

# Data exploration log for H1B LCA Data Disclosure

- Source: <https://www.kaggle.com/datasets/zongaobian/h1b-lca-disclosure-data-2020-2024/data>
- Overview

## What is H1B?

The H1B visa is a non-immigrant visa that allows U.S. companies to employ foreign workers in specialty occupations requiring theoretical or technical expertise. These roles typically include fields such as IT, engineering, finance, healthcare, and more. The H1B program is critical for addressing skill gaps in the U.S. workforce and supporting economic growth.

## What is LCA?

The Labor Condition Application (LCA) is a prerequisite for filing an H1B visa petition. Employers submit the LCA to the DOL to ensure compliance with wage and working condition requirements. The LCA process protects both U.S. workers and foreign employees by enforcing:

- Payment of prevailing wages.
- Assurance that hiring foreign workers will not adversely affect local labor conditions.

Each LCA disclosure contains information about the employer, job title, job location, wages, and visa classification.

## Why Analyze LCA Disclosure Data (2020-2024)?

The dataset spans a crucial period (2020-2024) characterized by:

- **Pandemic Impact:** Changes in employment patterns and visa policies due to COVID-19.
- **Remote Work Trends:** Shifts in work location dynamics for H1B visa holders.
- **Tech Layoffs and Restructuring:** Evolving job roles and industry demands, especially in tech.
- **Economic Recovery:** Insights into how industries and geographic regions rebounded post-pandemic.

Analyzing this data can provide:

1. **Employment Trends:** Discover trends in job roles, industries, and geographic locations hiring H1B workers.
2. **Wage Comparisons:** Compare wages across job titles, industries, and states.
3. **Policy Insights:** Assess the impact of government policies on foreign employment.
4. **Geographic Distribution:** Identify areas with the highest demand for H1B workers.
5. **Industry Insights:** Explore the reliance of various industries on foreign talent.

