Domain Background

The selected project is the **Starbucks Project**, one of the three default options available. This project focuses on determining which type of offer—Buy One Get One (BOGO), discounts, or informational messages—should be sent to a list of customers based on synthetic demographic and transactional data. The reason I picked this project is because I don't have experience building recommendation systems and could be a good chance to learn how to build one.

About historical information from Starbucks:

Company Background: Starbucks was founded in 1971 in Seattle, Washington, initially as a retailer of whole bean and ground coffee. Over the years, it has transformed into a global coffeehouse chain, renowned for its specialty coffee drinks and unique customer experience. By the late 1990s, Starbucks began an aggressive expansion strategy, opening stores across the United States and internationally (Michelli, 2007).
Market Trends: The coffee industry has undergone significant changes since Starbucks' inception. The rise of specialty coffee in the 1990s led to increased consumer interest in high-quality coffee. Additionally, the trend towards sustainability and ethical sourcing has influenced Starbucks' practices, resulting in initiatives such as the Coffee and Farmer Equity (C.A.F.E.) Practices (Starbucks, 2021).
Previous Research: Numerous studies have analyzed Starbucks' business model, focusing on its customer loyalty programs, brand positioning, and marketing strategies. Research indicates that Starbucks has successfully created a "third place" environment, fostering customer loyalty through a distinctive in-store experience (Michelli, 2007; Schmitt, 2010).
Technological Advances: Starbucks has embraced technology to enhance customer experience, introducing mobile ordering and payment systems in 2015. The Starbucks app has become a key tool for customer engagement, allowing users to earn rewards and customize their orders (Starbucks, 2021).
Competitor Analysis: Starbucks faces competition from various coffee chains and independent cafés. Its ability to adapt to changing consumer preferences, such as the demand for healthier options and plant-based products, has been crucial in maintaining its market leadership (Smith, 2020).
Customer Engagement: Starbucks employs a variety of marketing strategies, including seasonal promotions and community involvement initiatives. The company's emphasis on creating a personalized customer experience has been a significant factor in its success (Michelli, 2007).
The reason I chose this project is that I lack experience in building recommendation systems, and this presents a valuable opportunity to learn how to create one.

References:

- Michelli, J. A. (2007). *The Starbucks Experience: 5 Principles for Turning Ordinary Into Extraordinary*. McGraw-Hill.
- Schmitt, B. H. (2010). *Customer Experience Management: A Revolutionary Approach to Connecting with Your Customers*. Wiley.
- Smith, A. (2020). *The Coffee Market: Trends and Opportunities*. Journal of Business Research, 112, 123-134.
- Starbucks. (2021). *Global Environmental & Social Impact Report*. Retrieved from **Starbucks**.

Problem Statement

According to the project definition provided by *Udacity*: the basic task is to use the data to identify which groups of people are most responsive to each type of offer, and how best to present each type of offer.

Solution Statement

The problem can be solved using a recommendation system like the Matrix Factorization and would be developed under a standalone system like the provided workspace by *Udacity* or my personal computer.

Datasets and Inputs

The proposal uses three datasets provided by *Udacity*, it includes data about customers (profile.json), offers (portfolio.json), and transactions (transcript.json), described as follows:

profile.json: Rewards program users (17000 users x 5 fields)
- gender: (categorical) M, F, O, or null
- age: (numeric) missing value encoded as 118
- id: (string/hash)
- became_member_on: (date) format YYYYMMDD
- income: (numeric)

portfolio.json: Offers sent during the 30-day test period (10 offers x 6 fields)
- reward: (numeric) money awarded for the amount spent
- channels: (list) web, email, mobile, social
- difficulty: (numeric) money required to be spent to receive a reward
- duration: (numeric) time for the offer to be open, in days
- offer_type: (string) bogo, discount, informational
- id: (string/hash)

transcript.json: Event log (306648 events x 4 fields)
- person: (string/hash)
- event: (string) offer received, offer viewed, transaction, offer completed
- value: (dictionary) different values depending on event type
- offer id: (string/hash) not associated with any "transaction"
- amount: (numeric) money spent in "transaction"
- reward: (numeric) money gained from "offer completed"
- time: (numeric) hours after the start of the test

The importance of these datasets comes from understanding, 1) the offers and their components, 2) who are the target customers and 3) historical transactions, by combining these three aspects we would be able to understand who (customers) are more likely to react to a specific situation (a type of offer) according to historical behavior (transactions).

