	CPCS: TRIS case management positive youth outcomes
Contact	Ziyad Hopkins, Trial Attorney zhopkins@publiccounsel.net Youth Advocacy Division-Roxbury Office Committee for Public Counsel Services cpcs_TRIS project_code unofficial Team: Yuan Zhang, yuan0331@bu.edu Junyi Zhu, astalos@bu.edu You Peng, youpeng@bu.edu Renjian Zheng, rezheng@bu.edu
Organization	Committee for Public Counsel Services (CPCS), Youth Advocacy Division (YAD)
Organization Description	CPCS provides legal representation in Massachusetts for those unable to afford an attorney in all matters in which the law requires the appointment of counsel. This includes representation in criminal, delinquency, youthful offender, child welfare, mental health, sexually dangerous person and sex offender registry cases, as well as related appeals and post-conviction matters. The present project focuses on young people (up to age 18) accused of committing various crimes. This cohort is represented by CPCS' Youth Advocacy Division (YAD) who work within the Massachusetts juvenile justice system. YAD represents court-involved youth and partners with the Education Law Project (EdLaw) practice that focuses on youth entangled in school-based discipline.
Terminology	• PYD - positive youth development: Background on PYD • "Runsheet" (running sheet): the text notes for a case (ie, not parsed into sql structured fields). Typical term for legal cases

	• "Disposition" - legal term only: innocent or guilty (or something else), not the same as a "positive youth development" outcome • Trial attorney: just like on TV, the attorneys who go to court on a case.
Project Type	Data Science
Project Description	TRIS ("Transformational Representation Information System") is a web-based case management system created to centralize staff activity record-keeping with the hope that the data could then be used to improve our practice and our communities with evidence based analysis. The client would like to leverage the data actually captured in TRIS to further understand the client population and how effectively staff follow a Positive Youth Development ("PYD") model of representation. YAD theorizes that if the lawyers and social service advocates (trained social workers working as part of the defense team) apply the PYD based approach to advocacy, then youth will be empowered to better overcome the structural racism that pushed them into the court system in the first place. In order to better understand our client population, the team will create a heatmap visualization overlaying case client addresses with census tract information. The analysis will seek to create editable filters of various features such as: income levels, race, susceptibility to personal and family crises, educational and healthcare disparities, legal outcomes, charges, individual demographics of the client. Additional external data from the Massachusetts Department of Education could ideally also be included. This heatmap should offer interactivity similar to a Tableau dashboard where data can be filtered. This map should assist the advocates in understanding and addressing the various societal pressures leading to the over-representation of young people of color in the juvenile and criminal justice systems. Additional steps will involve performing text analysis, both natural language processing (NLP) and machine language learning on the case notes (text files called "run sheets" associated with the SQL database),

	including extraction of data from various fixed fields and open notes, to study the attitude of the lawyers and the social service advocates toward the client and the juvenile justice system. The hypothesis is that youth have better outcomes when YAD staff empathize with their clients, including youth anger toward the system-- or in Positive Youth Development terms act as "trusted and supportive adults."
Data Sets	The datasets will be provided by the client as a SQL database to TRIS (and or CSV extracts). The students will query from this as needed. (The data is SQL Server. In order to strip out personal information, CPCS IT will provide it as CSV files.) Census Tracts - MA website link MA Environmental justice dataset MA Department of Education data
Suggested Steps	**Step one:** Obtain the TRIS SQL data, which includes fixed field and runsheets. Case data includes charges, assignments, dispositions and outcomes, certain demographic data, and additional fixed fields that can be leveraged. "Run sheets" comprise free form typed case notes maintained by the assigned attorney and other involved CPCS staff. Also, acquire census tracts for all of MA (Spark has access to this data set). And acquire geographic assignments of the statewide CPCS offices. **Step two:** Preprocessing step. Clean and geocode any addresses in the TRIS data. Determine how to translate legal outcomes and other case information into outcome types for the purpose of this study (including demographic data of clients). **Step three:** Connect the address field for each client to corresponding census tracts. Transform the results onto a Tableau workbook and publish it.

	Step four: Determine the presence and success of social services provided by the social service advocates (SSAs) and/or attorney. Services may be mentioned in the case data and/or in the text notes. The text analysis will apply NLP based on close collaboration with the client on taxonomy and terminology. Also, extract prominent topics using topic modeling and contrast them to cases with limited or no SSA engagement.
Strategic questions	1. Are there disproportionate charge-types that vary based on various geographic areas (county, police department, school district, court area, etc.)? Are there racial or other biases which correlate with various charge types or disproportionate levels of specific charges? Is there a correlation of court involvement with demographic or sociological indicators such as income, housing density, schooling, environmental justice, demographics? Possible regression analysis. 2. What are prominent topics surrounding engagement of social service advocates provided? 3. What are success rates for social services provided in the course of a case? What are traits of cases with successful social service interventions? Identify key features here.
Additional Information	**Tools & Methods** *Data pre-processing:* Pandas, NumPy for processing and cleaning the data. *NLP & Machine Learning:* NLTK, Gensim, sci-kit learn, VADER, TextBlob for word pre-processing, tokenization, sentiment analysis, and topic modeling. *Data Visualization:* Matplotlib, Seaborn, Tableau for all kinds of interactive visualizations

Challenges	Completeness of data is unknown.
	What metrics for empathy with clients
	What metrics for positive outcome
	What keywords in run sheets or datafield data to find the metrics
	What do match with census data