**HBR – TD Bank Group: Building an Effective Enterprise Data Management Policy**

**Assignment 3.2**

June 18, 2021

Aaron Vo

BIA-678 Big Data Technologies

Stevens Institute of Technology – Summer 2024

Table of Contents

[1 HBR Case STUDY - Volkswagen Group: Driving Big Business With Big Data 3](#_Toc169627799)

[1.1 UVA CASE STUDY QUESTIONS 3](#_Toc169627800)

[1.1.1 What is the difference in the recommender system requirements between Bigbasket and other e-commerce companies such as Amazon and Flipkart? 3](#_Toc169627801)

[Unlike other e-commerce companies such as Amazon and Flipkart, Bigbasket’s system requirements requires them to accommodate customers’ high product counts. According to the document, “Bigbasket customers place order for several products, sometimes as high as 80 in one order depending on their purchase frequency” (Abraham, Pradhan, Iyer, Unnikrishnan). In addition, customers at Bigbasket typically bought the same product repeatedly since they were daily use products such as vegetables, bakery goods, and dairy products. This meant that they had to accommodate large orders of items where as their competitors did not. 3](#_Toc169627802)

[1.1.2 What are the different types of recommender systems? Which recommender system is more appropriate for Bigbasket? Different types of recommender systems include collaborative filtering, content-based filtering, or knowledge-based recommender system. However, for Bigbasket, it seems that the best system is a hybrid recommender system. This means that the recommender system uses a combination of both collaborative filtering and content-based filtering to identify the buying patterns of its customers and uses it to identify the large amounts of products that they buy on a regular basis. 3](#_Toc169627803)

[1.1.3 What are the possible data challenges in developing any data model? What approach should be taken to resolve these data challenges? 3](#_Toc169627804)

[Possible challenges in developing any data model includes context, requirements, and the quality of the data. With regards to Bigbasket, we saw that understanding the context behind customers’ purchases is important. Bigbasket had to learn from the data that customers have a high product count and that a lot of it were multiple counts of the same items. Another possible challenge is that the demands of customers are changing with respect to season and item. Analyzing these trends are useful to understand any business performance and for Bigbasket, it’s important to have to understand the full requirements of the model. 3](#_Toc169627805)

[1.1.4 Given the context of business carried out by Bigbasket, what basic tools can be used for understanding repeat purchases? 4](#_Toc169627806)

[Given the business context carried out by Bigbasket, the best tool for understanding repeat purchases would be clustering, or more specifically, time series clustering. Time series clustering can be used to discover commonly purchased items that are useful for pricing or formulating cross-selling strategies. However, since Bigbasket’s customers typically buy the same thing multiple time, we can see use support confidence and lift. Support indicates how frequently an item appears in the data. Confidence indicates the number of times the if-then statement is found to be true. Lift can be used to compare observed confidence with expected confidence, or how many times an if-then statement is expected to be found true. 4](#_Toc169627807)

[1.1.5 How do we find similarities between products based on what customers buy in different baskets? Can collaborative filtering be used to find similarities? What similarities are appropriate in this context and why? 4](#_Toc169627808)

[To find similarities between products based on what customers buy in different baskets, we can use several data mining techniques, including collaborative filtering, association rule mining, and clustering. In the context of finding product similarities based on customer purchases in different baskets, item based collaborative filtering is particularly useful. It works by examining the items that are frequently purchased together and identifying patterns in these co-occurrences. The choice of similarity measure—cosine similarity, Jaccard similarity, or others—depends on the specific characteristics of the transaction data and the business goals. Using these measures helps businesses understand product affinities, improve product recommendations, and ultimately enhance customer satisfaction and sales. 4](#_Toc169627809)

[1.1.6 How do we build a Smart Basket for a customer? Can we rank the products customers buy based on what they keep buying in different baskets and how do products appear together in different baskets? 4](#_Toc169627810)

[A Smart Basket is a personalized shopping experience that suggests products to customers based on their previous purchase behavior and the collective buying patterns of similar customers. Building a Smart Basket involves analyzing customer purchase behavior, identifying frequent purchases, and leveraging collaborative filtering and association rule mining to suggest relevant products. Ranking products based on purchase frequency and co-occurrence patterns ensures personalized and relevant recommendations, enhancing the customer shopping experience and driving sales. 5](#_Toc169627811)

