Novel Dataset: Trade Statistics of India

Group 24 — Data Analytics (CS61061)

November 13, 2025

Instructor: Dr. Abhijnan Chakraborty

Narendra Kumar (25CS60R18)
Prithvirajsinh G. Jadav (25CS60R62)
Manav Chaniyara (25CS60R87)

Contents

1 Abstract 3

2 Introduction 3

3 Problem Statement 3

4 Tools and Technologies Used 3

5 Data Collection Methodology 4

6 Example JSON 5

7 Data Analysis and Visualization 6

8 Challenges Faced 8

9 Conclusion 8

10 References 8

1 Abstract

India's trade ecosystem generates millions of records annually across imports and exports, distributed over hundreds of commodities and trading partners. However, the data is fragmented and unstructured across multiple government web portals.Using Python-based web scraping, a clean, structured dataset of over **6 million entries** was created. The processed data was later analyzed using Power BI and Tableau to visualize trends in import and export dynamics.

2 Introduction

Understanding trade flows is essential for economic planning, forecasting, and identifying emerging markets. While India's Ministry of Commerce publishes monthly import and export statistics, these datasets are not readily available in a single consolidated format.

The project **"Trade Statistics of India"** focuses on building a standardized data pipeline that automates:

- Scraping monthly trade data (country-wise and commodity-wise)
- Cleaning and organizing data into JSON format
- Providing insights using data visualization tools

3 Problem Statement

Existing data on Indian trade is:

- Scattered across multiple HTML-based reports.
- Inconsistent in structure across months and years.
- Lacking in automation for bulk download or analysis.

Thus, manual extraction is time-consuming and error-prone. The goal is to produce a reliable dataset suitable for research and analytics.

4 Tools and Technologies Used

Languages and Libraries

- **Python 3.11+**
- **BeautifulSoup4** – for parsing HTML content.
- **Requests** – for maintaining session and handling CSRF tokens.
- **Pandas** – for cleaning, structural, and converting CSV to JSON datasets.
- **Power BI / Tableau** – for interactive data visualization.
- **Jupyter Notebook** – for data exploration and experimentation.

Data Source

- Official Ministry of Commerce Portal:
  https://www.commerce.gov.in/trade-statistics/

5 Data Collection Methodology

The data scraping process involves:

1. Accessing the base URL and extracting CSRF tokens.
2. Iterating through all available countries.
3. Requesting monthly trade data (Jan–Dec) for each year.
4. Parsing each HTML table into a Pandas DataFrame.
5. Appending all records into a unified CSV file.

To ensure server friendliness, random delays and retry mechanisms were implemented. The dataset includes columns such as:

```
HS Code, Commodity Name, Month, Year, Country, Unit, Quantity,Type
```

6 Example JSON

A simplified sample of the processed data is shown below.

JSON Format

```
{
 "Export":{
        "2022":[
        {
            "HS Code": "10063010",
            "Commodity Name": "Basmati Rice",
            "Month": "Jan",
            "Year": 2022,
            "Country": "United Arab Emirates",
            "Unit": "Tonnes",
            "Quantity": 14250
        },
  ...
  "Import":{
        "2022":[
        {
            "HS Code": "27090000",
            "Commodity Name": "Crude Petroleum Oil",
            "Month": "Jan",
            "Year": 2022,
            "Country": "Saudi Arabia",
            "Unit": "Tonnes",
            "Quantity": 982341
        }
  ...
]
```

7 Data Analysis and Visualization

After building the complete dataset, Power BI and Tableau were used for visualization. Key insights included:

- **Top 15 Import and Export Countries** – showing India's main trading partners.
- **Yearly Comparisons** – analyzing trade changes from 2022 to 2024.

