CHAPTER 6

Data Science and Subscription Services

Fashion is part of the daily air and it changes all the time, with all the events. You can even see the approaching of a revolution in clothes. You can see and feel everything in clothes.

—Diana Vreeland, former editor, *Vogue* magazine

Data is shaping the way we experience retail by enabling customized experiences. In the fashion industry, the technology being built around subscription services provides an example of how these custom experiences can be applied to e-commerce.

While for most fashion brands data is important for driving sales and producing styles that customers want, subscription services often rely on data-driven custom curation as the only method of product discovery. There may be no online catalog or search bar to make purchases from these businesses. Because of this, subscription services have a higher cost of failure and are deeply motivated to develop methods to deliver exactly what a customer wants on an individual basis. Using data to keep customers captivated has become part of the DNA for companies like Stitch Fix, Rocksbox, and Le Tote, just to name a few.

L. Luce, *Artificial Intelligence for Fashion*,
https://doi.org/10.1007/978-1-4842-3931-5_6

Subscription Models

> *Women could have a closet on rotation, have unlimited possibilities of what to wear…*
>
> —Jennifer Hyman, CEO, Rent the Runway

Numerous mechanisms are being used to build subscription-based business models. Even before diving into the details of how customization and data are being applied, this type of business is new to many industries and requires a bit of a backstory.

Subscription business models can take many forms. Some might be pay-as-you-go, while others might have an annual fee that's paid in one lump sum. Services offer customers varying degrees of choice and surprise in what they receive. Customers might be totally surprised every time, or they might select some of the products they want in advance.

In many subscription services, information is collected about which items are accepted or rejected in a given shipment. The user is usually prompted to answer a brief survey about why they rejected a given garment. They could answer quantitative questions such as how they would rate their satisfaction on a scale of 1–10, or they could answer questions with text that give them an opportunity to express themselves. Specialized techniques in natural language processing and machine learning can help analyze and quantify free-text responses to learn from them at scale. With insights that address specific areas of satisfaction or dissatisfaction, these companies grow an understanding of which garments and which customers are well matched and various other information about both the product and the customer.

In these businesses, brands can learn about what customers think of a given product. They might collect responses from thousands that uncover problems with the garment, shown in Figure 6-1.

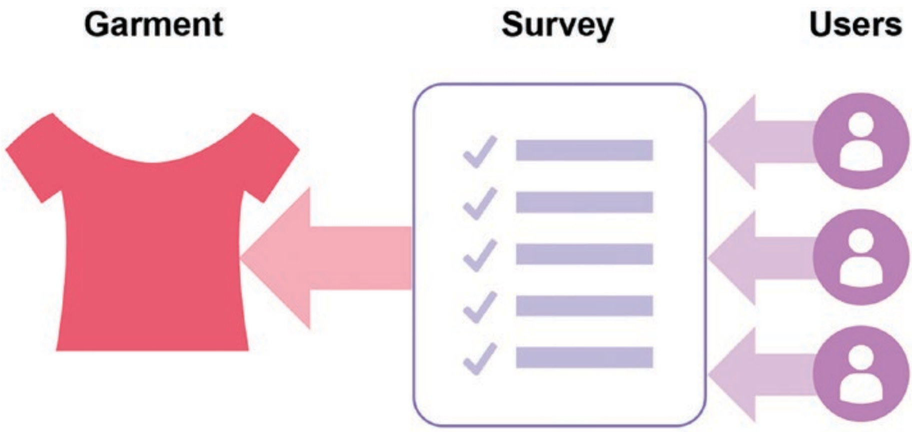

Figure 6-1. *The network of users giving feedback about a garment*

They also collect data about the user's preferences based on the garments they select, shown in Figure 6-2.

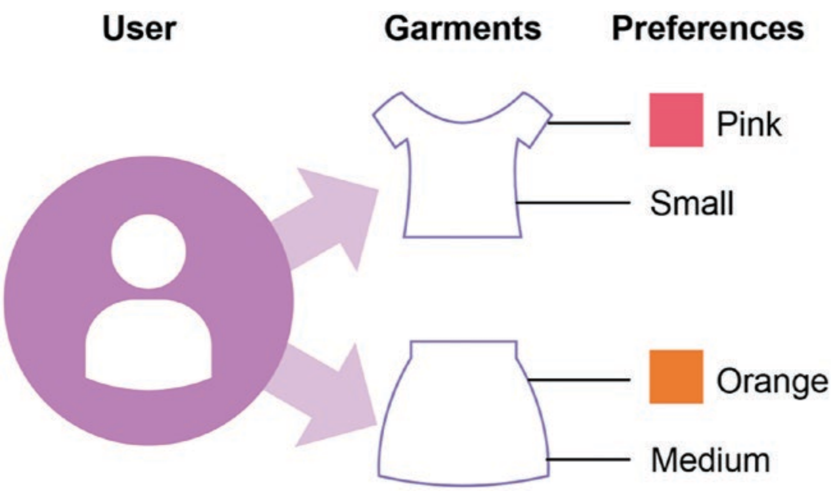

Figure 6-2. *The user's preferences can be "learned" through the choices the user makes about which items to keep and what they say about those items in a survey*

