Nora Keenan

December 8, 2018

**Springboard Capstone Project 1 Proposal**

**Pew Research Center - Libraries 2016**

**Description of Dataset**: This dataset includes 1601 responses to a nationwide telephone survey conducted by Princeton Survey Research Associates for the Pew Research Center in March and April 2016. The purpose of the survey was to assess library utilization across the United States. This survey is one of a series of periodic assessments on the question of library usage conducted for the Pew Research Center.

Surveyors asked, among other things, whether the respondent had visited a library or bookmobile in the previous twelve months, what they did at the library, their opinions on fundamental library services, their internet access/usage, as well as demographic information on age, education, ethnicity, dwelling type, zip code and income.

The quality of the dataset is high. Few data values are missing and the survey was designed to minimize response bias. Respondents are spread across geographic, income, age and education classifications.

**Problem**: Analysis of the questionnaire results by the Pew Research Center found that libraries do a good job of appealing to patrons with higher education and income levels. The intent of my project is to determine what aspects of libraries appeal to those with a high school or lower level of education. By identifying these features libraries can build and expand these programs to improve the education levels and, by extension, incomes of those in their communities. Libraries could also use the results to prioritize funding or approach schools, businesses or other education partners with specific projects.

To determine these features I will begin with an analysis of basic statistics (mean, median, mode, standard deviation) and feature correlation. I will use the data in a predictive model that seeks to accurately categorize the respondents by education level based on demographics and questionnaire responses. I will then more closely look at the features correlated with lower education level respondents and compare them to the typical features for those with a higher education level.

**Possible Project Extension Ideas:** An interesting additional option for this project would be to overlay United States census data with the model results. This analysis could pinpoint the counties and states that would most benefit from specialized library outreach and education programs. I could also try find survey results from other years and determine if the results could be combined. This would provide a larger dataset on which to build the model.