GBH INVESTIGATIONS: DEBT COLLECTION COURT

Final Report

Fall 2024

Team

Benjamin Coleman	benbc@bu.edu
Caslow (Shih Wen) Chien	caslow@bu.edu
Hitaishi Hitaishi	hitaishi@bu.edu
Hyun Sung Park	phs@bu.edu
Kenji Wagner	kenjiwag@bu.edu

Table of Contents

I. Introduction

The GBH's Debt Collection Court project examines the judicial handling of student loan debt cases, addressing an area that has not yet been systematically investigated. As outlined by our client Jenifer B. McKim, the primary goal of this project is to analyze the frequency, scale, and key parties involved in these lawsuits, specifically investigating Massachusetts Educational Financing Authority. This project aims to provide robust research, factual data, and statistical insights for the investigative story on debt collection practices in Massachusetts.

To achieve this, we analyzed historical court case rosters and identified major debt collectors involved in student loan cases. Our findings revealed key players such as the Massachusetts Educational Financing Authority (MEFA), National Collegiate Student Loan Trust (NCSLT), Sallie Mae, and others, with MEFA alone accounting for 5.8K cases, followed by NCSLT at 4.9K. This information was visualized through an Interactive Looker Studio dashboard and an Interactive Map of debt collection cases, offering the client a valuable tool for continued analysis and reporting.

Additionally, a Natural Language Processing (NLP) model was developed to standardize spelling variations of debt collectors' names in the database, ensuring data accuracy. This enhanced our ability to create detailed, accurate visualizations of the most active debt collectors.

Through these efforts, the project provided a comprehensive examination of debt collection practices in Massachusetts, contributing to a deeper understanding of the student loan debt crisis and offering actionable insights for public awareness and further investigative reporting.

II. Data Collection and Processing

We utilized two data sources to complete the tasks during the project:

- The MassCourtsPlus Database: We primarily used the MySQL database powering MassCourtsPlus, an online portal that allows individuals to access court information digitally. The database was developed by Adam Friedman (Civera) and consists of case data such as the names of plaintiffs, defendants, and attorneys, case filed dates, current case status, assigned courts, and judges.
- External Data Sources: We reviewed external data sources to investigate some tasks further. The external data varied from the annual bank division report to debt collector licensee lists

A. Handling Data Discrepancies of Previous Works

We started the project with the client's request to identify and resolve the discrepancy issues in the deliverables from the previous project. The numbers of 'Small Claims' and 'Civil consumer debt cases from the previous project differed from the report from the Massachusetts Trial Court Department of Research and Planning[1], which led us to focus on improving the credibility of the data as our initial task. We reviewed and analyzed the outcomes and methodologies used by the previous project. As a result, we were able to find three possible causes for the discrepancy.

The first cause was related to the "method of handling Unicode encoding." We discovered that, in the Python code used in the previous project, some data was excluded while handling Unicode decode errors that occurred when reading .csv files. The first code uses the default encoding, typically UTF-8, and skips files that raise a 'UnicodeDecodeError,' logging an error message for those files. This approach is simple but lacks flexibility in handling files with different encodings, so it skipped some files as it raised errors. Therefore, we modified the code that first attempts to read files using the latin1 encoding and, if that fails, retries with utf-8. Files are only skipped if both encoding attempts fail, making this approach more robust and suitable for datasets with mixed encodings. After implementing our modification to their existing code, we found 1,484 more data (from 5,358 rows to 6,842 rows) by improving the flexibility of reading files.

The second reason was the difference in the scope of the data. Unlike the Massachusetts Trial Court's report, which included all types of "Civil" and "Small Claim" cases, the previous project focused on collecting data specifically related to selected Debt Collectors included in debt collector licensee lists. This data was filtered through a separate data-cleaning process before producing the outcomes, and the difference in scope might have contributed to the discrepancy in the number of cases.

Finally, we suspected using the 'cleancourt' package may have caused the issue. According to the developer, Logan Pratico, the 'cleancourt' package is "a tool that uses AI and string similarity algorithms to clean, standardize, and link legal party names" (Pratico, 2024)[2]. Using the package in this project may seem appropriate, but the developer also mentioned that it is developed and tested specifically for 'Civil' cases; thus, it cannot guarantee validity when used in other contexts, such as 'Small claim' cases. Since the number used in this project includes

[1] https://www.mass.gov/doc/review-of-consumer-debt-cases-filed-and-disposed-2024

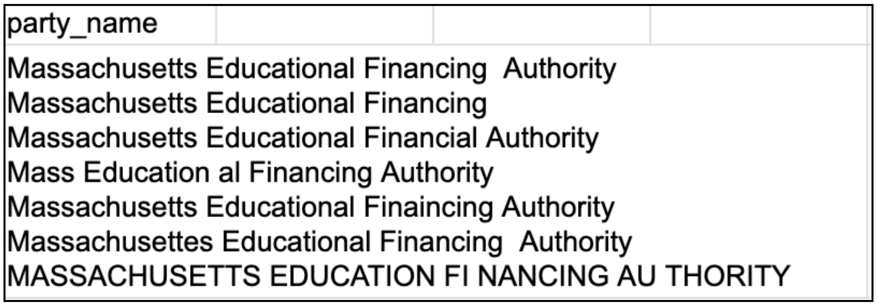
[2] https://pypi.org/project/cleancourt/

'Small claim' cases, the 'cleancourt' package might not have properly generated outputs as intended.

After sharing our findings on discrepancies in previous works, the client and our team agreed to start from the beginning, from fetching and cleaning the data to analysis.

