Analysis of United States Department of Labor Guest Worker Visa Data

What Has Transpired between 2016 and 2020?

SAVE AMERICAN INFORMATION TECHNOLOGY JOBS (SAITJ)

Wednesday, July 8th, 2021

Table of Contents

Table of Figures

Table of Tables

Definitions and Acronyms

TERM	MEANING
DOL	The United States Department of Labor
OPT	Optional Practical Training program

WORK VISA STATUS	MEANING
Certified	The visa application was approved.
Certified-Withdrawn	The visa application was approved but was subsequently withdrawn.
Denied	The visa application was not approved.
Withdrawn	The visa application was in the vetting process but was withdrawn. Perhaps because the candidate left the country or decided not to pursue obtaining the work visa.

SKILL LEVEL	MEANING
Level 1	The applicant is considered “Entry-Level”
Level 2	The applicant is considered “Qualified”
Level 3	The applicant is considered “Experienced”
Level 4	The applicant is considered “Fully Competent”
Level 5	This is a generic unspecified level

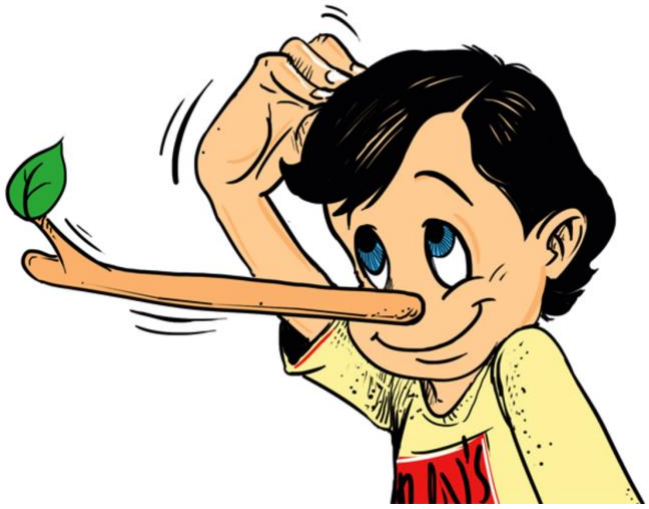
Come With Us And Learn How Deep The Rabbit Hole Goes

The various corporate lobbyists and congress and the executive branch of the US Government has stacked the deck against its own citizens - those best and brightest IT engineers are forced into unemployment under the terms of training their own replacements - foreign H-1B and OPT workers from half way around the globe. A form of humiliation that is unacceptable and cruel to any one, in any profession, Anywhere in this county; indeed, anywhere in the world. I cannot think of any other country in the world that treats its own citizens in such a horrible fashion, a fashion of cruelty that the AT&T, Verizon, Intel and many other companies to long to list and count has so well crafted.

A cruelty that the U.S. government and its so-called Labor Department wrecking crew has maintained and encouraged, a cruelty that digs deep due to promises made and not kept to American workers by administrations that has lied to its citizens and voter base for decades.

IT Pros are struggling to stay in technical positions, but they are not hired, some are force to leave the profession and attempt to branch into business analysts and other roles; that is if you have the architectural and writing and presentation skills. Some of us are born to code and are deprived of the inalienable right.

The Famous Industry Lie



Industry continues to propagate the claim that there are not enough high-skilled workers to fill demand. Hence, industry states it is "forced" to seek high-skilled workers through the United States governments work visa program(s). It seems there are just not enough competent workers in the United States workforce.

The available data indicates this claim is absurd.

The number of high-skilled (H1-B) applications certified by the United States Department of Labor between 2016 and 2020 inclusive highlights the absurdity of industries claim.

SKILL LEVEL	2016	2017	2018	2019	2020	MEAN
Level 1	NO DATA	188079	87506	80945	70565	106773.75
Level 2	NO DATA	156754	321	312	362	39437.25
Level 3	NO DATA	80	183	195	206	166
Level 4	NO DATA	59	2596	2886	2780	2080.25
Level 5	NO DATA	0	0	0	0	0
Total Records	-	344972	90606	84338	73913	-

TABLE 1. NUMBER OF CERTIFIED H-1B VISA APPLICATIONS BY YEAR BY SKILL LEVEL

Skill level for certified high-skilled H-1B visa holders falls into five (5) groups. Level 1 is Entry, Level 2 is Qualified, Level 3 is Experienced, Level 4 is Fully Competent. Interestingly, Level 5 is an undefined skill level. It is currently unclear what factors determine candidate placement in the Level 5 skill level category.

Table 1 above shows that the majority of all certified H-1B visas were granted to individuals classified at skill level 1 (Entry Level) and skill level 2 (Qualified). H-1B visas granted to individuals considered experienced (Skill Levels 3-4) is shockingly low

relative to entry level and qualified individuals. Finally, no visas were granted to level 5 (undefined) individuals. This is questionable and requires further research.

The equation below gives the percentage of entry level and qualified individuals granted a certified H-1B visa during years 2017 through 2020:

$$\frac{(nL1 + nL2)}{Tn} * 100 = ELP$$

Let nL1 = sum of Level 1 (Entry Level) certified H-1B applications from 2017 through 2020
Let nL2 = sum of Level 2 (Qualified) certified H-1B applications from 2017 through 2020
Let Tn = the sum of all certified H-1B applications from 2017 through 2020
Let ELP = The Experience Level Percent

Where nL1 = (188079 + 87506 + 80945 + 70565) = 427095
Where nL2 = (156754 + 321 + 312 + 362) = 157749
Where Tn = (344972 + 90606 + 84338 + 73913) = 593829

The sum of nL1 and nL2 divided by Tn gives the percent of H-1B visas granted to individuals classified as entry level or qualified. Multiplying this value by one hundred (100) shifts the decimal point to the right by two (2) for readability.

((427095 + 157749) / 593829) * 100 = 98.486938158964955%

Basic subtraction can be used to show only 1.513061841035045% of the total certified H-1B applications from 2017 through 2020 were granted to applicants deemed experienced or fully-competent.

The sheer magnitude of the discrepancy between entry level and high-skilled (aka fully-competent) certified H-1B applicants is illustrated (Figure 1).

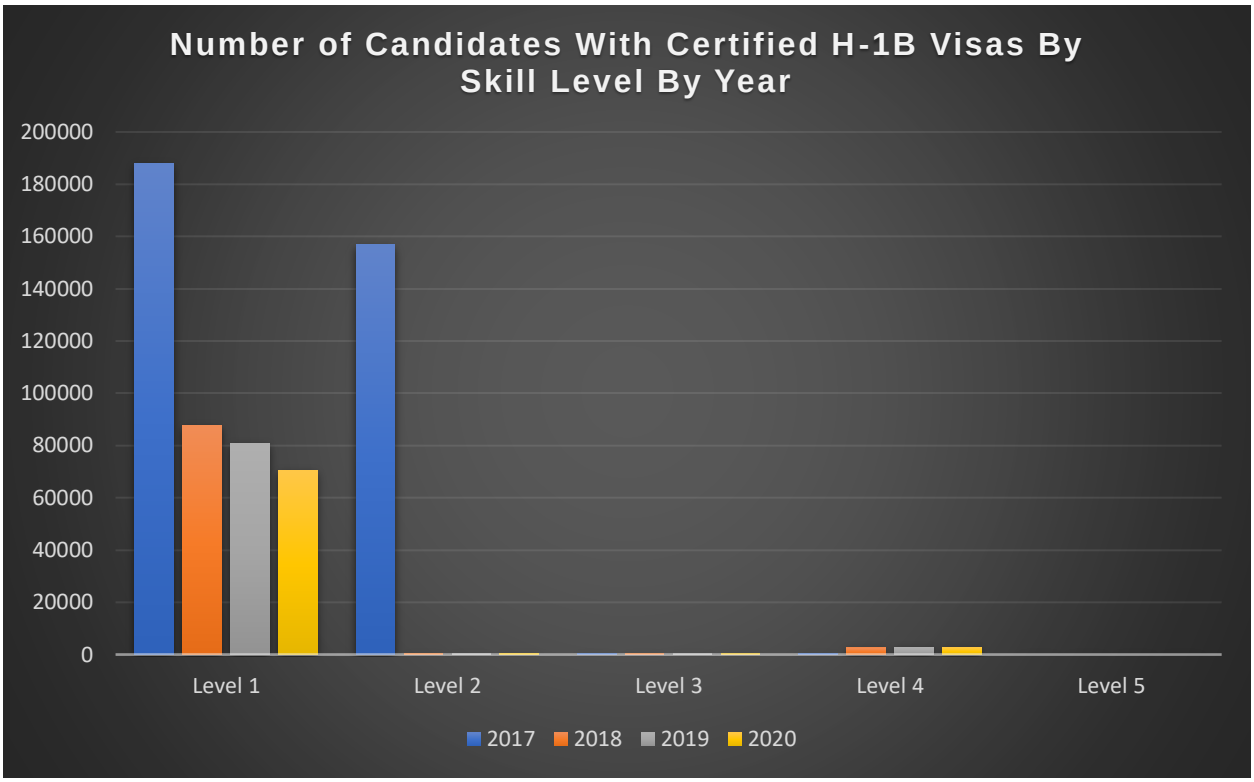


FIGURE 1. NUMBER OF CANDIDATES WITH CERTIFIED H-1B VISAS BY SKILL LEVEL BY YEAR

The magnitude of the disparity in certified H-1B visas by skill level is apparent when one considers the quantity of certified H-1B visas. Specifically, the volume of certified H-1B visas granted in 2017 for Level 1 (Entry) and Level 2 (Qualified) applicants dwarfs the volume of certified H-1B visas granted for Level 2 (Qualified) applicants in 2018, 2019 and 2020.

The number of Level 3 (Experienced) applicants is so small the values are not displayed. There were no Level 5 (undefined) applicants. As previously stated, the fact that there were no level 5 (undefined) applicants is questionable and requires further research.

