DATA SCIENCE TOOLBOX: PYTHON PROGRAMING

PROJECT REPORT

(Project Semester January-April 2025)

"Uncovering Patterns in Global Trade Data"

Submitted by **: Harsh**

Registration No: 12320578

Lovely Professional University,

Phagwara , Punjab

CERTIFICATE

This is to certify that Harsh bearing Registration no. 12320743 has completed INT375 project titled, **"Uncovering Patterns in Global Trade Data"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Dr. Tamina Thakur
Faculty
School of Computer Science
Lovely Professional University
Phagwara, Punjab.

Date: 12-April-2025

DECLARATION

I Harsh, student of B.Tech CSE under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-April-2025

Harsh

Reg no.- 12320743

Table Of Content

1. Introduction

Global trade plays a vital role in the development of economies and the interconnectedness of nations. Tracking the import and export of goods across countries helps in understanding market demands, production capabilities, consumption trends, and international competitiveness. Such analysis is essential for governments, businesses, and policy makers to make informed decisions, identify trade opportunities, and manage resources efficiently.

In this project, we are analyzing a real-world trade dataset that contains records of imports, exports, production, and consumption across multiple countries, products, and years. Each record includes important information such as the country name, product type, year of record, import/export values, and other key metrics. Using this data, we can extract meaningful insights about international trade patterns.

The main objective of this project is to clean, process, and analyze this dataset using Python. We aim to uncover patterns through data visualization and identify trends in global trade. Our focus areas include:

- Identifying the top exporting countries.
- Analyzing the trend of exports over the years.
- Comparing imports and exports for leading nations.
- Highlighting the most exported products globally.
- Evaluating trade balance over time.

This type of analysis promotes data-driven decision making in trade policy, market strategy, and economic planning. For example, if a particular country consistently shows a high trade surplus, it could indicate a strong export economy. Similarly, identifying top-performing products can help industries and governments focus on high-demand commodities. Overall, this project demonstrates how data and analytics can provide valuable insights into international commerce and contribute to economic growth.

2. Dataset Overview

The dataset used in this project contains records of import, export, production, and consumption of various agricultural and livestock-related products across different countries and years. The data is collected from reliable global trade sources and offers valuable insight into international trade activities.

Each record in the dataset includes:

- **Country (Area):** The name of the country reporting the trade activity.
- **Product (Item):** The type of commodity being imported, exported, or produced.
- **Year:** The calendar year in which the trade activity occurred.
- **Import:** Quantity of the product imported by the country.
- **Export:** Quantity of the product exported by the country.
- **Production:** Total production volume of the commodity in that country.

- **Consumption:** Estimated domestic consumption based on production and trade.

This structured information allows us to conduct a variety of real-world economic and trade analyses, helping to uncover trends in global commodity flows, identify trade surpluses or deficits, and support data-driven policy decisions.

Source of Dataset : https://github.com/HarshYdv07/Dataset

3. Data Preprocessing

Before performing any analysis, it was essential to clean and prepare the dataset for accuracy and consistency. The preprocessing steps included:

- **Checking for missing or null values** in key columns such as import, export, production, and consumption, and handling them appropriately.
- **Reviewing and correcting data types**, especially ensuring that the Year column is treated as an integer for proper time-series analysis.
- **Cleaning column names** by removing leading/trailing spaces and renaming them where necessary for easier coding and visualization.
- **Removing duplicate entries** to avoid redundancy and ensure the integrity of insights.
- **Standardizing entries** in categorical fields like country names and product types for uniform grouping.

These preprocessing steps were crucial to ensure that the dataset was reliable and ready for insightful analysis. Clean data helps avoid errors, improves visualization quality, and supports valid decision-making.

4. Data Visualization

To make the data more insightful and easier to interpret, we used a variety of graphs and charts. These visualizations help identify patterns and trends in global trade, making the report both informative and presentation-ready.

