Trends in T-Visa Application

Kehan Wang
Xiaoxin Gan
Yansong Bai

1. Introduction & Problem Statement

This project will look at T-Visas, a visa program for victims of human trafficking. There has been an increase in denials in T-Visa applications (there are 5,000 available per year). Never have more than 1,000 been granted in a year, and the past few years have seen both an increase in applications and an increase in denials. There is a hypothesis that the increase in denials is related to procedural barriers that immigration is constructing to make it more difficult for applications to find success. The data will look at administrative appeal decisions (AAO) to examine these trends in terms of appeals and the result of these appeals. The goal is to understand appeal T-visa case trends within the decisions by analyzing the presence of specific keywords within the texts, and to conduct topic modeling. In this project, we analyzed the following questions. How many cases were decided (separated by year); how many cases were dismissed (separated by year to demonstrate trends); how many cases were granted (indicated by language including grant, sustain, or withdraw and remanded back to the agency); what grounds were cases denied (separated by year to demonstrate trend); how has the trend changed over time; and what are the patterns indicating by the T-Visa Application. For data analysis, there are many duplicates between two datasets, and it is extremely hard to distinguish them accurately. And we ask our client for suggestions about that. She suggests we only focus on the LEXIS dataset. So our report mainly focuses on the analysis of the LEXIS dataset.

2. Methodology & Algorithm

2.1 Data Sources

The data sources include AAO Decisions from the website and LEXIS files from our client. All AAO Decisions are found on the website here. In order to search within the AAO decisions for Applications for T nonimmigrant status. For AAO data source, we have two options to search for what we need, which are “application for nonimmigrant status” and “form I-914,”; and we choose to use the “Form I-914” query because it is the more accurate one to use. Within this search query, we see a range of decision types, such as H1-B, L-1. Figure 1 is an example of the type.

Figure 1: Example of **Application for T Nonimmigrant Decision**.

2.2 Data Collection

For AAO websites, we first need to download decisions with the heading Application for T Nonimmigrant Decision. To do so, we collect all data from the above-mentioned data sources.

Figure 2: Search Page pattern

We found the URL has a specific pattern on the search result page, as Figure 2 shows. Hence, the task should be relatively easy —— we need to loop through all the pages by varying the page number, crawl all the links of applications, and download them into a single directory by sending HTTP requests (request package). During the data processing, we will need to exclude derivative (family member) applications and other visa types from this dataset. For the LEXIS data source, thanks to our client, we have had all PDFs downloaded. We just need to save them into the same directory so that we can do the data extraction.

2.3 Data Extraction & Data Cleaning

As we mentioned in the introduction, **we mainly used LEXIS Data sources.** In order to extract information from these downloaded files, we first convert them into texts by applying packages like pdfminer, PyPDF2. After a series of attempts, we use regular expression operations (re package) to specify a set of strings that match a specific pattern, then extract them and save them into a CSV file. As a result, we come up with data sets that consist of 12 features and 502 samples. You can observe what each column represents in Appendix A. Furthermore, we input the dataset using the python package Pandas, and implement some algorithms to unite the format of each column.

