**Analyzing New York State library data for future use**

By Sean Flanagan

**I. Abstract** – The goal of this project was to perform an initial round of exploratory data analysis with the theoretical goal of developing a useful machine learning model to assist libraries and librarians. In a world of limited resources, libraries that provide free services have to make choices about how to allocate those resources. However, without a profit motive, their barometers for success are slightly different; successful libraries are those that are used by their communities. An ML model trained on years of data could forecast trends in usage data to help librarians make decisions on how to allocate resources. This could be as small as recommending which books to acquire, or as big as making staffing decisions for the coming fiscal year. **Impact Hypothesis**: **building an ML model trained on historical data will assist libraries and library systems in allocating resources.**

**II. Design** – The project took two parts: first, I imported the data to Excel and cleaned the data there by removing NaNs, and calculating secondary data columns. I also aggregated the data using Pivot Tables to create summary statistics, and began making basic visualizations. Next, I created a Tableau worksheet to make more sophisticated visualizations. These included histograms, bar charts, scatter plots, and map charts.

**III. Data** – The data consists of ~1,500 rows with about 20 columns. These included measures like Total Operating Revenue, Total Circulation, Visits, and many more. All data was taken from Bibliostat Connect, which collects and aggregates library-collected data from Annual reports going back to 1991: https://www.nysl.nysed.gov/libdev/libs/biblcnct.htm.

**IV. Algorithms**

**-** Data sorting, aggregating, sorting, and filtering with Excel

**V. Tools**

**-** Excel for data cleaning, analysis, and visualization

-Tableau for data visualization

**VI. Communication**

In addition to the visuals and slides, this project will expand on my [personal GitHub page.](https://github.com/SeanMartinFlanagan/EDA-Project)