Brand Subscriptions

For some businesses, the subscription service itself is branded, but the contents are kept a secret until they arrive at your door. These businesses rely heavily on their brand to communicate that their products are aligned with the customer. Once the customer subscribes, they receive an assortment of items from the brand. Examples of this are Causebox or Unbound. These businesses might deliver their own brand or brands they work with as an advertisement or promotion. The contents of the box are not tailored to the individual in a brand subscription.

Targeted Subscriptions

Other subscriptions ask customers a series of questions that help them pinpoint exactly what kinds of products they're looking for. This targeted subscription sometimes takes the form of a personal stylist service, as in the case of Stitch Fix or Dia&Co. Ipsy and Birchbox are examples of similar businesses. In each of these, the customer does not know which products they will receive, and a major part of the service is predicting what they will like. The products are recommended to them based on their answers to survey questions and feedback about products they have received. Fashion services typically employ specialized algorithms alongside humans to bring the experience of a personal stylist to many users at scale.

User-Selected Subscriptions

Trunk Club and ShoeDazzle are examples of targeted subscriptions with a user-selected approval process. The stylist or a similar mechanism provides options to the user, and if the user confirms that they want them, the items are delivered. At that point, the user still has a choice to either purchase or return the items. This often comes with a styling fee, which likely is used to cover two-way shipping costs on orders when no purchase was made.

Consumables Subscriptions

Dollar Shave Club and Billie, a women's version of a similar service, both fall into a consumables category of subscription services. The user receives the same product at a set frequency—once per month, for example. These subscriptions allow people to automate the procurement of items they use every day.

What we culturally consider "consumables" has evolved over time. Items like socks and underwear, which are replaced more frequently than other apparel items, might fall into this category. Consumer habits around fast fashion have turned fashion items into consumables as well, creating new opportunities for subscription services like Stitch Fix.

Rental Subscriptions

The concept of renting items in the fashion space is not new. Suits and tuxedos, historically expensive purchases for men, have been widely available for rent since the 1970s. Rent the Runway got started by bringing this concept to women's fashion. They have since introduced subscription services around their rental garments, making it possible to rotate items in and out of your wardrobe every month.

What's different about a rental subscription is that it is recurring. These services allow the user to choose what they'd like to rent and ship it to them for a monthly fee. There are usually some rules around the duration of the rental and other constraints to the service. Unlike many other subscriptions, this model also allows the user to choose what they want.

Rent the Runway has created a unique community of women sharing a wardrobe with each other. They are incentivized to also behave in ways that are dissimilar to other e-commerce platforms. Rent the Runways customers are willing and excited to share images of themselves in the garments even when they look bad, provide data about their body types and the fit of the garments, and explain the wearing occasion.

The engagement of these customers allows Rent the Runway to constantly learn about items that are popular and the many properties that make those garments desirable. The information allows other customers to make more informed opinions about their rentals. Even negative experiences end up being valuable to everyone in the process.

Digital Personalization

Subscription services, especially targeted subscriptions, are important because they are able to embrace the benefits of digital commerce through personalization.

The emergence of digital products has made personalization an expected experience for consumers. They interact with digital products that remember what they watched and liked and that infer preferences they didn't even know they had. As the amount and accessibility of information and content has grown quickly over the past several decades, this type of personalization has gone from being a convenience to a necessity.

Hyperpersonalized recommendations for music, movies, digital products, and other digital services have become expected. Netflix, for example, doesn't show you every movie in existence when you log in. Instead, Netflix makes personalized recommendations based on other movies you've watched and specific interests relevant to your profile. As this has become the new normal, consumers are demanding it in every aspect of their lives. It gives them a sense of control and reduces the paralysis of choice. All of these experiences are driven by collecting data about what that user is doing on the site and then turning it into actionable customization.

For a consumer, going back into the fashion retail environment after being exposed to this level of personalization is like entering a world of complete and utter chaos. Suddenly, someone who is used to doing very little to get to a product they want has to sift through hundreds or thousands of unwanted items.

