Title: "Capstone Project 2 : Book bundles recommendation from book readers reviews and book sales data"

Proposal

1. What is the problem you want to solve?

I will try to create a book recommendation system which can recommend books to a reader on the basis of the reading history with the help of reviews and sales data.

2. Who is your client and why do they care about this problem? In other words, what will they do or decide based on your analysis?

The recommendation system can be deployed as a web app where people can use to get recommendations based on reading history. Schools can use the system to increase students reading interest, adapts to student needs, moving at a slower or faster pace to help students with different strengths and learning styles to reach their full potential. Online book stores can offer book bundles to their customers.

3. What data are you going to use for this? How will you acquire it?

I will use datasets from:

- Book database: http://www2.informatik.uni-freiburg.de/~cziegler/BX/
- Recommender system datasets: http://cseweb.ucsd.edu/~jmcauley/datasets.html
- Open library: https://openlibrary.org/dev/docs/restful_api
- Worldcat: https://www.oclc.org/developer/develop/web-services/worldcat-search-api.en.html
- Goodreaders dataset: https://www.goodreads.com/api
- Amazon product sales dataset

4. Outline your approach to solving this problem.

- Leverage Python's rich libraries to pull the data via API (in JSON format) calls for analysis.
- Understand the variables in the data

- Clean data to exclude redundant entries
- Exploratory data analysis
- Leverage machine learning methodologies, test multiple algorithms for prediction approach. Select best algorithm.
- Identify challenges with approach / results and provide recommendations on how to improve for future analysis.

- **5. What are your deliverables?**
- Report/Paper
- Codes in Jupiter notebook
- Publish recommendation system in a blog
- Build my portfolio