Michael Hissey

11/2/19 – UCB Data Science

ETL Project Proposal

**Data Source:** Education Statistics from World Bank Open Data

* Found via Kaggle – link: <https://www.kaggle.com/theworldbank/education-statistics/download>

**Project Proposal:** Using two CSVs that were part of this download (EdStatsCountry.csv & EdStatsCountry-Series.csv) I will join the two datasets using country code as the primary key. During this process, I will rename many columns from the EdStatsCountry.csv, in addition to dropping many columns. I am joining the data sets so that I can connect SeriesCode (column) from EdStatsCountry-Series.csv to the rest of the data from EdStatsCountry.csv.

**Purpose:** I’ve chosen to perform these functions and operations on education data to see how the United States compares with other countries in terms of educational attainment across demographics & wealth status.