B. Fetching and Cleaning Data

The data is fetched from the MassCourtsPlus database using SQL queries in Python codes. BU Spark! provided the credentials to access both MassCourtsPlus and the database, and we cross-checked both the online portal and the database to improve the reliability of our data. At the beginning of the project, our team fetched all 'Civil' and 'Small claim' cases to find major student loan debt collectors. After deciding on target debt collectors, we focused only on collecting all debt collection cases related to those debt collectors.



party_name
Massachusetts Educational Financing Authority
Massachusetts Educational Financing
Massachusetts Educational Financial Authority
Mass Education al Financing Authority
Massachusetts Educational Finaincing Authority
Massachusettes Educational Financing Authority
MASSACHUSETTS EDUCATION FI NANCING AU THORITY

Figure 1. Name Variations of MEFA (Massachusetts Educational Financial Authority)

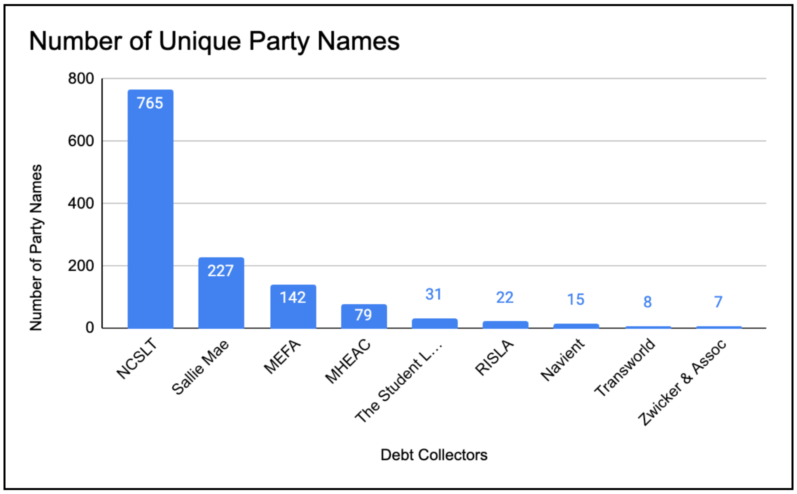


Figure 2. Number of Unique Party Names within the Database

The data was structured, but the names of debt collectors were so varied that we could not easily include all variations of names (Figure 1&2). To solve the issue, in addition to the keyword search in SQL queries, we developed an NLP model specifically designed to find all name variations of targeted debt collectors. Consequently, we got confirmation from a credibility test that our data exactly matched the state data.

Additionally, we cleaned and processed data for visualization and analysis. We standardized the varied names, excluded cases unrelated to debt collection (e.g., real property, housing), and categorized courts by Boston Municipal Courts (BMC), District Courts, and Superior Courts. We also cleaned some of the other terms that varied in several names.

C. NLP Model

The data, having been fetched from a government database, was bound to have data cleaning issues. As previously mentioned, the party column had numerous minor spelling errors within party names, which was somewhat problematic, given that our investigation relied on having an accurate number of cases per organization. Our solution to check for these spelling errors was to create an NLP model to find missing organization names that had different/incorrect spellings from our expectations. To account for the possibility of error in our NLP model, we also leveraged a regex query to account for all name variations. Therefore our solution was to create an NLP model and a regex query model to do an exhaustive search for all the potential names and eventually combine their correct results together.

The NLP model itself:

The NLP model we leveraged was the cosine similarity model, this model would find us party names that were similar to the official organization name. How it works: The model inputted the party names as vectors, compared their vectorization to the official organization name’s vectorization and then mapped the cosine angle to measure the similarity between said vectors. We lowered this threshold to relatively low values (0.5 to 0.6) so we could find all possible name variations (we would filter this down later, with the help of the regex querying model).

Regex querying:

We leveraged regex querying to find names that were similar to the official organization name. In this, we would take substrings of the name and query our data to only find party names that contained all queried substrings (ex. Massachusetts Educational Financing Authority looked for versions of the name that had mass, edu, fin, and auth as substrings in their party names) or abbreviations of the party (ex. Massachusetts Educational Financing Authority, was also known as MEFA).

Combining the model results:

To combine the model results, we needed to observe the disjoint sets of each model, that is we needed to see the unique results of each model. These sets were printed to csv files so we could evaluate which specific names in the sets were incorrect and manually filter out incorrect results accordingly. None of the regex sets included incorrect names but many of the NLP model names were incorrect, so we included certain restrictions on names and reran the NLP model to find the correct results. We then merged the sets together to have an exhaustive set of names. This led us to finding approximately 1000 new unique party name variations amongst the top debt collection agencies, with National Collegiate Student Loan Trust and Sallie Mae having the highest amount of name variation within the database, as shown in Figure 2. This also increased the number of cases found for our top 9 organizations by approximately 33%, highlighting the importance of this data cleaning task.