## 1. Cleanness/structure check

- a. We set the data types as below to import 96 columns/5 files of each year into Azure data studio (see sql create table clause)

```
CREATE TABLE [dbo].[LCA_20] (
    [CASE_NUMBER] NVARCHAR (50) NOT NULL,
    [CASE_STATUS] NVARCHAR (50) NULL,
    [RECEIVED_DATE] DATE NULL,
    [DECISION_DATE] DATE NULL,
    [ORIGINAL_CERT_DATE] DATE NULL,
    [VISA_CLASS] NVARCHAR (50) NULL,
    [JOB_TITLE] NVARCHAR (150) NULL,
    [SOC_CODE] NVARCHAR (50) NULL,
```

```

[SOC_TITLE] NVARCHAR (150) NULL,
[FULL_TIME_POSITION] NVARCHAR (50) NULL,
[BEGIN_DATE] DATE NULL,
[END_DATE] DATE NULL,
[TOTAL_WORKER_POSITIONS] INT NULL,
[NEW_EMPLOYMENT] INT NULL,
[CONTINUED_EMPLOYMENT] INT NULL,
[CHANGE_PREVIOUS_EMPLOYMENT] INT NULL,
[NEW_CONCURRENT_EMPLOYMENT] INT NULL,
[CHANGE_EMPLOYER] INT NULL,
[AMENDED_PETITION] INT NULL,
[EMPLOYER_NAME] NVARCHAR (150) NULL,
[TRADE_NAME_DBA] NVARCHAR (150) NULL,
[EMPLOYER_ADDRESS1] NVARCHAR (150) NULL,
[EMPLOYER_ADDRESS2] NVARCHAR (150) NULL,
[EMPLOYER_CITY] NVARCHAR (100) NULL,
[EMPLOYER_STATE] NVARCHAR (50) NULL,
[EMPLOYER_POSTAL_CODE] NVARCHAR (50) NULL,
[EMPLOYER_COUNTRY] NVARCHAR (100) NULL,
[EMPLOYER_PROVINCE] NVARCHAR (50) NULL,
[EMPLOYER_PHONE] NVARCHAR (50) NULL,
[EMPLOYER_PHONE_EXT] NVARCHAR (50) NULL,
[NAICS_CODE] NVARCHAR (50) NULL,
[EMPLOYER_POC_LAST_NAME] NVARCHAR (50) NULL,
[EMPLOYER_POC_FIRST_NAME] NVARCHAR (50) NULL,
[EMPLOYER_POC_MIDDLE_NAME] NVARCHAR (50) NULL,
[EMPLOYER_POC_JOB_TITLE] NVARCHAR (100) NULL,
[EMPLOYER_POC_ADDRESS1] NVARCHAR (150) NULL,
[EMPLOYER_POC_ADDRESS2] NVARCHAR (150) NULL,
[EMPLOYER_POC_CITY] NVARCHAR (100) NULL,
[EMPLOYER_POC_STATE] NVARCHAR (50) NULL,
[EMPLOYER_POC_POSTAL_CODE] NVARCHAR (50) NULL,
[EMPLOYER_POC_COUNTRY] NVARCHAR (100) NULL,
[EMPLOYER_POC_PROVINCE] NVARCHAR (50) NULL,
[EMPLOYER_POC_PHONE] NVARCHAR (50) NULL,
[EMPLOYER_POC_PHONE_EXT] NVARCHAR (50) NULL,
[EMPLOYER_POC_EMAIL] NVARCHAR (100) NULL,
[AGENT_REPRESENTING_EMPLOYER] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_LAST_NAME] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_FIRST_NAME] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_MIDDLE_NAME] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_ADDRESS1] NVARCHAR (150) NULL,
[AGENT_ATTORNEY_ADDRESS2] NVARCHAR (150) NULL,
[AGENT_ATTORNEY_CITY] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_STATE] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_POSTAL_CODE] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_COUNTRY] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_PROVINCE] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_PHONE] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_PHONE_EXT] NVARCHAR (50) NULL,
[AGENT_ATTORNEY_EMAIL_ADDRESS] NVARCHAR (100) NULL,
[LAWFIRM_NAME_BUSINESS_NAME] NVARCHAR (100) NULL,
[STATE_OF_HIGHEST_COURT] NVARCHAR (100) NULL,
[NAME_OF_HIGHEST_STATE_COURT] NVARCHAR (100) NULL,
[WORKSITE_WORKERS] NVARCHAR (50) NULL,
[SECONDARY_ENTITY] NVARCHAR (50) NULL,
[SECONDARY_ENTITY_BUSINESS_NAME] NVARCHAR (150) NULL,
[WORKSITE_ADDRESS1] NVARCHAR (100) NULL,
[WORKSITE_ADDRESS2] NVARCHAR (100) NULL,
[WORKSITE_CITY] NVARCHAR (50) NULL,
[WORKSITE_COUNTY] NVARCHAR (50) NULL,
[WORKSITE_STATE] NVARCHAR (50) NULL,
[WORKSITE_POSTAL_CODE] NVARCHAR (50) NULL,
[WAGE_RATE_OF_PAY_FROM] FLOAT (53) NULL,
[WAGE_RATE_OF_PAY_TO] FLOAT (53) NULL,
[WAGE_UNIT_OF_PAY] NVARCHAR (50) NULL,
[PREVAILING_WAGE] FLOAT (53) NULL,
[PW_UNIT_OF_PAY] NVARCHAR (50) NULL,
[PW_TRACKING_NUMBER] NVARCHAR (50) NULL,
[PW_WAGE_LEVEL] NVARCHAR (50) NULL,
[PW_OES_YEAR] NVARCHAR (50) NULL,
[PW_OTHER_SOURCE] NVARCHAR (50) NULL,
[PW_OTHER_YEAR] NVARCHAR (50) NULL,
[PW_SURVEY_PUBLISHER] NVARCHAR (150) NULL,
[PW_SURVEY_NAME] NVARCHAR (150) NULL,
[TOTAL_WORKSITE_LOCATIONS] NVARCHAR (50) NULL,
[AGREE_TO_LC_STATEMENT] NVARCHAR (50) NULL,
[H_1B_DEPENDENT] NVARCHAR (50) NULL,
[WILLFUL_VIOLATOR] NVARCHAR (50) NULL,
[SUPPORT_H1B] NVARCHAR (50) NULL,
[STATUTORY_BASIS] NVARCHAR (200) NULL,
[APPENDIX_A_ATTACHED] NVARCHAR (50) NULL,
[PUBLIC_DISCLOSURE] NVARCHAR (150) NULL,
[PREPARER_LAST_NAME] NVARCHAR (100) NULL,
[PREPARER_FIRST_NAME] NVARCHAR (150) NULL,
[PREPARER_MIDDLE_INITIAL] NVARCHAR (150) NULL,
[PREPARER_BUSINESS_NAME] NVARCHAR (150) NULL,
[PREPARER_EMAIL] NVARCHAR (150) NULL,
CONSTRAINT [PK_LCA_20] PRIMARY KEY CLUSTERED ([CASE_NUMBER] ASC)

