

# Early Insights Report

## Debt Collection Team A

Andrew Woska, Patrick Meade, Sumatra Dhimoyee, Quan Ho, Yiyang Cai

GitHub: <https://github.com/BU-Spark/ds-wgbh-debt-collection/tree/team-a/sp24-team-a>

### Insight 1

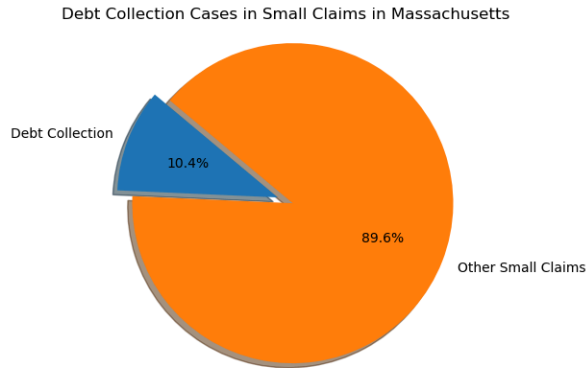
**Key Question 2:** What percentage of cases filed in small claims and district court are filed by debt collectors?

**Key Question 3:** Who are the debt collectors?

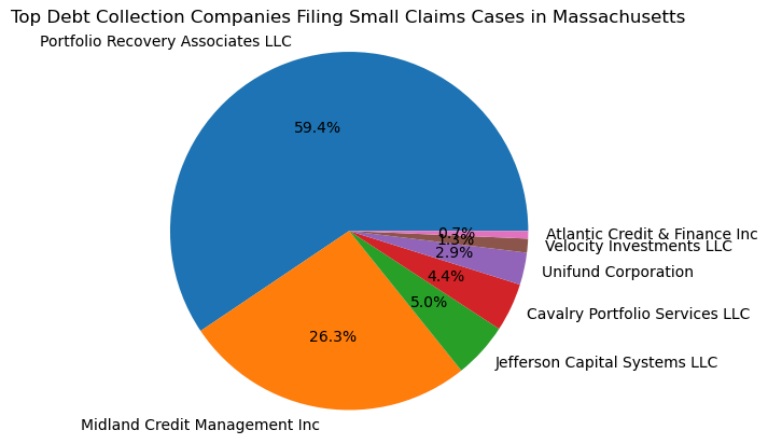
The *wp\_courtdocs* and *civica\_courtdocs* databases from MassCourtsPlus.org were analyzed to determine the number of debt collection cases. Both databases show nearly the same amount of small claims and district court cases. It was determined that the Civica database was easier to read the raw database columns. From this, a SQL query was created to get all small claims and district court cases.

Additionally, an Excel spreadsheet is located with all the registered debt collectors for Massachusetts on mass.gov. Comparing the names of the debt collection companies with the database gives the number of debt collectors who have brought a case to the small claims courts. Multiple companies may be missing since the current data set accounts for only the debt collectors registered in Massachusetts for the last quarter of 2023. There are likely some issues comparing debt collection company names with the listed plaintiff names. This means that there may be even more cases that have not been accounted for in this analysis. This analysis also does not account for the dates or status of each case.

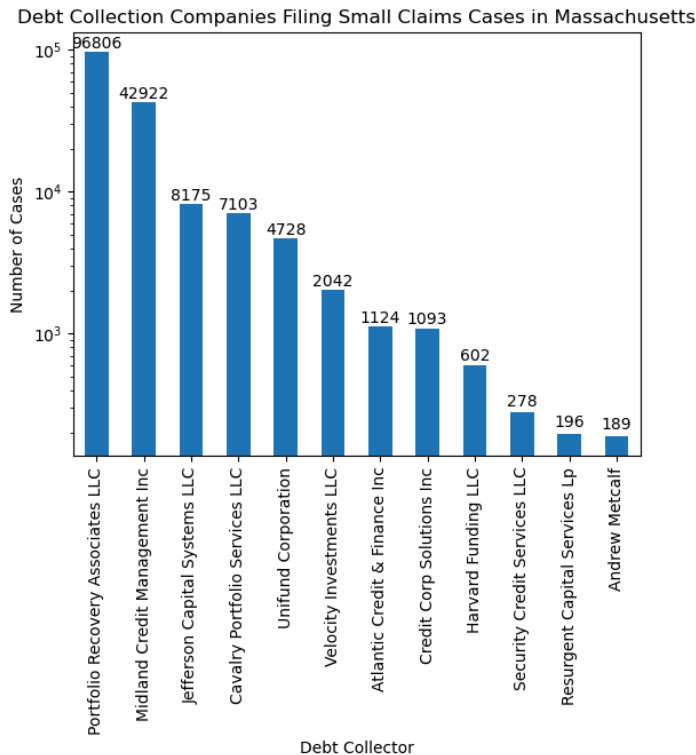
The following tables visualize the small claims cases as they refer to debt collection.



