

Execution

This case is split into 3 parts.

- **Reading the Case**
First off, this document contains the introduction to the case, the presentation of the business, a presentation of the data and 2 questions that you have to answer.
- **Reading the Theory**
To answer these questions correctly, you might need to consult a reference work. The 2nd part is reading that reference work, especially the recommended sections (presented later in this document).
- **Answer Individually and in Small Groups**
Before class, you will have to submit your answer to question 1 in Zone Cours. The answer has to contain (a) the model(s) that you suggest using and (b) a detailed justification for your choice (between 5 and 15 lines). During class, you will first be placed in small groups to discuss your solutions to question 1. Together, you will choose the most appropriate model(s) and then prepare a short presentation on the selected models. A few randomly chosen teams will then present their solutions. We will discuss the proposed solutions and the ones that were actually deployed by the company, and then we will move on to possible solutions to question 2.

Introduction

- A recommendation system!
- Are you sure?
- It will help personalize the user experience, and it'll drive on-line sales, adds Samuel, the data science team leader.
- Ok, I trust you!, approves Alexandre, technology director at Décathlon Canada.

Spring 2018, Montréal. Samuel, the data science team leader at Décathlon Canada, came up with a list of priorities for his department, which was brand new. At the top of his list, there was personalizing the client experience, through the use of recommendation systems.

For Samuel, this priority was self-evident. Numerous studies have shown the usefulness of a recommendation system, which drives sales and improves the client experience.

Recommendations systems have had profound effects on several fields, notably e-commerce and digital entertainment. “Already, 35 percent of what consumers purchase on Amazon and 75 percent of what they watch on Netflix come from product recommendations based on such algorithms. “[1]

And so, Samuel wanted to implement a recommendation tool on Décathlon’s digital platforms which would show a personalized choice of products each user might find interesting.

How best to go about this?

Presentation of Décathlon

Décathlon, A Giant in Sporting Goods

Above all, Décathlon is a general sporting goods store. The company wants to reach as many people as possible and aims to democratize the practice of sports. Décathlon stands out with its “lower prices everyday” approach, made possible by a high level of vertical integration. In fact, Décathlon operates across the entire value chain: research and development, design, manufacturing, logistics and sales. The product offer is enormous, with a catalogue covering over 65 sports with items for beginners and experts alike, totaling around 7,500 products on offer. All products sold are from in-house brands, meaning private brands owned by Décathlon, like Quechua for mountain hiking or Kipsta for team sports like soccer and rugby.

Like many retail giants, Décathlon started with in-store sales but now owns a large digital ecosystem. Consumers can buy products directly through Décathlon’s website or mobile app, and they can stay connected to the company through its newsletter and its sports promotion platform.

All in all, Décathlon processes over 30 million transactions a month, with about 100,000 in Canada [2]. Founded in France in 1976, the company has since expanded internationally: there are now about 1700 points of sale in about 60 countries, serving more than 110 million members. Décathlon is a private company owned notably by the founder’s family and by a number of its employees, who can become shareholders. In its home market of France, Décathlon dominates: there are more than 320 stores nationally, and the company is an integral part of the culture. Décathlon recently established itself in Canada, where it will soon have 8

stores with already nearly 300,000 members.

Well-Established Informational Resources

Over the years, Décathlon has developed a corporate culture based on data. It has many of its teams dedicated to collecting and storing quality data. This lets the data science team, which is based in Canada and is comprised of about ten experts, make use of the data with few logistical obstacles. Décathlon's vertical business model lets the company have a very complete overview of its product offer: data is collected and used across the value chain, without having to go through computer systems from third parties.

Improving the Offering Through Data: The Recommendation System

With all this data available for use, Décathlon is now looking for ways to improve the experience it offers its clients. Personalizing the on-line experience is one such improvement, which can be done notably through recommendation systems. These systems look to find products to recommend that might be of interest to each consumer. This recommendation simplifies the client's experience by letting them quickly find interesting products from the entire catalogue, thereby driving higher sales. Recommendation systems have been successful in the digital world, where they are already omnipresent.