```

```

);

```

- b. **The issue of duplicated case numbers**—primary keys couldn't be set for the year of 2022 and 2024. So, ideally, the case number for each H1B petition should be treated as the unique key for data analysis purposes; however, we noticed that there are duplicated case numbers when importing the data. (see screen shot below)

```
import status

X Microsoft.SqlServer.Prose.Import.BcpProcessException: Error inserting data into table.
----> Microsoft.Data.SqlClient.SqlException (0x80131904): Violation of PRIMARY KEY constraint 'PK_LCA_22'. Cannot insert duplicate key in object 'dbo.LCA_22'. The
duplicate key value is (I-200-20281-867624).
The statement has been terminated.
at Microsoft.Data.SqlClient.SqlConnection.OnError(SqlException exception, Boolean breakConnection, Action`1 wrapCloseInAction)
at Microsoft.Data.SqlClient.SqlInternalConnection.OnError(SqlException exception, Boolean breakConnection, Action`1 wrapCloseInAction)
at Microsoft.Data.SqlClient.TdsParser.ThrowExceptionAndWarning(TdsParserStateObject stateObj, Boolean callerHasConnectionLock, Boolean asyncClose)
at Microsoft.Data.SqlClient.TdsParser.TryRun(RunBehavior runBehavior, SqlCommand cmdHandler, SqlDataReader dataStream, BulkCopySimpleResultSet bulkCopyHandler,
TdsParserStateObject stateObj, Boolean& dataReady)
at Microsoft.Data.SqlClient.TdsParser.Run(RunBehavior runBehavior, SqlCommand cmdHandler, SqlDataReader dataStream, BulkCopySimpleResultSet bulkCopyHandler,
TdsParserStateObject stateObj)
at Microsoft.Data.SqlClient.SqlBulkCopy.RunParser(BulkCopySimpleResultSet bulkCopyHandler)
at Microsoft.Data.SqlClient.SqlBulkCopy.CopyBatchesAsyncContinuedOnSuccess(BulkCopySimpleResultSet internalResults, String updateBulkCommandText,
CancellationTokens cts, TaskCompletionSource`1 source)
at Microsoft.Data.SqlClient.SqlBulkCopy.CopyBatchesAsyncContinued(BulkCopySimpleResultSet internalResults, String updateBulkCommandText, CancellationTokens cts,
TaskCompletionSource`1 source)
at Microsoft.Data.SqlClient.SqlBulkCopy.CopyBatchesAsync(BulkCopySimpleResultSet internalResults, String updateBulkCommandText, CancellationTokens cts,
TaskCompletionSource`1 source)
at Microsoft.Data.SqlClient.SqlBulkCopy.WriteToServerInternalRestContinuedAsync(BulkCopySimpleResultSet internalResults, CancellationTokens cts,
TaskCompletionSource`1 source)
```

- c. We want to check if the duplicated data is of rare case or prevalent—looks like there are a lot.

The screenshot shows the SQL Server Enterprise Manager interface. On the left, the 'Databases' tree is expanded to show the 'master' database. In the center, a SQL query is executed, and the results are displayed in a table. The query is as follows:

```
1 with agg as (
2 select case_number, count(CASE_NUMBER) as case_number_count
3 from LCA_24
4 group by CASE_NUMBER)
5
6 select *
7 from agg
8 where case_number_count > 1
9 order by CASE_NUMBER
```

The results table has two columns: 'case\_number' and 'case\_number\_count'. It contains 18 rows of data. The first row, with case number 'I-200-19276-068690' and a count of 2, is highlighted in red. The other rows show case numbers with counts of 2, indicating duplicated data.

case_number	case_number_count
I-200-19276-068690	2
I-200-19276-069830	2
I-200-19276-070288	2
I-200-19276-070645	2
I-200-19277-071859	2
I-200-19277-072274	2
I-200-19280-074372	2
I-200-19282-077831	2
I-200-19282-078268	2
I-200-19282-079239	2
I-200-19282-079378	2
I-200-19283-080184	2
I-200-19283-080216	2
I-200-19283-081003	2
I-200-19283-081191	2
I-200-19284-083570	2
I-200-19288-086636	2
I-200-19289-090062	2

- d. Let's further dive into the patterns of duplicates (we tested and learned that the max of duplicated case number is 3)

