Debt Collection

Team A

Client: WGBH

CS 506

Spring 2024

Andrew Woska, Team Rep, 2024, agwoska@bu.edu

Patrick Meade, 2025, pmeade@bu.edu

Sumatra Dhimoyee, 2027, sumatrad@bu.edu

Quan Ho, 2024, hohuq@bu.edu

Yiyang Cai, 2025, yycai@bu.edu

Introduction

The group was approached by a trusted news organization, WGBH, to investigate debt collection practices in Massachusetts by analyzing and disseminating information provided by historical court records. These records, provided through databases, gave an unfiltered lens into the Massachusetts court filing system. The group was tasked with exploring these databases over the past ten years to answer initial questions provided by the client, followed by more extensive questions later. To efficiently navigate the large volumes of data, we employed various SQL queries tailored to each specific question posed by the client.

Base Analysis

Insight 1: How many cases have been filed yearly over the last ten years in each court?

The initial process involved looking for the number of cases filed each year over the last ten years in each court using the wp_courtdocs from the MassCortsPlus.org database. By writing a query to filter out the unneeded data and restricting the court to just the District Court, the query returned 1,167,365 cases over the last ten years. Then, the cases were grouped by district, and a trendline was added for further analysis, as shown in the graph below.

The graph depicts the average or above-average number of cases per year while not showing the below-average districts (note that 2024 is incomplete). This data capture gives a scope of the number of cases to be looked into for potential debt.

Extension 1.1: How many debt cases were in the database in the past 10 years?

After talking to the clients about the initial percentage received for debt collection cases in small claims courts, it became clear that there needed to be a better understanding of who the debt collectors were and how to find them. The clients asked for a few key components: only gathering ten years' worth of data and finding a more accurate way to determine who is a debt collector. According to one of the clients, small claims courts dominate more than half of all cases. The previous percentages range from 20% to 40%, so an updated method and list will produce a more accurate solution. The solutions for small claims are found in the base analysis section.

The two databases were looked at from MassCourtsPlus.org: *wp_courtdocs* (WP) and *civica_courtdocs* (CA). The WP database uses a WordPress version of the database, which is what the clients asked us to look into. The CA database uses a different version of the same database that displays data substantially differently. The WP and CA databases both use duplicate case and post IDs, which may be used to connect the databases. CA was ultimately used to determine which court cases were taken from the database due to the structure of WP. WP is structured in many tables that need to be connected, while CA is confined to a single table. This method was only used to get the court cases rather than for the actual data analysis of the exceptional cases that appear in the other parts of the database.

A filter was created that analyzes parts of the CA database and compares the information found to information about known debt collector companies outside the SQL database. The assumptions used in this methodology are listed below:

- The companies found in the debt collector licensee filtered list are known debt collectors
- Debt collectors do not only have debt collection cases
- Small claims and civil court cases are the only courts in which debt collection cases occur (to solve the previous assumption)
- Company names are not consistent throughout the databases and tables as a result of clerical errors
- Only ten years' worth of data is requested