The number of Level 4 applicants does provide visual representation of the disparity between level 1 (Entry) and level 2 (Qualified) and more experienced applicants (Level 3-5).

Industry has presented a false narrative to lawmakers and the American public. Industry and Congress want cheap foreign workers on H-1B visas because Industry can pay significantly less in wages to those workers and facilitate indentured servitude. Which leaves American workers out on the street.

How Long Does It Take To Get A Guest-Worker Visa Certified?

The United States Department of Labor (DOL) is tasked with vetting visa applications. This section of the summary is intended to provide readers with an understanding of the time (in days) the DOL processed guest-worker visas where the final decision was to certify the visa application.

VISA TYPE	2016	2017	2018	2019	2020
H-1B American	6	5.74488823	5.7730185	5.7865716	7.1510488
H-1B1 CHILE	5.850605653	5.82323232	5.7867903	5.8564307	7.2167019
H-1B1 SINGAPORE	5.777429467	5.85215794	5.810585	5.8519179	7.18911439
E-3 AUSTRALIAN	5.835179219	5.80779042	5.827491	5.8451363	7.18569969

TABLE 2. TIME IN DAYS TO CERTIFY GUEST WORKER VISAS BY VISA TYPE BY YEAR

Table 2 shows that for years 2016 through 2019 the average daily rate for visa certification was relatively consistent. The average daily rate did not significantly very for years 2016-2019. Additionally, the average daily rate also did not very by visa type during these years. Only in 2020, during the pandemic and under the Trump administrations watch, does a minor change in the average daily rate appear. Based on those conditions one might speculate the daily rate would be higher.

The H-1B1 Chile, H-1B1 Singapore and E-3 Australian visas are listed for comparison. A discussion on those visa types is outside the scope of this document.

How Long Does It Take To Get A Guest-Worker Visa Denied?

This section of the summary is intended to provide readers with an understanding of the time (in days) the DOL processed guest-worker visas where the final decision was to deny the visa application.

VISA TYPE	2016	2017	2018	2019	2020
H-1B American	4	4	3.222369643	4.0195509	3.1443539
H-1B1 CHILE	4	4	2.97104439	3.2096774	3.95
H-1B1 SINGAPORE	3.30612245	4	2.589285714	2.972973	4.13043478
E-3 AUSTRALIAN	3.5826859	4	2.859067358	3.4394904	3.65086207

TABLE 3. TIME IN DAYS TO DENY GUEST WORKER VISAS BY VISA TYPE BY YEAR

Table 3 illustrates for years 2016 through 2020 the average daily rate for visa denial varied significantly but notably for 2016 the average daily rate for visa denial remained consistent for both H-1B and H-1B1 Chile visa types. Surprisingly, in 2017 the average daily rate for visa denial did not vary regardless of visa type. Only in 2020, during the pandemic and under the Trump administrations watch, does a minor change in the average daily rate appear. Based on those conditions one might speculate the daily rate would be higher.

The H-1B1 Chile, H-1B1 Singapore and E-3 Australian visas are listed for comparison. A discussion on those visa types is outside the scope of this document.

All Visa Cases Per Year By Case Status

CASE STATUS	2016	2017	2018	2019	2020	Mean	Median
CERTIFIED-WITHDRAWN	47092	49704	45004	46946	16738	41069.8	46946
WITHDRAWN	2894	20772	21280	19673	10992	18922.2	20772
CERTIFIED	569646	545694	579449	592102	545621	566502.4	569646
DENIED	9220	8480	8627	5893	3983	7240.6	8480
TOTAL	647852	624650	654360	664614	577334	633762	647852
% CERTIFIED-WITHDRAWN (ROUNDED)	7.27	7.96	6.88	7.07	2.90	6.48	7.25
% WITHDRAWN (ROUNDED)	3.38	3.33	3.25	2.96	1.90	2.99	3.21
% CERTIFIED (ROUNDED)	87.93	87.36	88.55	89.09	94.51	89.39	87.93
% DENIED (ROUNDED)	1.42	1.36	1.32	0.89	0.69	1.14	1.31

TABLE 4. NUMBER OF VISA CASES BY CASE STATUS BY YEAR

Table 4 lists the total number of visa cases by case status by year with percentages for each case status also listed by year. Figures 2 through 4 below illustrate notable findings.

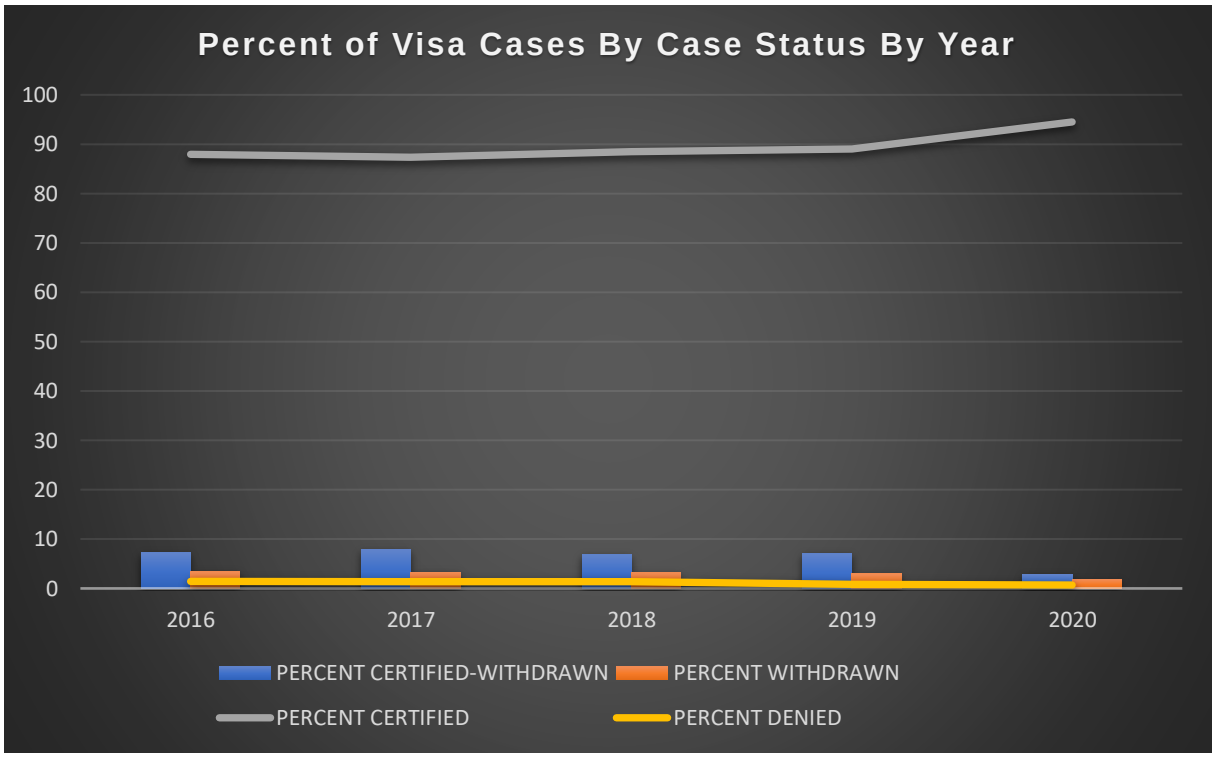


FIGURE 2. PERCENT OF VISA CASES BY CASE STATUS BY YEAR

Figure 2 shows the magnitude of the percentage of visas by case status by year during the years 2016 through 2020. Notably, the percentage of certified visas is exceptionally high relative to the percentage for all other cases status values.

An increase in the percentage of certified visas in 2020 indicates that a global pandemic neither caused a reduction or maintaining of the trend seen in 2016 though 2019. The Trump administrations so called “crackdown” on H-1B visas was a failure.

Discounting other visa case status values allows for bringing the magnitude of the percentage of certified visas cases into sharp relief.

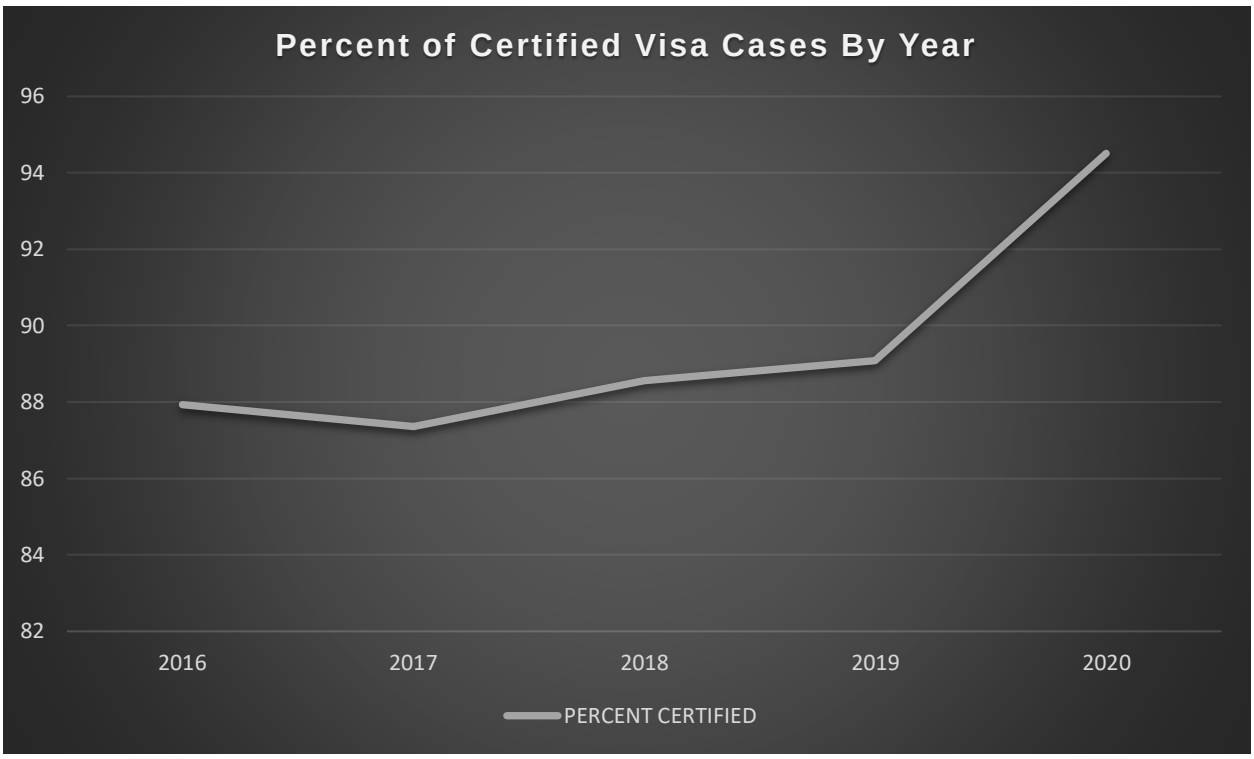


FIGURE 3. PERCENT OF CERTIFIED VISA CASES BY YEAR

Figure 3 illustrates that although a slight decline in the percentage of certified visas occurred in 2017 a gradual increase occurred during 2018 and 2019. In 2020 an exceptionally large increase was revealed. From 2017 through 2020 the percentage of certified visas cases has increased despite the Trump administration’s policy on immigration.

