TEAM BREAKDOWN

Caslow Chien
Master in DS
2025
caslow@bu.edu

Kenji Wagner
Team Leader
MSDS 2025
kenjiwag@bu.edu

Hitaishi Hitaishi
Master in DS
2025
hitaishi@bu.edu

Ben Coleman
Master in DS
2025
benbc@bu.edu

Evan Park
Master in DS
2025
phs@bu.edu

PROBLEM INTRODUCTION

- There hasn't been a systematic investigation into debt collection practices in Massachusetts courts.
- Focus on analyzing historical rosters of court cases related specifically to student loan debt.
- Goal: Investigate the frequency, magnitude, and parties involved with these lawsuits.
- Collaborating with Jenifer B. McKim, Deputy Investigative Editor at GBH, helping provide additional research, facts, statistics, and visualizations to assist with her story.

PROJECT OVERVIEW
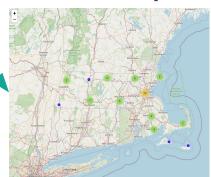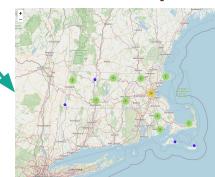
- Created and gathered accurate data, visualisations, and statistics on the most relevant debt collectors to assist GBH with their investigative story on student loan practices in MA.
- Created Interactive Looker Studio dashboard and Interactive Map for debt collection cases that the client can utilize moving forward.
- Developed a Natural Language Processing model to help us find different spelling variations of debt collection parties in the database.

DATA OVERVIEW

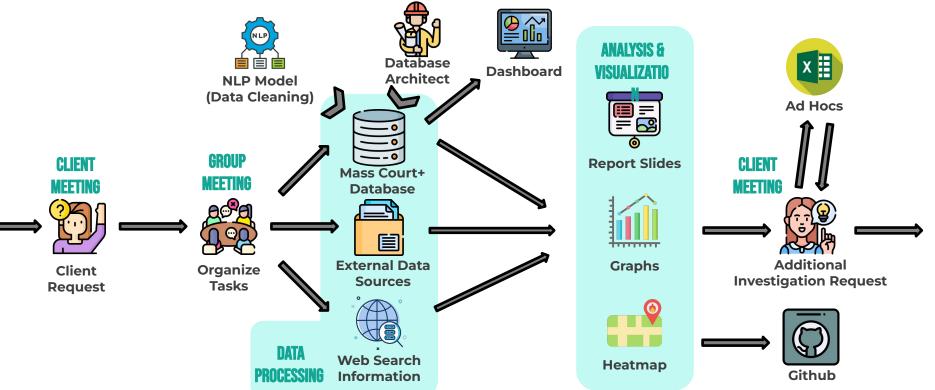Example Query:

```
sql_ncslt = """
SELECT DISTINCT
    cmi.post_id AS case_id,
    pi.post_id AS party_id,
    pi.party_name,
    pai.party_type,
    cmi.file_date,
    cmi.case_type,
    cmi.case_status,
    wt.name AS court_name
FROM
    wp_courtdocs.cdocs_party_index pi
JOIN
    wp_courtdocs.cdocs_party_assignment_index pai ON pi.post_id = pai.party_id
JOIN
    wp_courtdocs.cdocs_case_meta_index cmi ON pai.case_id = cmi.post_id
JOIN
    wp_terms wt ON cmi.court_term_id = wt.term_id
WHERE
    (pi.party_name LIKE '%nation%'
    AND pi.party_name LIKE '%colleg%'
    AND pi.party_name LIKE '%stud%')
    AND (LOWER(pai.party_type) LIKE 'plaintiff'
    OR LOWER(pai.party_type) LIKE 'trustee'
    OR LOWER(pai.party_type) LIKE 'other interested party');
"""
```

Relational Database Diagram:

- Source of Data: The database primarily tracks case details, party involvement, and court assignments in Massachusetts.
- Structure:

- Case Details: Stores case numbers, types (e.g., Civil, Small Claims), and status updates.

- Party Information: Tracks plaintiffs, defendants, and their attorney IDs.

- Court Metadata: Links cases to court names and terms for organizational purposes.

HOW WE WORKED WITH CLIENT?