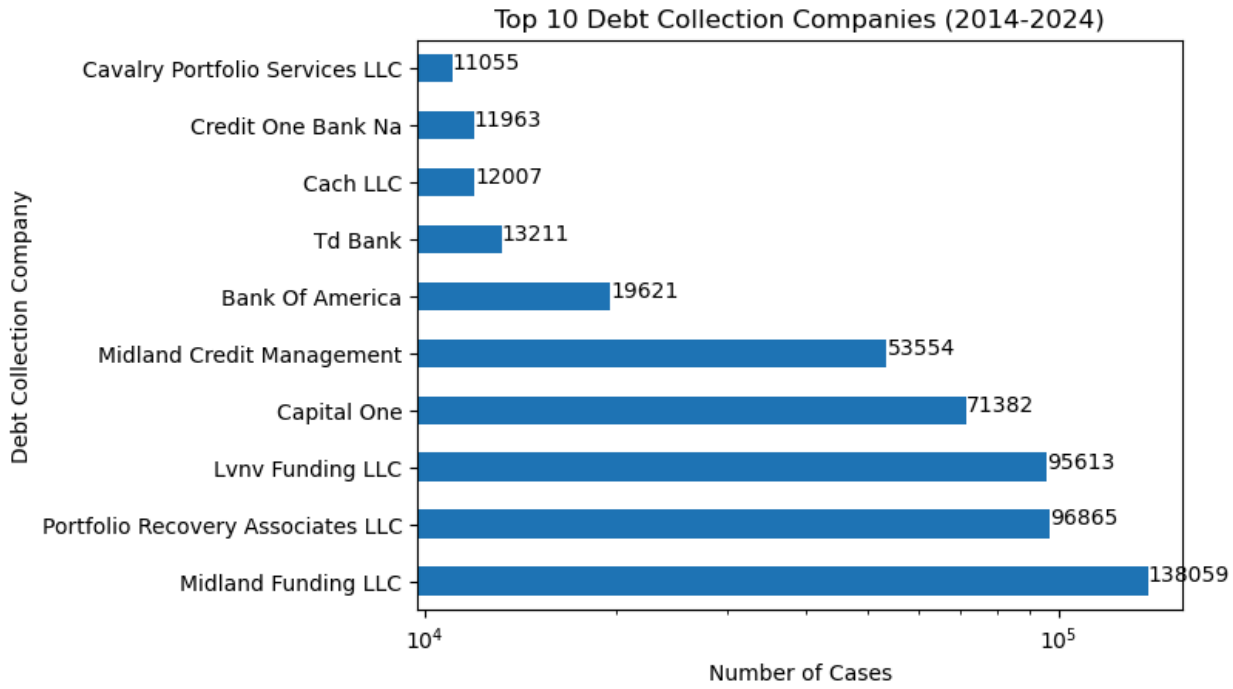
Following these assumptions, all debt collection cases were found using these steps:

1. Get the following columns from the cases_masscourts_org table in CA, ensuring that you receive only civil and small claims cases.
    - Case_type
    - Case_number
    - File_date
    - parties
2. Add additional columns by removing the plaintiff and defendant names from the party's column.
    - The plaintiff will be the company that launched the lawsuit
3. Filter out cases older than ten years old.
4. Get the company names from the filtered list of debt collector licensees.
    - Change all names to lowercase
5. Add an additional column to identify all cases with a known debt collector.

    - All punctuation and letter cases are removed from the names for consistency
6. Convert all debt collectors to the proper case as best as possible.
    - i.e., "llc" to LLC, inc to Inc
7. Remove all rows where the debt collector is not identified.

The following graph shows the top 10 debt collection companies with cases on a logarithmic scale. As can be seen, five companies make up most of the court cases.

The following graph shows the top 10 debt collection companies with cases on a logarithmic scale. As can be seen, five companies make up a majority of the court cases



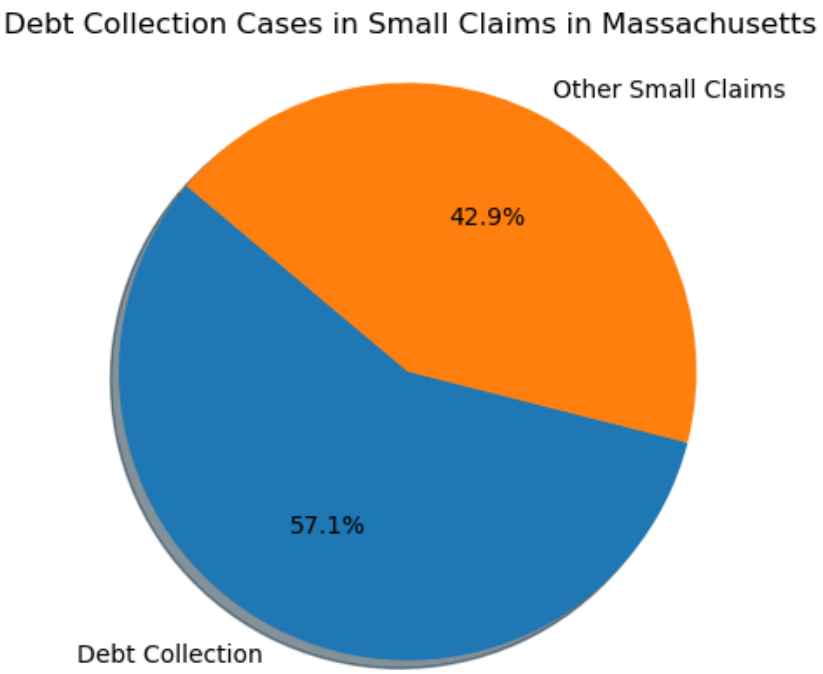
Some issues have come up while making this filter to ensure that the number of cases in debt collection is consistently underestimated. Filtering by name has many nuances that are not easily solved. Since four companies seemingly dominate the industry, the calculation presented in the report will be virtually unaffected because the change will be minuscule.