III. Objectives and Data Analysis

A. Finding Major Student Loan Debt Collectors

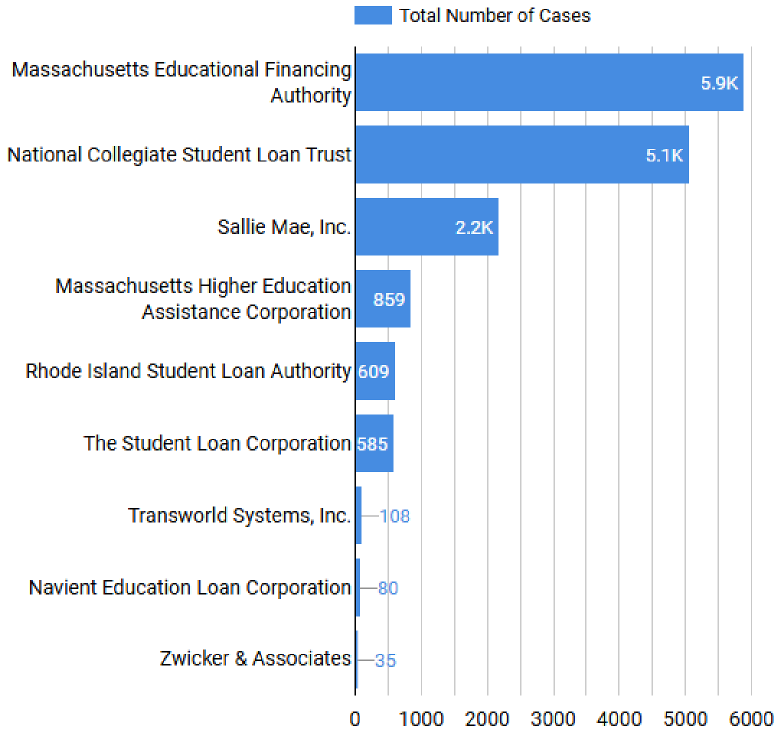


Figure 3. Student Loan Debt Collection Cases of Each Major Debt Collector

One of our main objectives in the project was to find major student loan debt collectors other than the *Massachusetts Educational Financing Authority (MEFA)* and the *National Collegiate Student Loan Trust (NCSLT*). Jenifer B. McKim has investigated and interviewed debtors of *MEFA* and *NCSLT*, but she wanted to seek other student loan debt collectors that she should look at.

We first started investigating major student loan debt collectors within the MassCourtsPlus database. We sorted student loan debt collection cases from the database and found the top plaintiffs who filed the most cases. From this process, we discovered *Sallie Mae, Massachusetts Higher Education Assistance Corporation (MHEAC), Rhode Island Student Loan Authority (RISLA),* and *The Student Loan Corporation (SLC)* as major student loan debt collectors.

According to Adam S. Minsky, a student loan lawyer, *Navient* is also one of the major student loan debt collectors, and its cases have gradually increased recently. He also mentions that *MEFA, NCSLT, Sallie Mae*, and *Navient* cases are handled by *Zwicker&Associates* (*MEFA*), *Ratchford Law Group* (*NCSLT*), and *Law Offices of Howard Lee Schiff* (*Sallie Mae* and *Navient*). We searched those law offices, and their names barely showed up in the database, except for *Zwicker&Associates*.

We identified nine major student loan debt collectors, including *Transworld*, which McKim requested to investigate. Subsequent analysis and data collection focused specifically on these nine debt collectors.

Since 2011, *MEFA* and *NCSLT* have been the most prominent debt collectors filing student loan debt collection cases, significantly outpacing other major collectors such as *Sallie Mae, MHEAC,* and *RISLA*, which have reduced their case filings over time. These two top debt collectors have filed thousands of cases, underscoring their dominant role in student loan debt collection cases. To enhance the data, a Natural Language Processing (NLP) model was developed to identify and standardize variations in the spellings of debt collectors, ensuring accurate and comprehensive data mapping. With efforts to improve the credibility of our data, we achieved more precise tracking of court filings and trends, offering valuable insights into the activities of key players in student loan debt recovery and the broader patterns of debt collection litigation in our analyses.

B. Case Status Analysis

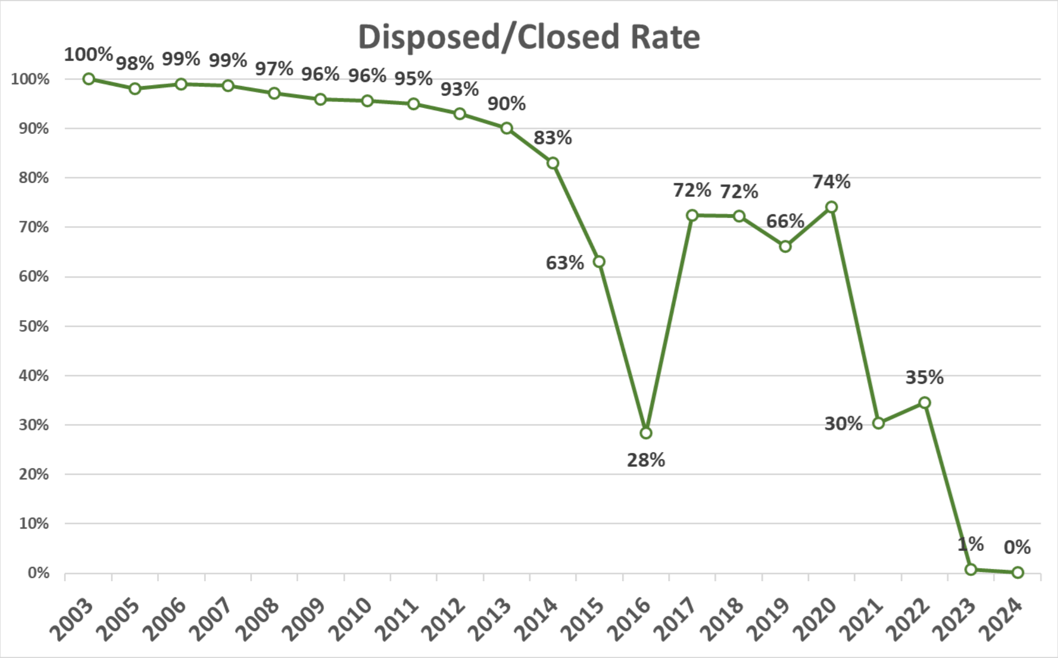


Figure 4. Disposed / Closed Rate of Nine Major Debt Collectors

(x-axis: Disposed / Closed Rate, y-axis: Years of Cases Filed)

