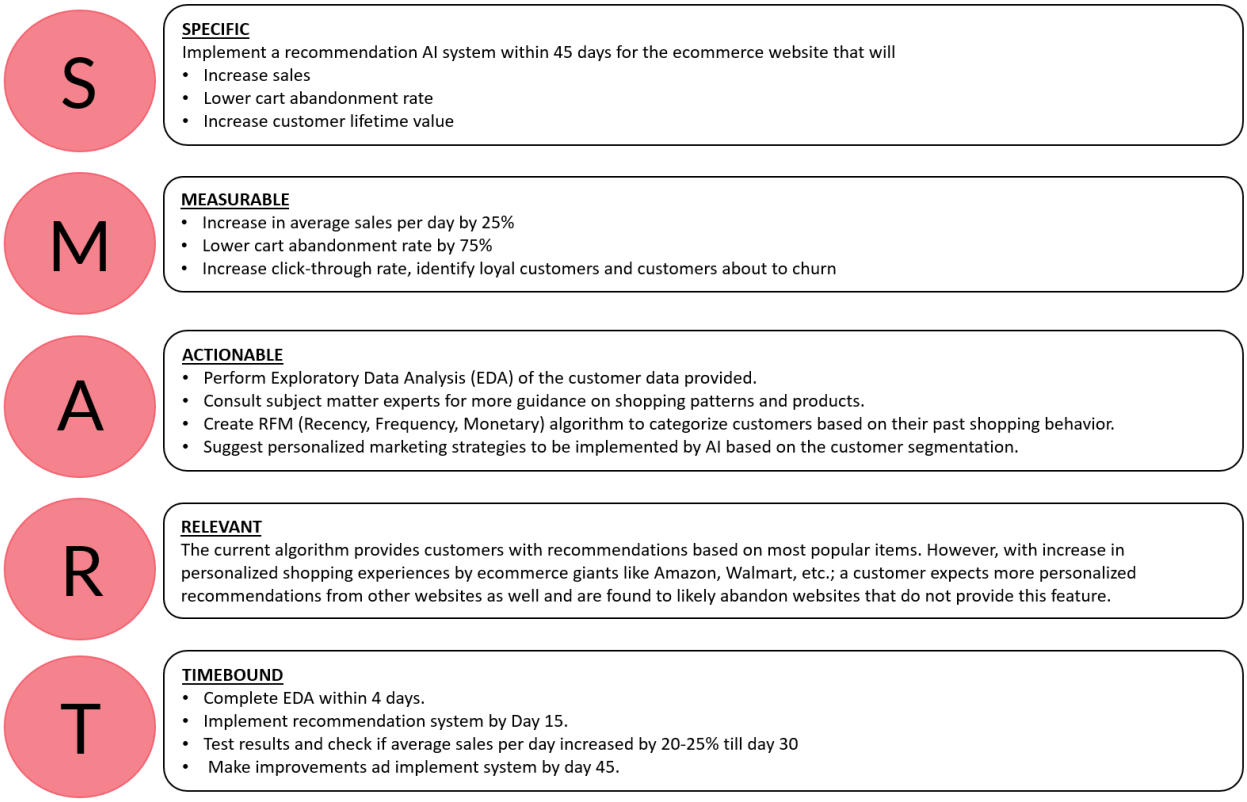
Final Assignment Part I

Please refer to the .html file for answers to questions 1 & 2.

3. ***Document the problem you will address using the SMART framework – use hypothetical data and goals for the improvement your Recommendation Engine will deliver.***

Business Problem: The consulting team is to design and build a Recommendation system that uses AI to categorize customers based on their past shopping patterns and accordingly market products to them to increase sales and customer lifetime value on the ecommerce website. This system is to be tested and proved to show more revenue against the current algorithm used by the website that doesn't provide a personalized shopping experience.



4. ***Select from a pool of titles and roles within the company to create a core team (Maximum 6-8 members) to perform the analysis and develop the pitch to your CMO. (Pool: VP Finance, Data Engineer, SVP Data Scientist, Financial Analyst, Customer Satisfaction Manager, Website Analyst, Webmaster, VP Marketing, Performance Marketing SEM/SEO Analyst, Customer Retention Manager, Marketing Manager, Customer Research Analyst, Data Visualization Specialist, Sr. Data Scientist, Marketing***

Analytics Manager, IT Manager for Ecommerce Data Storage, Marketing Messaging/Creative Designer, Ecommerce Financial Manager)

The team will consist of the following roles:

- SVP Data Scientist
- Customer Satisfaction Manager
- Marketing Analytics Manager
- Financial Analyst
- Sr. Data Scientist
- Data Engineer
- Marketing Messaging/Creative Designer

5. ***Describe the team strengths and discipline focus areas that justify your need for each person and why this talent will be necessary for your success. This justification is intended to both motivate the team members and gain support from their managers to join your team.***