Discovery and curation is changing in a number of ways. It's shifting away from the fashion brands and retailers to new categories of sales channels. Online marketplaces like Amazon and social media marketing platforms like Instagram are improving their ability to recommend products to users. The methods they use to generate recommendations are based on some foundational concepts: recommendation engines, databases, and statistics.

Recommendation Engines

On a high level, services like Stitch Fix built their entire service around mastery of **recommendation engines**, also known as **recommender systems**. Recommendation engines help users filter out tons of items that users don't want. They're used in virtually every popular service to reduce the burden of choice, from Instagram ads to Amazon product suggestions. Recommendation engines increase the chance of a conversion by returning the right results to a user.

Recommendation engines aren't new at all, and in fact they have their roots in e-commerce. The two most commonly used types of recommendation engines are collaborative filtering and content filtering. A third, hybrid filtering, is a combination of both methods.

With the amount of data consumers are exposed to online, these engines need to be able to learn from and adapt to taking in new information from the user in real time.

Collaborative Filtering

Collaborative filtering uses information from a large dataset of user purchases and other behavior to predict what another customer is looking for. There are two basic approaches to collaborative filtering: a user-based approach and an item-based approach.

In the user-based approach, collaborative filtering returns recommendations based on a user's similarity to other users, as shown in Figure 6-3. Examples of this include product recommendations based on what other users have purchased, which you can find on e-commerce sites like Amazon.

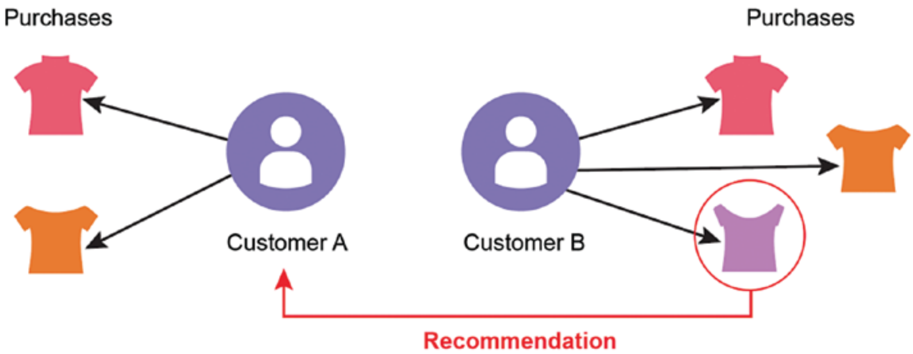

Figure 6-3. *Information about similar customers can be used to recommend products*

An item-based approach recommends products based on how users rated similar items. Item-based collaborative filtering is useful because it can give relevant recommendations even if it doesn't know anything about a given user.

Content-Based Filtering

Content-based filtering methods are based on user actions and preferences. If a user is exploring a site and likes and buys only red dresses, then more red dresses will be revealed to them during their search. However, one problem with this method is that it may continue to recommend products in the same category, potentially even after the user has lost interest in searching for items in that category. An example of content filtering is shown in Figure 6-4.

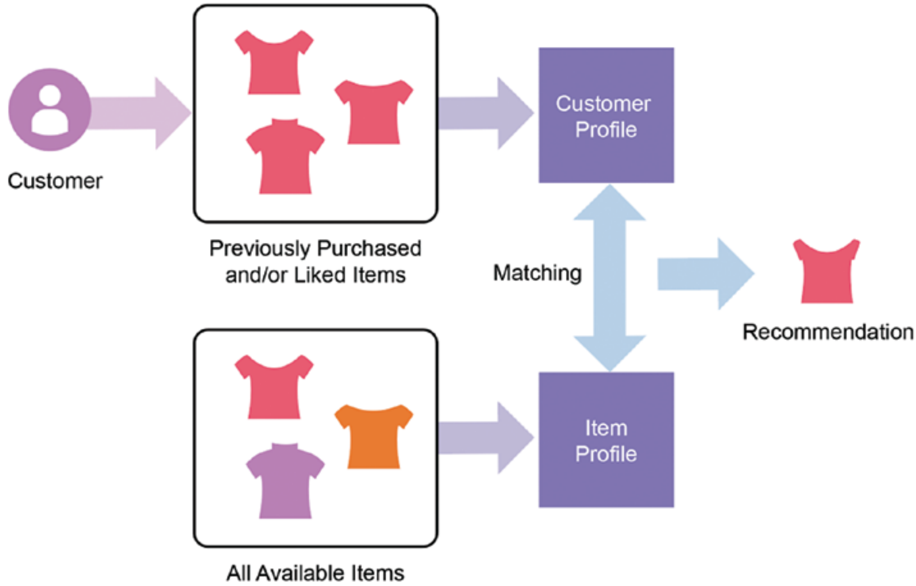

***Figure 6-4.** A customer who made a purchase in the past being recommended similar items*

Data Science

> *Data is a precious thing and will last longer than the systems themselves.*
>
> —Tim Berners-Lee, inventor of the World Wide Web

Data science is a multidisciplinary field that uses statistics and software engineering to extract business and product-relevant insights from data. Data science does not necessarily rely on machine learning, but it's becoming more and more relevant as a tool to people working in this field.