Figure 4 below brings the percentage of non-certified visas cases illustrated into sharp relief by discounting the percentage of certified visas.

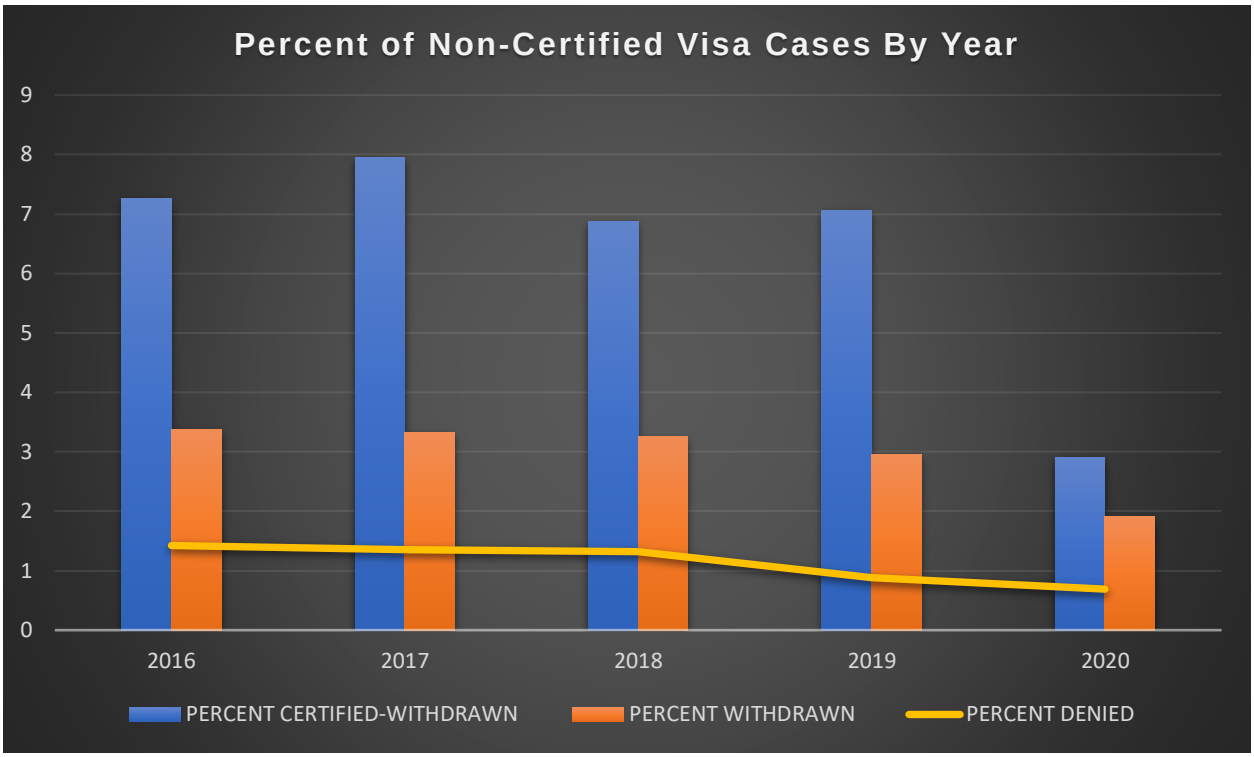


FIGURE 4. PERCENT OF NON-CERTIFIED VISA CASES BY YEAR

Figure 4 illustrates that there was a gradual decline in the percentage of visas denied and withdrawn between 2016 and 2020. The percent of visas certified-withdrawn varied between 2016 and 2020 with a drastic reduction observed in 2020.

It is clear the trend is to increase the number of guest-workers while simultaneously reducing the number of guest-worker non-certified visa cases during the Trump and pandemic era.

All Certified Cases Per Year By Visa Type

VISA TYPE	2016	2017	2018	2019	2020	Mean	Median
H-1B American	558014	533621	567011	578639	533801	554217.2	558014
H-1B1 CHILE	743	792	863	1003	946	869.4	863
H-1B1 SINGAPORE	957	1089	1077	1121	1084	1065.6	1084
E-3 AUSTRALIAN	9932	10192	10498	11339	9490	10350.2	10192
TOTAL	569646	545694	579449	592102	545621	566502.4	569646
% CERTIFIED H-1B American (ROUNDED)	97.96	97.79	97.85	97.73	97.83	97.83	97.96
% CERTIFIED H-1B1 CHILE (ROUNDED)	0.13	0.15	0.15	0.17	0.17	0.15	0.15
% CERTIFIED H-1B1 SINGAPORE (ROUNDED)	0.17	0.20	0.19	0.19	0.20	0.19	0.19
% CERTIFIED E-3 AUSTRALIAN (ROUNDED)	1.74	1.87	1.81	1.92	1.79	1.83	1.79

TABLE 5. NUMBER OF CERTIFIED VISA CASES PER YEAR BY VISA TYPE

Table 5 lists the total number of certified visa cases by visa type by year with percentages for each certified visa type also listed by year. Figures 5 and 6 below illustrate notable findings.

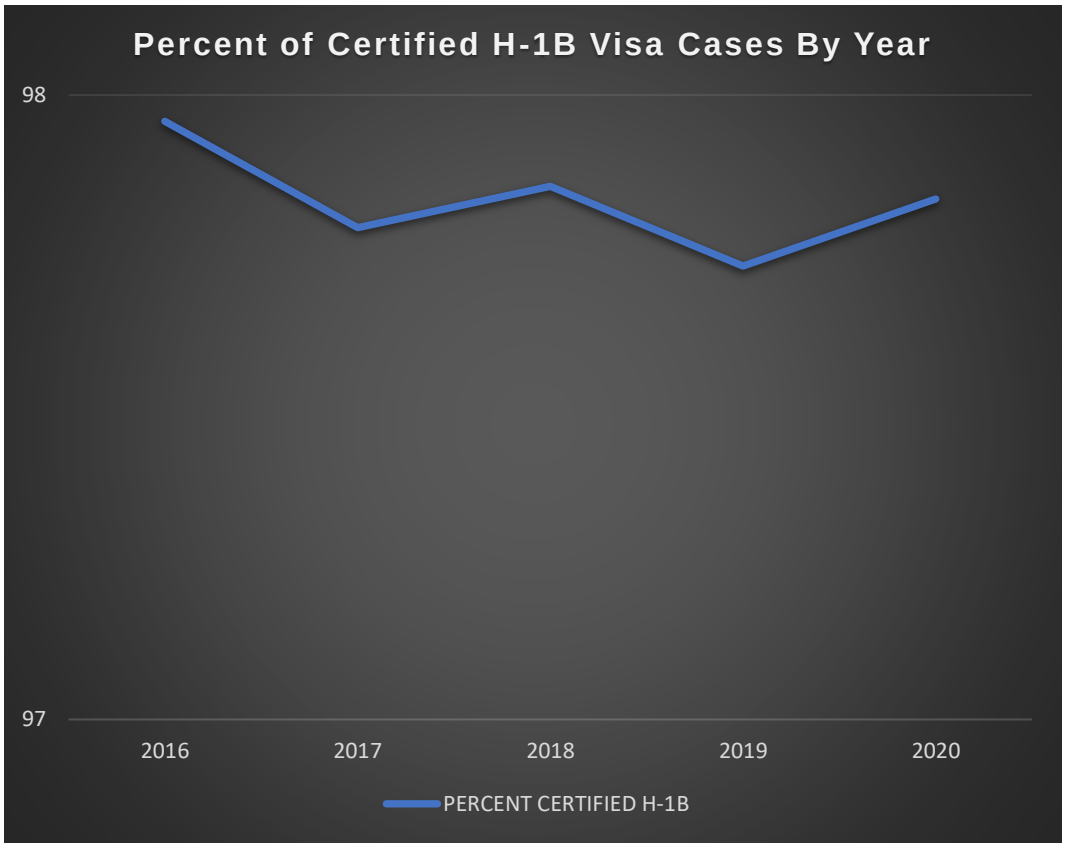


FIGURE 5. PERCENT OF CERTIFIED H-1B VISA CASES BY YEAR

Figure 5 illustrates that between 2016 and 2020 the percent of certified H-1B visas granted has declined slightly.