Additional consideration should be given to debt buyers in future project iterations.

Insight 2: What percentage of cases filed in small claims and district court are filed by debt collectors?

The civica_courtdocs and wp_courtdocs databases supplied by MassCourtsPlus.org were analyzed to determine the number of debt collection cases. The exact method of determining how to find debt collectors in the database is found in the extension about getting debt collectors from the database of this report.

The updated filtering methods determined that 57.1% of all small claims cases were filed by known debt collectors in the past ten years. At least 844,219 court cases are in the small claims courts, and as of May 1st, 2024, at least 482,461 cases were determined to have been created by debt collectors.



Insight 3: How many cases go to default?

A default judgment is an order issued by a judge or court to support the plaintiff in a case where the defendant does not show up or reply to a summons. We analyzed the last 11 years of data to understand the default cases. We search for default cases in the “cdocs_case_action_index” table from the wp_courtdocs database for our required timeline using the 'default' as the action. This allowed us to locate all of the default cases during the 11 years. From the “cdocs_case_meta_index” table, we got the post_id and the case_numbers of all cases during that time. The post_id in “cdocs_case_meta_index” maps to the case_id in “cdocs_case_action_index.” Using the known list of debt cases, we retrieved them from our

earlier analysis. We filtered out only the case_ids for the debt cases among all the default cases. The total number of debt cases that result in default is 79025.

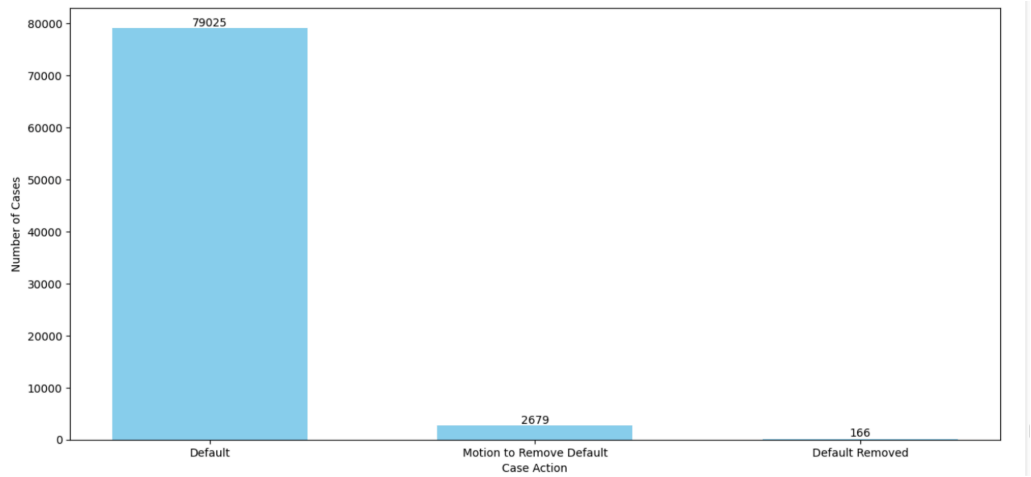
Extension 3.1: When there is a default, how often is there a judgment?

A query search in the MassDocs database, focusing on fields such as case_number, case_type, case_status, and attorney. This search also counted cases that resulted in a judgment, enabling us to analyze the frequency with which judgments follow defaults. Our query meticulously sifted through each recorded case, gathering the necessary data for our analysis.

Once the data is fetched, the analysis will proceed by calculating the ratio of cases with a judgment of the total cases in default related to debt. This involves assessing the unique cases where both 'default' and 'judgment' descriptions appear and comparing this to the total number of default cases retrieved.