One of the advantages of such systems is that they cut the amount of time that consumers spend looking for products and how long they spend considering purchasing them. Purchase decisions when it comes to sporting goods are not trivial and products are often hard to evaluate beforehand: an average order is around \$100 and the items are often technical in nature. Understanding which product from the entire catalogue best suits one's needs can be tiresome; recommendation systems preselect products considered relevant and present them to the client. These systems also create opportunities for cross sales by offering complementary products to the clients, like Amazon and other digital giants do with "clients who bought this product also bought..." sections, thereby increasing the number of products sold per client.

Décathlon is thinking of offering recommendations in a few different contexts: on landing pages for the website and the mobile app and in emails sent to members, notably. Examples of recommendations are shown in annex A.

Available Data: Diversified and Large-Scale

To feed the recommendation system, Décathlon already has a large quantity of data. Interactions between each user and each product is recorded (including the user's identity). The database also contains characteristics for each Décathlon member, like the sports they practice, and for the company's products, like their physical characteristics and the images and text associated with each product. An example of a dataset is shown in Annex B.

The number of users on Décathlon's platforms—almost 300,000 unique users each month [2], to which we can add 7,500 unique products—creates a technical challenge. To create relevant and personalized recommendations for such a large number of users, a large amount of data is necessary. This can complexify and slow down the prediction process of certain models, but these models need to be available in real time across the company's digital ecosystem.

Recommendation System Theory

To learn more about recommendation systems, we suggest the following reference: Aggarwal, Charu C. *Recommender Systems: the Textbook*. Cham: Springer, 2016.

Available in digital format from the HEC Montréal Library through the following link: <https://hecmontreal.on.worldcat.org/oclc/946011635>

Recommended sections to read

Chapter 1. Introduction

- 1.1 Introduction
- 1.2 Goals of Recommender Systems
- 1.3 Basic Models of Recommender Systems
 - 1.3.1 Collaborative Filtering Models (except 1.3.1.2, 1.3.1.3)
 - 1.3.2 Content-Based Recommender Systems
 - 1.3.6 Evaluation of Recommender Systems

Chapter 2. Neighbourhood-Based Collaborative Filtering

- 2.1 Introduction
- 2.3.1 User-Based Neighbourhood Models (except 2.3.1.1, 2.3.1.2, 2.3.1.3, 2.3.1.4)
- 2.3.2 Item-Based Neighbourhood Models
- 2.3.4 Comparing User-Based and Item-Based Methods
- 2.3.5 Strengths and Weaknesses of Neighbourhood-Based Methods

Chapter 3. Model-Based Collaborative Filtering

- 3.1 Introduction
- 3.6 Latent Factor Models (only 3.6.2, 3.6.3)

Chapter 4. Content-Based Recommender Systems

- 4.1 Introduction
- 4.2 Basic Components of Content-Based Systems
- 4.3 Preprocessing and Feature Extraction (except 4.3.4)
- 4.4 Learning User Profiles and Filtering (only 4.4.1, 4.4.5)
- 4.5 Content-Based Versus Collaborative Recommendations

Questions

Samuel is exploring the options available to his team. He has to build from scratch and so he would like to start with less complex systems, and then test the industry's best-performing models. The system will first recommend products to existing members on their landing page. Eventually, Samuel would like to extend the system as much as possible so that recommendations are given to new members.

Question 1

For the first prototype, there are no constraints on computing time needed to train the system or to offer recommendations. Which model(s) would the data science team need to choose for their recommendation system, and why?

Question 2

Once an initial model has been implemented, what improvements could be made, or what more advanced models could be prioritized for testing?

Reminder: Before class, you will have to submit your answer to question 1 in Zone Cours. The answer has to contain (a) the model(s) that you suggest using and (b) a detailed justification for your choice (between 5 and 15 lines). During class, you will first be placed in small groups to discuss your solutions to question 1. Together, you will choose the most appropriate model(s) and then prepare a short presentation on the selected models. A few randomly chosen teams will then present their solutions. We will discuss the proposed solutions and the solutions that were actually deployed by the company, and then we will move on to possible solutions to question 2.