Run | Cancel | Disconnect | Change | Database: master | A Estimated Plan | Enable Actual Plan | Parse | Enable SQLCMD | To Notebook

```
1 select *
2 from LCA_24
3 where CASE_NUMBER in ('1-200-2229-43062')
```

duplicated rows

CASE_NUMBER	CASE_STATUS	RECEIVED_DATE	DECISION_DATE	ORIGINAL_CERT_DATE	VISA_CLASS	JOB_TITLE	SEC_CODE	SEC_TITLE	FILE_TIME_POSITION	REC_DATE	END_DATE	TOTAL_WORKER_POSITIONS	NEW_EMPLOYMENT	CONTINUED_EMPLOYMENT	CHANGE_PREVIOUS_EMPLOYMENT	NEW_CONCURRENT_EMPLOYMENT	CHANGE_EMPLOYER
1-200-2229-43062	Certified - Withdrawn	2022-09-26	2022-10-01	2022-10-26	H-1B	Senior Software Engineer	15-1212-00	Software Developers, Applications	Y	2022-03-25	2022-03-24	1	0	1	0	0	0
1-200-2229-43062	Certified - Withdrawn	2022-09-26	2022-10-01	2022-10-26	H-1B	Senior Software Engineer	15-1212-00	Software Developers, Applications	Y	2022-03-25	2022-03-24	1	0	1	0	0	0

Results Messages

Run | Cancel | Disconnect | Change | Database: master | A Estimated Plan | Enable Actual Plan | Parse | Enable SQLCMD | To Notebook

```
1 select *
2 from LCA_24
3 where CASE_NUMBER in ('1-200-2229-43062')
```

duplicated rows + duplicated case numbers

CASE_NUMBER	CASE_STATUS	RECEIVED_DATE	DECISION_DATE	ORIGINAL_CERT_DATE	VISA_CLASS	JOB_TITLE	SEC_CODE	SEC_TITLE	FILE_TIME_POSITION	REC_DATE	END_DATE	TOTAL_WORKER_POSITIONS	NEW_EMPLOYMENT	CONTINUED_EMPLOYMENT	CHANGE_PREVIOUS_EMPLOYMENT	NEW_CONCURRENT_EMPLOYMENT	CHANGE_EMPLOYER
1-200-2229-43062	Certified	2022-09-26	2022-10-01	NULL	H-1B	Senior Software Engineer	15-1212-00	Software Developers	Y	2022-03-25	2022-03-24	1	0	1	0	0	0
1-200-2229-43062	Certified	2022-09-26	2022-10-01	NULL	H-1B	Senior Software Engineer	15-1212-00	Software Developers	Y	2022-03-25	2022-03-24	1	0	1	0	0	0
1-200-2229-43062	Certified - Withdrawn	2022-09-26	2022-09-01	2022-10-01	H-1B	Senior Software Engineer	15-1212-00	Software Developers	Y	2022-03-25	2022-03-24	1	0	1	0	0	0

Results Messages

- e. We then discovered that there is total 3.5M rows, with only around 2.2M rows with unique case numbers. We will have to address this table structure when selecting the data we want to analyze with.

Run | Cancel | Disconnect | Change | Database: master | A Estimated Plan | Enable Actual Plan | Parse | Enable SQLCMD | To Notebook

```
1 select
2 count(distinct case_number) as case_count_2020_to_2024
3 from
4 (
5     select case_number
6     from LCA_20
7     union all
8     select case_number
9     from LCA_21
10    union all
11    select case_number
12    from LCA_22
13    union all
14    select case_number
15    from LCA_23
16    union all
17    select case_number
18    from LCA_24
19 )
```

distinct case numbers

case\_number (duplicate not removed)

case_count
3500633

Results Messages

Run | Cancel | Disconnect | Change | Database: master | A Estimated Plan | Enable Actual Plan | Parse | Enable SQLCMD | To Notebook

```
1 select
2 count(distinct case_number) as case_count_2020_to_2024
3 from
4 (
5     select case_number
6     from LCA_20
7     union all
8     select case_number
9     from LCA_21
10    union all
11    select case_number
12    from LCA_22
13    union all
14    select case_number
15    from LCA_23
16    union all
17    select case_number
18    from LCA_24
19 )
```

Results Messages

case_count_20_...
2222331

- f. We also learned that case numbers are generated by job position+company. For example, Samsung filed a case for 25 SDEs in on case, which consisted of 5 new employment, 5 continued employment, 5 change of previously employment, 5 change of employer, and 5 amended petitions.

SEARCH

Home | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL Query | SQL

Now we are clear on the whole picture of how this data was modeled, let's dive into some high level stats for us to understand the data before we decide our goals of analysis and the approaches to sift/clean the data.

