# Paritosh Sinha

# Answer 1

**Machine Learning Unsupervised Learning - 2 Individual Assignment**

List of problems which Bigbasket is trying to address are: -

Bigbasket being a customer centric company it is looking for maximising customer satisfaction which will in return benefit them as well.

The list of issues in Bigbasket’s basket are as following

1. Bigbasket being an online platform of providing household daily needs items. So, when customers are trying to access Bigbasket using phone its interface is posing challenges to the user. Since it is a website driven platform because of which it is tedious to explore the same desktop version website on phone.
2. Browsing website on phone to order from Bigbasket is very challenging as of 2016 it is providing range of 18000 products which might exponentially grow in its offering by now. Considering large variety to explore from and then selecting what to buy is a very herculean task.
3. Bigbasket being a grocery selling platform customers place order for several products, sometimes as high as 80 in one order depending on their purchase frequency and size and demands of individual household’s needs.
4. A few customers buy all their groceries frequently like every other day to once a week and there are customers who would place order once a month. For frequent buyers who might order same thing over and over again like milk, bread etc. As their current interface on logging in shows all the variety of products which will make it tough to again navigate the same set of products again and again.

Monthly buyers who buy their ration once in a month, who tend to have a big shopping list which can range from 80-100 products for a medium to large sized household. And using a phone to make such an order can become very challenging.

1. Customers frequently forgetting to order certain items, resulting in either placing additional orders or purchasing the forgotten items from local stores. Frequent ordering impacts logistics and supply chain costs of Bigbasket.

Customers may start preferring to buy from nearby stores which might impact loyalty and may lead to smaller basket size and worst-case may lead to customer churn.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Answer 2

Fundamental differences in recommender system requirements between Bigbasket and other eCommerce companies:

* 1. **Repeat purchases of daily use items**: *Bigbasket* focuses on providing daily use items like vegetables, bakery and dairy products, daily use household products and high-end branded food items. The purchasing behaviour of customers in this category involves repeated purchases of the same products overtime, rather than exploring the wide variety of different items So here recommender system needs to balance a lot between suggesting the frequently ordered product and suggesting niche or new range of products.

For *Amazon and Flipkart* purchases here are less frequent for customers. As it caters to different needs of the customers so here recommender system can experiment and suggest products which are similar to customer’s buying patterns as well as some niche product based on customers browsing pattern, searches on social media etc.

* 1. **Delivery timelines:** for *Bigbasket* being dealing with perishable goods it has to provide shortest possible delivery date and time options to the customer so recommender system should take that into account availability of the product before suggesting a particular product to the customer.

While customer ordering from *ecommerce* site has a relaxed mindset while considering buying the products so the recommendation engine will have wider variety of options to suggest from.

* 1. **Type of product difference**: *Bigbasket* being a platform providing grocery items which are necessity it has to stick with what its major product range is so it can experiment in the realm into which it current market demands revolves in so recommender system should consider that very well whereas range of *Amazon and Flipkart* being more diverse in their offerings which can be considered as non-

essential at times as well so experimenting and introducing new range might not harm them.

* 1. **Basket size and mobile ordering challenges:** *Bigbasket’s* customers often have huge basket, sometimes resulting up to 80 items in a single order. This poses unique challenges when customers use mobile handsets to place orders due to limited screen space and potential usability issues so recommendation system should provide suggestions which it find most important for the customer in such a small space is what will give Bigbasket cutting edge advantage in the market, better the predictive analysis higher the customer satisfaction whereas basket of *Amazon and Flipkart* will not the that varied as will range from as low as 1-2 products may be maximum up to 15 products and Customer tend to browse a lot before making purchases on ecommerce sites as customer tend to read, explore options, reviews, compare prices

etc before making purchases on these site so recommendation engine should consider exploratory nature of customer in mind and then suggest.

* 1. **Context-specific recommendation**: *Bigbasket's* "Did you forget?" feature is specific to the situation where customers have already filled their baskets and are ready to check out. The product recommendations provided at this stage are based on the customer's basket contents and purchase history, rather than considering buying history or patterns of other similar customer profiles. The focus is on addressing the immediate need of the customer during the checkout process.

For *Amazon and Flipkart* it will be more of recommending based on user-to-user collaborative filtering or based on other factors like temporal, seasonal or on the basis of browsing history.

* 1. **Logistical constraints:** speaking of recommendation system doesn’t need “did you forget” for *Amazon & Flipkart* because the item may not be dispatched from same place and the delivery date might differ where in *bigbasket* we should not recommend something which is not present in the nearest hub also or going to get out of stock before the customer orders.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Answer 3

**Recommendation technique(s)**

Technique varies from user type to user type so following can be two broad categories of customer.

## New customer

When a new customer gets onboarded, we have to present the customer with an initial product suggestion. In order to do that we can follow the following approach.

At the beginning we can launch a questionnaire 8-10 radio button set of questions like eating preference, guilty pleasures, sweet tooth, age group, family size etc. and on filling the questionnaire Bigbasket can give an introductory offer or coupons based on their preferences as filled in the questionnaire.