Notes and References

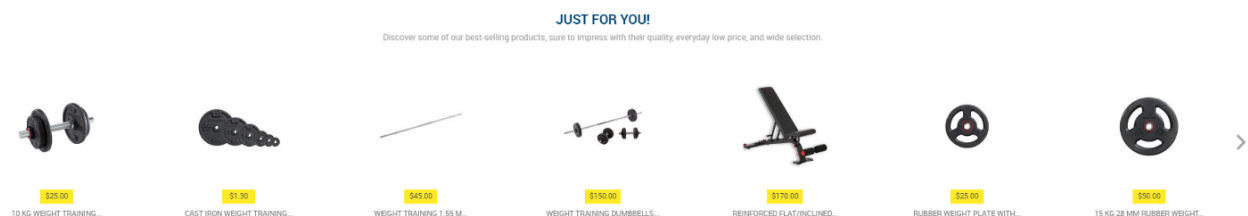
1. MacKenzie, I., Meyer, C., and Noble, S. (2013, October 1). How Retailers Can Keep up with Consumers | McKinsey. Retrieved from <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>

2. The number of visitors and sales numbers have fluctuated a lot since January 2020, because of the global health crisis. The statistics presented give a sense of scale, but they could vary greatly from month to month.

Annex A: Examples of recommendations

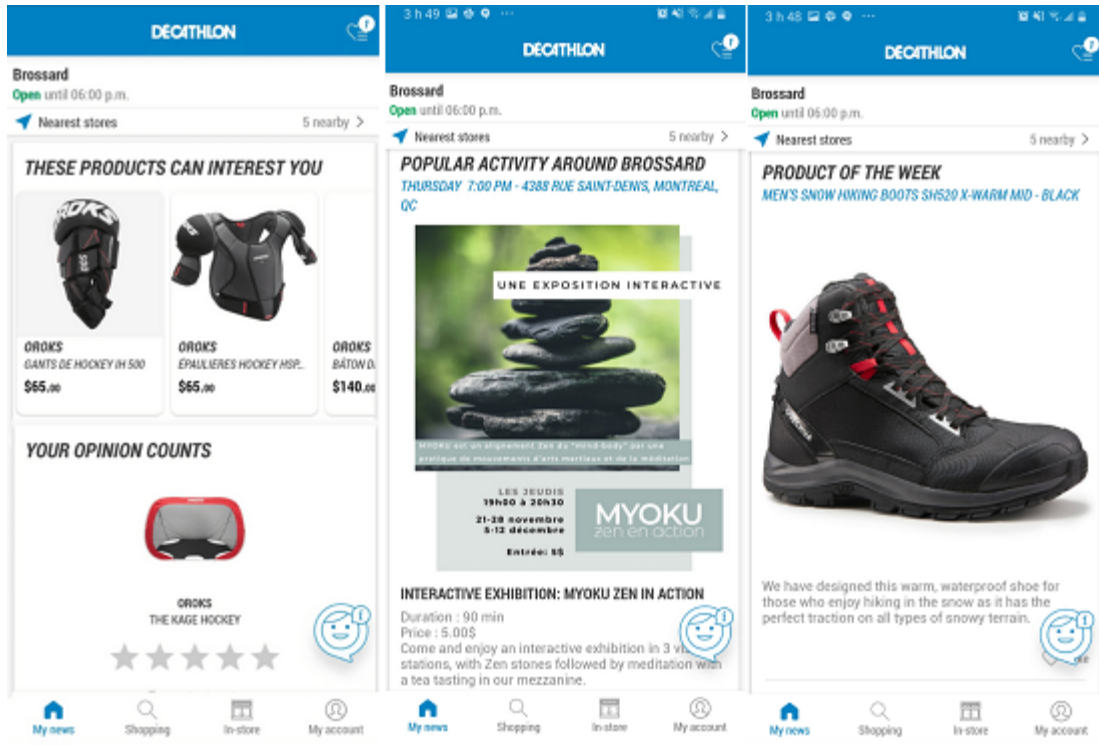
1. Website

Example of a landing page from the Décathlon website personalized for a member interested in weightlifting.



