Enhanced Data Report: Evaluating Microsoft's H1B Hiring Impact on Local Wages in Washington

A Comprehensive Analysis Using Automated Data Pipelines

1 Question

This investigation examines whether Microsoft's H1B hiring practices lower local tech wages in Washington. The study uses rich statistics and an automated pipeline to compare wages offered to H1B employees with local averages. This comparison analyzes Microsoft's impact on local pay structures in Washington, with a focus on certified H1B applications and occupational roles. .

2 Data Sources

2.1 Description of Data Sources

The H1B Labor Condition Application (LCA) dataset is sourced from the U.S. Department of Labor and contains 186,389 rows and 33 columns. This dataset provides details about H1B applications, including employer names, job roles, worksite locations, and wages. The analysis focuses on applications filed during Q2 2023 (April–June) in Washington. Key columns include `CASE_STATUS` for application status, `EMPLOYER_NAME` filtered for Microsoft, and `SOC_CODE`, a standardized occupational code used for merging with other datasets.

The Occupational Employment and Wage Statistics (OEWS) dataset, collected from the Washington State Employment Security Department, has 809 rows and ten columns. It gives aggregated pay and employment data for several occupational jobs in Washington across many years (2020-2023). The emphasis is centered on annual average earnings and occupational titles that match to H1B job codes. .

2.2 Licensing Details

The H1B dataset is provided to the public under the Open Data Policy of the United States Department of Labor. This policy permits study, analysis, and educational use, as long as full acknowledgment is supplied. This report's attribution involves referencing the dataset's source. its URL: U.S. Department of Labor. H1B Labor Condition Application Dataset.

The OEWS dataset is available under Washington State's Open Data Policy, which permits noncommercial usage with attribution. Redistribution of the data is allowed as long as it is not altered or misrepresented. The dataset is referenced by its source: Washington State Employment Security Department. OEWS Dataset.

2.3 Structure and Quality of Data

The H1B dataset exhibits a well-defined tabular format but includes inconsistencies, such as varying wage units and null values in critical columns. These challenges required extensive cleaning and transformation. In contrast, the OEWS dataset is concise and aggregated, offering high-quality data on local wages but limited in temporal granularity.

Compliance with licensing obligations is ensured by providing clear attribution for each dataset, adhering to non-commercial use requirements, and maintaining transparency in derived insights and visualizations.

3 Data Pipeline

3.1 Overview and Transformations

The data pipeline automates the extraction, transformation, and merging of datasets. The H1B dataset was filtered to retain only Microsoft roles with certified applications in Washington. To ensure consistent comparisons, wages were standardized to annual equivalents. The SOC codes were cleaned by removing decimal points, and null values were removed from both datasets to maintain completeness.

The OEWS dataset underwent transformations to align with the H1B data. Only job roles relevant to Microsoft's SOC codes were retained, and wages were filtered to exclude unrealistic values outside the $20,000–$300,000 range. The datasets were then merged on the standardized SOC code, and each job code was reduced to a single row for clarity.

3.2 Expanded Transformations and Column Renaming

In the H1B dataset, columns were renamed for consistency. `SOC_CODE` was renamed to `job_code`, `WAGE_RATE_OF_PAY_FROM` became `annual_wage`, and `SOC_TITLE` was renamed to `occupation_title`. Similarly, in the OEWS dataset, `SOC_CODE` was renamed to `job_code`, and `ANNUAL_MEAN_WAGE` was renamed to `avg_local_wage`.

4 Results and Limitations

4.1 Output Data

The final dataset contains 24 unique rows, each representing a distinct job code present in both datasets. The merged data includes columns for job titles, Microsoft H1B wages, local average wages, and the calculated wage difference. This output is stored in a SQLite database with three tables: `h1b_microsoft_roles`, `oews_microsoft_roles`, and `h1b_oews_combined`.

4.2 Limitations and Challenges

A key limitation of the analysis is the temporal misalignment between the datasets. While the H1B data represents Q2 2023, the OEWS data spans multiple years, introducing potential discrepancies. This issue could be mitigated by adjusting OEWS wages for inflation or incorporating historical H1B data for alignment. Additionally, the OEWS

dataset includes roles beyond the tech sector, which were carefully filtered to focus on relevant job codes.

5 Conclusions and Future Recommendations

The analysis reveals that Microsoft's H1B wages generally exceed local averages for similar roles, particularly in technical occupations. However, the temporal discrepancies and aggregation in the OEWS data suggest that further refinement, such as inflation adjustments or expanded data coverage, would strengthen the insights.

Future analyses could incorporate historical trends to contextualize Microsoft's hiring practices over time. Additional comparisons with other employers in the region could also provide a broader perspective on the impact of H1B hiring on local wages.