Post that we can give suggestions on the basis of–

1. **Temporal recommendations –** Recommendation engine must suggest product to the new user with respect to the time and ordering pattern of other user at the time of the day when this new user is trying to browse. Like if the customer has logged in at the start of the month than the chances are customer is looking forward to stock the ration.
2. **Seasonal recommendations –** Recommendations based on time of the year like holidays festivals etc. if it a Diwali season then probably recommending sweets, gift hampers are a good option.

**Cluster based recommendations –** Once the customer fills in the questionnaire on the basis of that along with various demographic features we can assign the customer to a cluster and make default list of products which can be recommended with respect

to the cluster. Suppose on filling the details of the form Customer falls in the age range of 25-28 years, Single and a calorie watcher then putting him in the appropriate cluster and suggesting him/her sprouts, fruits etc will be a sensible choice as he tends to stay fit.

1. **Browsing history –** if possible, we can access browsing history of the customer and then form the recommendations based on that.

Once the user put some product in search then the recommendations can be modified according to what has been put in search so like suppose if user is looking for an Ice-cream, then using Item based collaborative filtering recommendation system can suggest the variety of ice-creams which can be categorised on the basis of dietary preference, budgeting etc.

## Old Customer

For old customer of Bigbasket, one technique will not suffice for recommending so a combination of recommendation techniques would be more appropriate to cater to the specific needs of their customers.

1. **Collaborative Filtering**: Collaborative filtering is a commonly used technique that analyses user behaviour and their preferences to make recommendations.

In the context of Bigbasket, collaborative filtering can be effective in suggesting products based on the purchase history and preferences of similar customers. It can help customers discover new products that align with their tastes and preferences. This can be used for product-to-product collaborative filtering as well, when one product is out of stock then system can recommend a similar product to the customer. Challenge in this is huge dataset where product range is high – so attributes for each product should be filled correctly and missing data in this case might lead to incorrect mapping

Whereas User to User Collaborative Filtering is computationally very heavy considering a huge customer base so all these computation and similarity score calculation done before hand not real time. Better way is to cluster users on the basis of characteristics and demographics of the customer.

*For instance,* recommending Bread of different brands to the same customer whereas for User-based suggestions, if User1 buys new variety of niche product then User2 who is similar to User1 will also get suggested that product.

1. **Content-Based Filtering**: Content-based filtering focuses on analysing the attributes and characteristics of products to make recommendations. In the case of Bigbasket, content-based filtering can leverage attributes like product category, brand, nutritional information. It can provide personalized recommendations by considering a customer's previous purchases and suggesting similar or related products. While the customer is searching for a product then this can help in giving customer variety of options from the catalogue where customer can select from a wide range of options based on price, nutritional aspects etc.

*For instance* - Suggesting a health-conscious person who ordered protein shake then maybe suggest them sprouts, cottage cheese etc.

1. **Association Rules**: Association rules mining can identify patterns and relationships between different products. Bigbasket can utilize association rules to suggest complementary or frequently co-purchased items to customers. This technique can help in cross selling, placement of items on the website/interface. Like Red bell pepper being sold with yellow bell pepper

*For instance*, if a customer adds milk to their basket, the system can recommend bread or cereal as associated products.

This also help in bundling the products, forming promotional schemes, discount coupons etc.

1. **Hybrid Filtering**: A hybrid approach combines multiple recommendation techniques to provide more accurate and diverse recommendations. For Bigbasket, a hybrid approach can leverage both collaborative filtering and content-based filtering. By considering a customer's purchase history, preferences, and attributes of products, the hybrid approach can offer a well-rounded set of recommendations that balance personalization and diversity.

For giving the recommendations no one, technique can be implemented it is an amalgamation of all the techniques put together. And every technique serves its specific purpose to improve accuracy of prediction.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Answer 4

Data Challenges which we can encounter for Bigbasket are as follows: - 1.

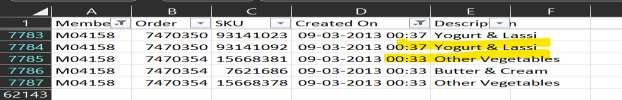
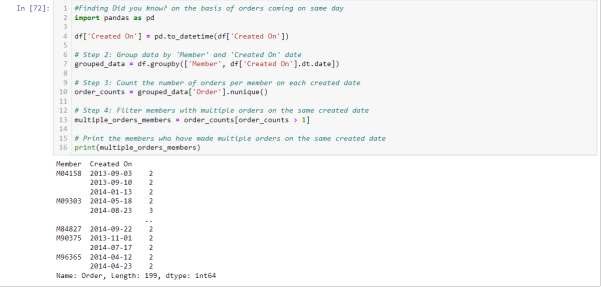
**Cold Start Problem:** When new customers join Bigbasket, they may have no purchase history available for recommendation algorithms to utilize for suggesting. This creates a cold start problem, making it challenging to provide personalized recommendations for new customers until sufficient data is collected to make meaningful suggestions.