Our analysis of student loan debt collection cases revealed some features in their status trends. Based on our investigation, the primary case statuses include *disposed*, *closed*, *open*, *pending*, *suspended*, and *suspended due to COVID-19*. Of the 15,405 cases, 11,055 have been disposed of or closed, while 3268 remained open, 1010 were pending, and 72 were suspended, including COVID-19-related delays.

The rate of disposed and closed cases is one of the key indicators of how those cases were resolved successfully. The trend showed a consistent peak of nearly 100% from 2003 to 2006 but experienced a sharp decline from 2013, reaching a record low of 28% in 2016. The decline was caused by the drastically increased number of pending cases, which increased between 2013 and 2016 and suddenly recovered from 2017. This trend highlights that many cases have not yet been resolved, but until the end of the project, the specific reason remained unknown.

After 2016, pending cases disappeared, and most cases have remained open. Considering the impact of the COVID-19 pandemic and the average years required to resolve cases, the recent decrease in the disposed or closed rate is reasonable. Analyzing these patterns was essential in understanding the effectiveness and timeliness of debt collection cases.

C. Case Type Analysis

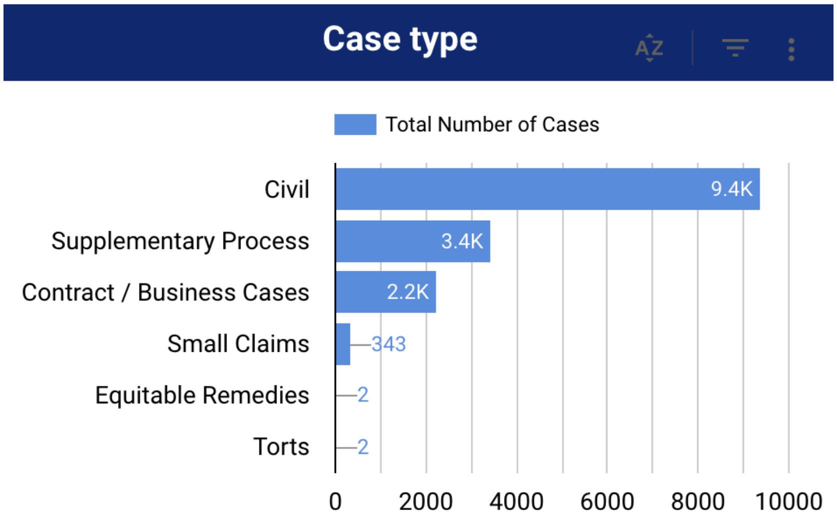


Figure 5. Number of Nine Major Debt Collector Cases by Case Types

The analysis of case types not only presents the proportion of how cases are categorized but also provides insights into the dollar amounts of debts involved with these cases. In Massachusetts, contract/business cases are assigned to superior courts, which deal with claims of $50,000 or more, while small claim cases involve $7,000 or less. Based on the dataset, civil cases account for 9,371 cases, followed by supplementary process cases at 3,440. Contract/business cases represent 2,247 cases, while small claims represent only 343 cases, about 2% of the total.

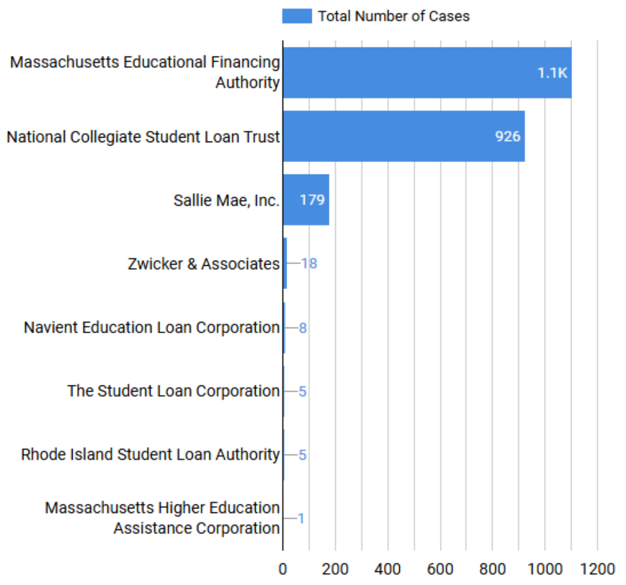


Figure 6. Number of Contract / Business Cases

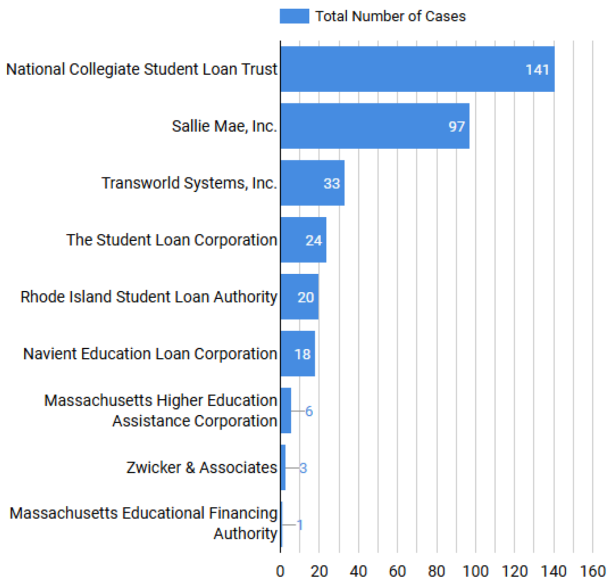


Figure 7. Number of Small Claim Cases

19% of MEFA and NCSLT cases were Contract / Business cases which is an above-average (average: 15%) proportion, while MEFA barely filed small claim cases. This indicates that those two debt collectors filed cases of generally higher dollar amounts of debt than others.