2.4 Data Analysis

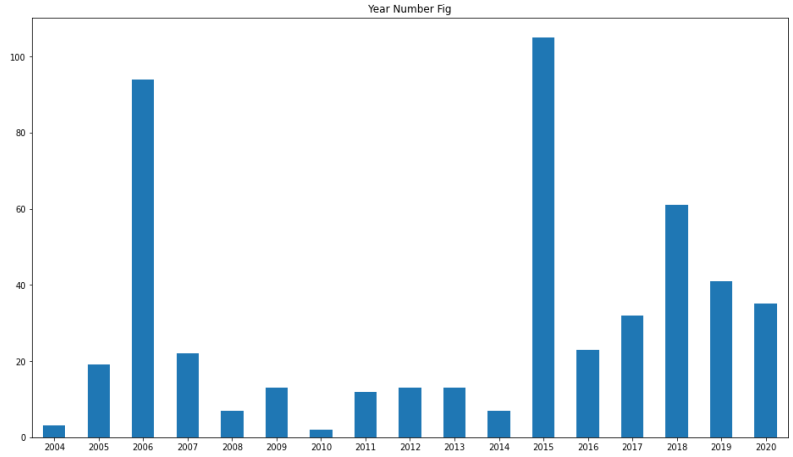
After getting all the data stored in CSV files, we want to find out what is the trend and what causes the difference. **Again, we only did the analysis of the appeal cases on LEXIS data source**. We used matplotlib.pyplot to visualize the data by years, months, days, and weekdays where the trend can be easily observed. In order to find out what topics are mentioned most frequently in applications, we did some analysis based on word count by using the CountVectorizer method. We also tried to use Latent Dirichlet Allocation to uncover hidden structure in the collection of texts as the package provides many good APIs, including dimensionality reduction, where results and patterns can be clearly evaluated. However, the method still didn't offer us good results because there are too many professional words.We will mention more details in Section 3 and Section 4. In addition, we also use regular expression operations to evaluate the appearance of some important keywords given by our client. You can observe how we design the extraction plan for strategic questions in Appendix B.

3. Results & Observations

3.1 How many cases were decided?

There are a total of 502 decided cases from LEXIS data sources [Appendix A]. Among them, there are 468 appeal cases and 34 motion cases. By excluding 34 motions from data analysis, we can observe a total of 421 dismissed cases, 37 granted cases, and 10 denied cases. Also, there are a total of 267 "not-represented" cases, 222 "represented" (with a representative) cases and 13 "self-represented" cases. Our dataset contains applications from 2004 to 2020. The shapes for decided cases from 2004 to 2010 and 2011 to 2020 are bell-shaped. They reached a peak in 2006 and 2015. The average decided cases from 2007 to 2014 are low. But the decided cases after 2015 are higher than before.

Figure 3. Trends for all decided cases, by years



3.2 How many cases were dismissed?

After evaluation, we found there are a total of 421 dismissed cases. In the dismissed cases, there are 207 not-presented cases, 203 presented cases, and 11 self-presented cases. From the graph above, we can see that the year 2006 and 2015 have the highest number of dismissed cases. The shape for dismissed cases is also two bell shapes from the year 2004 to 2012 and 2013 to 2020. The cases are more likely to be dismissed in the middle third of a month. And January to July is the case with a high probability of getting dismissed. The average dismissed case number each weekday is neat to 80. No obvious fluctuation is shown each weekday. Comparing represented counsel and non-represented counsel, they have almost equal amounts of cases. "Self-represented" cases are relatively small compared to those two types of cases.

Figure 4. Trends for all dismissed cases

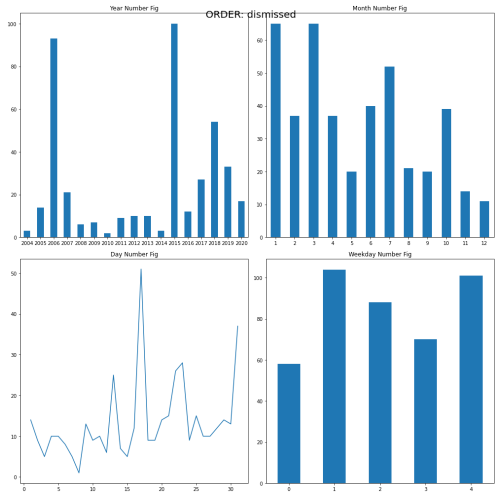


Figure 5. Counsel for all dismissed cases

3.3. How many cases were granted?

There are a total of 37 granted cases. In granted cases, there are 25 "not-presented" cases and 12 "present" cases. From the graph above, we can see the year 2020 has the highest number of granted cases. The shape of granted cases is a cosine function shape from the year 2005 to 2020. The cases are more likely to be granted in June.
June has the highest number of granted cases. Later in each month from 25th to 30th, the cases are more likely to be granted.