1. **Sparse Data:** The dataset for Bigbasket may suffer from sparsity due to the vast number of products of similar kind as well as and the diverse preferences of customers. Sparse data can make it difficult to identify meaningful patterns and associations between products and customers, leading to less accurate recommendations.
2. **Regional influence:** on the preference of products. Many of the products and their demands can be highly influenced by the region in which the customer is geographically present in. Some high demand product of one region might not be demanded or even present at other location.
3. **Seasonal Variations:** Basket patterns can be influenced by seasonal variations and specific events. For example, customers may have different purchasing behaviour during holidays or festivals. To understand and include these seasonal variations into the recommendation engine requires extensive data collection and extensive pattern analysis of the data.
4. **Dynamic Inventory**: As the customer base is huge, we need to make sure inventory is updated Bigbasket's inventory is subject to frequent changes as multiple people tend to order same thing along with that new products are added or discontinued periodically. Keeping the recommendation engine updated with the latest product information and aligning it with customer preferences can be a data challenge. Ensuring timely data updates and handling product substitutions in case of products going out of stock or replacements are important thing to consider.
5. **Data Quality and Noise**: Data quality issues, such as missing values, incorrect labels, biased reviews or inconsistent data, can affect the accuracy and quality of the recommendations. Additionally, noisy data, such as outliers or irrelevant information, can introduce biases and affect the performance of recommendation algorithms.

### Exploration of the dataset -

1. Multiple SKU has same description so if we went ahead with considering only description for Apriori algorithm might suggest only products and not suggest products of different companies we have considered SKUs for forming our algorithm.
2. Data not available in the dataset and also attributes of the products are missing like price, ratings, nutritional value etc of the products which serves as the base for Item-Item collaborative filtering.

Finding the orders by the customers of the same day which indicates the customer forgot to order the first time



Above customer ordered within 5 minutes of placing order.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Answer 5

Implementation and deployment challenges of a recommendation engine for Bigbasket:

* 1. **Scalability:** As mentioned every month 30% new customers get onboarded and balancing that with the already existing large customer base and handles a large volumes data. It is crucial to ensure that the recommendation engine can scale efficiently to handle increasing data volumes and user demands. Implementing a scalable infrastructure, utilizing distributed computing, and optimizing algorithms for performance are essential considerations.
  2. **Real-Time Recommendations:** Providing real-time recommendations to the customer based on time of the day, season, month, demographics etc. The recommendation engine needs to process and analyse data at a rapid speed to generate personalized recommendations in real time.
  3. **Personalization and Diversity:** Recommendation system must strike the right balance between personalized recommendations and offering new products to customers is important.

Setting up recommender system which keeps suggesting the product which customer is most likely to along with introducing new product which are niche and might interest customers is difficult to built and implement.

As over suggestion of new products might lead to customer dissatisfaction. We need to analyse the pattern whether the customer is willing to experiment or not and then only make such suggestions. Ensuring that recommendations are not overly focused on a narrow set of items requires careful tuning and experimentation.

* 1. **Data Privacy and Security Issues:** Recommendations are based on analysing customer order history data and browsing patterns, which raises privacy and security concerns to customer data. To ensure compliance with data protection regulation, implementing secure data storage and processing practices, and adopting privacy- preserving techniques like anonymization are crucial.
  2. **Continuous Learning and Adaptation:** Recommendation system need to adapt to changing customer preferences, new products, and evolving market equations. Implementing mechanisms for continuous learning, model retraining, and incorporating real-time feedback into the recommendation engine is essential. Making sure what is being searched often is available in the inventory, that also needs to be monitored and implemented carefully. This involves establishing data pipelines, scheduling regular updates, and monitor model performance over time.

Change in buying patterns should be analysed more diligently like addition of an infant to the family.

* 1. **Data Availability:** Labelling of data must be done properly as new items will be onboarded and they should be properly mapped with their attributes properly so that collaborative filtering algorithms can work efficiently and give accurate suggestions.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Answer 6

Note: For enhancing the readability of products I have added SKUs with the description.

Five consumer-agnostic “good-quality” association rules sorted by their lift ratios - These rules are being derived at