Data is often a raw set of information that requires some processing in order to have meaning. Digital data usually exists in databases, which can be structured or unstructured in nature.

What Is a Database?

A **database** gives a machine a way to know where to recall certain information. For instance, a product catalog might contain an item like a skirt, which has properties of color variations and sizes. To know where the size property is located, that information is made machine-readable in a database.

Databases are used to efficiently store large amounts of data and perform computations on that data. A database is not a machine learning concept but is important for all kinds of programming. Databases make information easy to access using machines and provide the foundation for structured data, which is data that is organized in a way that a machine can understand it. This is in comparison to unstructured data, which is more challenging to work with on a larger scale.

A basic database is a called **flat-file database**. In a flat-file database, the data exists in a single set of related data records called a **table**. A flat-file database is a lot like a spreadsheet. In a spreadsheet, you can do things like reorganize rows based on various sorting mechanisms. You can also set columns to contain specific kinds of data or apply mathematical equations to data in that column. Like spreadsheets, database tables also contain rows and columns, but there are a few important differences. In a table, columns have a set **data type** that describes what kind of data the column contains, such as a number or string of characters. The **schema** is the set of rules that defines the table. Though there may be more rules ascribed to a database table, the basic mechanisms are very similar to a spreadsheet.

In a flat-file database, each **record** in the database has its own row in the table. Each column stores some piece of data about the record. A flat-file database could be implemented as a text file that contains one row per line and has data from each column separated by commas. Figure 6-5 shows an example in which each SKU has multiple attributes associated with it. For example, “boyfriend jeans” has a category of “pants” and season “fall winter 2018.”

SKUs ← Name

id	style_description	category	season
1	silk bomber jacket	outerwear	ss_18
2	boyfriend jeans	pants	fw_18
3	slouchy tee	knit tops	ss_18
4	l/s button down	woven tops	fw_18
5	denim jacket	outerwear	ss_18
6	cardigan	sweaters	fw_18
7	fit and flare dress	dresses	ss_18
8	tank top	knit tops	ss_18

Attributes

Figure 6-5. *A visual representation of a flat-file database*

If you were to list a brand's entire repertoire this way, your database would quickly grow large and difficult to manage. **Relational databases** are useful for managing clean data by linking related tables together. Rather than having redundancies in the table entries, there can be a separate linked table with more detailed information about that attribute.

Figure 6-6 shows how you might use a relational database to define seasons. In its own table named "Season," each season can be assigned to other attributes such as "delivery dates."

SKUs

id	style_description	category	season
1	silk bomber jacket	outerwear	ss_18
2	boyfriend jeans	pants	fw_18
3	slouchy tee	knit tops	ss_18
4	l/s button down	woven tops	fw_18
5	denim jacket	outerwear	ss_18
6	cardigan	sweaters	fw_18
7	fit and flare dress	dresses	ss_18
8	tank top	knit tops	ss_18

Season

id	season	delivery_date
1	ss_18	03-15-2018
2	fw_18	10-19-2018
3	ss_19	03-17-2019
4	fw_19	10-14-2019

Figure 6-6. *A visual representation of a relational database*

Structured and Unstructured Data

Structured data refers to data that is organized. When data is organized, machines are more easily able to parse that data and make sense of it.

On the other hand, **unstructured data** is unorganized. Generally, this is more difficult for a machine to understand. This is, as mentioned, things like blog posts, e-mails, text messages, and other types of **free-form data**. As we make advances in natural language processing, there is more opportunity for machines to understand human languages. From this, we're able to posit both the content of messages written and the context that makes it relevant.

A good example of NLP in action is using Gmail. Google uses NLP to parse your inbox for information about flights and sends you reminders that you have an upcoming flight.

Data from Words

In Chapter 2, several natural language processing concepts were introduced, including sentiment analysis and word vectors. These techniques can be used to analyze natural language–based reviews, feedback, or comments left by customers. This kind of data is unstructured data; a machine cannot immediately understand what it means.