Figure 6. Trends for all granted cases, time

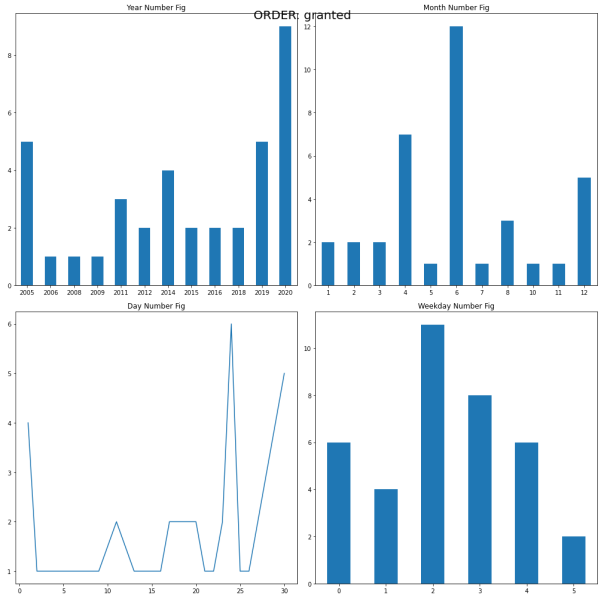
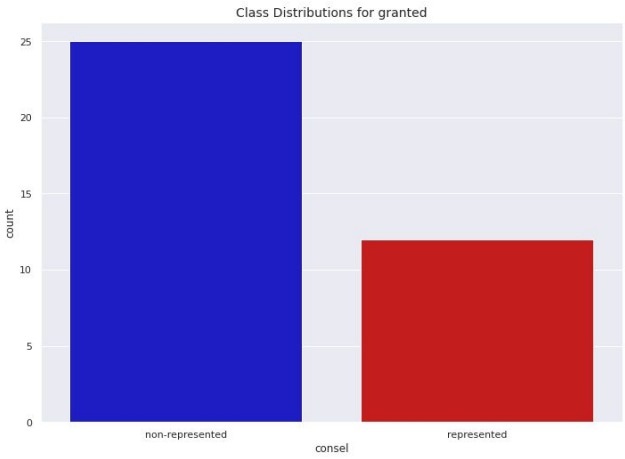


Figure 7. Counsel for all granted cases



3.4. What grounds were cases denied?

The visualizations are shown in the following two figures. In denied cases, there are 8 “not-represented” cases, 1 “represent” case, and 1 “self-represented” case. The cases are most likely to be denied if the applicant is not represented. In the Year Number figure, we see that 2009 has the highest number of denied cases and 2007, 2017 and 2018 all have the same number of denied cases. In the Month Number figure, we can see June is more likely to be denied compared to other months. In the Day Number figure, we see that the trend is shown to be declining from the start. In the Week Day figure, we can see that Monday is the most likely to be denied weekday.

