***Web analytics assignment***

**1. An ecommerce client wants to have a better understanding of the combinations of products that their users are buying. This is to allow them to better plan for promotions or product bundles. This information could also be used in their supply chain for product sourcing / procurement. How can you help them to analyze their customers' "basket"?**

*- How you would solve the problem?*

Market Basket Analysis technique can be used.

- *What data points do you need?*

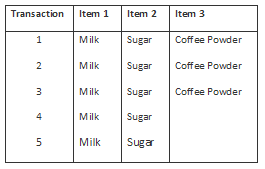
For example, in a grocery store we need the variables like Invoice number, Stock Code, Description, Quantity, Invoice Code, Unit price, Customer ID, Country.

*- What algorithm(s) you would use and why?*

Apriori algorithm and Association rules is the algorithm to do Market basket analysis and help in analyzing buying patterns of customers.

*- How the client can use your solution in their business?*

The client can use this for Cross selling. Consider the below table.



For this dataset we can write the following association rules: (Rules are just for illustrations and understanding of the concept. They might not represent the actuals).

**Rule 1**: If Milk is purchased, Then Sugar is also purchased.

**Rule 2**: If Sugar is purchased, Then Milk is also purchased.

**Rule 3**: If Milk and Sugar is Purchased, Then Coffee powder is also purchased in 60% of the transactions.

A Marketer can build crossing selling strategies for wooing the remaining 40% of the customer (who did not buy coffee powder) to buy Coffee powder along with Milk and Sugar. He might bundle all these products and give at discounted price.

A retail shop manager can stock same amount of sugar and coffee powder in his shop. He can place these items optimally in a store for the ease of picking by the customer.

*- How you would explain your solution in "layman's terms"*

Wikipedia defines Cross-selling as “an action or practice of selling an additional product or service to an existing customer”. While acquiring new customers has become a costliest way (and often deadliest) to increase the revenues, it has become imperative for the businesses to enhance their growth potential from their existing clients.

**2. A publisher wants to increase their advertising revenue by having different ad rates based on a user's age or gender. But they are not able to measure how many users they have for each age/gender bracket. Of all of their website's users, only about 1-2% of users are logged in, from which the publisher can reliably know their age/gender (from their membership database). How can you help them to better identify the age/gender of the remaining 99% of users, such that the publisher can target each group with different advertisements at different ad rates?**

*- How you would solve the problem?*

a) Clustering of User engagement pattern and Pattern Matching or

b) Recommendation Engine - Collaborative filtering

*- What data points you need?*

Age, Gender, other variables available in web log data

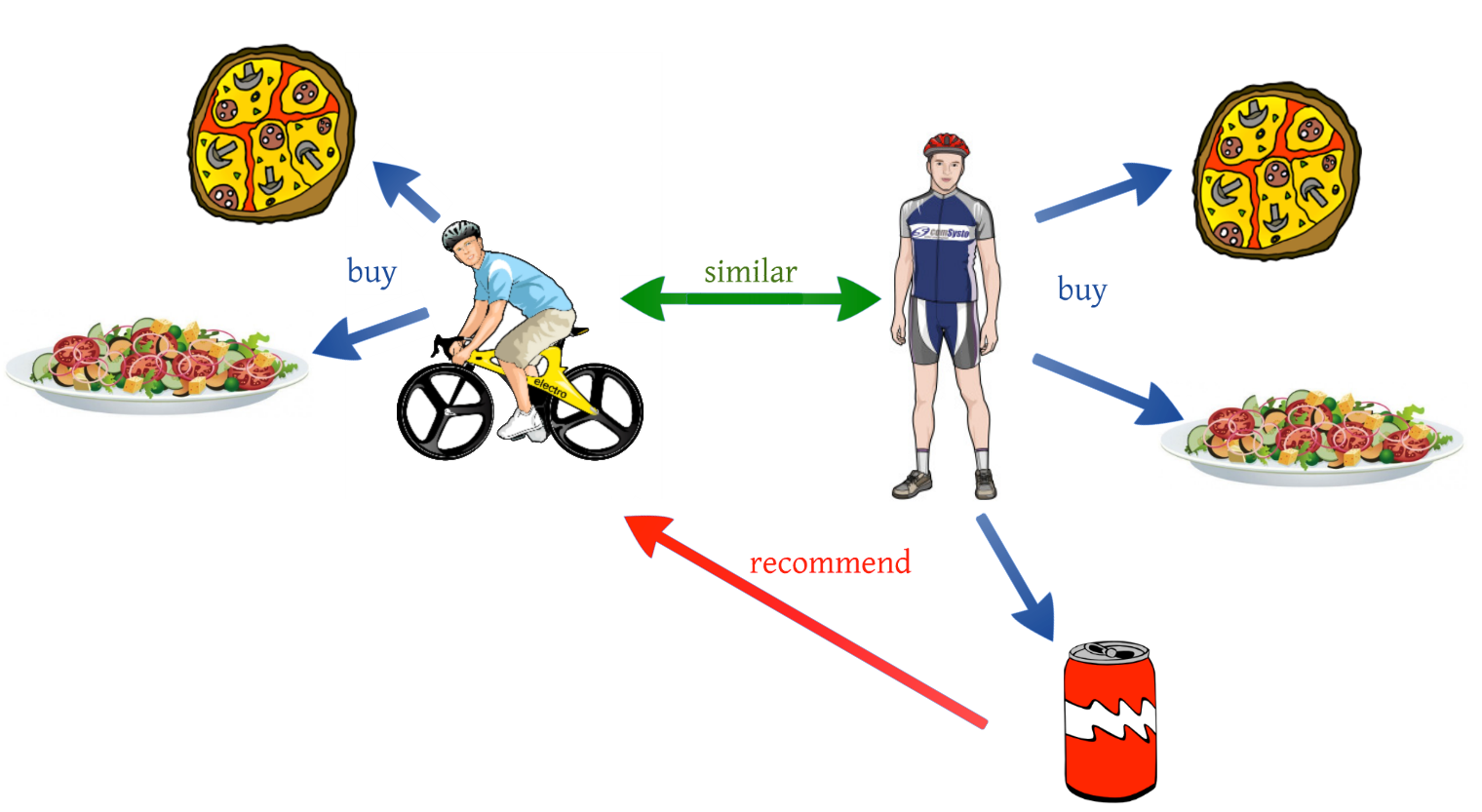
*- What algorithm(s) you would use and why*

1. Collaborative Filtering – It considers both the users and product for finding similarity whereas Content based filtering uses only the products to find similarity. So better to use Collaborative filtering here since we are going to find similarity between users based on age or gender.
2. Agglomerative Clustering and Pattern Matching – Since we have only 2% of target data available, it is better to go with unsupervised learning technique(clustering). Once all the users are clustered, for example let us say we have 5 cluster. We already know the reliable age and gender for 2% data with target variable. Now we do pattern matching between record with target variable and the record without target variable (records in clusters). In this way we can predict the age and gender of all records inside the clusters.

*- How the client can use your solution in their business*

Once the age group or gender of all the users visiting the website is predicted. They can be targeted with relevant advertisements.

- How you would explain your solution in "layman's terms"



Let us consider boy on cycle as Person A and boy standing as Person B. From the diagram, Person A likes Pizza and Pasta and Person B likes Pizza, Pasta and Cool drink. What they like commonly is Pizza and Pasta. So they are similar users and now we can recommend Cool drink to Person A since he is similar to Person B.