Benchmark Model

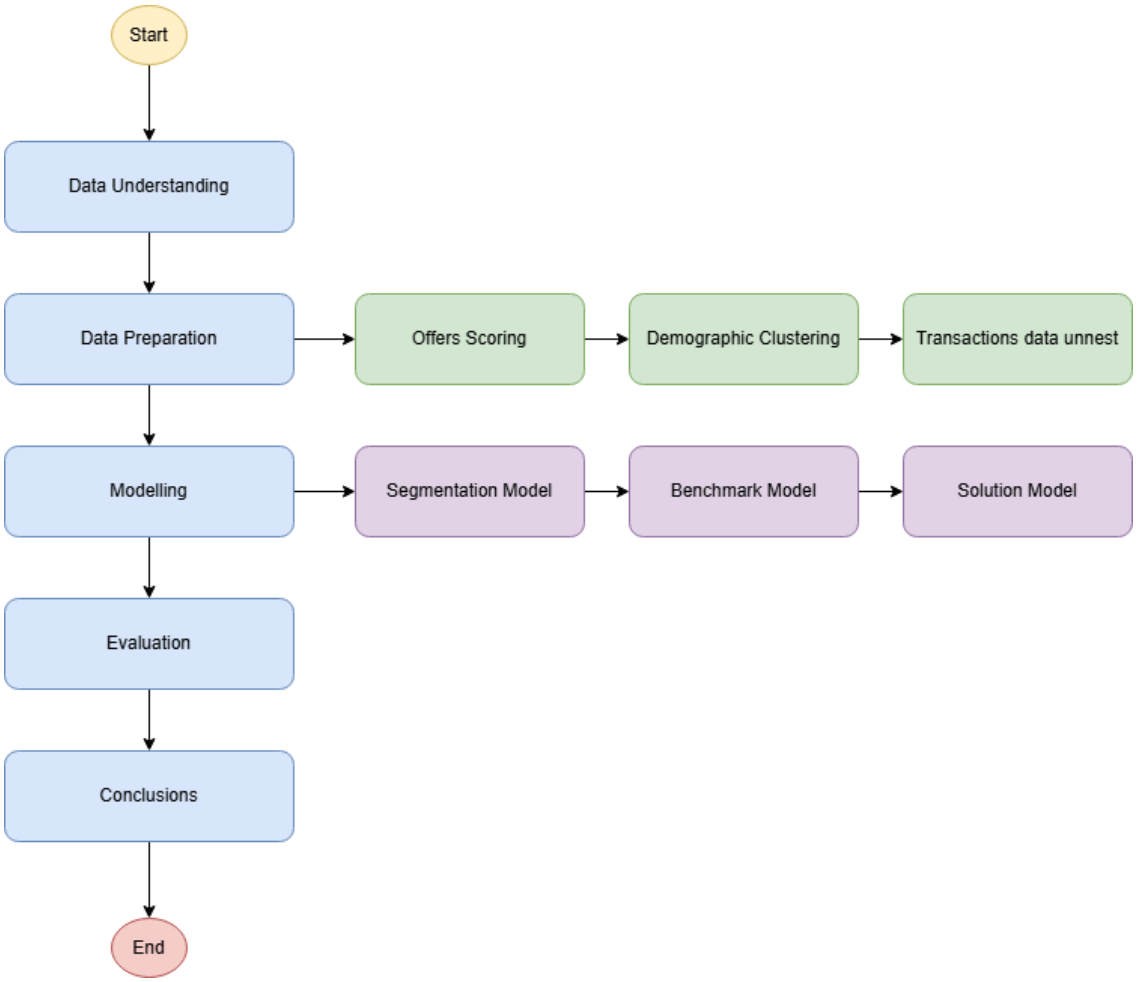
A basic collaborative filtering model.

Evaluation Metrics

Accuracy Metrics: Precision, Recall and F1-Score
Error Metrics: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)
Ranking Metrics: Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG)

Project Design

It is a five-step process that incorporates select elements of the CRISP-DM framework, with specific adaptations outlined as follows:



Data Preparation: one field to represent the data for offers and the demographics data in one variable, like the score for the offers in terms of how the benefit is for the client with a 0 - 100 score and the demographics with a cluster using a k-means model.

Modelling: 3 models, 1) a segmentation model to understand patterns 2) the benchmark, and 3) the solution model.