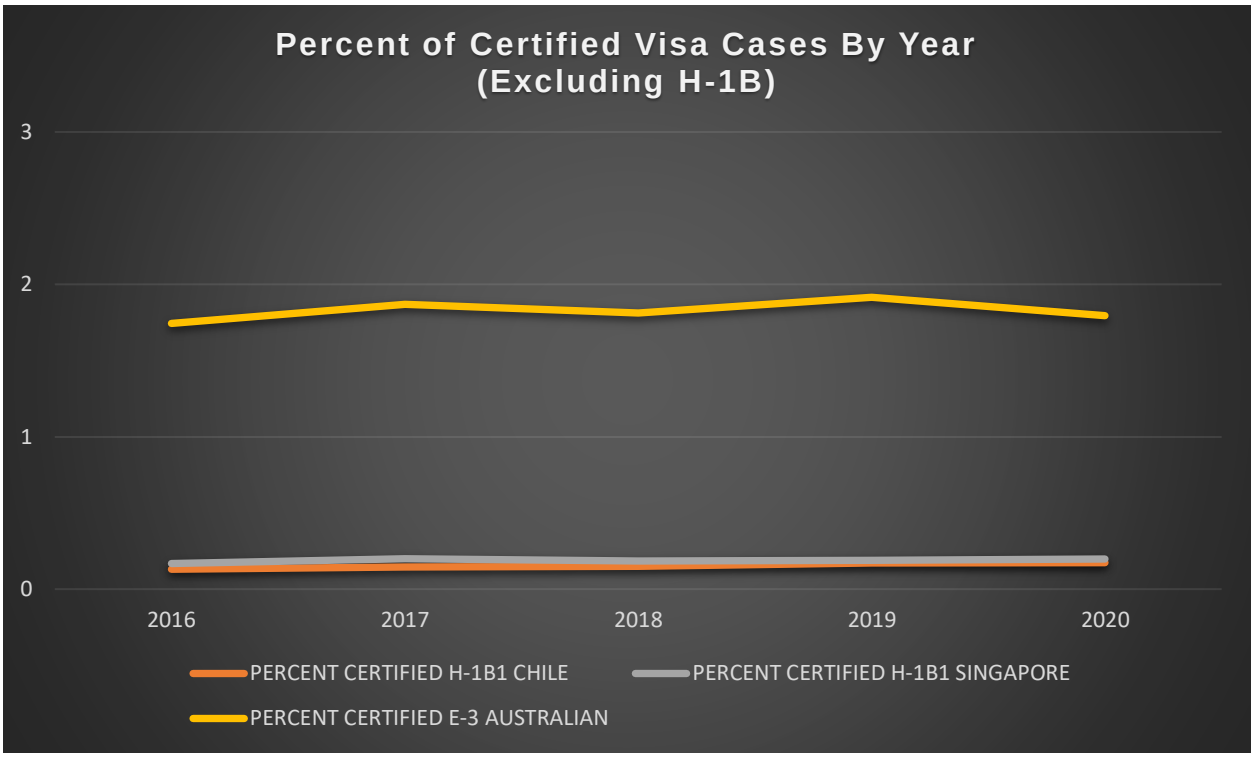


FIGURE 6. PERCENT OF CERTIFIED VISA CASES BY YEAR (EXCLUDING H-1B)

Figure 6 illustrates that the percentage of certified H-1B1 Chile, H-1B1 Singapore and E-3 Australian visas is both inconsequential when compared with the percentage of certified H-1B visas and the percentage of certified visas remains consistent from 2016 through 2020.

All H-1B Certified Cases Per Year By Worksite State / Territory

This section illustrates where these so called “high-skilled” guest workers are prominent geographically. This allows the capability to infer which states are the worst for American IT workers.

STATE / TERRITORY	2016	2017	2018	2019	2020	Total	Mean	Median
California (CA)	100902	99770	109367	113643	106367	530049	106009.8	106367
Texas (TX)	59642	57069	56167	58681	55147	286026	57205.2	57069
New York (NY)	47902	44568	47585	46469	40806	227330	45466	46469
New Jersey (NJ)	42857	36787	36291	34492	29431	179858	35971.6	36291
Illinois (IL)	30915	28125	28010	28918	25983	141951	28390.2	28125
Washington (WA)	19714	21907	24175	26789	30569	123154	24630.8	24175
Massachusetts (MA)	21325	21382	22480	22609	21173	108969	21793.8	21382
Pennsylvania (PA)	20978	18679	22948	20298	18605	101508	20301.6	20298
Georgia (GA)	22068	19824	50551	20101	18514	101058	20211.6	20101
Florida (FL)	18340	18659	19871	19131	15792	91793	18358.6	18659
Michigan (MI)	18253	17363	18770	17867	15732	87985	17597	17867
North Carolina (NC)	16395	14889	16506	19182	19709	86681	17336.2	16506
Virginia (VA)	16503	15449	16389	16823	15787	80951	16190.2	16389
Ohio (OH)	14647	13600	14314	14801	13228	70590	14118	14314
Maryland (MD)	9320	9042	9247	9293	8612	45514	9102.8	9247
Arizona (AZ)	8611	8353	5540	10165	9453	45422	9084.4	8840
Minnesota (MN)	8870	9167	9620	9376	8358	45391	9078.2	9167
Connecticut (CT)	9024	7076	7796	9221	8225	41342	8268.4	8225
Missouri (MO)	6823	6958	7775	8164	7529	37249	7449.8	7529
Tennessee (TN)	4910	6005	6480	6432	6017	29844	5968.8	6017
Colorado (CO)	5356	5558	6357	6529	5878	29678	5935.6	5878
Wisconsin (WI)	6233	5717	5940	6156	5111	29157	5831.4	5940
Indiana (IN)	5380	5228	5915	6587	5596	28706	5741.2	5596
Oregon (OR)	4039	4023	3934	4513	4063	20572	4114.4	4039
District of Columbia (DC)	3447	3507	3585	3610	3354	17503	3500.6	3507
Arkansas AR	3042	2429	3170	4008	3948	16597	3319.4	3170
Delaware (DE)	3065	2997	3505	3076	2735	15378	3075.6	3065
Utah (UT)	2445	2557	2746	3285	2749	13782	2756.4	2746
Iowa (IA)	2496	2652	2966	2872	2728	13714	2742.8	2728
South Carolina (SC)	2489	2488	2808	2924	2424	13133	2626.6	2489
Kentucky (KY)	2331	2363	2480	2631	2503	12308	2461.6	2480
Kansas (KS)	2434	2378	2419	2340	2006	11577	2315.4	2378
Rhode Island (RI)	2233	2148	2041	2315	2128	10865	2173	2148
Nebraska (NE)	1757	1847	2175	2037	1770	9586	1917.2	1847
Alabama (AL)	1725	1471	1678	1699	1561	8134	1626.8	1678
Louisiana (LA)	1847	1613	1516	1591	1291	7858	1571.6	1591
Oklahoma (OK)	1745	1497	1699	1534	1354	7811	1562.2	1534
New Hampshire (NH)	1665	1250	1450	1505	1472	7342	1468.4	1472
Nevada (NV)	1093	1189	1411	1446	1186	6325	1265	1189
New Mexico (NM)	914	735	862	910	566	4287	857.4	566
Idaho (ID)	755	759	796	753	686	3749	749.8	755
Mississippi (MS)	695	637	650	750	676	3408	681.6	676
Maine (ME)	616	590	581	470	476	2733	546.6	581
West Virginia (WV)	467	392	443	434	390	2126	425.2	434
North Dakota (ND)	450	383	429	375	296	1933	386.6	383
Hawaii (HI)	390	412	416	392	307	1917	383.4	392
Vermont (VT)	308	355	384	315	200	1562	312.4	315
Guam (GU)	290	291	254	323	261	1419	283.8	290
South Dakota (SD)	290	251	314	234	266	1355	271	266

Puerto Rico (PR)	151	525	124	100	78	978	195.6	124
Northern Mariana Islands (MP)	110	263	256	90	77	796	159.2	110
Montana (MT)	151	153	188	147	130	769	153.8	151
Alaska (AK)	139	162	148	93	75	617	123.4	139
Wyoming (WY)	91	100	152	98	71	512	102.4	98
Virgin Islands (VI)	51	44	35	40	51	221	44.2	44
Marshall Islands (MH)	4	0	1	0	0	5	1	0
Palau (PW)	1	2	0	1	1	5	1	1
Micronesia (FM)	0	1	0	0	0	1	0.2	0

TABLE 6. NUMBER OF CERTIFIED H-1B CASES BY STATE / TERRITORY BY YEAR

Table 6 shows the number of certified H-1B cases by state / territory by year in descending order by the number of certified H-1B cases per year.

