Query Meisters - Project 2

**Project Proposal:**

 To create a consolidated database with information on books, i.e., bestsellers, bestselling authors, highest rated books, and year.

With over a million new releases each year, the book industry can seem daunting when looking for a good read. However, finding the bestsellers, ratings, and book size can help customers decide if they should start that lengthy novel over their short weekend.

**List of Data Sources:**

* Goodreads https://www.kaggle.com/jealousleopard/goodreadsbooks?select=books.csv
* New York Times API https://developer.nytimes.com/docs/books-product/1/routes/lists.json/get

**ETL Steps:**

* Extract book titles, ratings, genre, and the number of book pages from the Goodreads CSV file.
* Extract data from NYT API based on ISBN
* Create an ERD table by defining primary keys and relative information to join tables.
* Merge data into SQL database.
* Create a descriptive ReadMe file.
* Include images of some of the queries we made throughout the project.

**Team Members:**

* Josh Watson will be the repo master. He will oversee authorizing the Github pushes, organizing the repo, and solving any merge conflicts.
* Amy Castillon will be working on retrieving and organizing data from the NYT API.
* George Vallejo will be working on recovering and cleaning the data from the Goodreads CSV. Worked on merging the two data frames.
* Mindy Garcia will be working on components of PostgreSQL, corrected ISBN-13 errors in the NYT dataset, and the ReadMe file.

All members will have a role in working with the dataset.