Figure 8. Trends for all denied cases

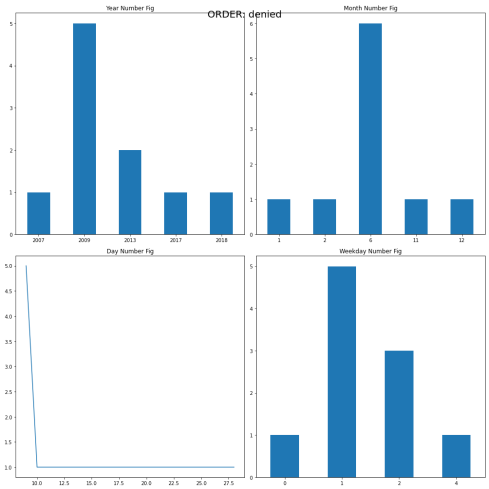


Figure 9. Counsel for all denied cases

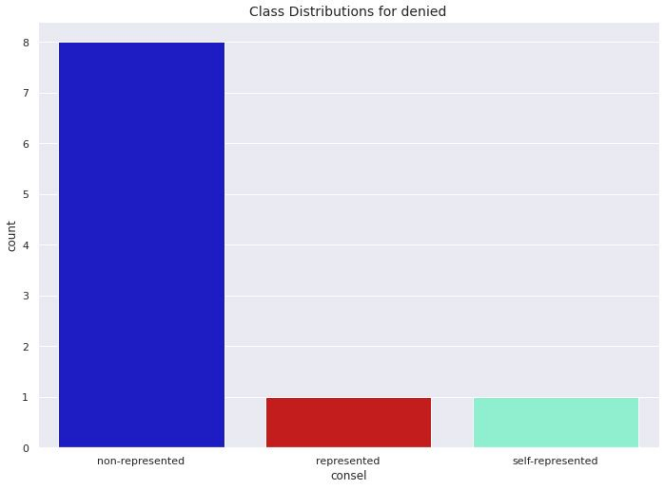
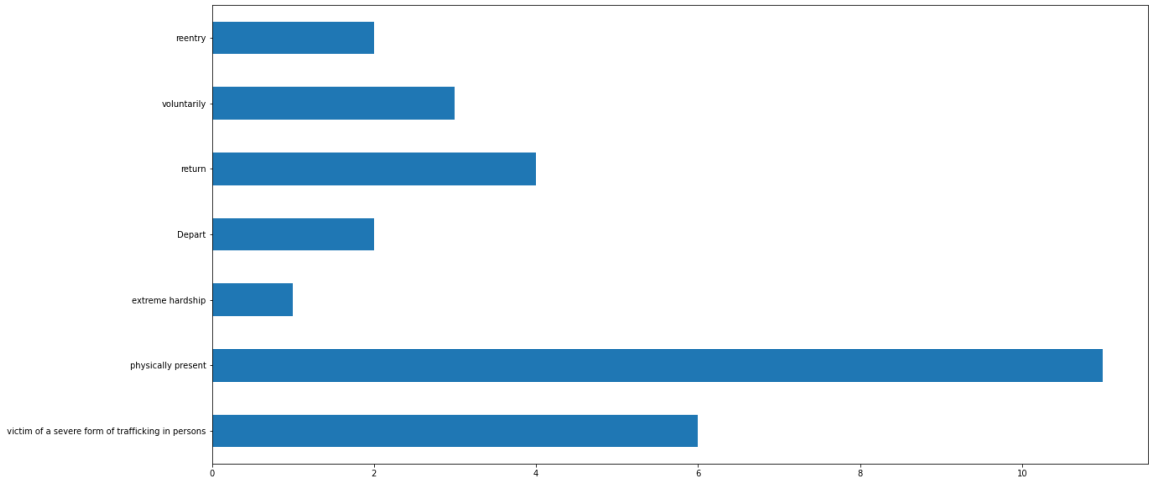


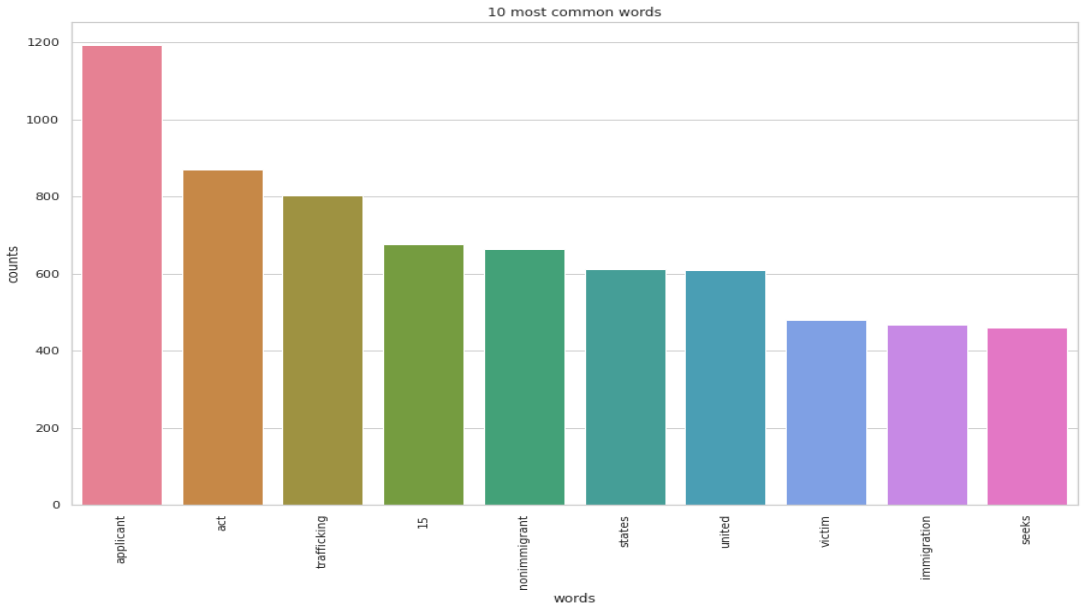
Figure 10. Frequency of keywords in denied cases