With the modified data, the analysis would proceed by using the pandas library in Python to manipulate and analyze the data. Here, the focus would be on assessing the prevalence of judgments in default cases, and this would be done by computing the percentage of cases with judgments out of those with defaults. The extracted result showed 133,691 debt cases that went into default(all time). Of all those 133,691 cases, only 12 default cases went into judgment.

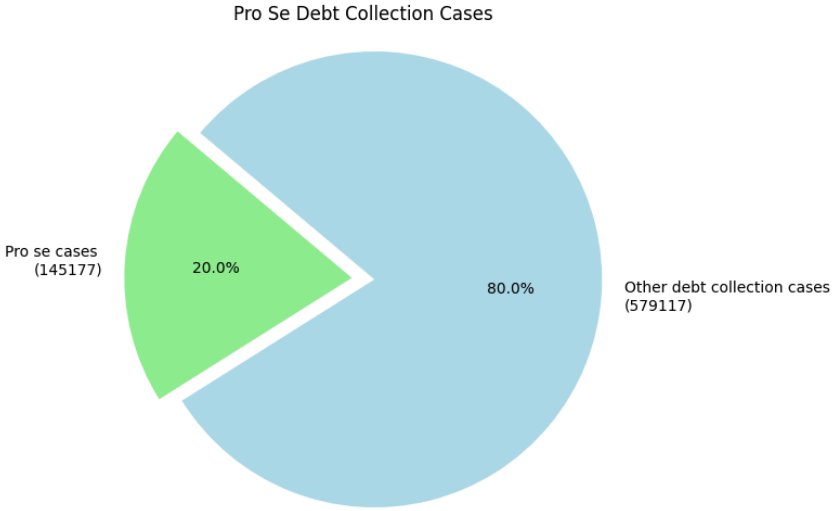
Extension 3.2: How many default cases are able to successfully remove default judgment?



Using a similar technique for locating default cases and actions such as 'Motion to remove default' and 'Default removed', we were able to find the number of cases that petitioned to remove default judgment and successfully removed the judgment. About 15% of debt cases result in default. As you can see from the figure, while the number of cases that result in default is significantly high, the number of cases that fight to get those default judgments is significantly low. Only 0.002% of the debt default cases successfully remove default judgment.

Insight 4: How many debtors are pro se?

The database was sourced from the internal SQL database through a comprehensive query that joined cdocs_case_meta_index, wp_terms, wp_term_taxonomy, docs_party_assignment_index, wp_posts, and cdocs_case_action_index. This approach was specifically designed to filter and extract data concerning cases with debt-related descriptions.



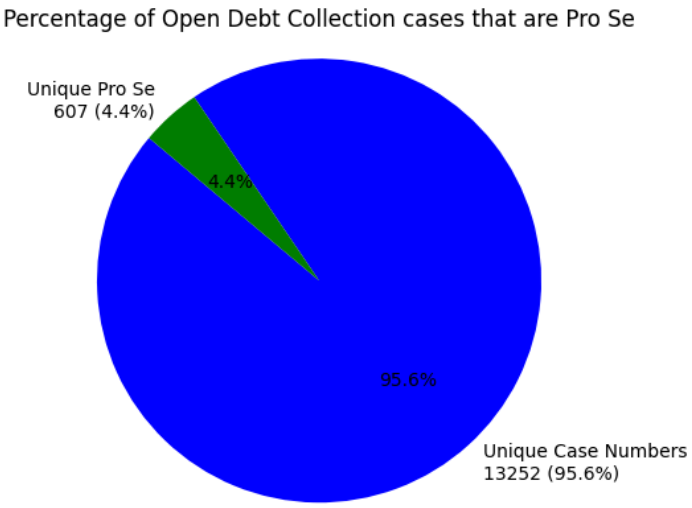
The executed SQL query was structured to select critical case details such as case number, type, status(open or close), and assigned attorney. The focus was on identifying debt-associated records by searching for the term 'debt' in the action descriptions. Lastly, a “WHERE description clause action_idx.description LIKE '%debt%'” was used to identify all the debt cases. The query was loaded into a data frame, and the Python line default_cases['Case Number'].nunique() was used to prevent repeat cases. Since the database did not include all the information on one line, each case had repeat case numbers in the rows(action, judgment, and any extra events in the case). Lastly, debt_cases[debt_cases['Attorney'].isin(['Pro Se'])] returned all the unique case numbers in which the attorney was Pro Se.

