For Project 6 and my Final Project, I will work on the Book Recommendation Dataset from Kaggle, which contains over one million user ratings for books. I have not worked with this dataset previously and it has a large amount of ratings. I will focus on recommendation algorithms that work off the ratings matrix in the interest of time; however, should I have success in implementing these algorithms, a potential further step would be to scrape book descriptions to use for content-based filtering or grabbing image covers and using the AI resources on Azure to extract features from images.

I will store this dataset on Azure via Blob Storage. I will then use an Azure Virtual Machine and Azure's built in machine learning capabilities to analyze the data and implement machine learning algorithms from the recommenders library to train and evaluate a recommendation system. Specifically, I will implement a global baseline model, an ALS model via Spark, and an SVD model from the recommenders library. Given the size of the dataset, using Spark will be necessary to avoid significant run-time and compute resources. I will practice on a small subset of the data before implementing on the full file. Lastly, for network security, the Azure Blob that stores the data will have a private endpoint. The ultimate goal is to create a recommender that can most accurately predict user ratings for items. Accuracy will be measured via error metrics such as RMSE and MSE.