However, we cannot get an accurate trend from the visualization since the sample size is too small. Therefore, it is more important to analyze how these cases got denied by evaluating the appearance of keywords (mentioned in Appendix B) given by our client. We successfully applied regular expressions operations to evaluate it, and the result is shown in the following figure. Some phrases from Appendix B are omitted since we didn't find them in the denial cases. Figure 10 is the frequency of keywords given by our client. We can see that the keywords "physically present" and "victim of a severe form of trafficking in persons" have the larger counts. The keywords including "reentry", "voluntarily", "return" and " Depart" have similar amounts. The keyword "extreme hardship" has the least count.

In addition to the keywords given by our client, we also use our own strategy to analyze the keywords in denied cases. We also find the 10 most common words shown in figure 11 by applying CountVectorizer to the denied cases. We have studied keywords like "trafficking", "nonimmigrant", "victim"... in Figure 10. The keyword "15" is most likely to represent the *Immigration and Nationality Act (the Act) § 101(a)(15)(T)(i), 8 U.S.C. § 1101(a)(15)(T)(i).*

Figure 11. 10 most common words



3.5. Any patterns?

According to visualizations, we can assume that June is a good month to apply for a T-Visa. The applications submitted on the 25th of each month are most likely to be granted, and Tuesday of each week is also a good day to start an application. Besides, during preprocessing, we found derivative applications (applications for family members) are most likely to be rejected/denied (from AAO data source, which we didn't provide analysis here). Hence, derivative applications are not recommended. In terms of representations, an applicant who doesn't have a representative nor self-represented is most likely to be rejected in the application. Also, based on the keywords given by the client (Appendix B), the phrases "physically present", "victim of a severe form of trafficking in persons" has the most appearance in denial cases. Conjugation of verbs like "reentry" and the "return" account for a large part. The act like "*Immigration and Nationality Act (the Act) § 101(a)(15)(T)(i), 8 U.S.C. § 1101(a)(15)(T)(i)"* are also most likely to be mentioned in the T Visa application. Studying the details of this act is strongly recommended.

4. Conclusion

By cooperating with BU Law school, we generated a T-Visa application analysis through web crawl, data analysis, and topic modeling. The whole project is relatively practical and challenging as the lack of data sources we have, and we have to spend a huge amount of time on data collection/processing. Additionally, the large number of professional phrases, such as "*sections l01(a)(15)(T)and 214(o) of the Immigration and Nationality Act (the Act), 8 U.S.C. §§ 1101(a)(15)(T) and 1184(o)*", make the topic modeling extremely difficult. Figure 12 shows a topic modeling on denied cases conducted by CounVectorizer and LDA. With a series of hyperparameter tuning, there are definitely important phrases that start to appear, but the whole sentence is still hard to read. Even though we can still do pattern matching with the keywords given by the client in this stage, in the future, we plan to collect more files from LEXIS data source and explore various natural language processing methods to come up with a more comprehensive and accurate analysis.