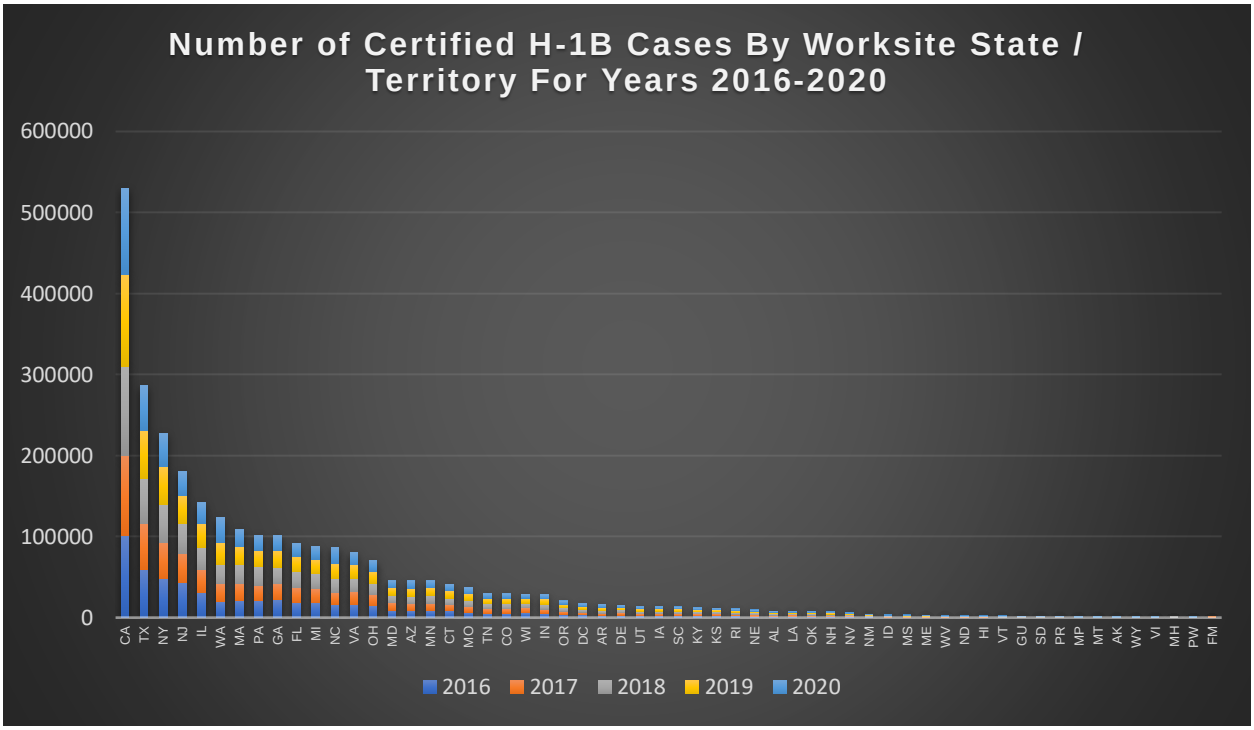


FIGURE 7. NUMBER OF CERTIFIED H-1B CASES BY WORKSITE STATE / TERRITORY FOR YEARS 2016-2020

Figure 7 provides a stunning representation of the data in Table 6 above. Beyond conveying which states / territories have the most H-1B workers the cumulative magnitude is also revealed.

The period from 2016 through 2020 shows that the top fourteen (14) states have consistently had an increase of over thirteen thousand (13000) H-1B workers each year with California having a stunning yearly average increase in H-1B workers of (106009.8). The top fourteen (14) states in descending order by cumulative H-1B workers are:

1. California (CA)
2. Texas (TX)
3. New York (NY)
4. New Jersey (NJ)
5. Illinois (IL)
6. Washington (WA)
7. Massachusetts (MA)
8. Pennsylvania (PA)
9. Georgia (GA)
10. Florida (FL)
11. Michigan (MI)
12. North Carolina (NC)
13. Virginia (VA)
14. Ohio (OH)

A 35.52% reduction in the cumulative yearly average increase in H-1B workers is observed when comparing the 14th and 15th states (Ohio and Maryland). This reduction was computed as follows:

((Mean (OH) – Mean (MD)) / Mean (OH)) * 100
((14118 – 9102.8) / 14118) * 100
(5015.2 / 14118) * 100 = **35.52%** (rounded to two decimal places)

The 15th through the 39th states (Maryland through Nevada) show a consistent gradual reduction in the cumulative yearly average increase in H-1B workers.

An additional 32.22% reduction in the cumulative yearly average increase in H-1B workers is observed when comparing the 39th and 40th states (Nevada and New Mexico). This reduction was computed as follows:

((Mean (NV) – Mean (NM)) / Mean (NV)) * 100
((1265 – 857.4) / 1265) * 100
(407.6 / 1265) * 100 = **32.22%** (rounded to two decimal places)

Additional significant reductions occur when comparing the cumulative yearly average increase in H-1B workers between Wyoming and the Virgin Islands and between the Virgin Islands and the Marshall Islands.

Finally, the period from 2016 through 2020 shows that for the top six (6) states:

1. California (CA)
2. Texas (TX)
3. New York (NY)
4. New Jersey (NJ)
5. Illinois (IL)
6. Washington State (WA)

The decision to exclude each year for these states beginning with 2020 has no effect on the cumulative magnitude of H-1B workers relative to the other six (6) states. Specifically, removing 2020, 2019, 2018 and 2017 data has no effect on the list of the top six (6) states where H-1B guest workers are located.

Figures 8-13 below provide further granularity of the cumulative magnitude of H-1B workers by splitting Figure 7 into six (6) figures representing a subset of the data in Figure 7.