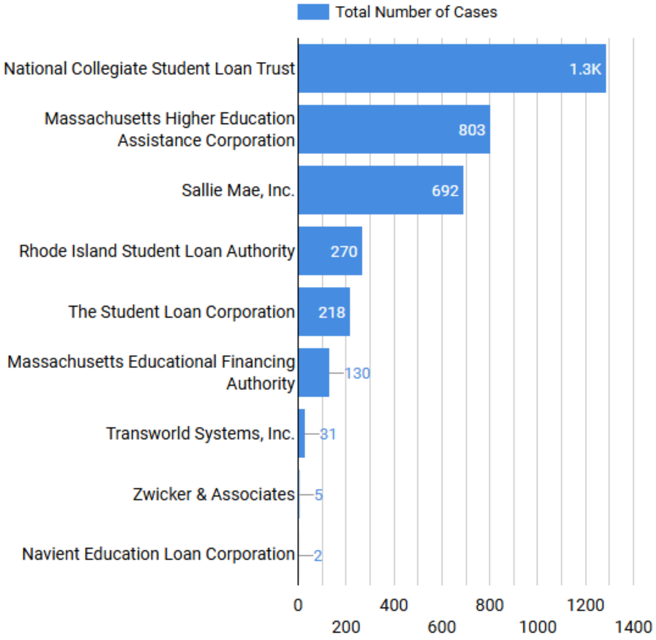


Figure 8. Number of Supplementary Process cases

The supplementary process is a legal procedure initiated after a creditor (debt collector) obtains a judgment against a borrower who has defaulted on their loan. This process is used to investigate the borrower's financial situation and enforce the debt repayment. In the database, NCSLT, MHEAC, and Sallie Mae filed most of the supplementary process cases. This implies that the debtors of those organizations had problems with paying debts, and debt collectors proactively utilized the process to collect debts from their debtors.

In summary, most student loan debt collection cases involve at least $7,000, and a moderate portion of cases involve $50,000 or more. MEFA and NCSLT cases average higher debts than others. Also, NCSLT, MHEAC, and Sallie Mae usually filed supplementary process cases in the debt collection, enforcing debtors to repay debts.

D. Court Analysis

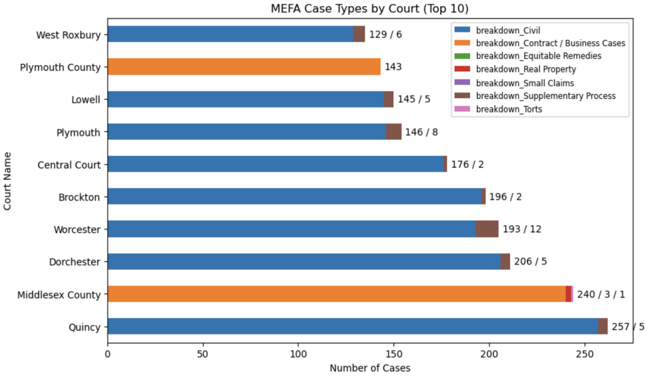


Figure 9. Top 10 Courts of MEFA Cases

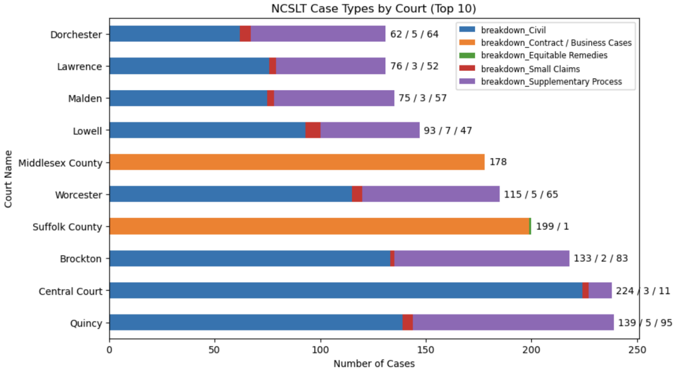
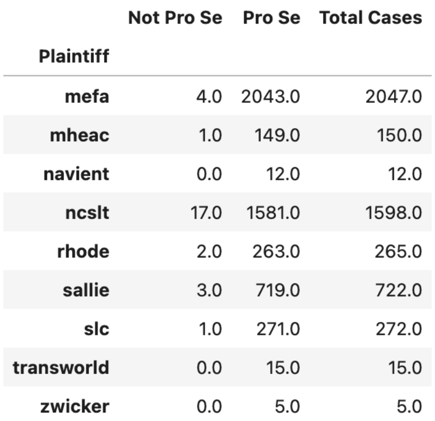


Figure 10. Top 10 Courts of NCSLT Cases

The court's analysis reveals how student loan debt collection cases were allocated to each court and provides geographical insights. In Massachusetts, the assignment of a case to a particular court is determined by factors such as the type of case, jurisdiction, the amount of monetary value involved, and the locations of the parties involved. As mentioned, contract/business cases involving $50,000 or more are assigned to superior courts. Since MEFA and NCSLT filed the most contract/business cases, a relatively large proportion of their cases were assigned to the superior court. MEFA's contract/business cases were mostly assigned to Middlesex County Superior Court, while NCSLT cases were also assigned to Suffolk County Superior Court. In most Boston Municipal Courts (BMC) and District Courts, supplementary process cases are evenly distributed with civil cases, but in Central Court, the ratio of supplementary process cases was relatively low. In common, Quincy District Court was assigned the most student loan debt collection cases among all Boston Courts. We assume that it is related to jurisdiction, but the specific reason is still unknown.

