**Value Proposition**

To justify the cost of initiating a machine learning project of this magnitude, a business should first evaluate the size of the market the project is proposed to enter as well as the potential return on investment it can deliver. These domains in tandem form the value proposition the project presents to the business.

1. Market Size

Goodreads’ consumer market is composed of book readers - both avid enthusiasts and casual participants - who share their reading activity with fellow readers. To understand how much value the recommendation project can capture for Goodreads, we first need to understand the size of the consumer market it is targeting - both in number of participants and dollar value. From there, we can estimate a monetary amount for a given share of the market that the project captures.

We begin by first calculating how many readers buy their books online. In 2021, the size of the US population by age bracket was as follows:

* 18-30 yr. Olds: 65.47M
* 30-49 yr. Olds: 86.29M
* 50-64 yr. Olds: 63.7M
* 65+ yr. Olds: 49.8M

In the same year, the proportion of people who stated they read as a pastime was as follows:

* 18-30 yr. Olds: 83% (54M)
* 30-49 yr. Olds: 77% (66M)
* 50-64 yr. Olds: 72% (49M)
* 65+ yr. Olds: 68% (39M)

We can thus surmise that the total number of Americans who read is 208 million.

In the same year, the distribution of online book sales by age category was as follows:

* 18-30 yr. Olds: 32.4% (17.6M online buyers in this age bracket)
* 30-49 yr. Olds: 32.4% (21.5M online buyers in this age bracket)
* 50-64 yr. Olds: 32.7% (15M online buyers in this age bracket)
* 65+ yr. Olds: 34.9% (11.8M online buyers in this age bracket)

Combining the reader age proportion data with the online book buyer age data, we can estimate the quantity of online book buyers by age as follows:

* 18-30 yr. Olds: 17.6M online buyers in this age bracket
* 30-49 yr. Olds: 21.5M online buyers in this age bracket
* 50-64 yr. Olds: 15M online buyers in this age bracket
* 65+ yr. Olds: 11.8M online buyers in this age bracket

In total, 65.8 million Americans bought their books online, accounting for 32% of all readers. To date, there are about 100 million users on Goodreads. If the proportion of users that buy their books online follows the proportion of the general public, then about 32 million Goodreads users buy their books online. We can then assume that the remaining 68 million users still buy their books in physical stores. This is the size of the Goodreads’ recommender’s target market.

1. Market Share Value

Now that we know how big the market Goodreads is entering is, we can layer on the data related to online books sales to measure how much a share of the market is worth.

In 2021, the revenue from online book sales was $8.8B USD in the United States. Hardcover and paperback books accounted for 29.6% and 26.7% of that revenue respectively. Amazon’s share of revenue from the Internet sale of printed books was about $4B USD.

That same year, the average annual spend on books was approximately $114 USD per capita. This means that 35 million buyers bought books from Amazon in 2021 ($4B/$114).

If the recommender manages to capture 1% of the GoodReads target market over the next 3 years and gets 680,000 users from the 68 million of physical-store book buyers to purchase online via Amazon (its parent company), then it would add $77M to Amazon’s print books revenue - an increase in revenue of 2%.

1. Return on Investment

In order to the true value a project can deliver to a business, its benefits need to be measure against its cost. This is what’s known as return on investment (ROI).

We know that the recommender can generate $77 million in additional revenue should it capture 1% of the market share. We must now estimate how much it would cost Goodreads to build a recommender that can reach that market share.

Goodreads’ cost of building and operating our recommender is based on the pricing model for the *Amazon Personalize Recommender System*. The first step is estimating how much data the recommender needs to store.

In a use case presented by Amazon, the company describes an example where an online retailer provides recommendations to customers based on their purchase history and interests.They state the retailer has 50,000 users, resulting in 10 GB of data to be stored. Although this is a hypothetical example, the business model described therein is highly comparable to that of our recommender project. Therefore, we’ll assume a linear relationship between the number of users whose history is stored and the size of the dataset in which it is stored, quantified by the ratio of 10GB per 50,000 users.

As previously mentioned, there are about 100 million users on Goodreads. Since the recommender project will be available to all users on the platform, it must be able to hold that many users. Extrapolating the ratio to 100 million users, we conclude that the database that serves as the foundation for the recommender must be able to hold 20,000 GB of data.

Now that we have a realistic approximation for the size of the recommender’s database, we can use the *Amazon Personalize Recommender System* pricing model to approximate a monetary cost to operate and maintain it.

**Cost of the recommender**

Based on pricing for Amazon Personalize Recommender System

Monthly cost of storing 100 million users: 20,000GB\*$0.05/GB=$1000

Monthly charges for 100 million users

First 10,000 users: $0.375 per hour \*732 hours per month= 274.50

Next 900,000 users: $0.045 per hour \* 732 hours per month=32.94

Next 9 million users: $0.018 per hour \*732 hours per month=13.17

Next 90090000 users: $0.005 per hour \*732 hours per month=3.66

Total= $324.27

**Total Cost**: 1000+324.27=$1324.27=~$16,000\*3 years= ~$50,000

With both the cost and revenue generated by the recommender know, we can now estimate an ROI metric to quantify the value proposition of our project.

ROI=($77M-$50,000)/$50,000= ~1500%