There are a total of about 1.6 million small claims cases according to the database. Analysis has discovered that there is about 165,000 debt collection cases brought before the small claims court. This means over 10% of all small claims cases brought before the district courts in Massachusetts are from debt collectors.



The top two debt collection companies based on cases filed for Massachusetts are responsible for nearly 86% of all debt collection cases. It should be noted that neither of these companies are located in Massachusetts. Portfolio Recovery Associates LLC is in Virginia, and Midland Credit Management Inc. is in California. No companies present in the above chart are located in Massachusetts either.



The chart above uses a logarithmic scale because of how many cases the top two companies have files cases. The first company located in Massachusetts is Harvard Funding LLC which makes up less than 0.4% of all cases filed for debt collection.

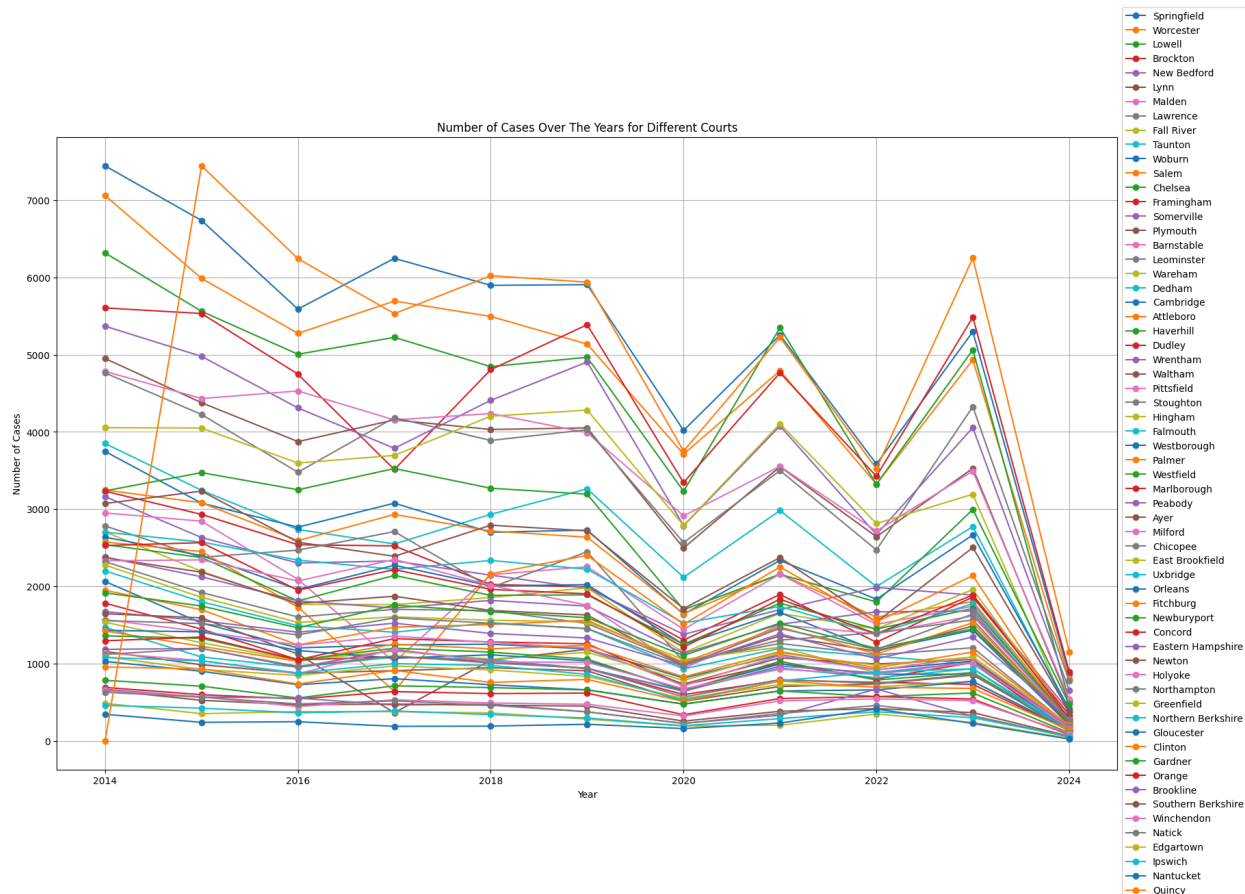
There are many small claims court cases filed by debt collection companies making up over 10% of all cases currently filed in the MassCourtsPlus.org database. Most of these companies are not located in the tri-state area. The Massachusetts government has a publicly available Excel spreadsheet that contains the registered debt collectors for the state which is used to identify debt collectors based on case plaintiff.