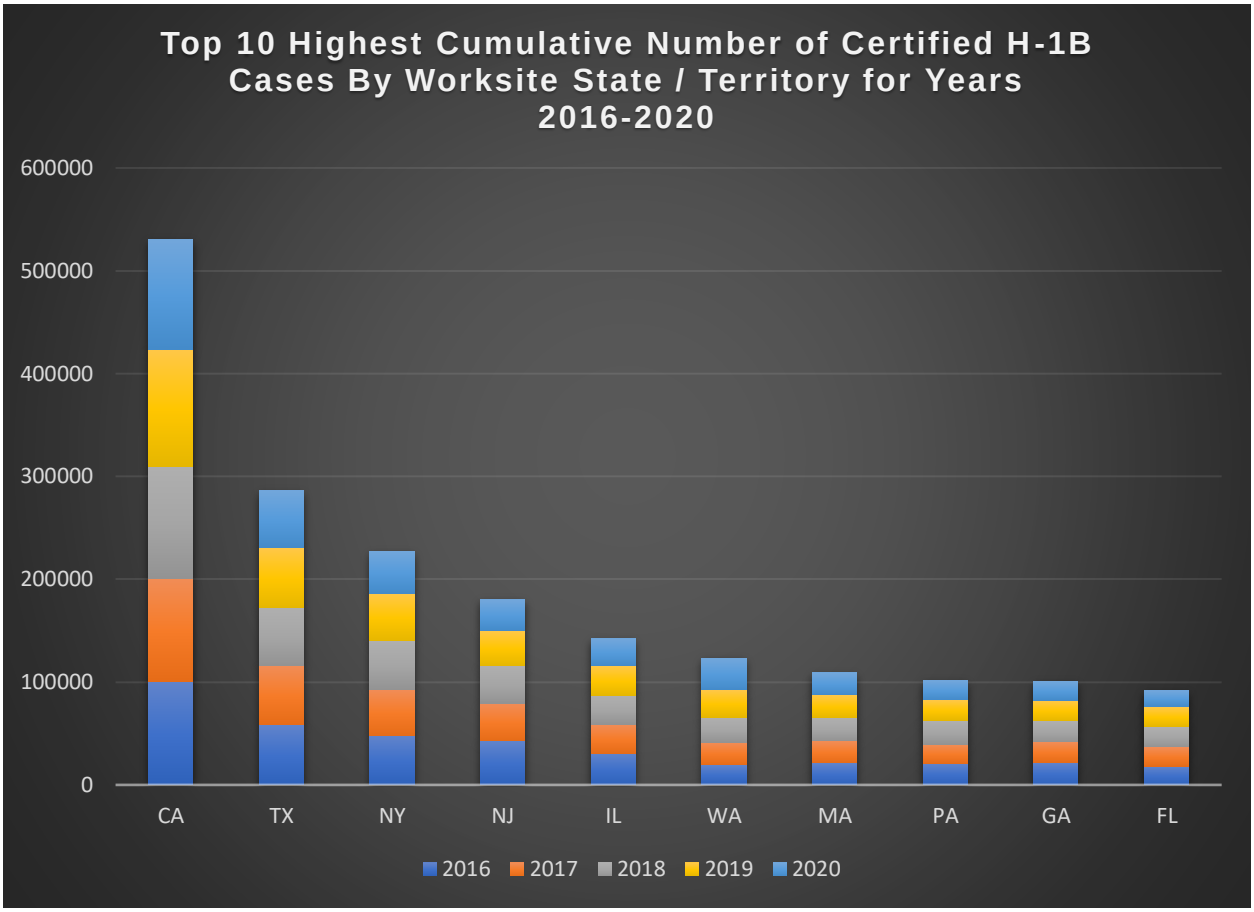


FIGURE 8. TOP 10 HIGHEST CUMULATIVE NUMBER OF CERTIFIED H-1B CASES BY WORKSITE STATE / TERRITORY FOR YEARS 2016-2020

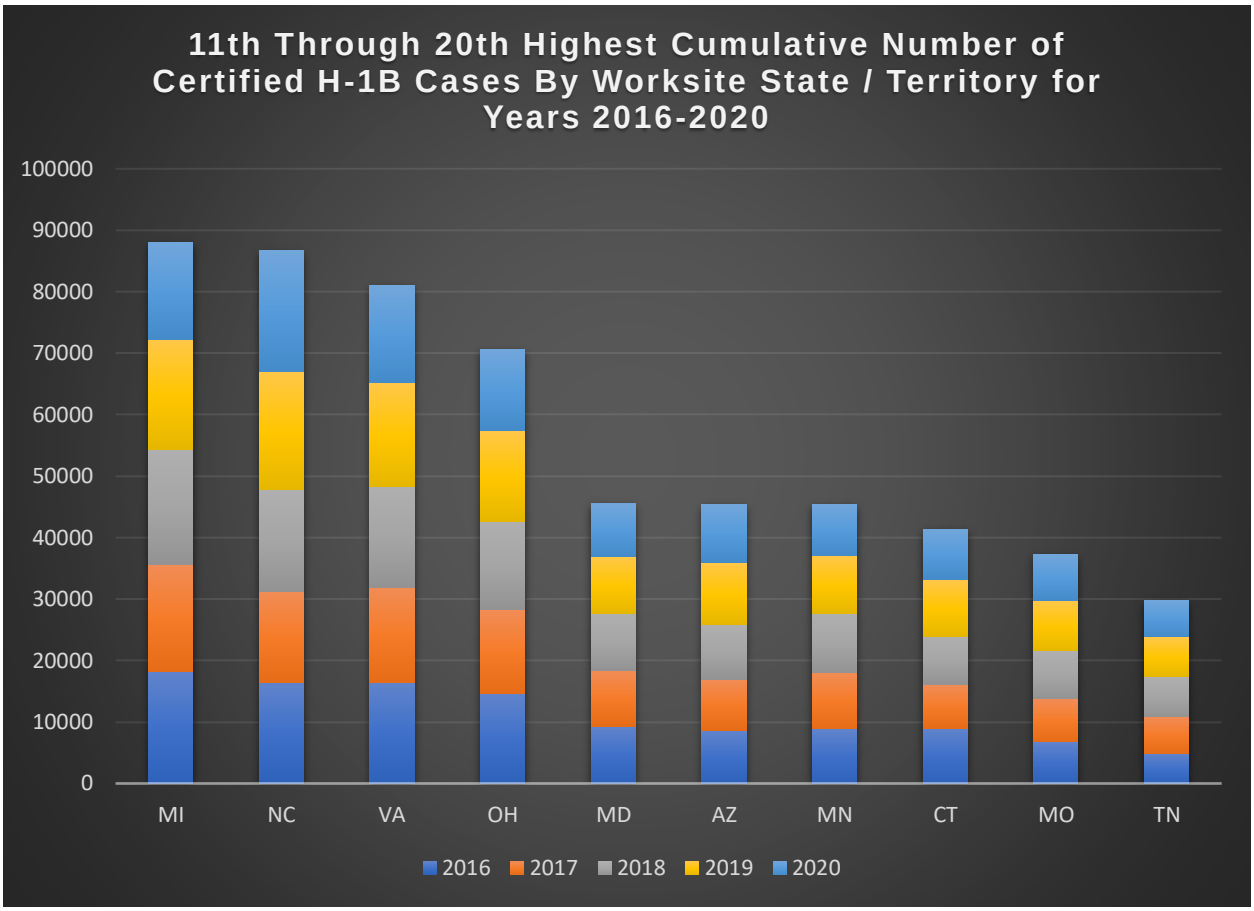


FIGURE 9. 11TH THROUGH 20TH HIGHEST CUMULATIVE NUMBER OF CERTIFIED H-1B CASES BY WORKSITE STATE / TERRITORY FOR YEARS 2016-2020

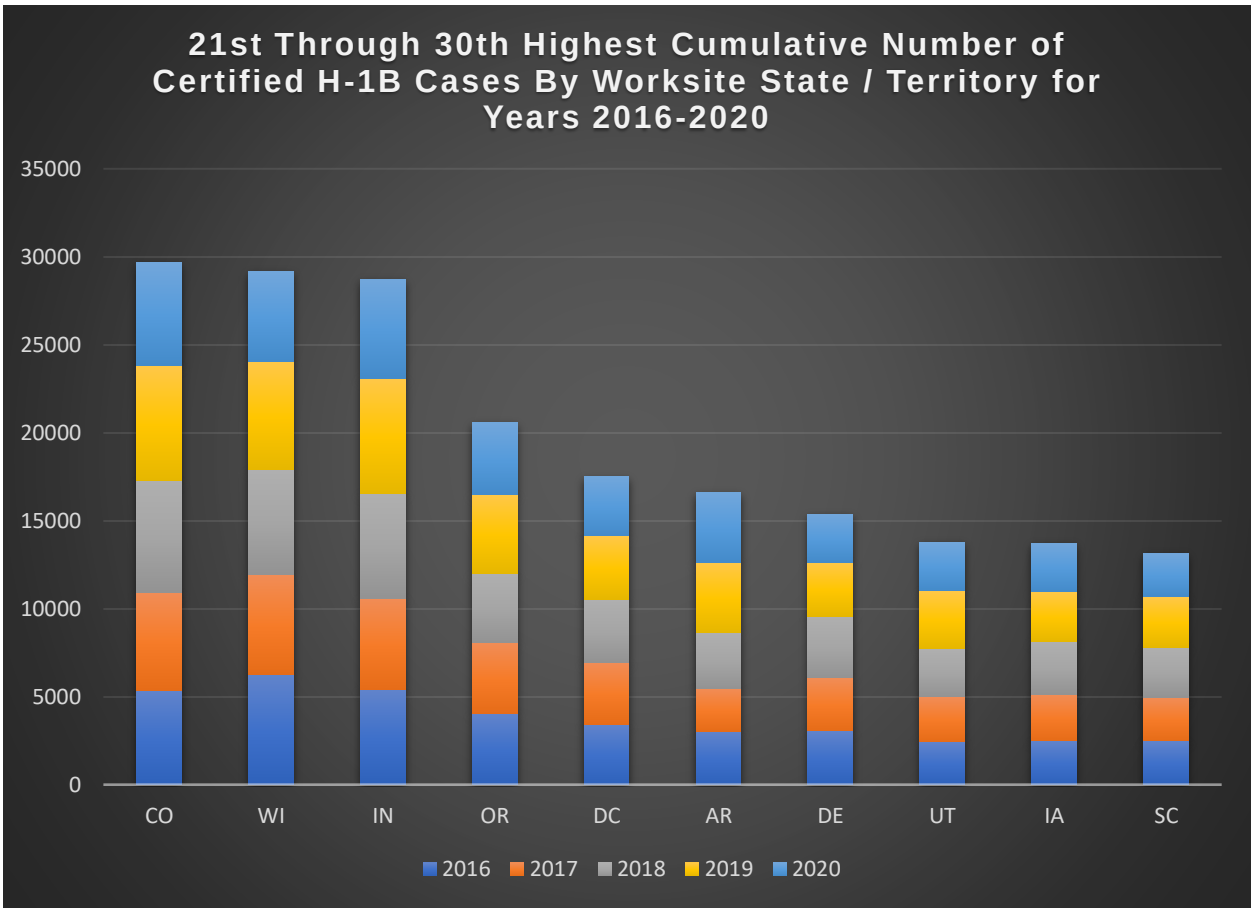


FIGURE 10. 21ST THROUGH 30TH HIGHEST CUMULATIVE NUMBER OF CERTIFIED H-1B CASES BY WORKSITE STATE / TERRITORY FOR YEARS 2016-2020

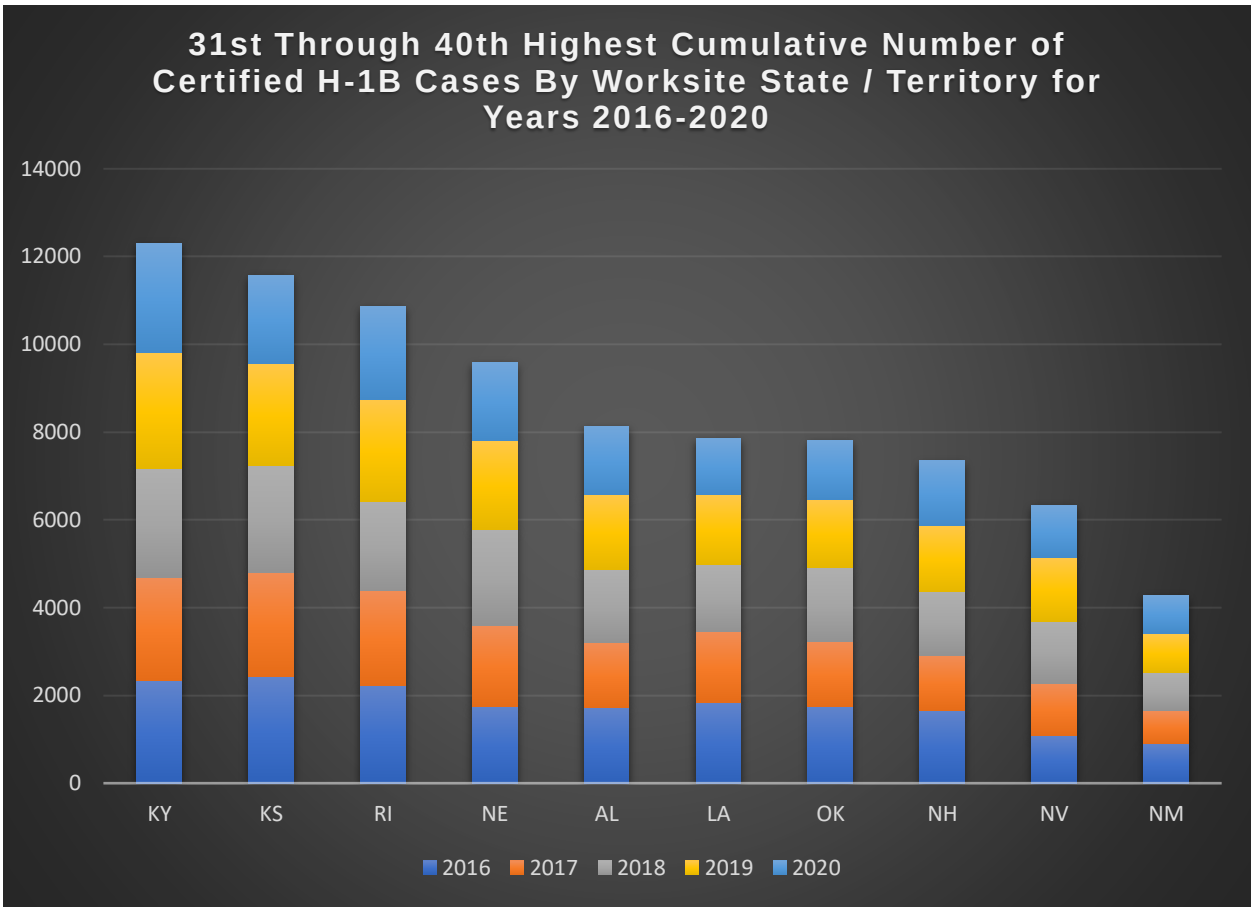


FIGURE 11. 31ST THROUGH 40TH HIGHEST CUMULATIVE NUMBER OF CERTIFIED H-1B CASES BY WORKSITE STATE / TERRITORY FOR YEARS 2016-2020