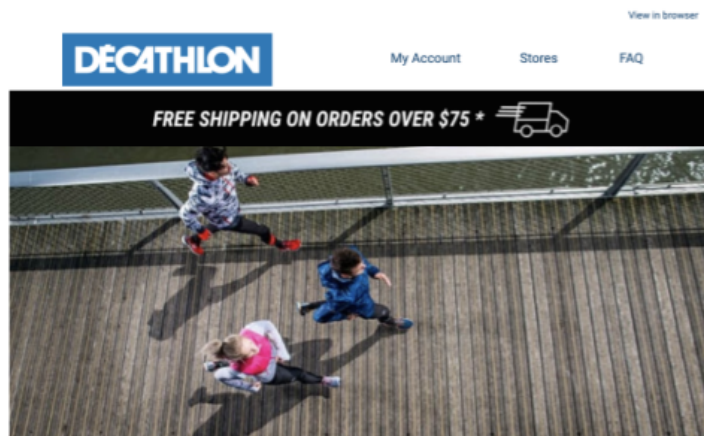
2. Mobile app

Example of a landing page from the Décathlon Mobile app personalized for a member interested in hockey.



3. Post-purchase email

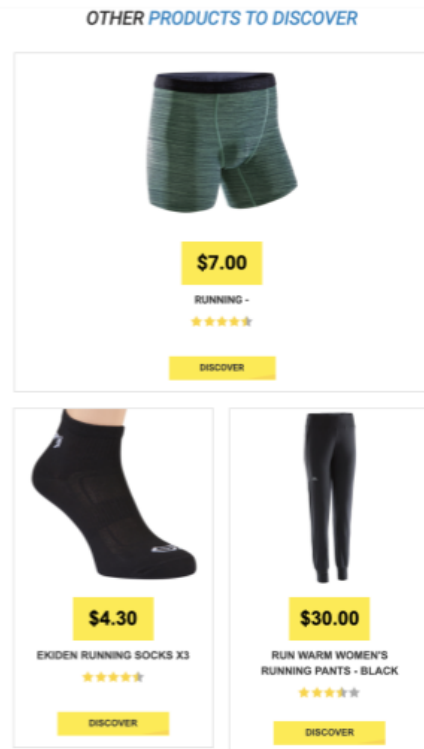
Example of a post-purchase email personalized for a member interested in running.



Hi Harlie,

Thank you for your recent purchase! Here are some more suggestions you might like.

Did you know? On top of products, **we also offer activities and accessible places to play (including in-store) for people of all levels.**



Annex B: Datasets

Note: The data has been modified to protect the confidentiality of the company and its members.

Table 1—Interactions between customers and items

customer_id	item_id	timestamp	click	purchase
imbvblxwvtiywunh	3384934262863770	2018-01-01	0	1
nzhrkquelkgflone	8321263216904593	2018-01-01	1	0
rgqgpqysetgoxdmz	7400694405709236	2018-01-01	0	1
xbnkvygodmlyfete	1769704982990886	2018-01-02	0	1
eqeefiaxvjxsyfxk	5460080147661416	2018-01-02	1	0
...
mkfqblfqwgzivfga	8625784510427076	2019-12-30	1	0
rzwykqujaznzmth	4673617174927844	2019-12-30	1	0
evbdnuhdfcpiyotj	9384258747220968	2019-12-31	1	0
cgatomzvjiizvctb	2756920171861146	2019-12-31	1	0
cnspkotxubxnztzk	5150255386059428	2019-12-31	0	1

Additional information:

- Each Décathlon member has a specific `customer_id`. The data only covers interactions done by members.
- Each Décathlon product has a specific `item_id`.
- A member can browse a product page (`click`) or buy the item (`purchase`). A member can have visited the page before buying (through an on-line purchase, for example), or not (through an in-store purchase, for example).
- On average, there are 600,000 clicks per month, from 31,000 different members. *Note: These statistics vary greatly from month to month.*
- On average, there are 80,000 purchases per month, from 20,000 different members. *Note: These statistics vary greatly from month to month.*