## Insight 2

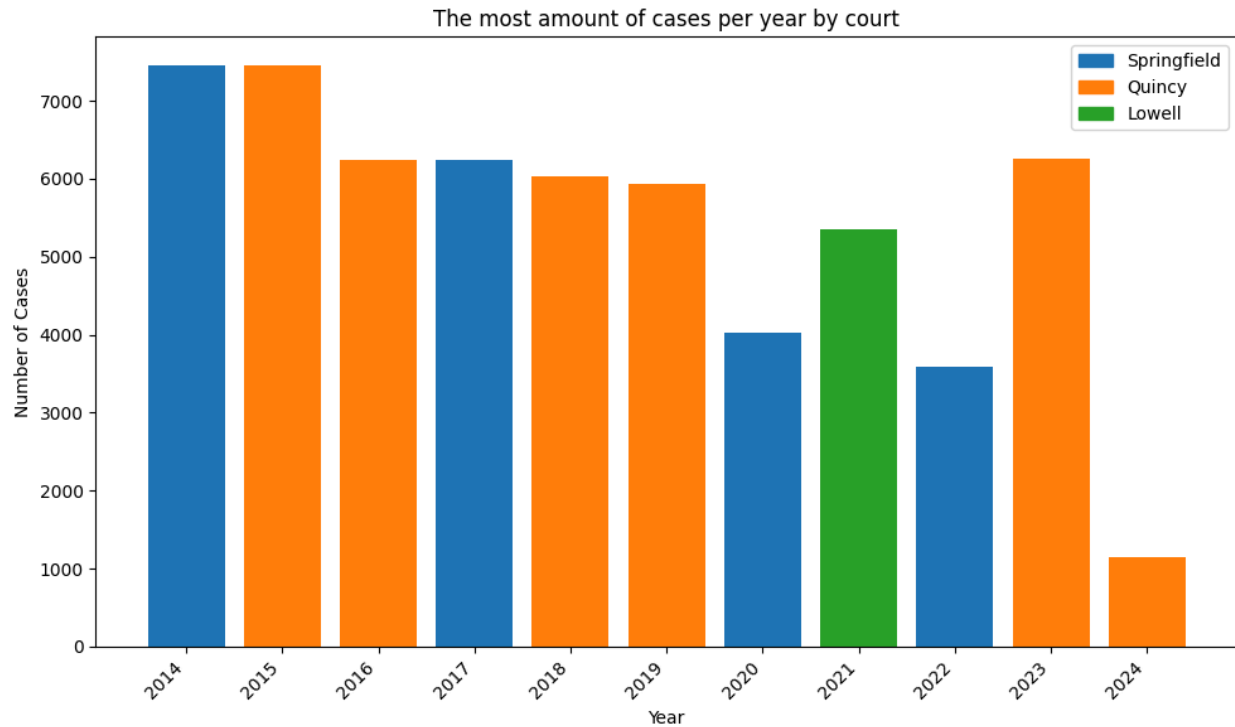
**Key Question 1:** How many cases have been filed each year over the last ten years in each court?