[1.1.7 What testing strategy should be applied to find out how the model works? 5](#_Toc169627812)

[To evaluate the effectiveness of a Smart Basket model, a comprehensive testing strategy should be applied. This strategy includes several steps: defining evaluation metrics, splitting the data, performing cross-validation, testing with different data scenarios, and conducting A/B testing. 5](#_Toc169627813)

[1.1.8 What are the challenges and recommendations in implementing a real-world solution of “Smart Basket” and “Did you forget?” use case? 5](#_Toc169627814)

[Implementing a real-world solution for a "Smart Basket" and "Did you forget?" use case involves several challenges and requires strategic recommendations to address them effectively. These challenges includes the Inconsistent, incomplete, or noisy data which can have a negative effect on the effectiveness of the recommendation system. Another challenge includes handling and processing large volumes of transaction data in real-time which as a result, requires substantial computational resources and efficient algorithms. Finally, recommendations is typically hard and can be biased, which affects accuracy to meet individual customer preferences and needs. 5](#_Toc169627815)

# HBR Case STUDY - Volkswagen Group: Driving Big Business With Big Data

## UVA CASE STUDY QUESTIONS

### What is the difference in the recommender system requirements between Bigbasket and other e-commerce companies such as Amazon and Flipkart?

### Unlike other e-commerce companies such as Amazon and Flipkart, Bigbasket’s system requirements requires them to accommodate customers’ high product counts. According to the document, “Bigbasket customers place order for several products, sometimes as high as 80 in one order depending on their purchase frequency” (Abraham, Pradhan, Iyer, Unnikrishnan). In addition, customers at Bigbasket typically bought the same product repeatedly since they were daily use products such as vegetables, bakery goods, and dairy products. This meant that they had to accommodate large orders of items where as their competitors did not.

### What are the different types of recommender systems? Which recommender system is more appropriate for Bigbasket? Different types of recommender systems include collaborative filtering, content-based filtering, or knowledge-based recommender system. However, for Bigbasket, it seems that the best system is a hybrid recommender system. This means that the recommender system uses a combination of both collaborative filtering and content-based filtering to identify the buying patterns of its customers and uses it to identify the large amounts of products that they buy on a regular basis.

### What are the possible data challenges in developing any data model? What approach should be taken to resolve these data challenges?

### Possible challenges in developing any data model includes context, requirements, and the quality of the data. With regards to Bigbasket, we saw that understanding the context behind customers’ purchases is important. Bigbasket had to learn from the data that customers have a high product count and that a lot of it were multiple counts of the same items. Another possible challenge is that the demands of customers are changing with respect to season and item. Analyzing these trends are useful to understand any business performance and for Bigbasket, it’s important to have to understand the full requirements of the model.

### Given the context of business carried out by Bigbasket, what basic tools can be used for understanding repeat purchases?

### Given the business context carried out by Bigbasket, the best tool for understanding repeat purchases would be clustering, or more specifically, time series clustering. Time series clustering can be used to discover commonly purchased items that are useful for pricing or formulating cross-selling strategies. However, since Bigbasket’s customers typically buy the same thing multiple time, we can see use support confidence and lift. Support indicates how frequently an item appears in the data. Confidence indicates the number of times the if-then statement is found to be true. Lift can be used to compare observed confidence with expected confidence, or how many times an if-then statement is expected to be found true. Based on the marketplace data, the lift and confidence can be reflected as such. We see that the antecedent and consequents makes a lot of sense as products such as After Shave, Dairy & Cheese, Organic Edible Oils & Ghee, and others show a very strong correlation, indicating that these items are almost always bought together. These insights can be leveraged for cross-promotion, product placement, and bundling strategies to increase sales and customer satisfaction.