We created the following types of visualizations:

- **Bar Charts:** Used to display the top exporting countries, top imported products, and comparisons between import and export volumes across countries.
- **Line Graphs:** Ideal for showing export and trade balance trends over time, helping to understand how trade dynamics evolve year by year.

- **Pie Charts:** Used to visualize the proportion of export volumes by product or country, providing a clear understanding of market share.
- **Grouped Bar Charts:** Help compare multiple metrics (such as import vs export) for the same country in a side-by-side manner.
- **Correlation Heatmap:** Although not always required in trade data, it can be used to understand how different numerical fields (e.g., production, export, consumption) relate to one another using color-coded intensity.
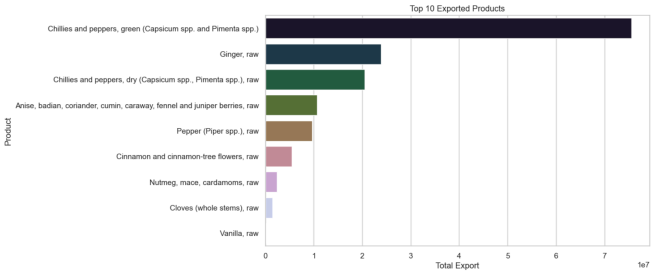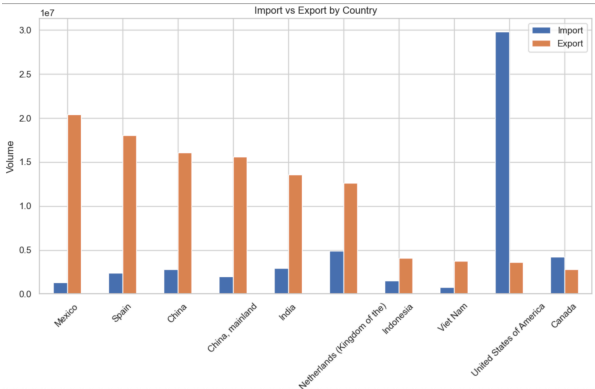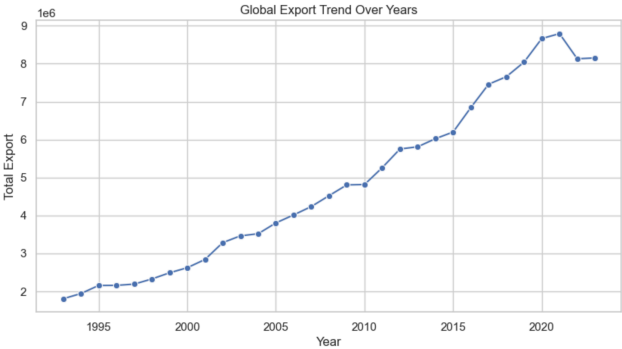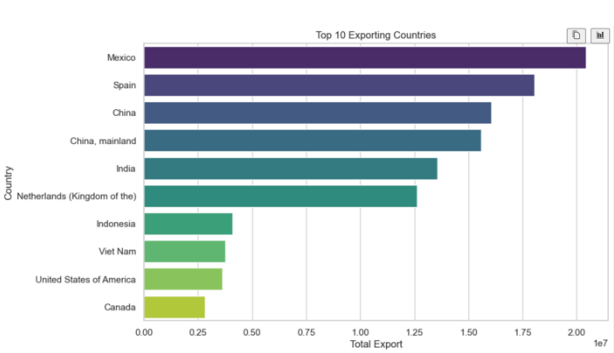
Objective 1: Top Exporting Countries

We analyzed which countries export the most agricultural and livestock-related products globally. For this, we:

- **Calculated the total export volume** for each country by summing up all export records.
- **Identified the top 10 exporting countries** based on cumulative export values.
- **Used a horizontal bar chart** to visually represent which countries lead in exports.
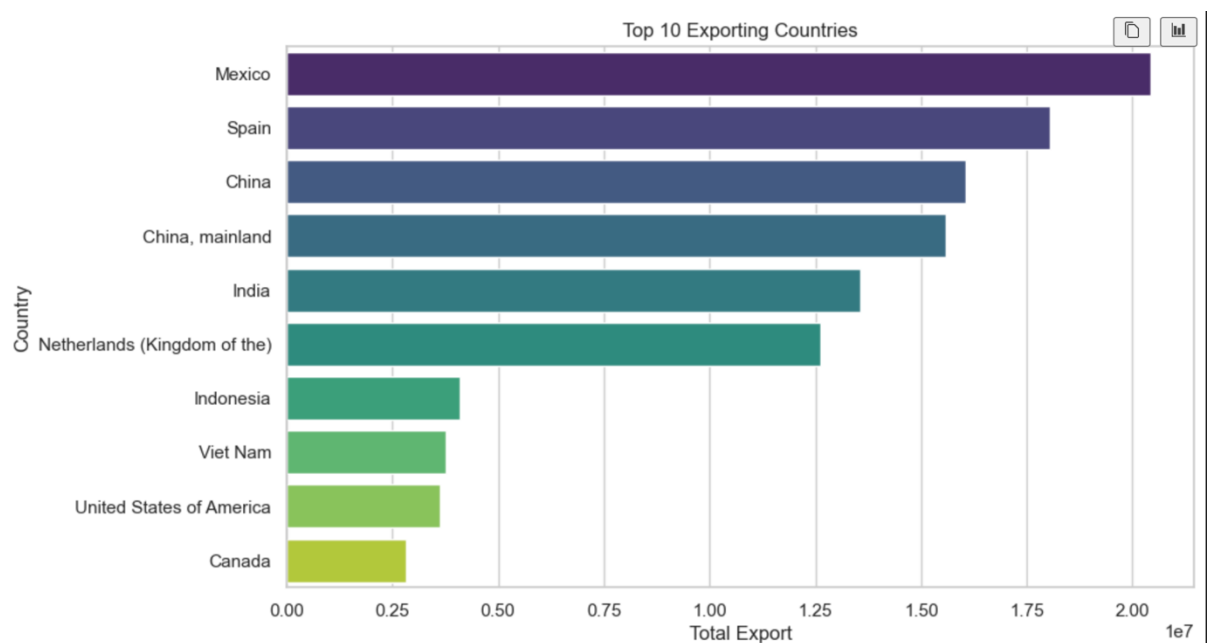

Objective 2: Export Trend Over Years

In this section, we analyzed how global export volumes have changed over time:

- **Grouped the data by year** and calculated the total exports across all countries and products.
- **Plotted a line graph** to observe year-by-year trends in global exports.