## 2. Numerical stats exploration (2020~2024)

- SQL snippet
- `SELECT`
- `COUNT(CASE_NUMBER) as total_row_count`
- `,COUNT(DISTINCT CASE_NUMBER) as unique_case_number`
- `,MIN(RECEIVED_DATE) as earliest_received_date`
- `,MAX(RECEIVED_DATE) as latest_received_date`
- `,MIN(DECISION_DATE) as earliest_decision_date`
- `,MAX(DECISION_DATE) as latest_decision_date`
- `,MIN(BEGIN_DATE) as earliest_begin_date`
- `,MAX(BEGIN_DATE) as latest_begin_date`
- `,ROUND(AVG(WAGE_RATE_OF_PAY_FROM),0) as avg_wage_rate_pay_from`
- `,ROUND(AVG(PREVAILING_WAGE),0) as avg_prevaling_wage`

Results	Messages	YEAR	total_row_count	unique_case_number	earliest_received_date	latest_received_date	earliest_decision_date	latest_decision_date	earliest_begin_date	latest_begin_date	avg_wage_rate_pay_from	avg_prevaling_wage
1		2020	577334	577334	2016-04-27	2020-09-30	2019-10-01	2020-09-30	2016-09-30	2021-03-25	102899	88579
2		2021	826305	528902	2019-10-01	2021-09-30	2020-10-01	2021-09-30	2019-10-03	2022-03-25	107538	94173
3		2022	826303	528902	2019-10-01	2021-09-30	2020-10-01	2021-09-30	2019-10-03	2022-03-25	107538	94173
4		2023	626084	626084	2019-10-01	2022-09-30	2021-10-01	2022-09-30	2019-10-04	2023-03-30	112562	95695
5		2024	644607	543580	2019-10-03	2023-09-30	2022-10-01	2023-09-30	2019-10-14	2024-03-30	115815	98649

- `FROM LCA_20` # change for each year
- Result for each year

## 3. Categorical stats exploration

- SQL snippet (change tables for each year)
- `select case_status, COUNT(case_status) as case_count, SUM(TOTAL_WORKER_POSITIONS) as head_count, ROUND(cast(SUM(TOTAL_WORKER_POSITIONS) as float) / COUNT(case_status),2) AS avg_hc_per_case from LCA_20 group by case_status order by case_count desc;`
- `select VISA_CLASS, count(VISA_CLASS) as case_count, sum(TOTAL_WORKER_POSITIONS) as head_count, ROUND(cast(SUM(TOTAL_WORKER_POSITIONS) as float) / COUNT(case_status),2) AS avg_hc_per_case from LCA_20 group by VISA_CLASS order by case_count desc;`
- `select TOP 10 SOC_TITLE, count(SOC_TITLE) as case_count, sum(TOTAL_WORKER_POSITIONS) as head_count, ROUND(cast(SUM(TOTAL_WORKER_POSITIONS) as float) / COUNT(case_status),2) AS avg_hc_per_case from LCA_20 group by SOC_TITLE order by case_count desc;`
- `select TOP 10 EMPLOYER_NAME, count(EMPLOYER_NAME) as case_count, sum(TOTAL_WORKER_POSITIONS) as head_count, ROUND(cast(SUM(TOTAL_WORKER_POSITIONS) as float) / COUNT(case_status),2) AS avg_hc_per_case from LCA_20 group by EMPLOYER_NAME order by case_count desc;`
- `select TOP 10 EMPLOYER_STATE, count(EMPLOYER_STATE) as case_count, sum(TOTAL_WORKER_POSITIONS) as head_count, ROUND(cast(SUM(TOTAL_WORKER_POSITIONS) as float) / COUNT(case_status),2) AS avg_hc_per_case from LCA_20 group by EMPLOYER_STATE order by case_count desc;`

- Result of each year
  - 2020

Results		Messages		
	case_status	case_count	head_count	avg_hc_per_case
1	Certified	545621	838627	1.54
2	Certified - Withdrawn	16738	22691	1.36
3	Withdrawn	10992	13413	1.22
4	Denied	3983	4754	1.19

  

	VISA_CLASS	case_count	head_count	avg_hc_per_case
1	H-1B	564251	865572	1.53
2	E-3 Australian	10867	11389	1.05
3	H-1B1 Singapore	1156	1239	1.07
4	H-1B1 Chile	1060	1285	1.21

  