To understand natural language using NLP, some engineers use word vectors. Investing in techniques to do this analysis means learning and processing more information about your customer than you would have been able to otherwise.

It requires more than just word vectors to make inferences about what customers write, but it is possible to interpret complex phrases that are relevant to fashion businesses. For example, if a customer writes that they are in their "first trimester," you can infer that they're pregnant. If they write that they're "going on their honeymoon," you might infer that they're about to take a trip. It might be obvious to a human, but to a machine, it takes training. This training can be used to help create tailored recommendations for the end customer.

Even if your business is not a subscription service, you can use this information to segment your customer base. With specific details about a customer's needs, left from their comments and feedback about products in the subscription model, you can build more effective ad targeting, e-mail campaigns, and other marketing materials. You can also adopt these insights to shape your approach to product development.

Applications

To take a step back, what would the motive be to understanding the context of words around a product? Let's look at a hypothetical product entry from a fashion brand web site. In this case, it's this image of a tan dress in Figure 6-7. The product in this image is broken into six parts: image, title, description, price, variants, and reviews. This information is limited, because without using techniques to understand the content of the six parts outlined, we are limited to tags, categories, and specific variants listed to narrow our search.

Figure 6-7. Information can be extracted from an example product entry

Each section gives different information about the product that can be used both for search and for product recommendations. Using computer vision, the image can be analyzed to provide more information about the garment and its fit. The product title and description express garment intention—or, more accurately, what the brand intended the garment to be like. The price and variants of the garment give more information about whether that garment suits the customer's needs. Is it available in their size? Are they looking for that color? Is it a price they're willing to spend? Product reviews provide information about how that garment is perceived in the real world. Does it fit the customers who bought it? What does it feel like? Is the quality what they expected?

This maps to a large matrix of characteristics that could be pulled out in a structured way to help pair a person with a product that suits them, as pictured in Figure 6-8.

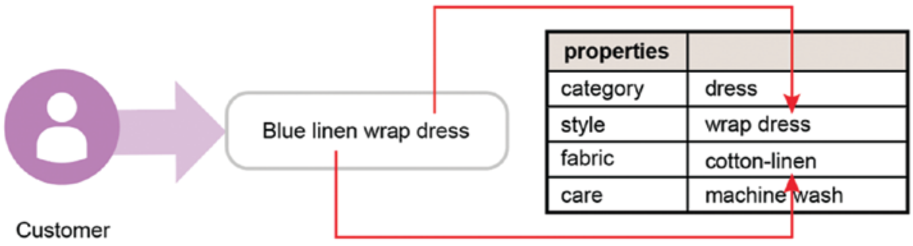

***Figure 6-8.** Product characteristics as correlated with an example customer's needs*

Think of search and recommendations as a matching game. How can we learn just enough information about what someone is looking for to deliver the right product to the right person at the time they're looking for it?

Summary

It is not fashion brands, but technology companies in this arena that are taking advantage of the insights data can provide about customers and turning them into personalized recommendations in an otherwise oversaturated market. Ironically, by building businesses this way, these services have often also found that it's more profitable and practical to create their own apparel brands. As they enter into the manufacturing realm, they're taking over market share in the fashion industry. In 2017 alone, Stitch Fix hovered somewhere around 1% market share.

Terminology from This Chapter

Collaborative filtering—A recommendation technique that makes recommendations based on the behavior of similar users.

Content-based filtering—A recommendation technique based on user actions and preferences.

Data science—A multidisciplinary field that uses statistics and software engineering to extract business- and product-relevant insights from data.

Data type—Data types help computers understand how a piece of data will be used by in a program. For example, strings contain words, text, and numbers as human language, whereas integers contain whole numbers that could be used for calculations.

Databases—A structured format which gives a machine a way to know where information is stored making it easier to recall as needed.

Flat-file databases—Store data in a plain-text file.

Free-form data—Unstructured text that is without a format. See also *unstructured data.*

Recommendation engines—A tool used to predict items that a customer or user might like.

Recommender systems—See *recommendation engines.*

Record—The basic unit of entry in a database. For example, your name, social security number, and birthday could be a record in a US Citizens database.

Relational databases—A collection of data broken into well-described tables. These can be challenging to set up but provide well-structured data.

Schema—Describes the data contained in a table of a database. Database schemas can be thought of as the blueprint for setting up the database.

Structured data—Data that is organized in a structured format, like a database.

Table—A structure in a database with columns and rows that holds records.

Unstructured data—Data that doesn't have a predefined data model or organizational structure. It is usually text heavy.