The *wp\_courtdocs* from MassCourtsPlus.org was used and analyzed to determine the number of cases that have been filed each year over the last ten years in each court. Through writing an SQL query, we're able to collect the necessary data needed. In order to get the data we need to restrict the department of the courts to just that of district court. When using the query we found out that there has been a total of

1,167,365 cases in the last ten years. The data obtained was also used to create an interactive data table based on year, court, count with the data with the year being the case filing date. The interactive notebook can be found in the associated notebook. With the data and the knowledge of the total number of cases in the last ten years the below graphs below are to help visualize overtime trends and percentages of case filings.

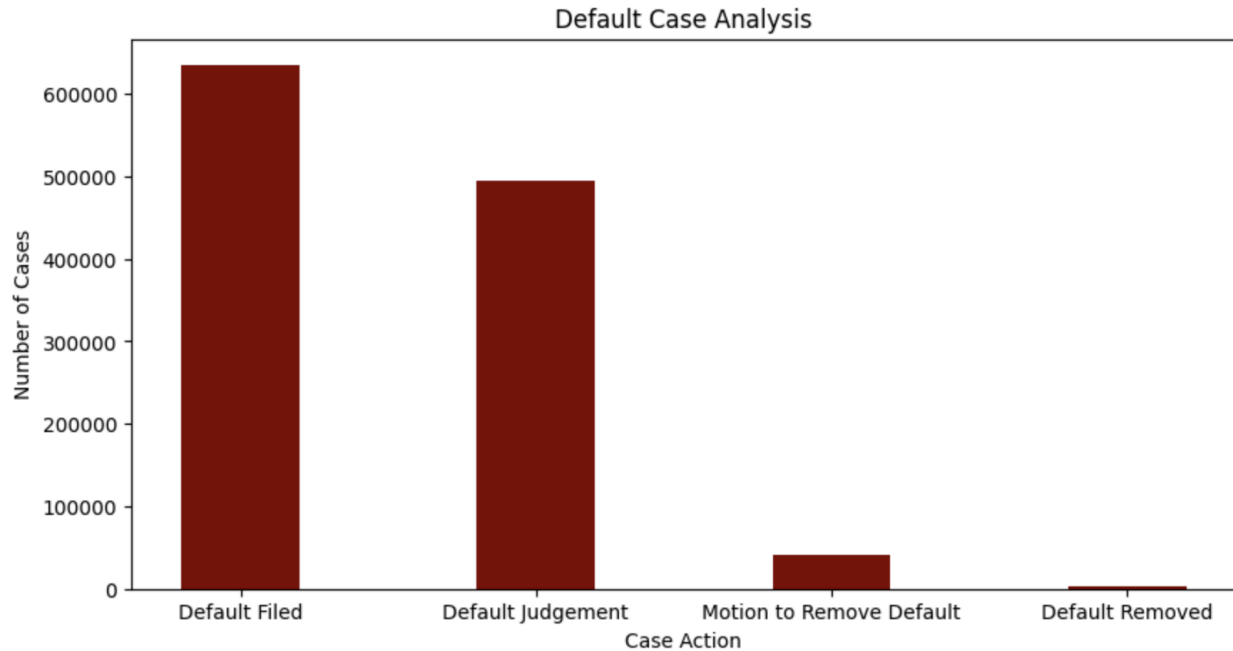


Even with the multitude of courts the graph shown does have repeating colors but through the use of the other graph located below and the data table we can see the general trends with all the courts. Through obtaining this overall data we can see how for some courts, there was a dramatic increase in court filing, with a sudden decrease in 2020, which is most likely due to the start of the pandemic. We can see the sharp increase around 2023 in most notability as coinciding with the rise of inflation during that year. With this data we can also look at the various trends that were obtained with this data.



