**Texas Public Libraries Statistics 2020 Analysis**

Phase One: Data Exploration

* Downloaded the full dataset from TSLAC <https://www.tsl.texas.gov/ld/librarydevelopments/2021/08/31/2020-texas-public-library-statistics-available/>
* Broke up the workbook into two separate worksheets
  + 2020CentralStatistics -> TX\_PL\_COVID\_Response\_2020 and TX\_PL\_AllStats\_2020
* Removed personal identifiable information and unnecessary information
  + From TX\_PL\_AllStats\_2020: “Phone”, “Telefax”, “Admin Email”, “Library Director/Head Librarian First Name”, “Library Director/Head Librarian Last Name”, “Board Chair First Name”, “Board Chair Last Name”, “Friends President First Name”, “Friends President Last Name”
  + From TX\_PL\_AllStats\_2020 removed “Mailing ZIP+4 Extension”, “Street ZIP+4 Extension”
* Removed branch libraries from TX\_PL\_AllStats\_2020, kept only city and county
* Removed two records from TX\_PL\_COVID\_Response\_2020 that were not present in the TX\_PL\_AllStats\_2020 worksheet: ID 253 Hidalgo County Federated Library System and ID 508 Karnes County Federated Library System
* Feature Engineering:
  + TX\_PL\_COVID\_Response\_2020 – created new columns for Q521.5 “Percentage of year closed”, Q522.5 “Percentage of year limited occupancy”, q526.5 Added E-Library Cards Service during pandemic”, Q531.5 “Average LV Child program attendance”, Q533.5 “Average LV YA program attendance”, Q535.5 “Average LV Adult Program attendance”, Q538.5 “Average viewings of recorded programs”
  + TX\_PL\_AllStats\_2020
* Saved each workbook as TX\_PL\_AllStats\_2020\_cleaned and TX\_PL\_COVID\_Response\_2020\_cleaned