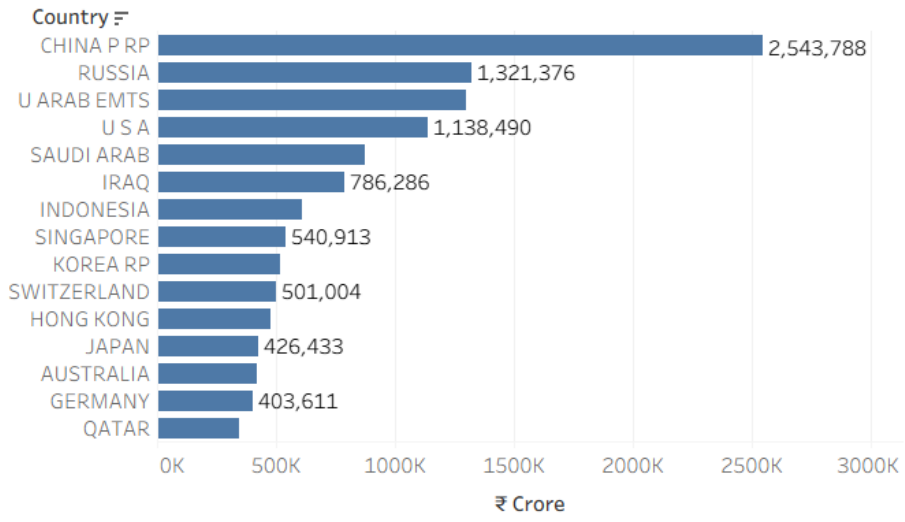

Figure 1: India's Top 15 Importing Countries

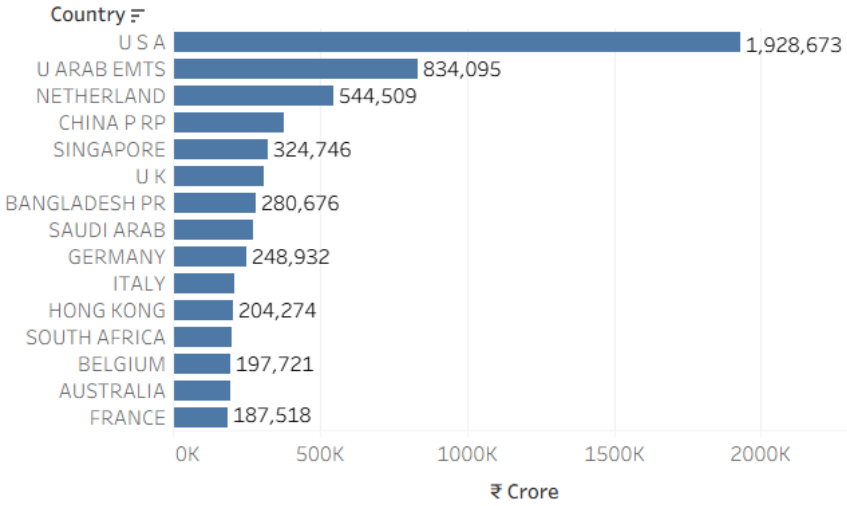

Figure 2: India's Top 15 Exporting Countries

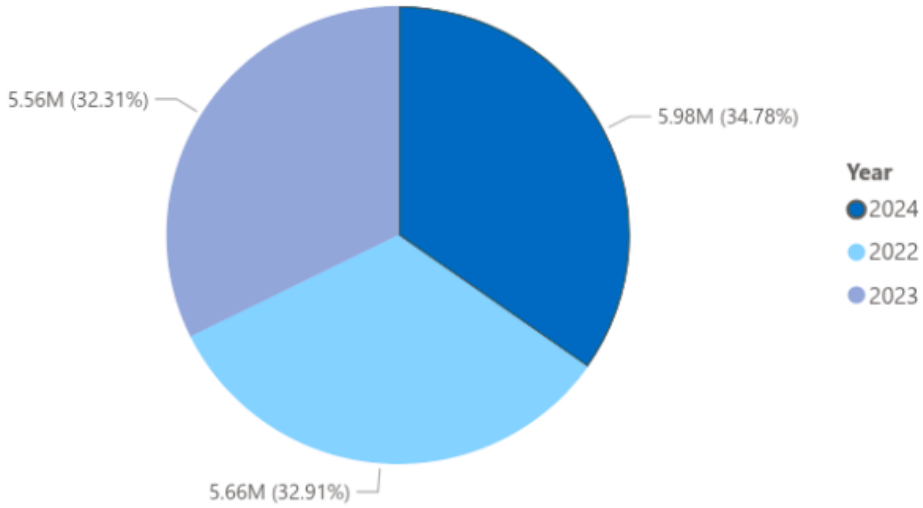

Figure 3: Year-Wise Import

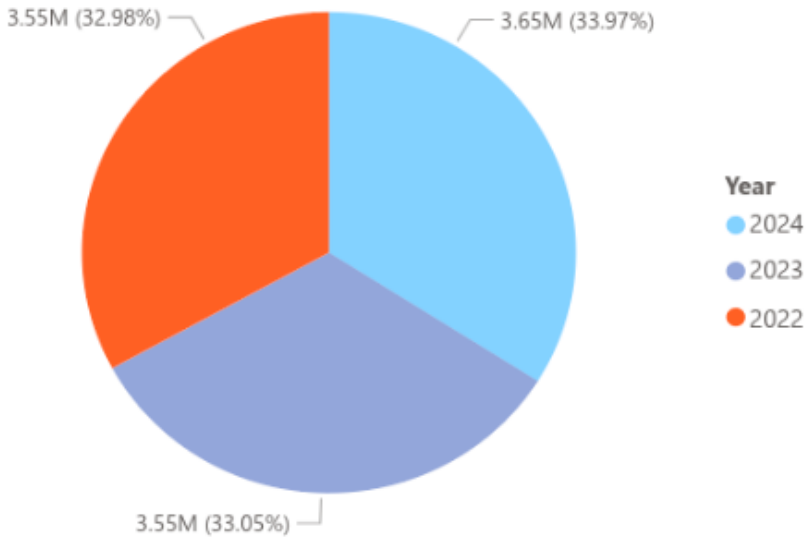

Figure 4: Year-Wise Export

8 Challenges Faced

- Frequent session expiry and CSRF token regeneration.
- HTML structure inconsistency across months.
- Handling large data volume efficiently without server blocking.

9 Conclusion

Our Dataset is successfully Create with 2022 to 2024 Import and Export Trade Data of India This dataset provides a foundation for deeper trade analysis, forecasting, and policy-making.

10 References

- Trade Statistics Portal – https://tradestat.commerce.gov.in/

— End of Report —