The extracted data was stored in a Python list using the fetch() method and subsequently manipulated using the panda's library. Our analysis primarily targeted filtering and reviewing data under specific categories relevant to debt cases, such as 'Housing Court Civil', 'Housing Court Small Claims', 'Small Claims', and 'Civil'.

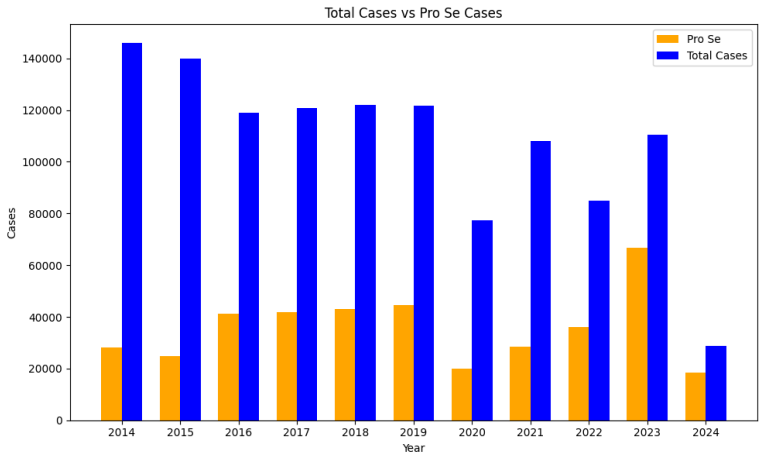
The analysis provided a clear overview of the distribution of debt cases across the identified categories. It revealed a relatively small fraction of debt cases are handled without formal legal representation, suggesting a potential gap in legal support for affected parties.

The data analysis offers valuable insights into the landscape of debt-related cases and highlights critical areas where policy intervention or support could be significantly beneficial.

Further steps will involve a more detailed analysis to develop targeted solutions to address the identified gaps in legal representation.



Further analysis by looking at the number of Pro Se cases per year and total cases shows that we can distinctly see the percentage of cases that were Pro Se compared to the total number of cases in that year. The graph below shows this relationship:



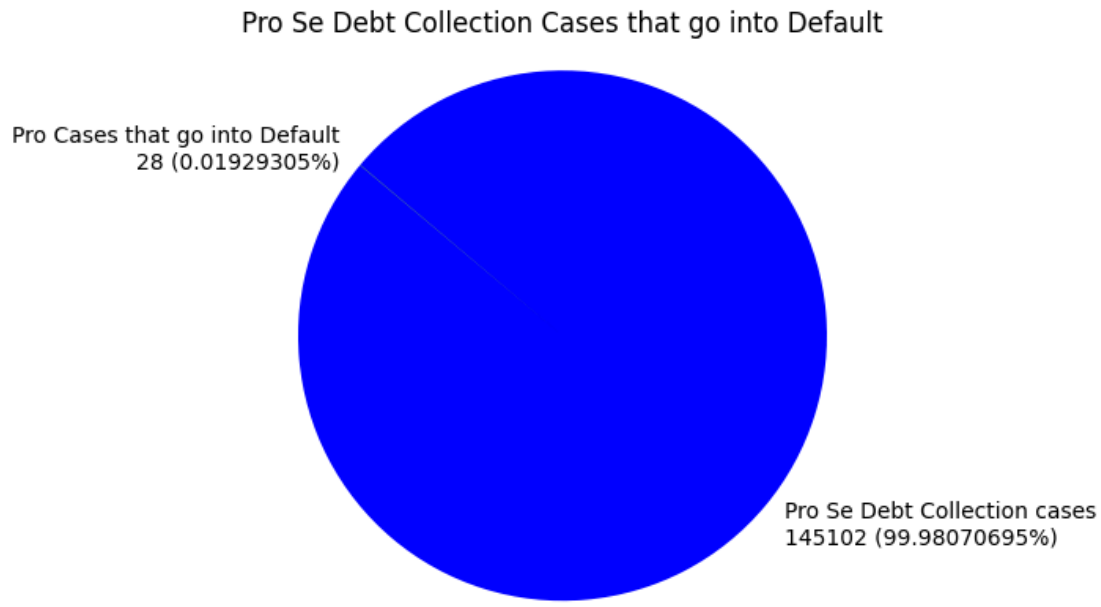
The graph above was obtained by combining the query from the first insight and then creating another query that searches for all Pro Se debt cases during the same period (the last ten years). Some interesting points are that over the last decade, there have been 393,817 cases and counting, with the most being in 2023 at 66,846 cases.