NLP Model (Data Cleaning)

Database Architect

Dashboard

ANALYSIS & VISUALIZATION

Ad Hocs

CLIENT MEETING

Client Request

GROUP MEETING

Organize Tasks

Mass Court+ Database

External Data Sources

DATA PROCESSING

Web Search Information

Report Slides

Graphs

Heatmap

CLIENT MEETING

Additional Investigation Request

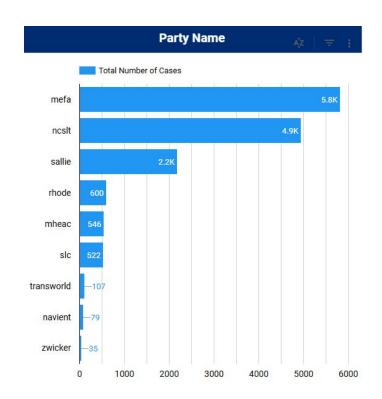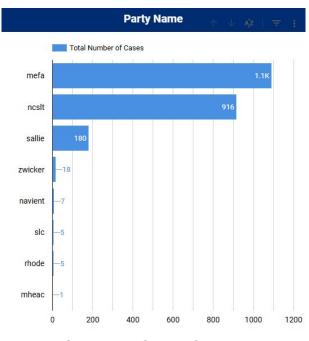Github

MAJOR DEBT COLLECTORS
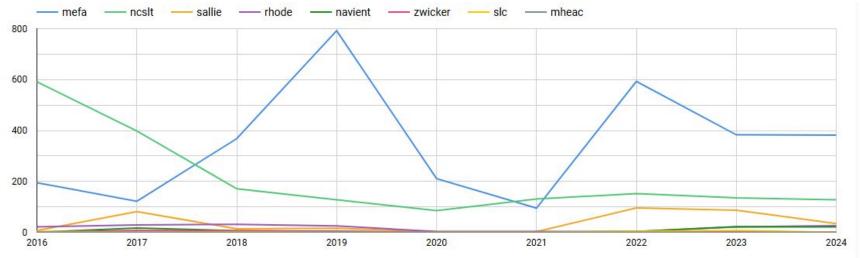
SUPERIOR COURT CASES

Superior Court Cases

- Superior court cases

  ☐ Civil cases involving more than $50,000

- Discovered from database analysis

- Those cases are called '**Contract/Business cases**' rather than '**Civil**'

- On average, 15% of total cases were '**Contract/Business cases**'

- About 19% of MEFA and NCSLT cases were involving more than $50,000

SUSPENDED DUE TO COVID-19

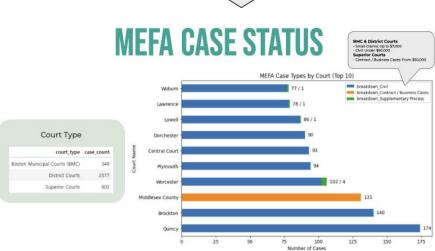Legend: mefa, ncslt, sallie, rhode, navient, zwicker, slc, mheac

- Filed cases were significantly reduced during the Covid-19 period

- Few cases were even suspended due to Covid-19

- However, MEFA and Sallie Mae increased filing cases on 2022, which is before the end of the Covid-19 pandemic

COURTS ANALYSIS

- For each debt collector party we gathered information about the courts the cases were taking place in.
- We gathered information about the court types (Boston Municipal Court, District Courts, Superior Courts). This helps us learn about the amount of money involved in cases.
- We also obtained information about the frequency of cases in the different courts.

Example of court information gathered for MEFA

INTERACTIVE COURT MAP

We created reproducible interactive html maps displaying the number of cases at different courts for different debt collectors.

Run Python Script Inputting:
- Case data from debt collector of interest (can be gathered from a simple query)
- File name of choice to store the html map file
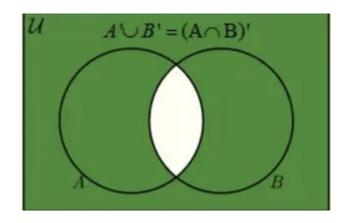
Screenshots from MEFA map

INTERACTIVE DASHBOARD

SCAN ME

NLP MODEL

- **Problem:** Non-standardized naming practices for the debt collectors.

  - Plaintiff party names had spelling errors and variations.

- **Solution:** Developed an NLP Model to find all the name variations for the debt collectors.

  - Used cosine similarity metric and regular expressions to find similar names
  - Leveraged manual filtering to find mutually exclusive sets

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			

CONCLUSION

- Aided GBH in their investigations into other student loan debt collectors such as MEFA, National Collegiate Student Loan Trust, Sallie Mae.
- Created interactive dashboard and court map.
- Developed NLP Model to discover more potential data points