Dataset- https://www.dol.gov/agencies/eta/foreign-labor/performance

H1-B Economic Impact Analysis Platform

Project Proposal

Group Members: Jayden Bai, Unnati Agrawal, Avery Mattoon
Tech Stack: PostGreSQL, Next.js

1. Introduction

In the current political climate, visa holders—particularly H-1B workers—are under increasing scrutiny. It is therefore crucial to examine whether there is a sound economic justification for recent and proposed policy changes, including significant visa fee hikes. By grounding the discussion in empirical data rather than rhetoric, this project seeks to assess the true financial and labor-market impact of H-1B employment.

Ultimately, our goal is to reinforce the understanding that immigrants are not only culturally enriching but also economically indispensable to the United States. Through rigorous data analysis, we aim to demonstrate that immigrant workers drive innovation, fill critical skill gaps, and contribute substantially to the nation's productivity and long-term growth.

2. Features

We aim to quantify the economic impact of H-1B visa holders on the U.S. economy through a comprehensive, data-driven analysis platform. Our project will explore where this value originates—by company, industry, and region—and how it connects to broader economic forces such as employment, technological innovation, and immigration trends.

By the end of the project, we will demonstrate several key features that bring these insights to life:

1. Identify where critical expertise is concentrated and analyze how shifts in U.S. immigration policy would affect economies in the United States.
    a. We need to know which industries most H-1Bs are working in
    b. Should pretty much just be comparing the growth trends in different industries - the challenge here is classifying the industries
        i. We might be able to use the role
        ii. Or the company name as well
    c. We can also do an analysis on the state, how different local economies will be affected by a complete removal of that workforce

    d. **WEIGHT - 25%**
2. Compare company behaviors—which firms are likely to absorb increased visa costs versus those that may scale back on international hiring.
    a. The immediate thought is that larger companies with more spending will be able to eat the 100k cost
    b. **WEIGHT - 25%**
3. Project the economic outlook across industries, distinguishing sectors that would be most vulnerable to reduced H-1B participation from those with minimal impact.
    a. Our work in the demographic information for which industries H-1Bs are coming from would be useful here. We should just try to quantify the dollar amount that each company/industry would be losing ie. if the employee's salary is 100k, we should subtract 100k from the industry's overall valuation?
    b. **WEIGHT - 25%**
4. Isolate impacts of COVID and analyze identify industry trends in H1-B participation across states.
    a. Use years 2019 (before), 2020 (during), and 2021 (after), looking at attributes such as total volume of visas, industry, state, etc
    b. **WEIGHT - 12.5%**
5. Compute and visualize average and median offered wages
    a. Compare attributes such as industry/state to look for trends
    b. **WEIGHT - 12.5%**

3. Database Design

Schema

- Employer info (<u>emp_id</u>, employer_name, industry)
    - Estimated 350K-700K tuples (H-1B visa acceptance rate is around 35%)
- Occupation info (<u>soc_code</u>, soc_title, major_group)
    - Estimated 1K-2K tuples (for every different job that is done)
- Location info for employer (emp_id FK, state, city, zip)
    - Estimated 1K-2K tuples (based on number of companies in the US that sponsor H-1Bs)
- Application info (<u>app_id</u>, emp_id FK, soc_code FK, site_id FK, case_status, wage_offer, wage_unit, decision_date, visa_class, fiscal_year)
    - Estimated 1M-2M tuples

FK- Foreign Key

Technology Stack:

Backend: PostgreSQL 16 for relational storage and query optimization.
Frontend: Next.js 15 with TypeScript for the dashboard and API routes.

Our system will be an interactive web dashboard with PostgreSQL powering the backend and Next.js handling the frontend, connected through RESTful API endpoints.

The backend consists of a normalized PostgreSQL database with indexed tables and materialized summary views.

Assumptions:

a) Each attribute is not changing much during the year or in between the years in between their last and most recent application (ie. salary is the same throughout a year until it is updated until the next year)
b) The size of the database is multiplied by the number of years and the size of one year of our dataset. We are assuming that each year has relatively the same amount of new and previous applications
c) We've made broader assumptions about the number of different types of employers, occupations, locations, and application information based on rough market sizing. The sizes of these would probably change as we move forward.

4. Task List

Data Preparation

1. Collect and merge OFLC LCA disclosure data from 2019–2021 to capture pre-, during-, and post-COVID periods.
2. Standardize wage units (hourly → annual) and normalize employer names.
3. Map occupations (SOC codes) to industry categories.
4. Clean and geocode worksite states to support state-level analysis.

Database Implementation

1. Design normalized PostgreSQL schema linking employers, occupations, worksites, and applications.
2. Create indexes and materialized views for median wage per industry and approval rate per state.
3. Load cleaned data and verify referential integrity across all entities.

Analytics & Modeling

1. Compute industry growth trends in H-1B filings and wages.
2. Estimate industry-level exposure to a hypothetical $100K visa cost increase (compare adjusted wages vs. median industry wages).
3. Quantify COVID-era impacts using 2019–2021 comparisons.

Frontend & Visualization

1. Build interactive dashboards for industry, company, and state visualizations.
2. Implement filters for year, industry, and state, and ensure REST API endpoints deliver precomputed metrics efficiently.