To visualize the distribution of courts, we prepared a court heat map to see the weight of allocation of student loan debt collection cases in Boston.

E. Pro Se and Default Judgement



Plaintiff	Not Pro Se	Pro Se	Total Cases
mefa	4.0	2043.0	2047.0
mheac	1.0	149.0	150.0
navient	0.0	12.0	12.0
ncslt	17.0	1581.0	1598.0
rhode	2.0	263.0	265.0
sallie	3.0	719.0	722.0
slc	1.0	271.0	272.0
transworld	0.0	15.0	15.0
zwicker	0.0	5.0	5.0

Figure 11. Default Judgements Data

The analysis of Pro Se cases within debt collection litigation highlights significant trends regarding representation and outcomes. Approximately 96% of the cases involving debt collector parties of interest were classified as Pro Se, where defendants represented themselves without legal counsel. Specifically, there were 14,459 Pro Se cases compared to only 564 Non-Pro Se cases. This substantial disparity underscores the prevalence of self-representation in such cases. The classification was determined using an attorney ID value of 0 within the database. Pro Se cases involve defendants representing themselves without legal counsel, and data shows that approximately 95% of these cases fall into this category. Among them, a striking 99.4% result in default judgments. These findings suggest systemic barriers to legal representation in debt collection disputes and provide a basis for further investigation into the implications of Pro Se status on case outcomes, access to justice, and the likelihood of default judgments.

F. COVID-19 Period

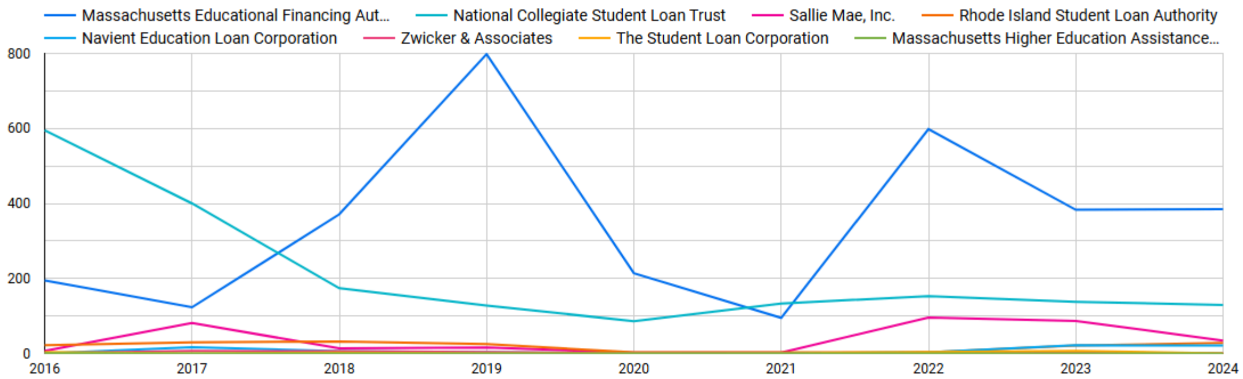


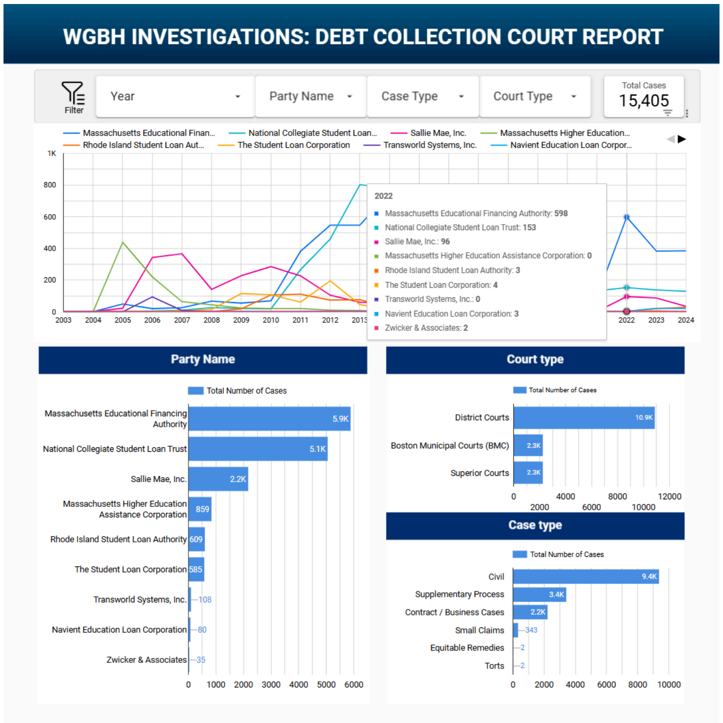
Figure 9. Decreased Number of Cases During COVID-19 Pandemic Period

During the COVID-19 pandemic, student loan debt collection filings significantly decreased, reflecting the pandemic's impact on legal and financial proceedings. Case filings were notably reduced across all major debt collectors, with some cases even being suspended due to pandemic-related disruptions. However, from 2022, MEFA and Sallie Mae showed an increase in filings, indicating a return to pre-pandemic operations even before the formal conclusion of the pandemic. The data highlights the temporary disruption caused by the pandemic and subsequent efforts of MEFA and Sallie Mae to resume debt collection activities.

