DATA COLLECTION

* Data was collected from kaggle.com using with Python in Jupyter notebook
* The specific Python libraries used were Selenium and Pandas
* Selenium was used in partnership with a webdriver. This allowed for the website to be rendered so that the html code could be scraped from it
* This html code was already in a table format so it was easily added to a Pandas DataFrame
* The DataFrame was cleaned of null values and then saved as a csv

DATA STORAGE

* Data was stored using MongoDB
* To use this online storage system, the Python libraries json and pymongo were used
* To do this, the client is defined (the specific MongoDB account) and Jupyter notebook connects to this account
* Then, a database and collection are defined and the information in the DataFrame is converted to a dictionary format
* Final, the data is sent to MongoDB where it is stored in a specific database and collection as defined by the client