	SOC_TITLE	case_count	head_count	avg_hc_per_case
1	Software Developers, Applications	190939	258689	1.35
2	COMPUTER SYSTEMS ANALYSTS	49512	64244	1.3
3	SOFTWARE DEVELOPERS, SYSTEMS SOFTWARE	29878	74161	2.48
4	COMPUTER SYSTEMS ENGINEERS/ARCHITECTS	16822	26760	1.59
5	SOFTWARE QUALITY ASSURANCE ENGINEERS AND TESTERS	14599	23199	1.59
6	COMPUTER PROGRAMMERS	12355	16822	1.36
7	Information Technology Project Managers	12281	24088	1.96
8	COMPUTER AND INFORMATION SYSTEMS MANAGERS	11485	17268	1.5
9	MECHANICAL ENGINEERS	10283	11487	1.12
10	OPERATIONS RESEARCH ANALYSTS	10074	15908	1.58

  

	EMPLOYER_NAME	case_count	head_count	avg_hc_per_case
1	COGNIZANT TECHNOLOGY SOLUTIONS US CORP	28735	28735	1
2	TATA CONSULTANCY SERVICES LIMITED	11998	12817	1.07
3	GOOGLE LLC	10038	10038	1
4	AMAZON.COM SERVICES LLC	9192	12577	1.37
5	INFOSYS LIMITED	9003	23161	2.57
6	Ernst & Young U.S. LLP	8949	8949	1
7	Deloitte Consulting LLP	8204	17959	2.19
8	Microsoft Corporation	8004	8091	1.01
9	ACCENTURE LLP	6202	6323	1.02
10	CAPGEMINI AMERICA INC	5366	5366	1

  

	EMPLOYER_STATE	case_count	head_count	avg_hc_per_case
1	CA	109200	277466	2.54
2	TX	77174	100610	1.3
3	NJ	65055	72651	1.12
4	NY	38807	55348	1.43
5	IL	33091	44169	1.33
6	WA	30609	35990	1.18
7	PA	25311	49487	1.96
8	MA	23352	25595	1.1
9	MI	20823	22148	1.06
10	MD	19196	20279	1.06

○ 2021

	case_status	case_count	head_count	avg_hc_per_case
1	Certified	772369	1708112	2.21
2	Certified - Withdrawn	33619	54869	1.63
3	Withdrawn	15948	25332	1.59
4	Denied	4369	5700	1.3

	VISA_CLASS	case_count	head_count	avg_hc_per_case
1	H-1B	803733	1765673	2.2
2	E-3 Australian	18310	23106	1.26
3	H-1B1 Chile	2314	3156	1.36
4	H-1B1 Singapore	1948	2078	1.07

	SOC_TITLE	case_count	head_count	avg_hc_per_case
1	Software Developers, Applications	265718	497481	1.87
2	Software Developers, Systems Software	51588	175152	3.4
3	Computer Systems Analysts	47970	77348	1.61
4	Computer Systems Engineers/Architects	25905	38396	1.48
5	Software Quality Assurance Engineers and Testers	20539	35236	1.72
6	Information Technology Project Managers	18101	35756	1.98
7	COMPUTER AND INFORMATION SYSTEMS MANAGERS	17479	43729	2.5
8	Mechanical Engineers	15230	25317	1.66
9	Business Intelligence Analysts	15193	34483	2.27
10	Operations Research Analysts	13512	33741	2.5

	EMPLOYER_NAME	case_count	head_count	avg_hc_per_case
1	COGNIZANT TECHNOLOGY SOLUTIONS US CORP	19525	19525	1
2	GOOGLE LLC	16174	16174	1
3	AMAZON.COM SERVICES LLC	14873	150566	10.12
4	Tata Consultancy Services Limited	12147	17168	1.41
5	Microsoft Corporation	11763	49065	4.17
6	Ernst & Young U.S. LLP	10284	10284	1
7	INFOSYS LIMITED	8077	128487	15.91
8	FACEBOOK, INC.	7849	20681	2.63
9	Apple Inc.	6878	87361	12.7
10	ACCENTURE LLP	6378	6424	1.01

	EMPLOYER_STATE	case_count	head_count	avg_hc_per_case
1	CA	184966	609136	3.29
2	NJ	86021	115141	1.34
3	TX	84591	228251	2.7
4	NY	61016	86760	1.42
5	IL	45832	51863	1.13
6	WA	45421	281105	6.19
7	MA	35081	55696	1.59
8	MI	29239	33808	1.16
9	PA	28985	45774	1.58
10	VA	22719	45382	2



	case_status ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	Certified	772369	1708112	2.21
2	Certified – Withdrawn	33617	54867	1.63
3	Withdrawn	15948	25332	1.59
4	Denied	4369	5700	1.3