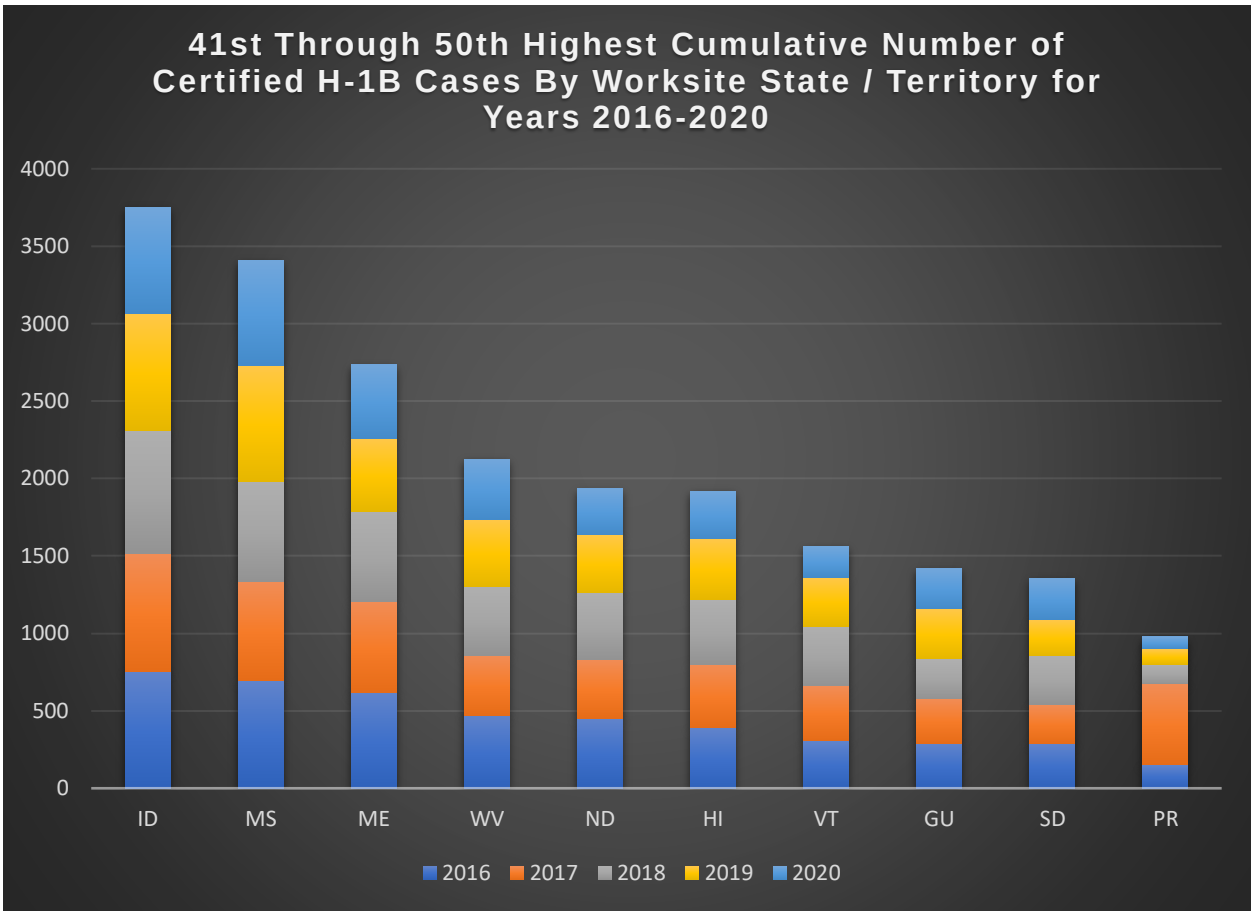


FIGURE 12. 41ST THROUGH 50TH HIGHEST CUMULATIVE NUMBER OF CERTIFIED H-1B CASES BY WORKSITE STATE / TERRITORY FOR YEARS 2016-2020

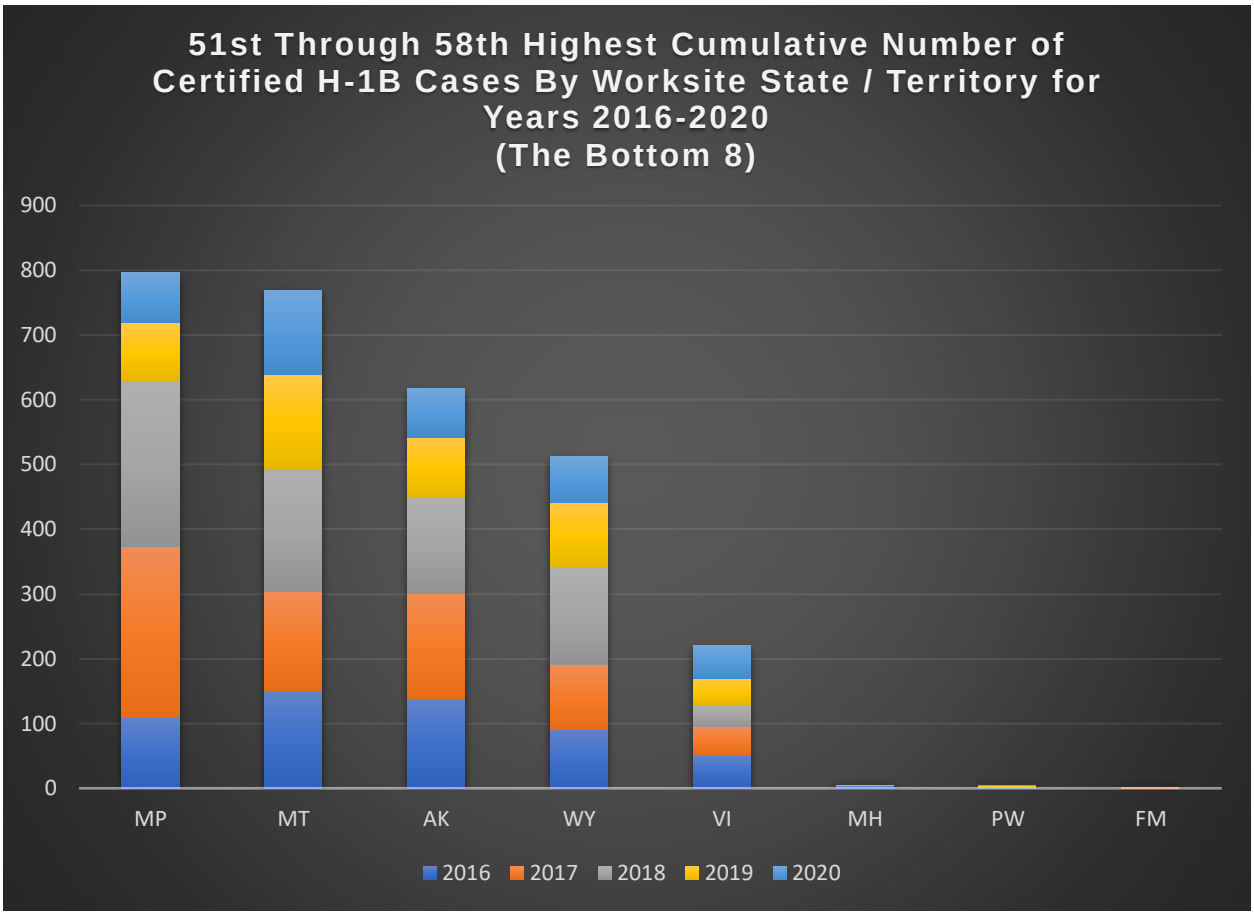


FIGURE 13. 51ST THROUGH 58TH HIGHEST CUMULATIVE NUMBER OF CERTIFIED H-1B CASES BY WORKSITE STATE / TERRITORY FOR YEARS 2016-2020

Figures 8 through 13 above illustrate that (with few exceptions) the decision to exclude each year for the states / territories beginning with 2020 has no effect on the cumulative magnitude of H-1B workers relative to the states / territories represented in each of these figures.

Conclusion

The trend in guest-worker employment in the United States continues to increase over time.

The average duration of time the DOL needed between 2016 and 2020 to certify (approve) a visa was between approximately six (6) to seven (7) days. Denying a visa required only three (3) to four (4) days on average.

The DOL is consistent. The low-end and hi-end delta between the number of days to approve vs. deny a visa is three (3) days. It is reasonable to conclude that it takes longer to approve rather than deny a visa. It is assumed the length of time to approve a visa is due to more rigorous review of the visa application. It is also assumed that if basic information on a visa application is incorrect and that error or omission is easily verifiable visa application denial would be prompt.

The metrics and trends continue to support the claim that United States workers continue to be squeezed out of their professions and Information Technology professionals feel the brunt.

Nothing has changed. The United States Federal Government in coordination with private industry continues to value foreign guest workers above US workers. The only argument both government and private industry can make has been proven false (see Section “The Famous Industry Lie”, p5).

Finally, the data clearly shows that while the Trump administration’s rhetoric on immigration policy was overt the Trump administration’s immigration policies had no effect on reducing H-1B guest workers. In fact, there was an increase in H-1B guest workers during the Trump administration. An administration based on Nationalism and an America First agenda failed to protect US workers.

SAITJ had no illusions of the real reason for the Trump administration to promote Nationalism and an America First agenda. The data shows that American voters who became caught up in the Trump rhetoric got screwed without being kissed.

Suggestions For Further Analysis And Data Refinement

1. An effort should be made to ensure the SOC_NAME values across each years dataset does not include redundant values or typos. The following is just one example of SOC_NAME values that are redundant and include typos.

    COMPUTER OCCUPATION, ALL OTHER
    COMPUTER OCCUPATION, ALL OTHERS
    COMPUTER OCCUPATIONS - ALL OTHER
    COMPUTER OCCUPATIONS - ALL OTHERS
    COMPUTER OCCUPATIONS, ALL OTHER
    COMPUTER OCCUPATIONS, ALL OTHER
    COMPUTER OCCUPATIONS, ALL OTHER*
    COMPUTER OCCUPATIONS, ALL OTHERS
    COMPUTER OCCUPATIONS, OTHER OTHER
    COMPUTER OCCUPATIONS,ALL OTHER
    COMPUTER OCCUPATIONS,ALL OTHERS
    COMPUTER OCCUPATIONS: ALL OTHER
    COMPUTER OCCUPATIONS/ALL OTHER
    COMPUTER OCCUPTATIONS, ALL OTHER
    COMPUTER OCUPATIONS, ALL OTHER

Removing redundant values and typos will allow for analysis that can answer questions like “How many employment opportunities were lost in a given profession to some guest-worker?”.

