Extract:

Downloaded Kickstarter dataset from Kaggle and downloaded S&P500 historical dataset from Yahoo! Finance with dates collected set as the oldest and most recent Kickstarter dates. Both datasets were in CSV format.

Transform:

Load datasets into Pandas database

* For S&P500 dataset drop extraneous columns and keep only Date, Open and Close columns
* Kickstarter data needed to have launch column split into just the date to match up with dates of the stock index
* Ran into null values due to stock data only being available for days the market was open
* Reindexed stock data to fill in missing dates and used method option to backfill the values
* Merged twice (once on the launched date and the other on the deadline date)
* Removed extra columns

Load:

Convert to list of dictionaries and load to database

* Used to\_json to convert and export to json format
* Used PyMongo and json to load to a MongoDB database