Project Description:

The Better Business Bureau (http://www.bbb.org/) helps consumers find trustworthy businesses. Its mission is to be the leader in advancing trust in the marketplace between businesses and consumers. The BBB sets standards for marketplace trust, encourages and supports best practices in business, celebrates marketplace role models, and calls out and addresses substandard behavior in businesses.

There are many (96) regional BBBs in the United States, Canada, and Mexico. Each operates with local autonomy and with a special focus on meeting the specific needs of their region's consumers and businesses. We will be working with the BBB that is based in the Twin Cities and serves a large portion of the midwestern United States. We will also be working with technical and organizational specialists from the International Association of Better Business Bureaus (IABBB) (https://www.bbb.org/local-bbb/international-association-of-better-business-bureaus). The IABBB is a network hub that supports all BBBs with knowledge, data, and expertise.

BBB is a data company. It is imperative that the data in the system be trustworthy and up to date. Unfortunately, collections of data (especially those as large and changeable as those managed by the BBB) can rapidly get out of date. Data elements such as business URLs, website links, primary contact information, and addresses can rapidly become inaccurate. People visiting the BBB's web pages are looking for information on businesses. If that information is not accurate and up to date, the BBB loses trust in the marketplace.

This is the fourth semester we've done CS projects with the combined client team from the Twin Cities chapter and the IABBB! Last semester two of our CS student teams were very successful in automating both the addition of a considerable amount of new company data to and the removal of "bad" data from the BBB's database.

This semester we'll run one team, which will focus on ensuring that the BBB's data is accurate, complete, and consistent. Data validation and matching against the existing database is a crucial step in maintaining data quality and integrity - so that's this semester's focus!

The project team will use a variety of **data engineering** and **data science** techniques to achieve the desired data validation.

The main client for this project will be Ryan Sharp of the BBB Twin Cities. The team will also work with technical and organizational specialists Eli Johnson and Rubens Pessahna of the AIBB, as well as other colleagues that they may choose to include.

Deliverables	Type of work	Activities	Resources	Tech Skills	Priority
Requirements analysis document that includes identification of desired data profiling, validation, cleansing, and error handling outcomes, as well as desired documentation and plans for long term maintenance.	Requirements analysis, initial potential solution itemization, documentation, approval of client	Interviews with client contacts, “hands on” access to data sources and systems, iterative cycles of client review and commensurate improvement in plans.	Client lead, client experts, client data systems, existing database, previous student project systems and results	SQL (and/or other similar database languages), Python, web scraping packages, regex packages, and possibly various data warehouse or data lake technologies	High
Data quality development/management environment that includes testing tools.	Implementation/improvement of an experiment and test environment that accesses client data and supports development and testing	Design, build, and use a setup that supports both the types of databases used by the client and the building of solutions for types of known challenges with the database	Client staff, client data systems, online documentation of database and programming language technologies	SQL (and/or other similar database languages), Python, and scripting languages to be agreed with client (TBD)	Medium
Core algorithms (in both document and working code forms) that solve targeted data validation and alignment issues	Software development and testing. Use of techniques drawn from data science, data engineering; approval of client	Design, implementation, and test of software modules. Iterative cycles of client review and making improvements.	Client staff, client data systems, online and other descriptions of a variety of data science, data engineering techniques	Python, Python packages for data management, statistical analysis, data connectivity, and data science	HIGH
Documentation of both core algorithms and the **final usable system** delivered both for reuse by client and future student teams	Documentation	Documentation	Word processing tools, access to client	Documentation	High
Final presentation and digital artifact handoverA playable demo and a comprehensive report detailing the development process,	Presentation and delivery of final work output	Presentation and delivery	Client, coaches	Technical presenting, documentation, communication	High