MAIS202 Deliverable 2

1. Problem Statement
   a. I will make a book recommendation generator using bag of words model and kaggle good books 10k data
2. Dataset
   a. I am using kaggle goodbooks-10k which has information on 10,000 books. I will preprocess data by importing csv file with pandas and accessing columns of information that i need. For example, I accessed title, author, book id, and book tags to generate recommendations.
3. Machine Learning Model
   a. I will use bag of words model. Specifically, I use TfidVectorizer from sklearn.feature_extraction.text and linear_kernel from sklearn.metrics.pairwise to vectorize my data and generate cosine similarity (to calculate numeric similarity) respectively.
4. Preliminary results
   a. The model works and currently has three functions that return a list of 20 book recommendations given a book title as input.
5. Next Steps
   a. I can try to incorporate sentiment analysis
   b. Create a function to recommend authors
   c. Find out how to test accuracy/error of predictions
   d. Integrate model into simple landing page web app