What is very useful about the provided bar chart above is that it shows where the maximum number of cases are filed. It's very interesting as we see the general trend of Quincy having the highest number of cases, with Springfield popping up every couple of years . The most surprising thing about this graph is Lowell in 2021, which could imply how pandemic restrictions during 2021 affect the court filings among other things.





The `cdocs_case_action_index` table from the `wp_courtdocs` database was analyzed for the entire time range starting from 2003 to the latest of 2023 to understand the condition of the total default cases. We investigated to find the number of total cases that file for judgment default (634977), receive default judgment (494663), the number of those cases that process to file a motion to remove default judgment (40885), and finally, the number of cases that get the default judgment revoked (2595).

As you can see from the bar plot and the numbers, while the number of cases that get filed for default judgment and receive the judgment are comparable, the number of instances that file motions to remove default judgment and eventually receive their default judgment revoked are significantly lower. The observations align with our assumption that most debtors might not have the resources to defend themselves legally. It opens up further questions that we would like to explore in the coming weeks, such as:

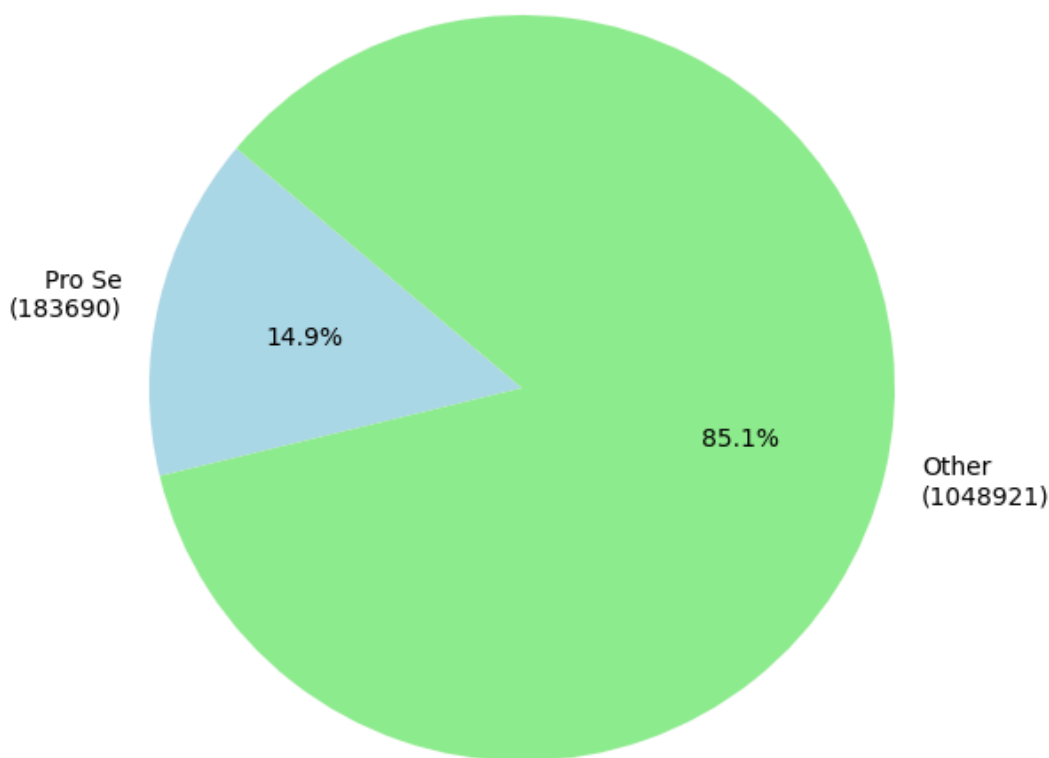
1. What are the mean, maximum, and minimum debt values that go into default?
2. How many of those debtors get default judgment files for bankruptcy?
3. Who are the debtors who can successfully get their default motion revoked?
4. Are the most debtors individuals or organizations?

## Insight 4

- **Key Question:** How many debtors are pro se?