- SVP Data Scientist – Use analytical data to develop a recommendation system. Analyze project risk, provide leadership for analysis and translate analytics into actionable marketing campaigns.
- Customer Satisfaction Manager – Come up with loyalty programs, interact with customers to understand their shopping journey and explain their issues. Come up with programs/deals to convert customers who will churn.
- Marketing Analytics Manager – Conduct research on current and future marketing campaigns and check their effectiveness.
- Financial Analyst – Provide information of current performance of the ecommerce website, identify budget and projections of sales for implementation of new marketing strategies.
- Sr. Data Scientist – Collating and cleaning data collected. Collaborate with SVP Data Scientist to develop recommendation system.
- Data Engineer – Deploy and maintain the system and the data pipelines.
- Marketing Messaging/Creative Designer – Design landing pages, social media and other marketing banners/advertisements to increase click-through rate.

6. ***Choose either a "supervised" or an "unsupervised" approach to segment/cluster current customers using the data provided. You may develop a rules base algorithm to ID a customer on each new transaction and assign them to one of your segments/clusters. Create sample data (5-10 records) with synthetic data to show how a new transaction will be scored and assigned to a segment/cluster. (You can use RFM<Recency, Frequency, Monetary Value>, K-Means or other rules based approached. Your segmentation must be MECE (Mutually Exclusive, Comprehensively Exhaustive) and any record of any incoming***

customer must receive one score and become assigned to a single segment – even if the transaction is incomplete or abandoned prior to purchase.

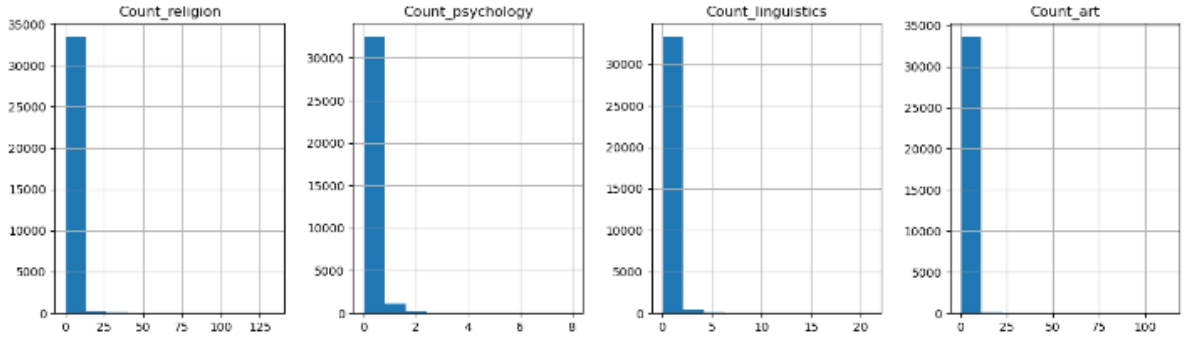
METHODOLOGY

1. DATA PRE-PROCESSING:
    a. There are 33713 rows of customer data and 69 columns.
    b. Dropped the columns which have null values - 'Unnamed: 66','Unnamed: 67', 'Dear Greg,'
    c. Dropped column 'logtarg' as it consisted of 75% null values and it was unsure what it represented. Didn't seem to be of significance for the model.
    d. Checked the data types of the remaining columns and found that all are numerical and none are categorical.
    e. Renamed the columns:
        i. Start with F – replaced 'F' with 'Count_' as it represents frequency of sales in each category.
        ii. Start with M – replaced 'M' with 'Sales_' as it represents monetary value of sales under each category.
        iii. Replaced the digits in each column name
        iv. Renamed 'r' – 'Recency', 'f'- 'Frequency', 'M'-'Monetary' and 'tof'- 'Time on File'
2. EXPLORATORY DATA ANALYSIS:
    a. Based on external research and the data given, the below variables were identified as the business drivers:

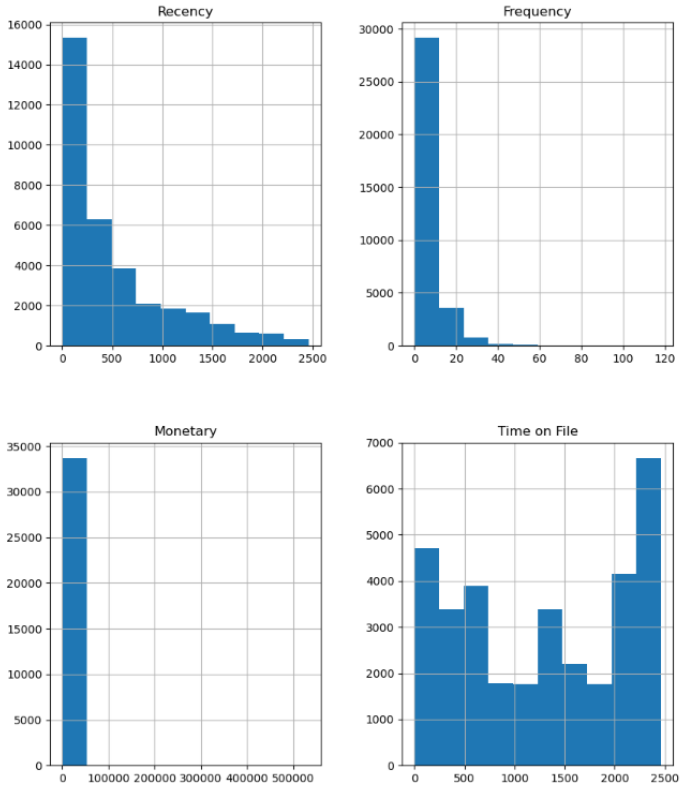
INTERNAL FACTORS	EXTERNAL FACTORS
Staff, merchants	Customer preferences
Technology used for website	Location
Marketing strategy	Device used for online shopping
Customer retention programs	Customer Age
Logistics of delivery	Time of transactions
	Access to advertisements