|  | **Left Hand Side** | **Right Hand Side** | **Support** | **Confidence** | **Lift** |
| --- | --- | --- | --- | --- | --- |
| **3** | After Shave | Dairy & Cheese | 0.009434 | 1.0 | 106.0 |
| **11** | After Shave | Organic Edible Oils & Ghee | 0.009434 | 1.0 | 106.0 |
| **98** | Baby Care Accessories | Foot Care | 0.009434 | 1.0 | 106.0 |
| **205** | Battery | Tea | 0.009434 | 1.0 | 106.0 |
| **239** | Biscuits | Lip Care | 0.009434 | 1.0 | 106.0 |
| **272** | Body Lotion | Talc | 0.009434 | 1.0 | 106.0 |
| **587** | Dairy & Cheese | Organic Edible Oils & Ghee | 0.009434 | 1.0 | 106.0 |
| **4** | After Shave | Hand Sanitizer | 0.009434 | 1.0 | 53.0 |
| **89** | Baby Care Accessories | Baby Cereal | 0.009434 | 1.0 | 53.0 |
| **90** | Baby Care Accessories | Beverages | 0.009434 | 1.0 | 53.0 |
| **126** | Baby Cereal | Foot Care | 0.009434 | 0.5 | 53.0 |
| **191** | Battery | Chocolate | 0.009434 | 1.0 | 53.0 |
| **195** | Battery | Festive Gift Packs | 0.009434 | 1.0 | 53.0 |
| **212** | Beverages | Foot Care | 0.009434 | 0.5 | 53.0 |
| **215** | Beverages | Magazine | 0.009434 | 0.5 | 53.0 |
| **236** | Biscuits | Ice Cream | 0.009434 | 1.0 | 53.0 |
| **292** | Bottle & Tin Openers | Ladles & Spatulas | 0.009434 | 1.0 | 53.0 |
| **432** | Chocolate | Tea | 0.009434 | 0.5 | 53.0 |
| **449** | Colours | Cutters, Peelers & Scrapers | 0.009434 | 0.5 | 53.0 |
| **468** | Colours | Pens | 0.009434 | 0.5 | 53.0 |

### How do we find similarities between products based on what customers buy in different baskets? Can collaborative filtering be used to find similarities? What similarities are appropriate in this context and why?

### To find similarities between products based on what customers buy in different baskets, we can use several data mining techniques, including collaborative filtering, association rule mining, and clustering. In the context of finding product similarities based on customer purchases in different baskets, item based collaborative filtering is particularly useful. It works by examining the items that are frequently purchased together and identifying patterns in these co-occurrences. The choice of similarity measure—cosine similarity, Jaccard similarity, or others—depends on the specific characteristics of the transaction data and the business goals. Using these measures helps businesses understand product affinities, improve product recommendations, and ultimately enhance customer satisfaction and sales.

### How do we build a Smart Basket for a customer? Can we rank the products customers buy based on what they keep buying in different baskets and how do products appear together in different baskets?

### A Smart Basket is a personalized shopping experience that suggests products to customers based on their previous purchase behavior and the collective buying patterns of similar customers. Building a Smart Basket involves analyzing customer purchase behavior, identifying frequent purchases, and leveraging collaborative filtering and association rule mining to suggest relevant products. Ranking products based on purchase frequency and co-occurrence patterns ensures personalized and relevant recommendations, enhancing the customer shopping experience and driving sales.

### What testing strategy should be applied to find out how the model works?

### To evaluate the effectiveness of a Smart Basket model, a comprehensive testing strategy should be applied. This strategy includes several steps: defining evaluation metrics, splitting the data, performing cross-validation, testing with different data scenarios, and conducting A/B testing.

### What are the challenges and recommendations in implementing a real-world solution of “Smart Basket” and “Did you forget?” use case?

### Implementing a real-world solution for a "Smart Basket" and "Did you forget?" use case involves several challenges and requires strategic recommendations to address them effectively. These challenges includes the Inconsistent, incomplete, or noisy data which can have a negative effect on the effectiveness of the recommendation system. Another challenge includes handling and processing large volumes of transaction data in real-time which as a result, requires substantial computational resources and efficient algorithms. Finally, recommendations is typically hard and can be biased, which affects accuracy to meet individual customer preferences and needs.

References

Artificial Intelligence and Machine Learning as tools to augment Human Intelligence. (2021, April 12).

*Market Basket Analysis, Collaborative Filtering- Unsupervised Algorithm - Part 4*.

Dominicm73.Blogspot. Retrieved April 13, 2022, from

<https://dominicm73.blogspot.com/2020/06/market-basket-analysis-collaborative.html>