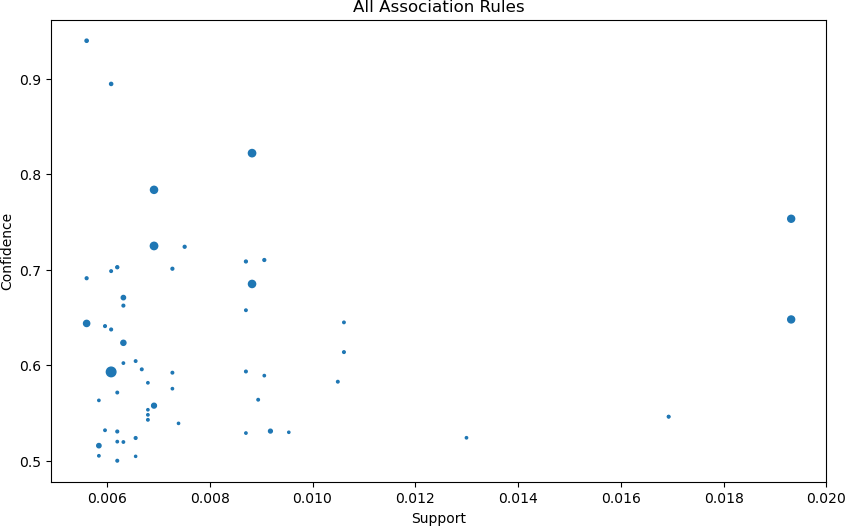
Support = 0.0055

Confidence = 0.5

Code Snippet of python



Plot of Association Rules: Support vs. Confidence



Rule Interpretation – goodness of rules

Rule 6: If a customer purchases "93141022 Yogurt & Lassi," there is a high likelihood (confidence of 59.30%) that they will also purchase "93141023 Yogurt & Lassi." The lift value of 47.37 indicates a strong association between these items.

Rule 35: If a customer purchases "15668688 Root Vegetables" and "15668594 Exotic Vegetables" together, there is a high likelihood (confidence of 72.50%) that they will also purchase "15668494 Exotic Vegetables." The lift value of 28.28 suggests a strong association between these items.

Rule 11: If a customer purchases "15668379 Other Vegetables" and "15668494 Exotic Vegetables" together, there is a high likelihood (confidence of 82.22%) that they will also purchase "15668594 Exotic Vegetables." The lift value of 27.58 indicates a strong association between these items.

Rule 12: If a customer purchases "15668379 Other Vegetables" and "15668594 Exotic Vegetables" together, there is a moderate likelihood (confidence of 68.52%) that they will also purchase "15668494 Exotic Vegetables." The lift value of 26.73 suggests a moderate association between these items.

Rule 34: If a customer purchases "15668688 Root Vegetables" and "15668494 Exotic Vegetables" together, there is a high likelihood (confidence of 78.38%) that they will also

purchase "15668594 Exotic Vegetables." The lift value of 26.29 indicates a strong association between these items.

Overall, these rules highlight strong associations between certain groups of items. For example, customers who purchase "Root Vegetables" and "Exotic Vegetables" are highly likely to purchase "Other Exotic Vegetables" in different varieties.

The high confidence and lift values suggest strong relationships between the antecedents and consequents, indicating potential cross-selling or bundling opportunities for these items.

**Rule: (93141022 Yogurt & Lassi) -> (93141023 Yogurt & Lassi)**

**Support**

**Support of first rule:** Support measures the frequency or popularity of a product in a dataset. It calculates the proportion of transactions in the dataset that contain the itemset. Support can be seen as an indicator of how frequently an itemset occurs. Higher support values indicate more frequent item sets.

**Antecedent**: The antecedent refers to the products or set of products that are present in the transactions and serve as the basis for making predictions or inferences that what is likely to be purchased following that products purchase. It represents the condition or event that precedes or occurs before another item or set of items.

Here antecedent is (93141022 Yogurt & Lassi)

**Consequent**: The consequent refers to the product or group of products that are predicted or inferred based on the presence of the antecedent. It represents the outcome or event that follows or occurs after the antecedent.

Here consequent is (93141023 Yogurt & Lassi)

**Antecedent support**: proportion of transactions containing the antecedent for the rule under study it is obtained as 0.010254

**Consequent support**: proportion of transactions containing the consequent for the above rule so it is calculated as 0.012519

**Support**: proportion of transactions containing both antecedent and consequent so the Support is 0.006081

## Confidence

**Confidence**: measures the reliability or certainty of the relationship between the antecedent and consequent of an association rule. It calculates the conditional probability of finding the consequent in a transaction given that the antecedent is present. Confidence can be interpreted as the likelihood of the consequent being purchased when the antecedent is already purchased or is in the state of getting purchased. Higher confidence values indicate a stronger relationship between the antecedent and consequent.

**Confidence**: probability of finding the consequent given the antecedent for the rule under study it is 0.593023

Confidence = Support (Antecedent and Consequent) / Support (Antecedent)

= 0.006081 / 0.010254

= 0.593023

## Lift Ratio

**Lift**: measures the strength of association between the antecedent and consequent of an association rule compared to their individual occurrences. It calculates the ratio of the observed support of the itemset to the expected support if the antecedent and consequent were independent. Lift greater than 1 indicates a positive relationship, where the presence of the antecedent increases the likelihood of the consequent. Lift less than 1 indicates a negative relationship, where the presence of the antecedent decreases the likelihood of the consequent. Lift equal to 1 indicates independence. It indicates how confident we are on our Confidence.

**Lift**: measure of how much more likely the consequent is given the antecedent compared to its independent occurrence it is calculated as 47.368439

Lift = Support (Antecedent and Consequent) / (Support (Antecedent) \* Support (Consequent))

= 0.006081 / (0.010254 \* 0.012519)

= 47.368439

For the first rule, the support indicates that the rule holds strong for in approximately 0.61% of all transactions.

The confidence suggests that when customers purchase (93141022 Yogurt & Lassi), there is a 59.30% chance they will also purchase (93141023 Yogurt & Lassi).