  

	VISA_CLASS ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	H-1B	803731	1765671	2.2
2	E-3 Australian	18310	23106	1.26
3	H-1B1 Chile	2314	3156	1.36
4	H-1B1 Singapore	1948	2078	1.07

  

	SOC_TITLE ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	Software Developers, Applications	265718	497481	1.87
2	Software Developers, Systems Software	51586	175150	3.4
3	Computer Systems Analysts	47970	77348	1.61
4	Computer Systems Engineers/Architects	25905	38396	1.48
5	Software Quality Assurance Engineers and Testers	20539	35236	1.72
6	Information Technology Project Managers	18101	35756	1.98
7	Computer and Information Systems Managers	17479	43729	2.5
8	Mechanical Engineers	15230	25317	1.66
9	Business Intelligence Analysts	15193	34483	2.27
10	Operations Research Analysts	13512	33741	2.5

  

	EMPLOYER_NAME ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	COGNIZANT TECHNOLOGY SOLUTIONS US CORP	19525	19525	1
2	GOOGLE LLC	16174	16174	1
3	Amazon.com Services LLC	14873	150566	10.12
4	Tata Consultancy Services Limited	12147	17168	1.41
5	Microsoft Corporation	11763	49065	4.17
6	Ernst & Young U.S. LLP	10284	10284	1
7	INFOSYS LIMITED	8077	128487	15.91
8	FACEBOOK, INC.	7849	20681	2.63
9	Apple Inc.	6878	87361	12.7
10	ACCENTURE LLP	6378	6424	1.01

  

	EMPLOYER_STATE ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	CA	184966	609136	3.29
2	NJ	86021	115141	1.34
3	TX	84589	228249	2.7
4	NY	61016	86760	1.42
5	IL	45832	51863	1.13
6	WA	45421	281105	6.19
7	MA	35081	55696	1.59
8	MI	29239	33808	1.16
9	PA	28985	45774	1.58
10	VA	22719	45382	2

## Results Messages

	case_status	case_count	head_count	avg_hc_per_case
1	Certified	576221	1023941	1.78
2	Certified - Withdrawn	35336	41685	1.18
3	Withdrawn	11431	15387	1.35
4	Denied	3096	4179	1.35

	VISA_CLASS	case_count	head_count	avg_hc_per_case
1	H-1B	610883	1067766	1.75
2	E-3 Australian	11628	13089	1.13
3	H-1B1 Chile	2106	2792	1.33
4	H-1B1 Singapore	1467	1545	1.05

	SOC_TITLE	case_count	head_count	avg_hc_per_case
1	Software Developers, Applications	167829	248598	1.48
2	Software Developers	35517	53846	1.52
3	Computer Systems Analysts	34227	56829	1.66
4	SOFTWARE DEVELOPERS, SYSTEMS SOFTWARE	33875	84261	2.49
5	Computer Systems Engineers/Architects	22574	28789	1.28
6	Information Technology Project Managers	15996	25410	1.59
7	COMPUTER AND INFORMATION SYSTEMS MANAGERS	15495	24908	1.61
8	Business Intelligence Analysts	13779	22659	1.64
9	Software Quality Assurance Engineers and Testers	12524	17725	1.42
10	Operations Research Analysts	11641	23610	2.03

	EMPLOYER_NAME	case_count	head_count	avg_hc_per_case
1	Amazon.com Services LLC	16307	81362	4.99
2	COGNIZANT TECHNOLOGY SOLUTIONS US CORP	13906	13906	1
3	Google LLC	11947	11947	1
4	Microsoft Corporation	11547	15665	1.36
5	Tata Consultancy Services Limited	11172	11919	1.07
6	Ernst & Young U.S. LLP	9551	9641	1.01
7	INFOSYS LIMITED	7695	104738	13.61
8	META PLATFORMS, INC.	5407	8379	1.55
9	Deloitte Consulting LLP	5219	7712	1.48
10	Apple Inc.	5053	10248	2.03

	EMPLOYER_STATE	case_count	head_count	avg_hc_per_case
1	CA	132751	306696	2.31
2	TX	70672	185646	2.63
3	NJ	61763	90021	1.46
4	WA	43639	142931	3.28
5	NY	43283	53207	1.23
6	IL	34931	37537	1.07
7	MA	24139	25204	1.04
8	PA	22397	33006	1.47
9	MI	19339	20566	1.06
10	MD	18362	19254	1.05

