-Imported Data into Spyder as a data frame

With Python:

-Capitalized the title column

-Remove duplicates

-Rows within the zip column that were “.” were changed to null

-Rows within EDU\_NO\_HS, EDU\_HS\_SOME\_COLLEGE, EDU\_BachelorsDegree, EDU\_GradutateDegree that were “.” were changed to null

-Updated the region rows to fit the right zip code if the region row was set to null (some regions are null because the zip code is null)

-Duplicated Title code column and named it title ID

-Two different books have the same title ID (we know their different because themes, genre, reading level, prices, and Lexile codes are different, most errors in channel 1)

-Added a letter to the title id (the duplicate column for Titlecode) to the end of the title id that represent potentially two different books (based off differences mentioned in previous bullet)

Eight changes based on Lexile code (wrote a SQL Query to figure that out)

-If there are two different codes for the same book, we merge the title code so the book only has one code (this is done on the duplicated column title id, we did about sixteen changes (wrote a query in SQL to figure out those changes)

-Created a new column called ID that uniquely identifies each row starting at 1

Made a dataframe called uniqueid that includes:

Index, ID, Title\_ID

-Using the title\_id, created the following dataframes:

Ch1\_Genre: Distinct TitleID( CLONED FIXED one), Genre

Ch1\_theme: Distinct TitleID( CLONED FIXED one), theme

Ch2\_Category: Distinct TitleID( CLONED FIXED one), category

Ch2\_Subcategory: Distinct TitleID (Cloned Fixed one), subcategory

Within each dataframe there are no nulls

-Get rid of clone column (titleid) for every data frame ID is the new unique identifier

-For each data frame, separates the spaces between each column (the whole separation by rows situation)

-Put all the data frames to a csv

-IN SQL, created a new database and added all the CSV files as tables and changing data types and title names.

-Created the schema in SQL Server of how the CSV files connect

-Connect SQL Server Database to tableau