Extension 4.1: How many pro se cases go into default

Extension insight 4.1 explores the characteristics and frequency of debt collection cases at risk of default, particularly focusing on cases represented by Pro Se.

The SQL query was adjusted to normalize attorney representation data, specifically converting entries labeled 'Pro Se (PROPER)' to 'Pro Se' for uniformity. It targeted open cases involving debt and flagged them with default risk descriptions. Once again, the WHERE clause was used to find debt cases and default cases:
WHERE
action_idx.description LIKE '%debt%'
AND action_idx.description LIKE '%default%'

In addition, an AND statement was used to extract cases with debt and default in their action and description. Similar to insight 4, .nunique() returned a unique case number, and ['Attorney'] == 'Pro Se'] returned only Pro Se cases. The findings reveal that many Pro Se cases have lower default risk, with only 28 of the 145130 cases going into default.



Insight 5: Who are the debt collectors?

All debt collectors in Massachusetts should be registered and licensed to operate as debt collectors. Working off this assumption, Mass.gov was analyzed, and a list of licensed debt collectors, updated quarterly, was located. A general list was created using government requests and archives with 4835 records.

Additional records were supplied from an annual debt collector report 2023 and the Receivables Management Association International (RMAI) Membership Roster. More records exist on the RMAI certifications and RMAI individual records. Not all the records supplied by RMAI have been considered due to timing and accuracy. Still, five companies dominate the records or may

already be present in the other records. Additionally, while the records on the debt buyers have been added, they cannot be processed due to the database server experiencing multiple interruptions while completing the project. The Excel list of RMAI certified individuals is mostly filled but is missing a good portion of the members which will need to be discovered in the project's future interaction.

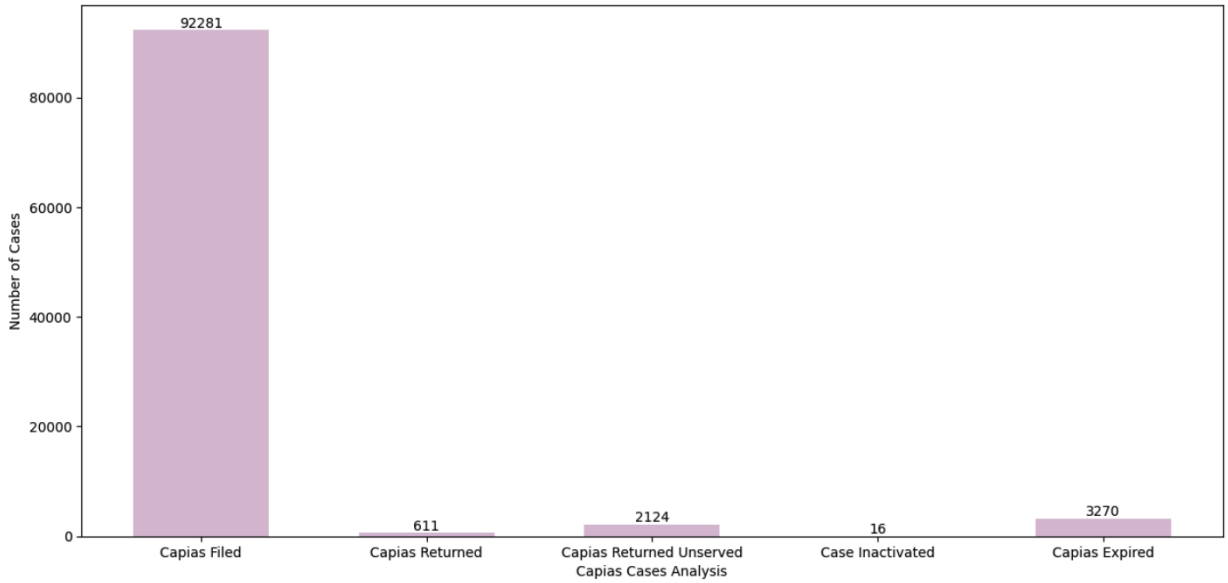
The list of companies collecting this information is in the data folder for Team A, with the names *debt-collector-licensee-all.csv* for the unfiltered dataset with duplicates and debt-collector-licensee-filtered for the filtered dataset. The extended analysis finding debt cases will discuss how this filtered dataset was created. All records are accessible on the repository for this section in the data directory.

