Documentation: Why These Datasets, Why This Analysis, and How They Were Collected

Motivation for the Analysis

The semiconductor industry is central to global competitiveness. Inflationary pressures, labor costs, tariffs, and international demand are reshaping supply chains and influencing strategic decisions for governments, CEOs, startups, and recruiters. My goal was to design an analysis that:

- **Connects trade flows, labor trends, and logistics data** into one coherent picture.
- **Identifies nearshoring opportunities** for the U.S. and its partners.
- **Quantifies risks** from tariff structures and concentrated value chains.
- **Demonstrates my ability** to collect, clean, and analyze complex datasets for actionable insights.

This analysis is both a professional debut and a showcase of my passion for data, analytics, and business strategy.

Why These Datasets

I selected datasets that are **official, legal, and publicly available**, ensuring credibility and reproducibility. Each source was chosen to answer a specific dimension of the problem:

Dataset	Source	Why It Was Chosen	Analytical Purpose
Trade Flows (Imports/Exports)	UN Comtrade (HS 85423100)	Global standard for trade statistics	Compare U.S. semiconductor imports/exports by partner country
Tariff Structures	WTO Tariff Data, USITC HTS	Authoritative tariff schedules	Quantify competitiveness

	Search, Section 301 Tariffs		impacts of MFN vs. Section 301 penalties
Labor & Employment	BLS (QCEW, ECI, SAE, OEWS)	Comprehensive U.S. labor statistics	Assess employment costs, wage trends, and talent availability
Logistics & Transportation	Rail Equipment Accident Data (Form 54)	Operational efficiency and risk	Evaluate cost savings and reliability in trade routes
Global Demand & Market Trends	OECD, SEMI.org, SIA	Industry-level demand projections	Forecast growth and identify emerging hubs

Why This Analysis

I designed the analysis to bridge **trade, labor, and logistics** because these three pillars jointly determine competitiveness:

- Trade flows show **where goods move**.
- Tariffs reveal **policy barriers and incentives**.
- Labor data highlights **costs and talent pipelines**.
- Logistics data uncovers **efficiency and risk in supply chains**.
- Demand projections provide **forward-looking context**.

By combining them, the analysis delivers insights that are **strategic, actionable, and relevant** to executives, recruiters, and policymakers.

How the Data Was Collected

- **Phase 1:** Located and extracted datasets from official portals (UN Comtrade, WTO, USITC, BLS, OECD, SEMI).
- **Phase 2:** Downloaded raw files (CSV, XLS, PDF) and API outputs, ensuring proper documentation of each source.

- **Phase 3:** Cleaned and standardized millions of rows, aligning HS codes, NAICS categories, and time series formats.
- **Validation:** Cross-checked figures across multiple sources (e.g., WTO vs. USITC) to ensure consistency.
- **Documentation:** Each dataset is accompanied by a PDF or README in the repository for transparency.

Why From These Sources

- **Credibility:** All sources are internationally recognized institutions (UN, WTO, OECD, BLS).
- **Reproducibility:** Publicly available datasets ensure anyone can replicate the analysis.
- **Coverage:** Together, they cover trade, tariffs, labor, logistics, and demand—the full ecosystem.
- **Strategic Value:** These sources provide the kind of data CEOs, recruiters, and policymakers rely on for decision-making.

Closing Note

This project is structured in **five phases** (data collection, cleaning, EDA, visualization, reporting). It is currently in **Phase 3 (EDA)**, where many discoveries are emerging daily. I will continue updating the repository with progress, insights, and documentation.

Explore the repository and my CV here:
GitHub – Semiconductor Trade & Labor Analysis – USA 2025

Thank you for supporting my debut project. Data tells stories—I'm here to turn those stories into strategies 😊