Figure 12. Topic Modeling on denied cases

```
Topic #0:
applicant states united act falcon pursuant agreed employ status

Topic #1:
15 classification trafficking act nonimmigrant victim 1101 immigration nationality

Topic #2:
trafficking applicant severe persons form victim act 15 director

Topic #3:
applicant nonimmigrant status act states united classification family redacted

Topic #4:
applicant application nonimmigrant trafficking form act appeal director 15
```

A separate folder storing different types (granted, dismissed, and denied) of pdfs can be found here: LEXIS. All works, including data collection/cleaning, visualization, topic modeling, can be found here: Github. You will need these files downloaded in the same folder with all codes in order to reproduce our result.

5. Appendix

Appendix A:

a. What does each column represent?

We made two datasets based on the given data sources, namely AAO Non-Precedent Decisions and LEXIS. We contacted them during the analysis stage.

'Url':	**LEXIS**
'File_name'	This column is for the file name
'title'	This column is for the title of each file
'ID'	This column is for the case number, a unique identifier to distinguish each case
'date',	This column is for the date of the application
'type'	This column is used to distinguish appeal cases or motion cases.
'order'	The column is used to describe the result of the application, e.g dismissed or granted.
'description'	This Column declares the description of the application.
'core_terms'	This column declares the description of the application.
'citation'	This column is for the head of the application record in the LEXIS system. The four/five-digit number can be used to distinguish each application.
'S-M-P'	This column describes whether the application contains "Matter of S-M-P"
'aao citation'	This column describes the aao citation of each case.
'consel'	This column is to illustrate the representation of the applicant. Presented: The applicant is presented with a representative. Not-Presented: The applicant is not presented. Self-presented: The applicant is presented by self.
'description'	This column contains a description of the application.

Appendix B:

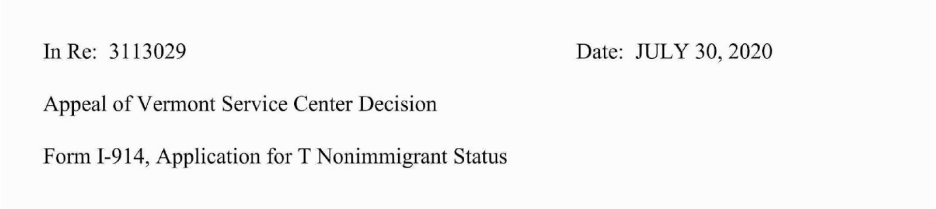
a. How many cases were decided? (separated by year)

This can be determined by downloading the decisions for "Application for T nonimmigrant status" and determining how many decisions there are by year (divide by precedent and non-precedent). We also exclude derivative applications.

These cases still fall into two categories -- appeals and motions to reopen and reconsider. We analyze the **appeals** category.

Appeals will say the following in the top left section of the decision: **Appeal of Vermont Service Center Decision**. Figure 3 illustrates an example of it.

Figure 9: Example of Appeal cases



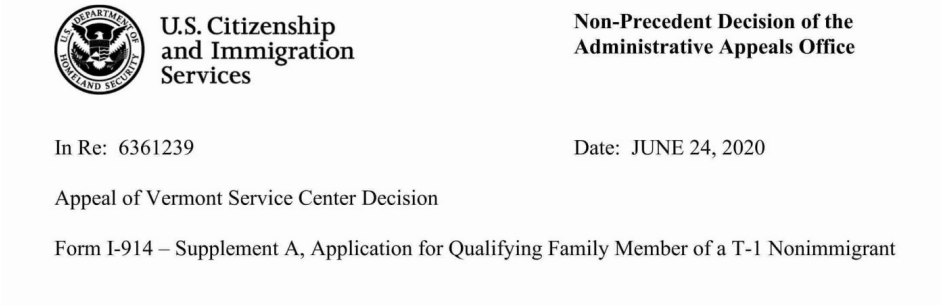
In Re: 3113029 Date: JULY 30, 2020

Appeal of Vermont Service Center Decision

Form I-914, Application for T Nonimmigrant Status

We **exclude derivative applications** (applications for family members). We search within the set for cases with the following heading in the PDF document and exclude them from our general analysis. See below for this heading. The search would be for "Supplement A, Application for Qualifying Family Member of a T-1 Nonimmigrant" In the heading area below.