The lift ratio of 47.37 indicates a strong positive association between the antecedent and consequent, suggesting that the presence of (93141022 Yogurt & Lassi) significantly increases the likelihood of (93141023 Yogurt & Lassi) being purchased compared to their independent occurrences.

### Suggested action plans on the basis of top 5 good rules:

All the rules can be benefitted from the below action plans

**Plan 1**: we can promote the sale of the two SKU’s by bundling them together and by giving discounts or personalised coupons specific to the user who most frequently purchase these two products together.

**Plan 2:** Encourage customers who have purchased "15668688 Root Vegetables" and "15668594 Exotic Vegetables" to also purchase "15668494 Exotic Vegetables". By showing the consequent item as the Did you forget? Option while checking out. This can be done by showcasing recipes or meal ideas that incorporate all three items.

**Plan 3:** Promote sale of antecedent and consequent by suggesting the products to the customers when browsing any of the product other one being shown in “You may also like” section.

**Plan 4:** Being a grocery store we can showcase the recipes for most frequently purchased products in our newsletters or in our dedicated links. And below that provide the list to items which user need to order to prepare that recipe.

**Plan 5:** Cross-promote the products like "15668594 Exotic Vegetables" to customers who have bought "15668379 Other Vegetables" and "15668494 Exotic Vegetables". Highlight the unique features, flavours, or health benefits of exotic vegetables to encourage their purchase.

**Plan 6:** Offer bundle deals or discounts to incentivize the purchase of both items together.

**Plan 7**: Highlight the complementarity of these items in terms of taste, nutrition, or culinary uses. Like promoting Gulab Jamun tin with Ice cream.

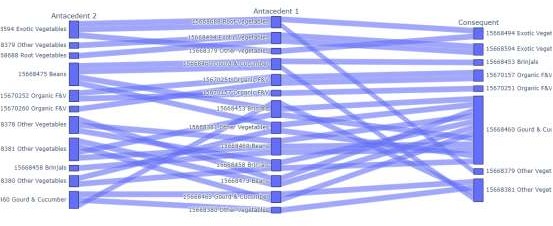
Connectivity Graphs as per the association rules obtained by apriori algorithm Graph 1

1-1 Association of the products.



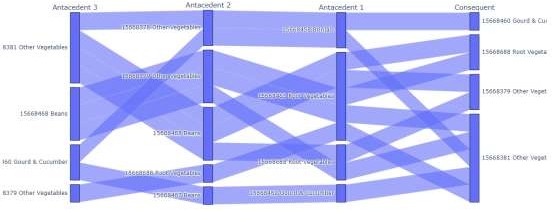
Graph 2

Product 1 and Product 2 resulting in purchase of Product 3



Graph 3

Multilayer Antecedent and Consequent products.



\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Answer 7

Bigbasket is interested in introducing a “Did you forget?” feature to identify items a customer may have forgotten.

Discuss how this feature can be created following data-driven approaches.

**Data Collection:** Collect and store historical transaction data, including customer orders, items purchased, and the frequency of purchases. This data will serve as the base for identifying patterns in purchases and making recommendations for “Did you forget?” functionality.

**Customer Purchase History:** Each customer's purchase history has to be understood and analysed thoroughly which brings out their preferences and common purchasing behaviour. Consider factors like frequently purchased items, purchase frequency, and recent purchases. If the customer purchases rice and pulses bi-weekly, so when the customer is making its basket and if rice and pulses are not added then it should mention rice and pulses in “Did you forget?” section after the interval of 15 days.

**Basket Analysis:** Perform basket analysis to identify frequently co-purchased items. Association rule (apriori algorithm) mining techniques can be applied to find relationships between different products. For example, if customers frequently purchase bread and milk together, the recommendation engine can ask the customer in “Did you forget?” milk to customers who have bread in their current order.

**Customer Similarity:** Utilize customer similarity metrics like assigning customers to a cluster on the basis of there shopping patterns if they are existing in the system since long then clustering can help identify customers who have similar purchase behaviours and preferences. This can be achieved through techniques such as collaborative filtering, where customers with similar purchasing patterns are grouped together. Recommendations made to one customer can be extended to similar customers who may have forgotten or might like to try and buy similar item.

**Contextual Recommendations:** Take into account the current items in the customer's basket. Analyse the basket contents and identify items that are commonly associated with the products already selected by the customer. For example, if a customer has added pizza base to their basket, the recommendation engine can suggest pizza toppings, chilli flakes and oregano or grated cheese.

**Machine Learning Models:** Employ machine learning algorithms to predict the likelihood of a customer forgetting certain items based on their past behaviour. Train models using historical data to identify patterns and create predictive models that can estimate the probability of a customer forgetting specific items and then make the suggestions accordingly.

If a person buys some items and then again order after an hour or so in order to mitigate the chances of reordering model should learn it should be used in future in order to make better predictions and improve Did you forget suggestions. This will improve customer satisfaction.

**Real-time Recommendations:** Implement a real-time recommendation engine that can provide on-the-spot recommendations as customers are checking out. Consider factors such as current basket’s contents, customer line of order like if a customer has ordered everything premium then probably suggest all premium products and similar customer behaviour to generate personalized recommendations for forgotten items.