    b. Based on external research, it was found that companies like Amazon increased their sales over the year 2021 by ~31% with the help of their recommendation system. Based on research performed by Accenture, a product is 75% likely to bought if it was a recommendation based on personal preferences and analysis of past shopping behaviour. As data, AI and technological solutions are being implemented, a customer's shopping journey is also rapidly changing but the traditional system of recommendations is a constant. For example, before ecommerce was popular, an item would be bought based on word of mouth reviews and recommendations from family and friends that would have reached a person based on their preferences. This logic continues to apply now with ecommerce websites and produces amazing results. Thus, the more personalized a journey, the more interesting it is for the customer.

c. Plotted histogram for each variable in the data. As can be seen, as each row is a representation of a customer across the various ecommerce categories, each histogram is showing the count for 0 to be maximum. It is observed that the top 5 book categories frequently bought are: maps, games & riddles, politics, economy and psychology. While the top 5 book categories with best sales are: games & riddles, economy, science, computer science and philosophy. This may be due to these books being more expensive and since these seem like books used by scholars they may be their general customers.



d. Plotting only the Recency, Frequency, Monetary and Time on file values in a histogram. It can be seen that recency, frequency and monetary are extremely skewed. The values with 0 may represent new customers that have just joined the website but haven't used it yet.



3. <u>RFM MODEL (taken up the unsupervised approach):</u>
   a. The recency, frequency and monetary values are categorized based on a quartile scale of 4 with 4 being the best on the scale.

b. The scores for the above three columns are then concatenated to give segments like 444 which represents the best customer for the website or X1X represents a non-frequent customer. However, on checking it is found that there are 61 such unique segments.
c. To make it easier for categorizing the customers, the score is summed up and are given the following level based on this website:

Score	Level	Definition
12	Best Customer	These customers have recently been checking the website, are frequent customers and regularly buy the products.
9-11	Can't Lose Them	Customers that are either frequent or spend a lot or been recently scrolling through.
7-8	Champions	Not the best customers but they are loyal customers who aren't spending much
6-7	Potential	High potential to enter loyal customer segments
5-6	Promising	Showing promising signs with quantity and value of their purchase but it has been a while since they last bought something
4-5	Needs Attention	Made some initial purchase but have not seen them since.
1-4	Require Activation	Poorest performers of our RFM model. They might have went with our competitors for now and will require a different activation strategy to win them back.

4. INSIGHTS:
   a. With the help of the RFM algorithm in combination with a Market Basket Analysis using Apriori Algorithm, the customer can have a more personalized

shopping experience. The RFM model will help the ecommerce website identify which kind of customer is currently being dealt with and using the apriori algorithm identify which book category they prefer and accordingly provide a discount or loyalty program r advertisement/notification as required. These can be shown on each page under the 'Recommended for You' tab.

b. While the original algorithm being used that suggests the most popular item to customers need not be scrapped, instead it can be recommended on each page under the tab 'Most Popular Now' or 'Can be bought with' to show bundles that may interest the user along with their preference.
c. The above recommendations can also be shown on the cart page to interest the customers even more.
d. Based on the RFM customer segmentation, the marketing strategy can be tuned to each customer.

Level	Marketing Strategy
Best Customer	Develop and send out loyalty programs like gold, silver and platinum customer programs that provide discounts or freebies.
Can't Lose Them	Frequently send out notification or mails with discounts based on their preference. Send advertisements of popular selling items within their preferred category.
Champions	Frequently send out notifications, mails, newsletters about benefits of being part of their loyalty programs or using their discounts.
Potential	Send out personalized advertisements and discounts over mails and notifications especially during times of sales or peak shopping periods like Thanksgiving.
Promising	Similar strategy to Potential customers but at a higher frequency.
Needs Attention	Send out a discounted loyalty program offer or free trial runs over mails/notifications.
Require Activation	Since they may have gone to competitors, perform an analysis on the general products liked by these customers and conduct a competitive market research on these categories to identify what is being offered to them by the other stores. Accordingly, develop

	advertisements for them to be sent by mails.