IV. Visualization

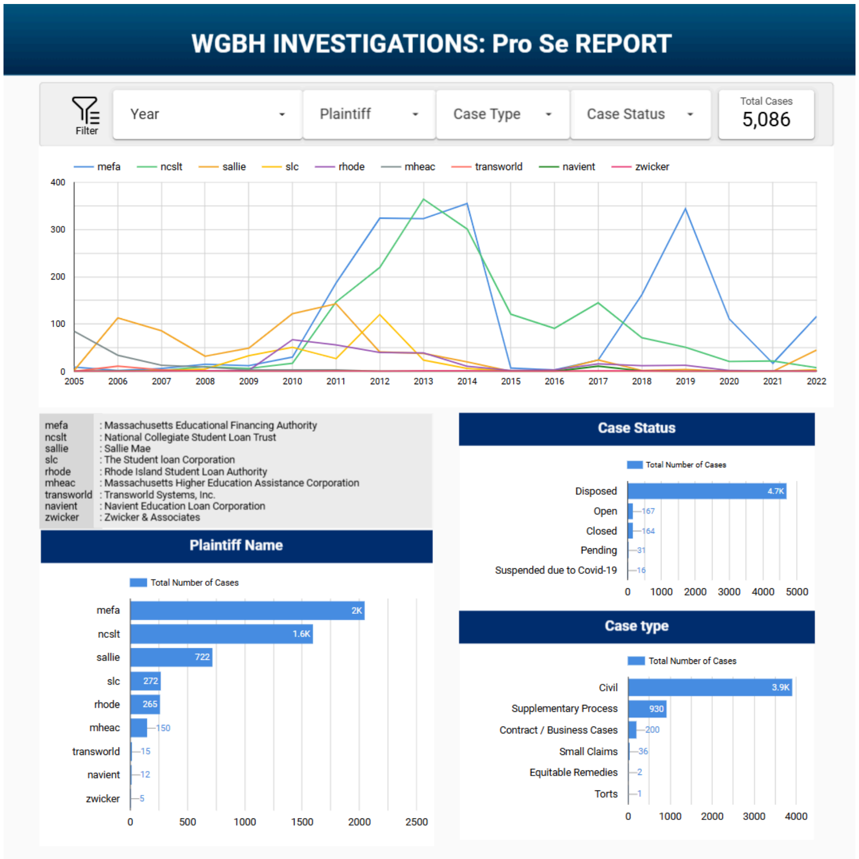
A. Looker Studio Dashboard

The dataset was generated using an IPython Notebook file, combining SQL queries and an NLP model to capture all relevant cases, as detailed in the “Data Collection and Processing” section above. A dashboard report was created to provide insights into student loan debt collectors identified throughout the project. Users can filter the data by year, party name, case type, and court type for more targeted analysis.


WGBH INVESTIGATIONS: DEBT COLLECTION COURT REPORT
Filter
Year
Party Name
Case Type
Court Type
Total Cases
15,405
Massachusetts Educational Finan...
National Collegiate Student Loan...
Sallie Mae, Inc.
Massachusetts Higher Education...
Rhode Island Student Loan Aut...
The Student Loan Corporation
Transworld Systems, Inc.
Navient Education Loan Corpor...
1K
800
600
400
200
0
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2022
2023
2024
2022
Massachusetts Educational Financing Authority: 598
National Collegiate Student Loan Trust: 153
Sallie Mae, Inc.: 96
Massachusetts Higher Education Assistance Corporation: 0
Rhode Island Student Loan Authority: 3
The Student Loan Corporation: 4
Transworld Systems, Inc.: 0
Navient Education Loan Corporation: 3
Zwicker & Associates: 2
Party Name
Total Number of Cases
Massachusetts Educational Financing Authority
5.9K
National Collegiate Student Loan Trust
5.1K
Sallie Mae, Inc.
2.2K
Massachusetts Higher Education Assistance Corporation
859
Rhode Island Student Loan Authority
609
The Student Loan Corporation
585
Transworld Systems, Inc.
108
Navient Education Loan Corporation
80
Zwicker & Associates
35
0
1000
2000
3000
4000
5000
6000
Court type
Total Number of Cases
District Courts
10.9K
Boston Municipal Courts (BMC)
2.3K
Superior Courts
2.3K
0
2000
4000
6000
8000
10000
12000
Case type
Total Number of Cases
Civil
9.4K
Supplementary Process
3.4K
Contract / Business Cases
2.2K
Small Claims
343
Equitable Remedies
2
Torts
2
0
2000
4000
6000
8000
10000

https://lookerstudio.google.com/s/pMLFsZAEAx4

The dashboard report highlights all student loan debt collector cases where the student represented themselves pro se, meaning without an attorney. Users can filter the data by year, plaintiff party name, case type, and case status for a more focused examination.