With the dataset given I have tried to build an individual customer purchase history based recommendation:

Code Snippet of Python for the same:

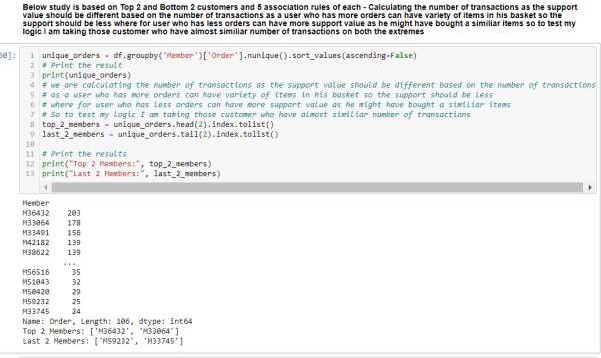


\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Answer 8

Below study is based on Top 2 and Bottom 2 customers and 5 association rules of each individual -

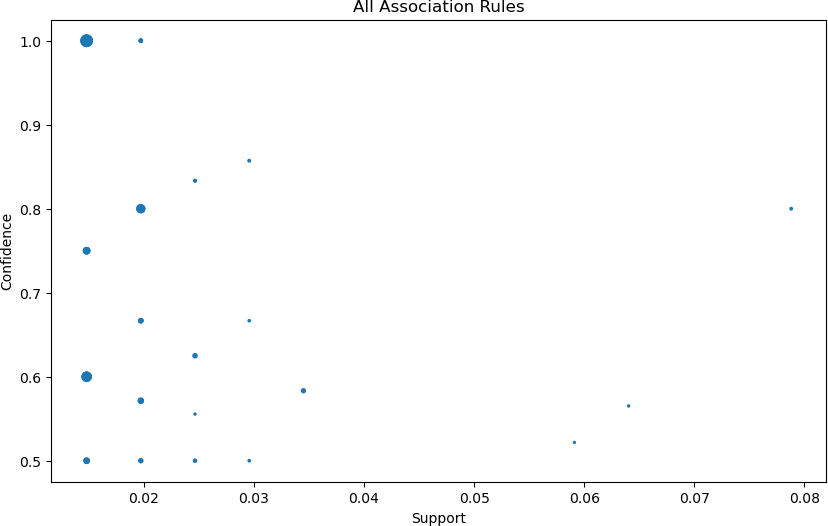
Calculating the number of transactions as the support value should be different based on the number of transactions as a user who has more orders can have variety of items in his basket so the support is less where for user who has less orders can have more support value as he might have bought a similar items so to test my understanding I am taking customer who have almost similar number of transactions on both the extremes.

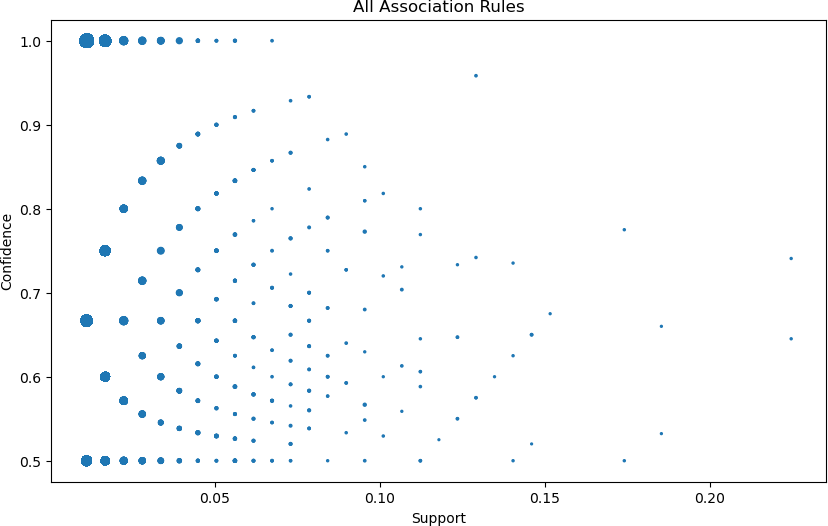


Assumption as a transaction created on as it might happen that the user forgot about the item and he reordered same day so we want to take it as one transaction that is why we used created on instead of Order Number