- **Highlighted increases or drops** in export volume, which may relate to global events, market changes, or policy shifts.
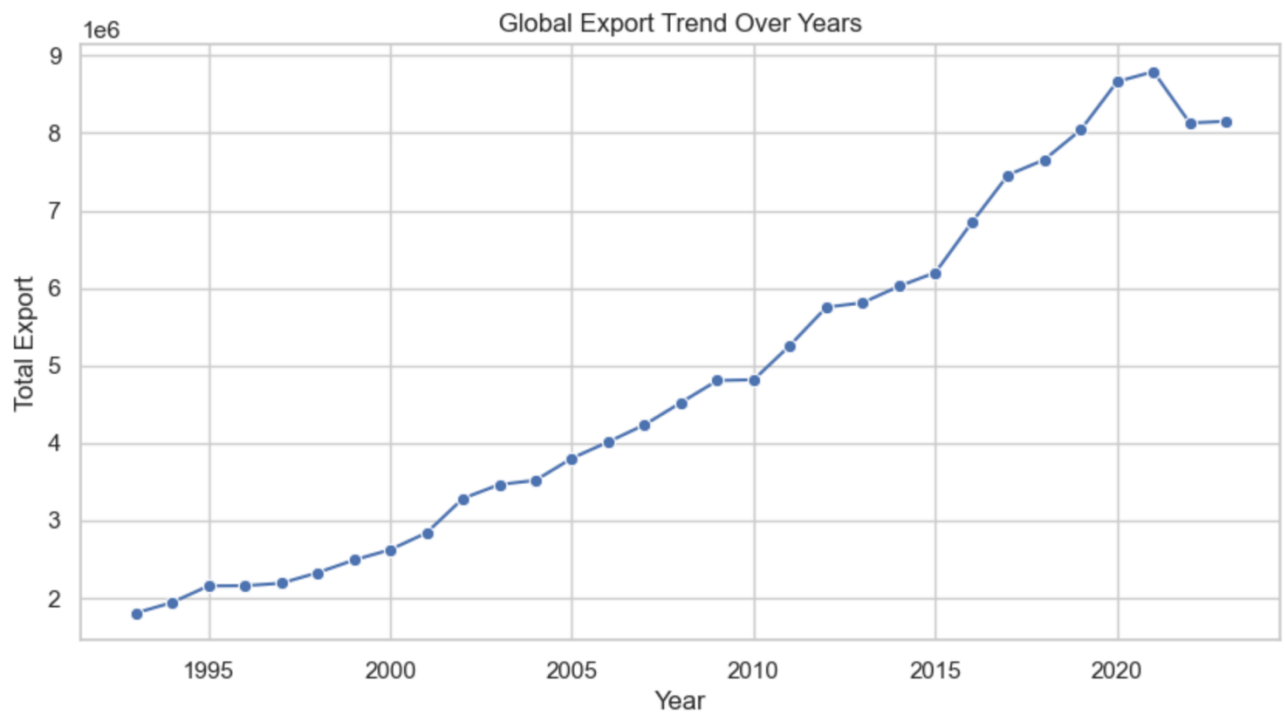
Objective 3: Import vs Export by Country

We compared the total import and export volumes across countries to evaluate their trade balance and economic roles:

- **Grouped the data by country** and calculated the total import and export volumes.
- **Identified the top 10 countries** with the highest export volumes and compared their import figures.
- **Used grouped bar charts** to show both imports and exports side by side for each country.

Import vs Export by Country
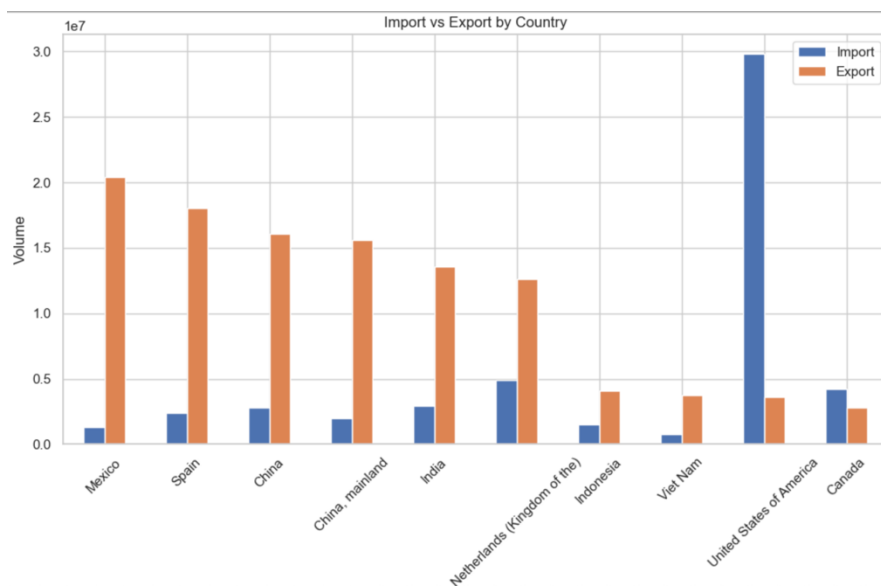
Objective 4: Top Exported Products

In this section, we identified which products were most commonly exported across all countries:

- **Grouped the dataset by product type** and calculated the total volume exported for each.
- **Sorted the products** to find the top 10 most exported items.
- **Created a bar chart** to clearly visualize which commodities are in highest demand on the global market.