Figure 10: derivative applications



U.S. Citizenship and Immigration Services

Non-Precedent Decision of the Administrative Appeals Office

In Re: 6361239 Date: JUNE 24, 2020

Appeal of Vermont Service Center Decision

Form I-914 – Supplement A, Application for Qualifying Family Member of a T-1 Nonimmigrant

b. How many cases were dismissed? (separated by year to demonstrate trends)

The information available at end of the decision, where it says ORDER. See below for an example. It will either say that the **motion or appeal was dismissed**.

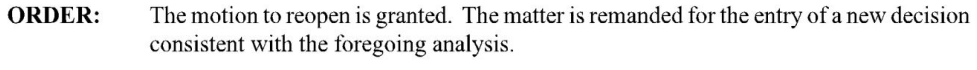
Figure11: dismissed cases

ORDER: The appeal is dismissed.

How many cases were granted (this can be indicated by language including **grant, sustain, or withdraw** and **remanded** back to the agency)?

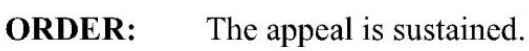
We analyze by year to demonstrate trends. The motion or appeals that were granted should state at the end (final section) under OUTCOME that they were "granted" or "sustained" or "withdrawn" and/or "remanded." Below are a few examples. The "remand" captures all of the terms because this means it is sent back to the agency.

Figure 12: granted cases



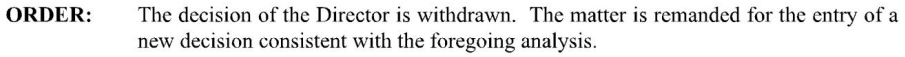
ORDER: The motion to reopen is granted. The matter is remanded for the entry of a new decision consistent with the foregoing analysis.

Figure 13: sustained cases



ORDER: The appeal is sustained.

Figure 14: withdrawn cases



ORDER: The decision of the Director is withdrawn. The matter is remanded for the entry of a new decision consistent with the foregoing analysis.

c. What grounds were cases denied? (separated by year to demonstrate trends)

In the introductory paragraph of the decision, AAO generally states the reasons for the denial. We then search in this first paragraph section for the prevalence of the following terms, and this is an indication of the grounds for denial:

i. "victim of a severe form of trafficking in persons"
ii. "physically present"
iii. "complied with reasonable requests for assistance"
iv. "extreme hardship"
v. "inadmissible" or "inadmissibility"

We conduct this search in **denied cases** as well as **granted cases** and then compare the results in each batch. Under the **ANALYSIS** section of the decision, we also search the terms below to determine how USCIS is deciding particular cases.

vi. "victim of a severe form of trafficking in persons"
vii. "physically present"
viii. "complied with reasonable requests for assistance"
ix. "extreme hardship"
x. "Inadmissible" or "Inadmissibility"

Other terms of interest in this section:

- "Sufficient evidence"
- "Any credible evidence"
- "Depart" and "return" or "voluntarily" or "reentry" (including conjugations of these verbs)
- "8 C.F.R. § 214.1 l(g)(l)(i)- (v)" (which is the "on account" of requirement
- "8 C.F.R. § 214.ll(g)(4)"
- "USCIS may consider when applicants escaped their traffickers, what activities they have since undertaken to deal with the consequences of having been trafficked, and their ability to leave the United States"
- "8 C.F.R. § 214.11 (g)(2)" (physical presence)
- "Preponderance"
- "Domestic violence"
- "Smuggling" (and variations -- "smuggle" / "smuggled")
- "Escaped a severe form of trafficking in persons before an LEA was involved" or "[e]scaped a severe form of trafficking in persons before an LEA was involved"
- "labor exploitation"

Not all decisions have subsections within the ANALYSIS section, but some do. We search for the prevalence of the following subsections within ANALYSIS:

xi. "The Applicant's Trafficking Claim"
xii. "The Applicant is Not Physically Present On Account of Trafficking"
xiii. "Extreme Hardship"
xiv. "Reasonable Response"
xv. "Inadmissibility"