The extension about getting debt collectors from the database includes an analysis of the debt collectors. This explains where this data generated was used. The CSV files are also located in the GitHub repository.

Insight 6: How many cases end up with a capias warrant?

A capias is a legal order or warrant directing law enforcement to take a person into custody. It is generally issued by a court to compel someone to appear before the court. In debt cases, it can be seen that the court files for capias in cases where the defendant is absent from court proceedings or not complying with the provided judgments. We analyzed the last 11 years of data to understand the capias cases. We searched for capias cases in the '"cdocs_case_action_index" table from the *wp_courtdocs* database, using "capias" as the action type to identify all capias cases over the past 11 years. From the "cdocs_case_meta_index" table, we got the post_id and the case_numbers of all cases during that time. The post_id in "cdocs_case_meta_index" maps to the case_id in "cdocs_case_action_index." Using the known list of debt cases we retrieved in our earlier analysis; we filtered out only the case_ids for the debt cases among all the capias cases. The total number of debt cases that result in capias is 92281.

Extension 6.1: How many capias warrants are served successfully?



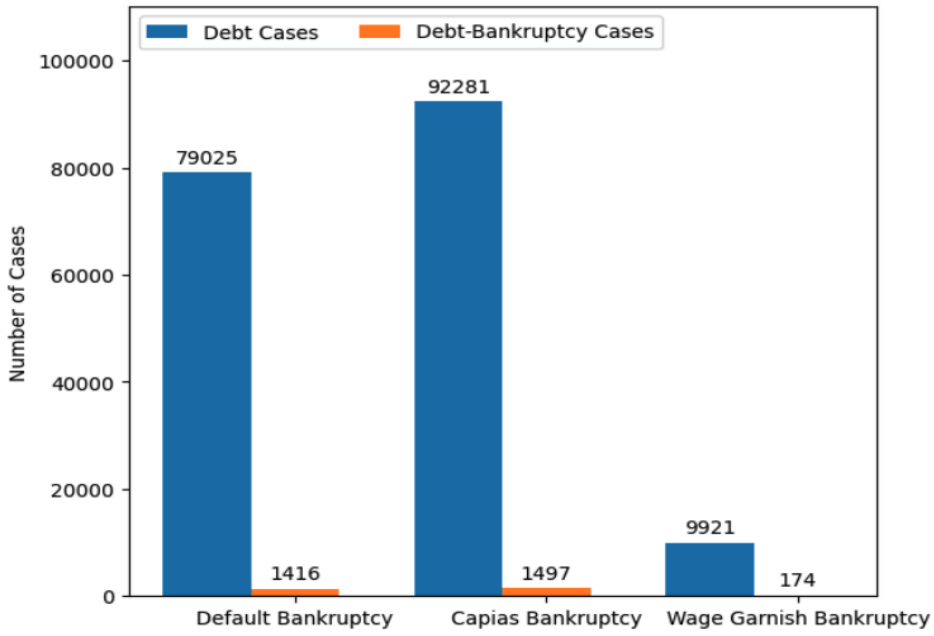
The figure above represents the progression of capias cases. It shows the number of debt cases that result in court filing capias, cases with capias returned (successful in serving capias), cases with capias returned unserved (unsuccessful in serving capias), inactivated capias cases, and expired cases (capias time limit lapsed). As you can see from the figure, while many debt cases have capias filed (92281), only a small number successfully served the capias(611). About 17.7% of debt cases result in capias warrants, but 31.8% of cases that end up in capias warrant cases are debt cases.