Table 2—User features

customer_id	is_male	is_female	declared_sports
imbvblxwvtiywunh	0	1	baseball; boxing
nzhrkquelkgflone	1	0	
rgqgpqysetgoxdmz	1	0	snowboard; golf; running

xbnkvygodmlyfete	0	0	
eqeefiaxvjxsyfxk	0	1	ski; rugby
...
mkfqblfqwgzivfga	1	0	soccer
rzwykqujaznzmth	1	0	
evbdnuhdfcpiyotj	1	0	
cgatomzvjiizvctb	0	1	
cnspktotxubxnxtzk	0	1	capoeira; beachtennis

Additional information:

- Each member has `user_features` an entry in the database . However, a member can leave their gender or their favourite sports blank.
- One member can have many different `declared_sports`.
- About 20% of members have stated their favourite sports.
- An overwhelming majority of members have stated their gender. *Note: Décathlon will progressively remove questions relating to demographics (age and gender) from its digital platforms. This data will not be available anymore.*
-

Table 3—Item features

item_id	is_male	is_female	is_junior	label	family	subdepartment	department	universe	linked_sports	image
3384934262863770	0	1	0	JACKET 920 SPACER HOOD W GREYNEPS	WOMAN PANT JACKET SWEAT	WOMAN APPAREL	PILATES, SOFT GYM	PILATES, SOFT GYM	musculatio n	3384934262863770.jpg
8321263216904593	0	0	0	GLOVE RAIN M PAIR	GOLF BALLS, GLOVES, TEES	GOLF EQUIPMENT ALL LEVELS	GOLF	GOLF	golf	8321263216904593.jpg
7400694405709236	1	0	0	SUPPORT M SHOES RED2	REGULAR MEN JOGGING SHOES	JOGGING SHOES	JOGGING	JOGGING	running	7400694405709236.jpg
1769704982990886	0	0	0	TARO #1 HOLOGRA PHIC	SPOONS TROUT PERCHE	TROUT PERCH LURE FISHING	PREDATOR FISHING	FISHING	fishing	1769704982990886.jpg
5460080147661416	0	1	0	SH100 X-WARM W TROUSERS GREY	WOMEN SNOW HIKING WARM PANTS & FLEECEs	WOMEN SNOW HIKING APPAREL, SHOES	SNOW HIKING	SNOW HIKING	hiking	5460080147661416.jpg
...
8625784510427076	0	0	0	BAR 0,35M	FREE WEIGHTS AND EQUIPMENT	BODYBUILDING	BODYBUILDING	CROSS-TRAINING, BODYBUILDING	musculatio n	8625784510427076.jpg
4673617174927844	0	0	1	GLOVE FH100 JUNIOR+ADULT BLACK	PROTECTION FIELDHOCKEY	FIELD HOCKEY	FIELD HOCKEY	FIELD HOCKEY	fieldhocke y	4673617174927844.jpg
9384258747220968	1	0	0	SHORT B300 AD TARMARK BLACK	MAN BASKETBALL OUTFIT	MAN BASKETBALL OUTFIT	BASKETBALL	BASKET	basketball	9384258747220968.jpg

27569201718 61146	0	0	0	TORCH LIGHT DYNAMO 5 0 WHITE COLLECTI ON	TORCH LAMPS HIKING/TREK	TREK BACKPACKS, POLES, HEADLAMPS, ACC	TREKKING SEVERAL DAYS	TREKKING SEVERAL DAYSniv	hiking	27569201718611 46.jpg
51502553860 59428	0	0	0	ARTENGO TR 560 LITE	ADULT BEGINNER/INTERM EDIATE RACKETS	TENNIS RACKETS/BALLS/ACCE SSORIES	TENNIS	TENNIS	tennis	51502553860594 28.jpg

Additional information:

- Each product has an entry in the database item features.
- A label belongs to a family, which belongs to a subdepartment, which belongs to a department, which belongs to a universe. However, some subcategories are identical. For example, both the department and the universe for item ARTENGO TR 560 LITE are TENNIS.
- Every product is tied to one or many sports.
- An image is available for each product.
- A product can be specifically for men ($is_male = 1$), for women ($is_female = 1$) or for children ($is_junior = 1$).