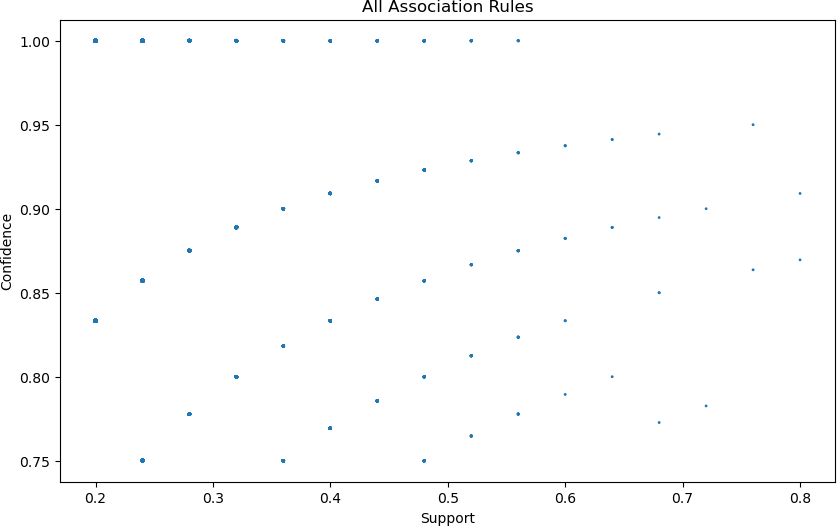
Graph Obtained: Top#1

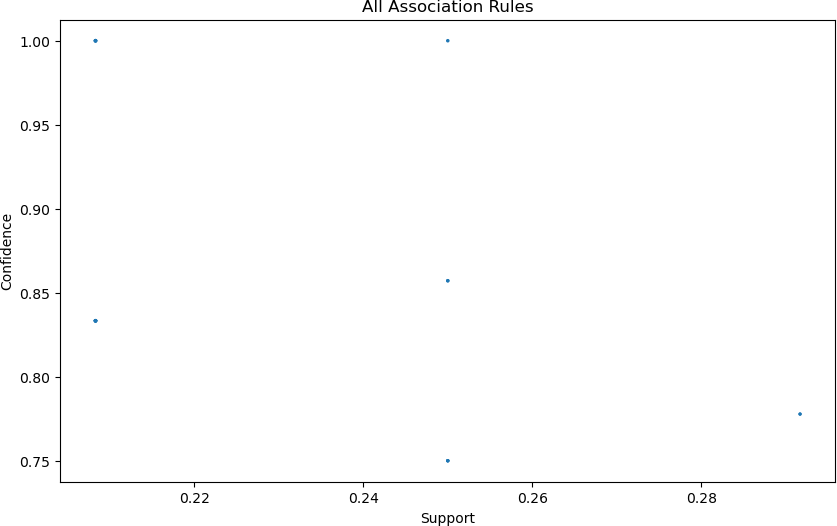


Top#2

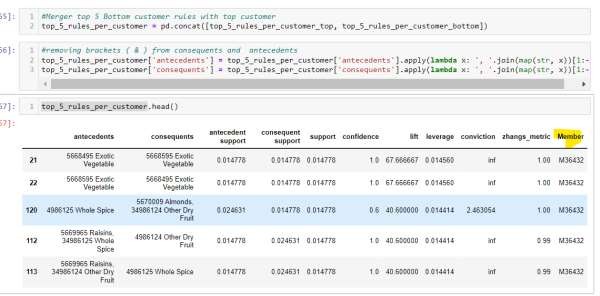
As per the above graph for Top#1 & Top#2 the support value as well as confidence looks fine and i kept it one lower side as member 1 needs lower support value otherwise the rules are not getting generated but as I want to use same support and confidence for both the member, I used these values.

I am keeping Top 5 association rules for the customer with Member ID in a dataframe.

Bottom 2



In the same dataframe top\_5\_rules\_per\_customer I am keeping customer and association rules of them in the same dataframe.



Here we need to give Member id and the product which is being ordered and in the output it will give suggestion based on the association rule of that member.



Above when customer M36432 ordered 4986124 Other Dry Fruit then it will get suggestion o f ‘4986125 Whole Spices, 15670009 Almond'

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Answer 9

To find similarity between users based on baskets on products being bought:

Bigbasket collects transaction data, including the list of products purchased by each customer in their order history so with the help of that multiple methods can be followed to find similar patterns between the users.

**User-Item Matrix:** Create a user-item matrix where each row represents a user and each column represents an item. The matrix is filled with binary values indicating whether a user has purchased a particular product or not.

**Similarity Calculation**: Apply a similarity metric, such as cosine similarity or Pearson correlation, to calculate the similarity between user baskets. The similarity value measures the degree of resemblance between two baskets based on their overlapping purchases.

**Finding Similar Users**: For a given user, calculate the similarity with other users and identify the top N users who have the highest similarity scores. These users are considered similar in terms of their purchase history.

**User-Based Collaborative Filtering:** Utilize the purchase history of similar users to recommend products to the target user. If a similar user has purchased a product that the target user hasn't, the recommendation engine can suggest that product to the target user.

**Generating Personalized Recommendations:** Combine collaborative filtering with other techniques, such as content-based filtering or item popularity, to provide diverse and personalized product recommendations. This ensures that the recommendations are not solely based on similarity but also take into account other factors like item attributes and overall popularity.

**Iteration and Feedback:** Continuously collect user feedback and iterate on the recommendation system to improve its accuracy and relevance. Incorporate mechanisms for users to rate or provide feedback on the recommended products, which can further enhance the recommendations in the future.

Yes, collaborative filtering can be used on existing dataset is possible as transactions of the users are given and each user’s transaction details can help us with buying patterns of the each.