Insight 7: How many cases end up with wage garnishes?

Wage garnishment is a legal process in which a court orders an employer to withhold a portion of an individual's earnings to pay off a debt that the individual owes. This mechanism is typically used when a debtor has not voluntarily met their payment obligations on debts. To understand the pay garnishment cases, we examined data spanning the previous 11 years. By searching for the term "wage garnish" in various permutations in the description column, we were able to discover the wage garnishment cases in the "cdocs_case_action_index" table from the wp_courtdocs database for our necessary period. All of the duplicate cases were eliminated. As a result, we were able to track down every case of wage garnishment during the previous 11 years. We obtained the post_id and case_numbers of each case that occurred during that period from the "cdocs_case_meta_index" table. The case_id in the "cdocs_case_action_index" corresponds to the post_id in the "cdocs_case_meta_index." We filtered out only the case_ids for the debt cases among

all the wage garnishment cases using the known list of debt cases that we had previously recovered. The total number of debt cases that result in wage garnishment is 9921.

Extension 7.1: How many of the default, capias and wage garnishment cases file for bankruptcy?



Bankruptcy is a legal process through which individuals or businesses who cannot meet their financial obligations can seek relief from some or all of their debts. Bankruptcy laws intend to give debtors a fresh start financially by discharging debts that are beyond their capacity to pay while also providing creditors an opportunity to obtain some repayment based on the assets available for liquidation.

We examined data spanning the previous 11 years to gain insight into the bankruptcy cases. Using the term "bankruptcy" in the description field, we found the bankruptcy cases in the "cdocs_case_action_index" table from the wp_courtdocs database for our necessary period. All of the duplicate cases were eliminated. This helped us find every bankruptcy case filed throughout the 11 years. We obtained the post_id and case_numbers of each case during that period from the "cdocs_case_meta_index" table. The case_id in the "cdocs_case_action_index" corresponds to the post_id in the "cdocs_case_meta_index." We filtered out only the case_ids for the debt cases among all the bankruptcy cases using the known list of debt cases we had previously recovered. The total number of bankruptcy debt cases is 7903.

The figure above shows the total number of debt cases for default, capias, and wage garnishment and how many of them filed for bankruptcy. The number of debtors filing for bankruptcy is significantly smaller for all cases.

Future Scope

In considering the project's future direction, the group proposed several enhancements. Firstly, we suggested restructuring the database to expedite query processing and creating new tables that consolidate all available information under a single case number and one singular row. Additionally, a GUI or application can be developed to display critical statistics of cases at a glance. Lastly, further research could explore trends and relationships between debt collection agencies and their debtors. There was a notable discrepancy in the number of cases resulting in wage garnishments; the wp_cdoc database recorded only 1,359 instances, with approximately 7% involving defendants. More comprehensive details are available in the "Defendant Yiyang" notebook. This issue should be investigated in future studies, especially if subsequent teams report similar findings, as insights 6 and 7 might reflect duplicate results due to cartesian products.

Individual Contributions

Andrew Woska

- Who are all the debt collectors from the last few years in Massachusetts?
- How many small claims cases are made by debt collectors?
- Find better ways of identifying debt collectors to get closer to the expected values
- Create a way to process files of debt collectors
- Team representative

Patrick Meade

- How many cases have been filed each year over the last ten years in each court?
- Pro Se Cases vs total amount of cases
- Helping to identify debt cases over the last ten years

Sumatra Dhimoyee

- How many cases have gone into default in the last decade?
- How many cases lead to capias warrants?
- How many cases end up filing for bankruptcy?
- How many cases end up with wage garnishes?

Quan Ho

- How many total debt cases go into default?
- When there is a default, how often is there a judgment?
- How many debtors are pro se?

- How many pro se cases go into default?
- Found column name of each file for all three databases

Yiyang Cai

- How many cases are Wage Garnishes?
- What are the percentages of defendants in Wage Garnishes and Capia Warrant?