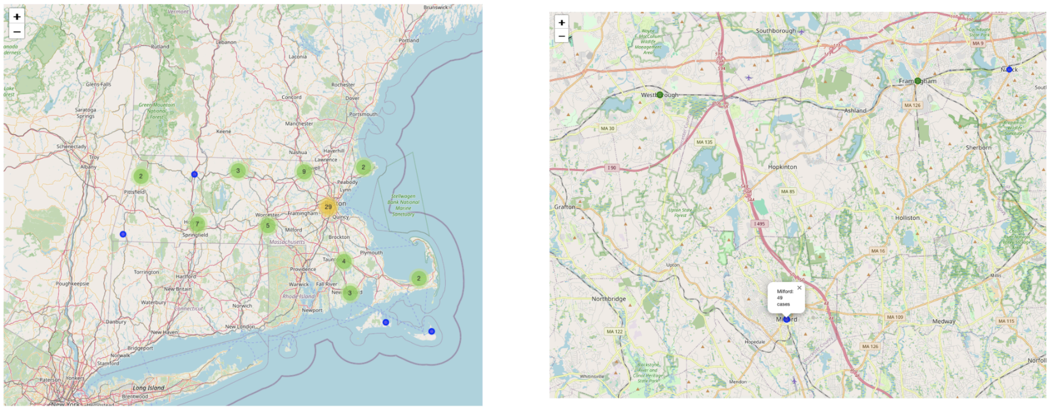
WGBH INVESTIGATIONS: Pro Se REPORT

mefa	: Massachusetts Educational Financing Authority
ncslt	: National Collegiate Student Loan Trust
sallie	: Sallie Mae
slc	: The Student loan Corporation
rhode	: Rhode Island Student Loan Authority
mheac	: Massachusetts Higher Education Assistance Corporation
transworld	: Transworld Systems, Inc.
navient	: Navient Education Loan Corporation
zwicker	: Zwicker & Associates

https://lookerstudio.google.com/s/gJpXiDX--y0

B. Court Heat Map

Screenshots from MEFA map



To add to our other visualizations regarding the court analysis, we created reproducible and interactive html court maps. The purpose of these maps was to present more of our findings on case counts in a more interesting and engaging manner. We specifically focused on MEFA and NCSLT, creating seaporate maps for both. The points on the map represented different courts and clicking on one would allow the user to see how many cases took place there involving the debt collector of interest. There was also a python file created “generate_html_courtmap_function.py” in the “courts_interative_map” folder where one can reproduce a map with an inputted dataset of cases.

V. Discussion

A. Challenges and Limitations

1) Name Variation

As mentioned above, the MassCourtsPlus database had no key indicators to differentiate specific parties, so party names were the only option for identification. However, the wide variation in these names caused challenges in fully capturing the targeting debt collector cases. These variations ranged from simple typos to differences in spacing and combinations with other stakeholders. To mitigate this issue, we used keyword searches and a specifically designed NLP model.

2) Hired third-party debt collectors

In this analysis, we included only cases in which major debt collectors were explicitly listed as plaintiffs. Therefore, cases related to their debt collection but not assigned to them as plaintiffs for various reasons might not be included in our dataset.

According to Jenifer B. McKim, these entities hire both registered debt collector companies and unregistered lawyers to collect their debts. In other words, the major debt collectors we focused on, such as MEFA and NCSLT, may have been associated with third-party debt collectors, leading to cases not revealed in the database. Other than a few law offices mentioned by Adam Minsky, we could not obtain detailed information about other third-party debt collectors. Consequently, cases where major debt collectors were not directly named could not be included in our dataset.

B. Future Work

1) Sharply increased pending cases between 2013 and 2016

There were drastic increases in pending cases between 2013 and 2016, which caused a decline in the disposed/closed case rate. However, pending cases have decreased since 2017. This can be investigated by examining whether it is an issue with the database, a transition of terms, or other reasons.

2) Hired third-party debt collectors

One of the challenges in our research was the inability to include cases in which the targeted major debt collectors were not mentioned in the case information but hired third-party debt collectors to collect their debts. Future studies may investigate hired debt collectors to include all possible cases related to targeted organizations.

3) Quincy District Court

Our court analysis reveals that Quincy District Court has been assigned the most student loan debt collection cases of major debt collectors. Future research could explore whether this phenomenon is due to jurisdictional factors, the concentration of student loan debtors in certain areas, or any other underlying reasons.

VI. Conclusion

In our efforts to understand the impact of major debt collectors in the state of Massachusetts, we investigated nine major debt collectors and tried our best to analyze their impact on debt collection cases as plaintiffs. Through leveraging an NLP model to find all known cases and conducting statistical analyses, three major debt collectors stood out, those being Massachusetts Educational Financing Authority (MEFA), National Collegiate Student Loan Trust (NCSLT), and Sallie Mae. MEFA, as originally theorized, had the biggest impact on student debt collectors in Massachusetts and was the primary organization that Jenifer highlighted in her story. With that being said this project allowed us to obtain more accurate data, investigate other major debt collectors, and set the framework for future investigations into other major debt collectors.

VII. Project Team Contributions

Benjamin Coleman: Supplemental research into debt collectors, analysis of annual debt collection reports, analysis of pro se cases, creator of html court maps, contributing report writer.

Caslow Chien: Created interactive dashboard, supplemental research into debt collectors, primary github repository manager and documentation lead, contributing report writer.

Hitaishi Hitaishi: Supplemental research into debt collectors, case status analysis, contributing report writer.

Hyun Sung Park: In-depth research into debt collectors, student loan debt collector investigation, case/court analysis, attorney analysis, client meeting preparation, contributing report writer.

Kenji Wagner: NLP model creation and application, supplemental research into debt collectors, contributing report writer.