From the masscourtsplus.org SQL database, there are at most 183,690 debt cases that are Pro Se. Since all debt collection cases either fall under civil or claims, a maximum amount of debt cases can be calculated. From this logic, there can be a maximum of 1,232,611 debt collection cases as debt collection cases would only fall under the case types: Housing Court Civil, Housing Court Small Claims, and Small Claims. The 1,232,611 cases can be filtered by attorney name, which includes Pro Se and Pro Se(ProPer). Pro Se and Pro Se(ProPer) are used interchangeably to mean Pro Se, which is the representation in court without an attorney. Pro Se and Pro Se(ProPer) are just different in which court the case was filed, but they both mean representation without an attorney. From the 1,232,611 cases, there are at most 183,690 debt collection cases that are Pro Se. This means that of the debt collection cases, at most, 14.9% are Pro Se.

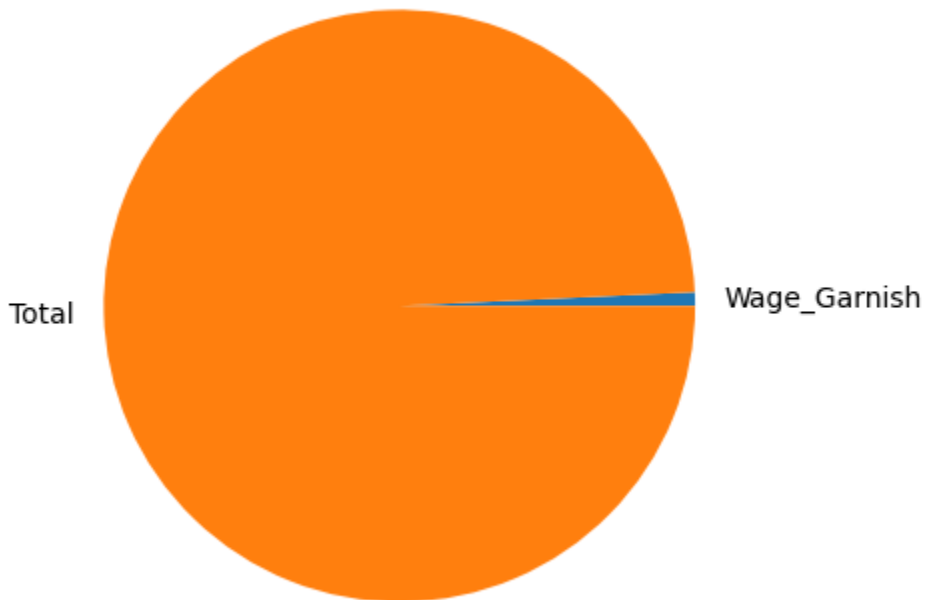
Maximum Percentage of Debt Collection cases that are Pro Se





## Insight 5

- How many cases end up with wage garnishes?



From the masscourtsplus.org SQL database, there are 1357 debt cases that are pertinent to Wage Garnishes. Since The process for employers usually begins with a garnishment notice or order, which generally comes from a court or government agency. Employers must start withholding the specified amount from the employee's paychecks and send it to the creditor. Hence This number is pretty crude considering such wage Garnishes can be paraphrased in various ways. However this number does suggest that wage Garnishes is only a very small percentage of all cases filed.

## Is the data enough to answer the question?

The supplied MySQL databases provide enough information for most questions. Further information is being gathered from mass.gov with all the debt collector's information.

## Next Steps

All team members have been assigned new tasks that either expand on their current key question or are working on new key questions. The current questions that the team is working on are:

Andrew

- Who are all the debt collectors from the last few years in Massachusetts?
- How can data be better filtered to identify debt collectors?

Patrick

- Looking at population density over the years and comparing that to the provided data.
- Can we do a breakdown about where the debtors live?

Sumatra

- Find the mean, maximum, and minimum debt values that go into default
- Find how many of those debtors that get default judgment file for bankruptcy

Quan

- Find a minimal number of debt collection cases and minimal number of Pro Se cases
- Find a way to extract information for each case number so it can identify debt better.

Yiyang

Find the actual number of Wage Garnish in all cases

Evaluate the possibility of using a language model to help classify the Wage

Garnish type