Similarity scores can be calculated and can be used as input for the recommendation engine to cross suggest these products to user with high similarity score.

If we have had ratings of the products in our dataset then it would’ve improved our suggestion pattern and recommendation quality would’ve improved too.

### Use Case

Two users, User1 and User2, User1 who is a new parent

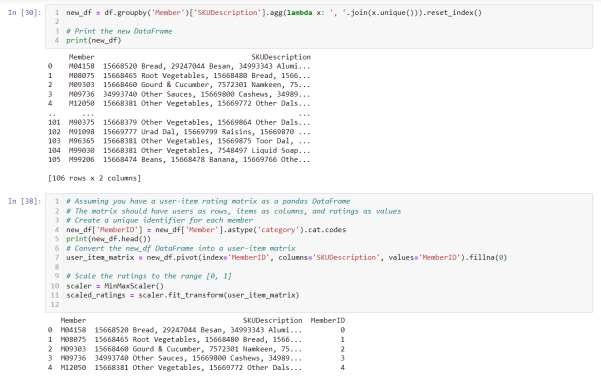
User2 who has idea about parenting needs and what to buy and what not

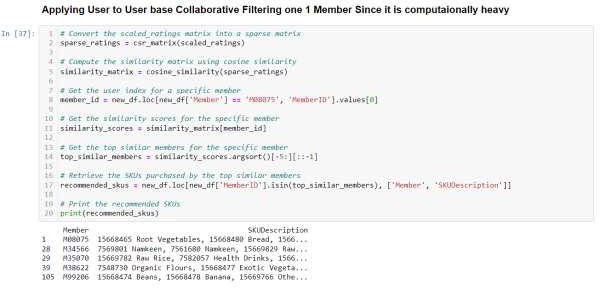
And their buying patterns, items bought by them are also almost very similar then on the user-user collaborative filtering their similarity score will be very high.

So in this case if recommendations are made on the basis of user-user collaborative filtering then there might be a chance that products bought by User1 and User2 are cross suggested in this case User1 will be benefitted because of the experience of the User2 and User2 might get suggested some products which are new in market and User2 might get suggested about products which they might not generally would’ve explored.

Keeping in mind the above use case not only User1 and User2 are benefitted but also Bigbasket as a platform will reap higher profits and will increase in customer satisfaction and as well as increase in loyalty.

Applied Collaborative filtering for users in the dataset:





\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Answer 10

Following are the major ideas for implementing Smart Basket:

**Customer Profile:** Build a comprehensive customer profile by collecting data on their demographics, preferences, and previous purchase history this has to be done diligently and information as and when required must be updated periodically. This includes analysing their past orders, saved shopping lists, ratings, and reviews.

**Preferences and Customization:** Allow customers to set their preferences, such as dietary restrictions, preferred brands, organic products, or specific categories they frequently purchase. All the preferences as added by the user must be incorporate into the Smart Basket algorithm to prioritizing relevant products.

**Machine Learning Algorithms:** Utilize machine learning algorithms to analyse the customer's purchase history and browsing behaviour. Apply techniques like collaborative filtering, content-based filtering, or hybrid approaches to identify products that are likely to be of interest to the customer this can be based on age group, demographics., size of family etc

**Real-Time Recommendations:** Provide real-time recommendations as the customer browses the website for products and as they add items to their cart. These recommendations can be based on the customer's current browsing behaviour, popular products, personalized preferences, temporal pattern and related items.

**Contextual Recommendations:** Consider the customer's context, such as the time of day, upcoming events, or seasonal factors, to make relevant product suggestions. For example, if it's close to a customer's regular purchase cycle for certain items, the Smart Basket can remind them to add those items to their cart. During festival time it should make suggestions accordingly

**Dynamic Updates:** Continuously update the Smart Basket as the customer interacts with the website or makes purchases. Incorporate machine learning models that adapt to changes in preferences, trends, or purchasing patterns over time. Purchasing pattern may vary depending on income of the household or inflation in the market like the price of tomatoes have hiked it must suggest alternatives to it like tomato puree etc.

**Cross-Selling and Complementary Products**: Recommend cross-selling and complementary products based on the customer's current selections. This can be done by analysing co-purchasing patterns and associations between different items in the customer's basket or purchase history. Like while ordering Noodles the person might be interested in buying soy sauce, vinegar etc as well so make relevant suggestions.

**User Feedback and Iteration:** Gather feedback from customers regarding the Smart Basket recommendations. Implement mechanisms for users to rate and provide feedback on the suggested items. Use this feedback to improve the recommendations and ensure they align with the customer's needs and preferences.

**Personalization and Flexibility:** Allow customers to customize the Smart Basket recommendations further. Provide options to refine or adjust the recommendations based on their specific requirements or shopping goals.

Providing user reminders on the basis of his purchasing patterns. User can add and ask Smart basket to notify the user if the products’ price has dropped or if the product was running out of stock and it is in stock now that can be done.

**Privacy and Data Security:** Prioritize data privacy and security while implementing the Smart Basket feature. Ensure compliance with data protection regulations and maintain robust data security practices to safeguard customer information.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*