○ 2024

	case_status ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	Certified	588259	1087496	1.85
2	Certified – Withdrawn	41957	44683	1.06
3	Withdrawn	11375	13930	1.22
4	Denied	3016	5658	1.88

	VISA_CLASS ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	H-1B	626887	1129672	1.8
2	E-3 Australian	12891	14758	1.14
3	H-1B1 Chile	2861	4963	1.73
4	H-1B1 Singapore	1968	2374	1.21

	SOC_TITLE ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	Software Developers	213166	409083	1.92
2	Computer Systems Analysts	35740	51950	1.45
3	Computer Systems Engineers/Architects	28692	41948	1.46
4	Information Technology Project Managers	21663	47709	2.2
5	Software Quality Assurance Analysts and Testers	18349	29837	1.63
6	COMPUTER AND INFORMATION SYSTEMS MANAGERS	15557	27434	1.76
7	Data Scientists	12688	19782	1.56
8	Computer Programmers	11661	16090	1.38
9	Business Intelligence Analysts	11542	23281	2.02
10	Accountants and Auditors	11268	16852	1.5

	EMPLOYER_NAME ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	Amazon.com Services LLC	16293	117923	7.24
2	COGNIZANT TECHNOLOGY SOLUTIONS US CORP	15729	15729	1
3	Ernst & Young U.S. LLP	13366	13366	1
4	Google LLC	11723	11723	1
5	Tata Consultancy Services Limited	10212	10235	1
6	Microsoft Corporation	9611	15818	1.65
7	INFOSYS LIMITED	9269	48599	5.24
8	Apple Inc.	5380	12891	2.4
9	Meta Platforms, Inc.	5162	7039	1.36
10	AMAZON WEB SERVICES, INC.	4510	27761	6.16

	EMPLOYER_STATE ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	CA	128931	338833	2.63
2	TX	79369	137521	1.73
3	NJ	64464	94531	1.47
4	NY	48247	60336	1.25
5	WA	40249	201723	5.01
6	IL	33816	36404	1.08
7	MA	24846	25921	1.04
8	PA	21831	31131	1.43
9	VA	20128	24355	1.21
10	MI	20115	20684	1.03

2020~2024

	case_status ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	Certified	3254839	6366288	1.96
2	Certified - Withdrawn	161267	218795	1.36
3	Withdrawn	65694	93394	1.42
4	Denied	18833	25991	1.38

	VISA_CLASS ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	H-1B	3409485	6594354	1.93
2	E-3 Australian	72006	85448	1.19
3	H-1B1 Chile	10655	15352	1.44
4	H-1B1 Singapore	8487	9314	1.1

	SOC_TITLE ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	Software Developers, Applications	896524	1509210	1.68
2	Software Developers	249082	463328	1.86
3	COMPUTER SYSTEMS ANALYSTS	215419	327719	1.52
4	SOFTWARE DEVELOPERS, SYSTEMS SOFTWARE	169174	510994	3.02
5	COMPUTER SYSTEMS ENGINEERS/ARCHITECTS	119898	174289	1.45
6	Information Technology Project Managers	86142	168719	1.96
7	COMPUTER AND INFORMATION SYSTEMS MANAGERS	77495	157068	2.03
8	Software Quality Assurance Engineers And Testers	69010	112271	1.63
9	Business Intelligence Analysts	63149	125267	1.98
10	MECHANICAL ENGINEERS	62285	87917	1.41

	EMPLOYER_NAME ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	COGNIZANT TECHNOLOGY SOLUTIONS US CORP	97420	97420	1
2	AMAZON.COM SERVICES LLC	71538	512994	7.17
3	Google LLC	66056	66056	1
4	TATA CONSULTANCY SERVICES LIMITED	57676	69307	1.2
5	Microsoft Corporation	52688	137704	2.61
6	Ernst & Young U.S. LLP	52434	52524	1
7	INFOSYS LIMITED	42121	433472	10.29
8	Deloitte Consulting LLP	28821	49582	1.72
9	Apple Inc.	27891	205469	7.37
10	Accenture LLP	26211	26588	1.01

	EMPLOYER_STATE ▾	case_count ▾	head_count ▾	avg_hc_per_case ▾
1	CA	740814	2141267	2.89
2	TX	396395	880277	2.22
3	NJ	363324	487485	1.34
4	NY	252369	342411	1.36
5	WA	205339	942854	4.59
6	IL	193502	221836	1.15
7	MA	142499	188112	1.32
8	PA	127509	205172	1.61
9	MI	118755	131014	1.1
10	MD	101450	118050	1.17