It is important to note that other fields include similar typos.

2. Obtaining the complete list of US Congressional Districts by zip code for each year would allow a granular level of analysis that was not possible in this report.

3. While Microsoft Excel proved reasonable for this report it was evident that more robust statistical analysis tools (e.g., SAS, SPSS or MATLAB) should be used for future analytical efforts.

    a. While software tools such as the R programming language could be used it is our recommendation that for the sake of ensuring the credibility of the results and limiting the risks of introducing bugs in custom built statistical software it is recommended to use industry standard analytic tools.

4. Once the complete 2021 data is available from the DOL it would be interesting to see a trend analysis that incorporates 2016-2021 data. Specifically, an annual trend report that integrates the data for the previous year.

Appendix A – Datasets Used

Data used in this report was obtained from the United States Department of Labor's public website. The website is accessible at the following URL at the time this report was published.

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

In the event the Department of Labor moves or removes the data from the URL above the data files have also been archived at the following URL:

https://www.saitj.org/supplemental-material/United-States-Department-of-Labor/

The files are defined below.

DATA FILE NAME	MD5 FILE NAME	DESCRIPTION
historical-case-disclosure-data-for-lca-h1b-h1b1-e3.tar.gz	historical-case-disclosure-data.md5	The original data obtained from the United States Department of Labor
SAITJ-Processed-Data.tar.gz	SAITJ-Processed-Data.md5	The Microsoft Excel files created by SAITJ based on the original data from the United States Department of Labor

TABLE 7. DATA FILES

Appendix B – Data Cleanup Process

The original data files require a fair amount of effort to clean the data. Removing, parameters with invalid characters is just one example. Efforts to clean the original data involved:

1) Opening each original data file in Excel and exporting the file to CSV format

2) Using custom software to read the csv formatted files, perform basic cleanup operations and write the cleaned data to a separate csv formatted file.

3) Cleaned csv formatted data files were opened in Excel for further cleanup operations.
    a. Excel's search and replace capability proved satisfactory

4) The 2020 data files were combined into a single 2020 data file.

The source code for the custom software mentioned in item 2 above is discussed in Appendix C – Custom Software.

The data cleanup process took more effort than was initially expected and the cleanup process described did not fully refine all the parameters in the data. Additional efforts should be taken to further refine the data.

Appendix C – Custom Software

The custom software used as part of the data cleanup effort was written in Java and requires the Apache Commons CSV API. The Apache Commons CSV API is accessible at the following URL:

https://commons.apache.org/proper/commons-csv/

The custom software source code is provided below.

```
package org.saitj.data.cleanup;

import java.io.File;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;
import java.io.Reader;
import java.util.Iterator;
import java.util.Vector;

import org.apache.commons.csv.CSVFormat;
import org.apache.commons.csv.CSVParser;
import org.apache.commons.csv.CSVPrinter;
import org.apache.commons.csv.CSVRecord;



/**
 * A dirty simple program that reads the Department of Labor
 * data files and does some initial cleanup of the data.
 *
 * @author SAITJ.org
 *
 */
public class CSVFileCleaner2 {

	private static CSVParser parser = null;
	private static Vector<String> invalidCellValues = null;
	private static CSVPrinter printer = null;

    public static void main(String[] args) throws IOException {

	//---------------------
	// Input Files
	//---------------------
//	String filePath = "/Documents/SAITJ-Processed-Data/2020/LCA_Disclosure_Data_FY2020_Q1.csv";   // 2020
//	String filePath = "/Documents/SAITJ-Processed-Data/2020/LCA_Disclosure_Data_FY2020_Q2.csv";   // 2020
//	String filePath = "/Documents/SAITJ-Processed-Data/2020/LCA_Disclosure_Data_FY2020_Q3.csv";   // 2020
//	String filePath = "/Documents/SAITJ-Processed-Data/2020/LCA_Disclosure_Data_FY2020_Q4.csv";   // 2020
//      String filePath = "/Documents/SAITJ-Processed-Data/2019/H-1B_Disclosure_Data_FY2019.csv";     // 2019
      String filePath = "/Documents/SAITJ-Processed-Data/2018/H-1B_Disclosure_Data_FY2018_EOY.csv"; // 2018
//	String filePath = "/Documents/SAITJ-Processed-Data/2017/H-1B_Disclosure_Data_FY17.csv";        // 2017
//	String filePath = "/Documents/SAITJ-Processed-Data/2016/H-1B_Disclosure_Data_FY16.csv";       // 2016
```

```
            //---------------------
            // Output Files
            //---------------------
//          String outPath = "/Documents/SAITJ-Processed-
Data/2020/LCA_Disclosure_Data_FY2020_Q1.updated.csv";   // 2020
//          String outPath = "/Documents/SAITJ-Processed-
Data/2020/LCA_Disclosure_Data_FY2020_Q2.updated.csv";   // 2020
//          String outPath = "/Documents/SAITJ-Processed-
Data/2020/LCA_Disclosure_Data_FY2020_Q3.updated.csv";   // 2020
//          String outPath = "/Documents/SAITJ-Processed-
Data/2020/LCA_Disclosure_Data_FY2020_Q4.updated.csv";   // 2020
//          String outPath = "/Documents/SAITJ-Processed-Data/2019/H-1B_Disclosure_Data_FY2019.updated.csv";
// 2019
            String outPath = "/Documents/SAITJ-Processed-Data/2018/H-
1B_Disclosure_Data_FY2018_EOY.updated.csv"; // 2018
//          String outPath = "/Documents/SAITJ-Processed-Data/2017/H-1B_Disclosure_Data_FY17.updated.csv";
// 2017
//          String outPath = "/Documents/SAITJ-Processed-Data/2016/H-1B_Disclosure_Data_FY16.updated.csv";
// 2016


        File csvData = new File(filePath);

        if (csvData.exists() && csvData.isFile() && csvData.canRead()) {

            invalidCellValues = CSVFileCleaner2.getInvalidCellValues();

            Reader in = new FileReader(csvData);

            CSVParser parser = CSVParser.parse(in,
CSVFormat.EXCEL.withNullString("").withIgnoreSurroundingSpaces(true).withDelimiter(',').withTrim(true));
            printer = new CSVPrinter(new FileWriter(outPath), CSVFormat.EXCEL);

            for (CSVRecord csvRecord : parser) {


                    if (CSVFileCleaner2.isRecordValid(csvRecord)) {

                        //-----------------------------------------------------------
                        // Now write the CSV parameters to disk
                        //-----------------------------------------------------------
                        Iterator<String> it = csvRecord.iterator();

                        while (it.hasNext()) {
                            String value = it.next();
                            if (value == null) {
                                printer.print("NULL");
                            } else if (value.compareToIgnoreCase("null") == 0) {
                                printer.print("NULL");
                            } else {
                                boolean invalidCharacter = false;
                                for (int i = 0; i <= CSVFileCleaner2.invalidCellValues.size() -1; i++) {
                                    if
(value.compareToIgnoreCase(CSVFileCleaner2.invalidCellValues.get(i)) == 0) {
                                        invalidCharacter = true;
                                    }
                                }

                                if (invalidCharacter) {
```

```
                                        printer.print("NULL");
                                        System.out.println("Found Invalid Character");
                                } else {
                                        printer.print(value.toUpperCase().trim().strip());
                                }
                        }
                }
                printer.println();


        }
    }

    printer.close(true);

  }
}

  public static boolean isRecordValid(CSVRecord record) {

        if (record.isConsistent() != true) {
                System.out.println(record.getRecordNumber() + " Does Not Match The Number of
Columns");
                System.exit(0);
        }

        return true;

  }

  public static Vector<String> getInvalidCellValues() {

        Vector<String> result = new Vector<String>();

        result.add(",");
        result.add(".");
        result.add(";");
        result.add("-");
        result.add("?");
        result.add(":");
        result.add("!");
        result.add("@");
        result.add("#");
        result.add("$");
        result.add("%");
        result.add("^");
        result.add("&");
        result.add("*");
        result.add("(");
        result.add(")");
        result.add("_");
        result.add("=");
        result.add("+");
        result.add("{");
        result.add("[");
        result.add("}");
        result.add("]");
        result.add("\\");
        result.add("/");
        result.add(",,");
        result.add("..");
```

```
            result.add(";;");
            result.add("--");
            result.add("??");
            result.add("::");
            result.add("!!");
            result.add("@@");
            result.add("##");
            result.add("$$");
            result.add("%%");
            result.add("^^");
            result.add("&&");
            result.add("**");
            result.add("((");
            result.add("))");
            result.add("__");
            result.add("==");
            result.add("++");
            result.add("{{");
            result.add("[[");
            result.add("}}");
            result.add("]]");
            result.add(",,,");
            result.add("...");
            result.add(";;;");
            result.add("---");
            result.add("???");
            result.add(":::");
            result.add("!!!");
            result.add("@@@");
            result.add("###");
            result.add("$$$");
            result.add("%%%");
            result.add("^^^");
            result.add("&&&");
            result.add("***");
            result.add("(((");
            result.add(")))");
            result.add("___");
            result.add("===");
            result.add("+++");
            result.add("{{{");
            result.add("[[[");
            result.add("}}}");
            result.add("]]]");

            return result;


    }

}
```

Appendix D – Database Table Column Definitions

The Department of Labor provides definitions for each column in the database by year. Additional columns were added between years 2016-2020. Notably, only the 2020 database table includes email addresses.

The complete list of database column definitions by year are provided below.

YEAR	FILE DOWNLOAD LINK
2016	Download
2017	Download
2018	Download
2019	Download
2020	Download

TABLE 8. DATABASE TABLE COLUMN DEFINITION FILES