Top 10 Exported Products
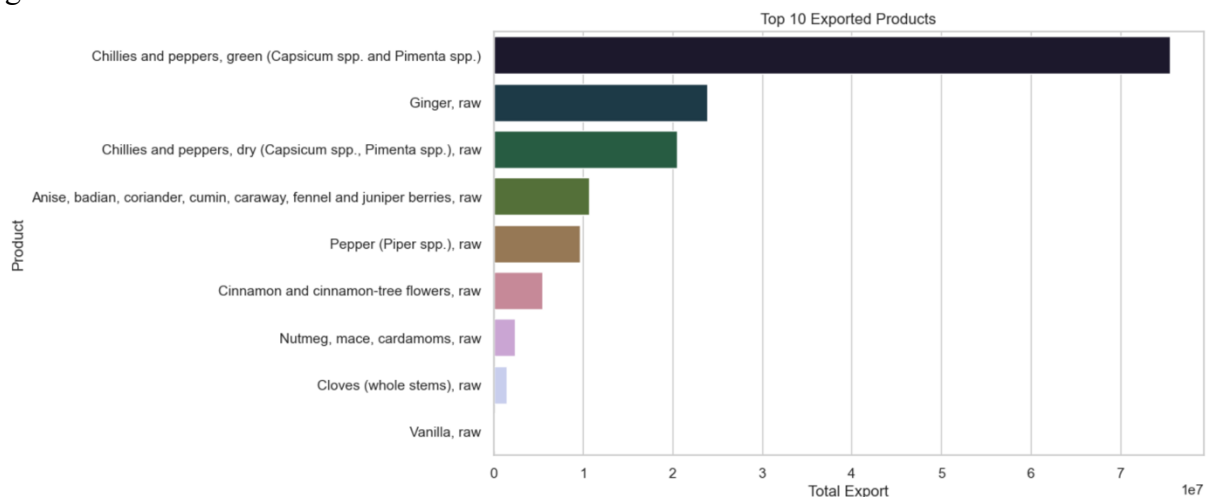
Objective 5: In this section, we evaluated the overall cleanliness of the dataset:

- Identified and counted missing, unknown, or invalid entries (e.g., incorrect values such as unrealistic ages).
- Calculated the proportion of these problematic values.
- Assessed the dataset's reliability based on the extent of these issues.

10. Conclusion

By analyzing the trade dataset, we achieved the following:

- Identified the most traded items and their associated regions.
- Examined how trade volumes vary by country, item type, and over time.
- Discovered key trends and patterns that can inform economic and trade-related decisions.

11. Future Scope

There's significant potential to expand this project further, such as:

- Incorporating price and quantity data to analyze trade value trends more deeply.
- Using machine learning models to forecast trade volumes based on historical data.
- Adding more recent years or expanding to additional countries or regions for broader insights.

12. Reference

- Pandas Documentation: https://pandas.pydata.org/docs

- Matplotlib Documentation: https://matplotlib.org/stable/contents.html

- Seaborn Documentation: https://seaborn.pydata.org

- Python Official Website: https://www.python.org

- YouTube: https://youtu.be/tjIWRqqMDaw?si=4sAZtKVb1d3kQSvp

13. LinkedIn Post link : https://www.linkedin.com/posts/harshydv07_dataanalysis-tradedata-python-activity-7319672139156992000-WNrl?utm_medium=ios_app&rcm=ACoAAEYvOPkBISbkJWWTqTlNp3Q9Rqkm2mhbrk0&utm_source=social_share_send&utm_campaign=copy_link

14. GitHub Link : https://github.com/HarshYdv07/Trade-Analysis