**Book Recommendations**

Project 4 Proposal

Elisabeth, Julia, Melissa, Ryan, Brittany

Aims:

* Describe basic statistics regarding books, book reviews and users (readers) captured within dataset.
* Perform exploratory data analysis by book genre (ave length of review, average ratings, #readers); generate visualizations to demonstrate relationships/correlations.
* Based on reader preferences (previous reviews or user inputs), predict whether the reader would like the book or not.
  + Additional output: suggestions for user to choose from
  + Additional output: create user profile to store user preferences
  + Interface to show book covers with book suggestions

Data source: [goodreads]

Other resources:

* kavitareader.com (for source code, API calls to retrieve book covers)
* pytorch.org (source code)

Process:

1. Identify and download csv.
2. Review and clean as necessary using Python Pandas.
3. Perform descriptive and exploratory analysis in Tableau. Create story.
4. Machine learning component:

Requirements:

Use at least 2 other data analytics tools to supplement machine learning.

#### Data Model Implementation (25 points)

* A Python script initializes, trains, and evaluates a model (10 points)
* The data is cleaned, normalized, and standardized prior to modeling (5 points)
* The model utilizes data retrieved from SQL or Spark (5 points)
* The model demonstrates meaningful predictive power at least 75% classification accuracy or 0.80 R-squared. (5 points)

#### Data Model Optimization (25 points)

* The model optimization and evaluation process showing iterative changes made to the model and the resulting changes in model performance is documented in either a CSV/Excel table or in the Python script itself (15 points)
* Overall model performance is printed or displayed at the end of the script (10 points)

\*\* site any source code used from external sites