is busy. Since there are multiple URLs for each year, they are placed in a configuration file (`app.config`) so that the URLs can be updated easily if they change. The function reads the URLs from this config file. Each year's file contains tariff data for each HTS. The Description column is not uniform and contains many product descriptions that change each year. The USA adds many goods with different descriptions under the same HTS. The description should not be empty, and I have ensured in the code to exclude those. The data is stored in SQLite year-wise with columns mentioned in the data source and arranged in increasing order of time source.



Sources - 2

- **USA Trade® Online**: This pipeline requires a user account to access the data. Since the website is interactive, I automated the clickable events with Selenium and ChromeDriver. This automation searches imports of HTS, selects country, products, and timeline, and downloads the data in ZIP format. Due to the large data size, I designed the process to download in batches. The data contains HTS, countries, import counts, units, and the timeline when it was imported. This was challenging as it involves automation of clickable events. The data is stored in SQLite format year-wise for only China and Taiwan.



Problems Encountered

- The pipeline from source 2 has huge data that requires batch downloads and implementing this was challenging as the website limits the data size and maximum batch size should be below thus data size.

- The automation using selenium included username and password. The clicking of imports tab in the website has to search html script of the website.

Dealing with Changing Input Data

- Placing URLs in Config file. This can be changed any time if URL changes without changing code.
- Time SQLite file updated is recorded. So if data is lost in website the data roll back to last time.
- Data Validation Checks has been done if does not match will not contaminate the existing data.

Error Handling

- Logs error entries into a separate error log table.
- Retries and batch retries have been implemented

4. Results & Limitations

Output Data

- **Format**: SQLite with tables for each data pipeline as integers for numerical columns, decimal for tariff rate. The time series of US government implemented tax rate on goods imported into USA mapped with quantity and countries it was imported gives an idea of whether US people used the goods with high import rates or moved to some other country products.

Quality Assessment

- Normalized values with minimal outliers.

Limitations

- Some datasets may have delays in updates.
- Economic data variability due to geopolitical shifts not captured in real-time.

Critical Reflection

- The pipeline is robust for current datasets, but new variables HTS may require future updates to transformation logic.