O'REILLY®

Fashioning Data: A 2015 Update

Data Innovations from the Fashion Industry

Liza Kindred
with Julie Steele

SAN JOSE

LONDON

NEW YORK

SINGAPORE

Strata+ Hadoop WORLD

Make Data Work

strataconf.com

Presented by O'Reilly and Cloudera, Strata + Hadoop World is where cutting-edge data science and new business fundamentals intersect—and merge.

- Learn business applications of data technologies
- Develop new skills through trainings and in-depth tutorials
- Connect with an international community of thousands who work with data

Job # 15420

Fashioning Data: A 2015 Update

Data Innovations from the Fashion Industry

Liza Kindred
with Julie Steele

Beijing • Boston • Farnham • Sebastopol • Tokyo O'REILLY®

Fashioning Data: A 2015 Update
by Liza Kindred with Julie Steele

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://safaribooksonline.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or corporate@oreilly.com.

Editor: Shannon Cutt
Production Editor: Dan Fauxsmith

Interior Designer: David Futato
Cover Designer: Randy Comer
Illustrator: Rebecca Demarest

September 2015: First Edition

Revision History for the First Edition
2015-09-02: First Release

978-1-491-93110-3

[LSI]

In order to be irreplaceable one must always be different.

—Coco Chanel

Vain trifles as they seem, clothes have, they say,
more important offices than to merely keep us warm.
They change our view of the world
and the world's view of us.

—Virginia Woolf

Table of Contents

Fashion: What Has It Done for You Lately?

When it comes to big data, maybe a lot.

Fashion is an industry that struggles for respect—despite its enormous size globally, it is often viewed as frivolous or unnecessary.

And it's true—fashion can be spectacularly silly and wildly extraneous. But somewhere between the glitzy, million-dollar runway shows and the ever-shifting hemlines, a very big business can be found. One industry profile of the global textiles, apparel, and luxury goods market reported that fashion had total revenues of $3.05 trillion in 2011, and is projected to create $3.75 trillion in revenues in 2016.

The majority of these purchases are made not out of necessity, but out of a desire for *self-expression* and *identity*, two remarkably difficult things to quantify and define. Yet somehow myriad different businesses are finding clever ways to use big data to do just that—to turn fashion into bits and bytes, as much as threads and buttons.

In these shrewd applications of big data are lessons for industries of all types. From the lessons of a complex lifecycle to the methods of new startups, and from merging humanity with machine learning to improving visual search, the information in this report will change the way you think about the applications of big data.

So how can we turn the emotional aspects of fashion into actionable data? What can be learned from the fashion industry that differs from what is already familiar to the Strata audience? How can humans and machines work together to help solve problems that are

at once sentimental and pragmatic? We aim to address these questions in this report.

What's Inside

This updated report takes a look at the important ways that fashion has used big data to address the complications of the industry, the importance of algorithms, and one of the biggest technical challenges in fashion and beyond: visual search. We also explore the complexities of natural language processing and its implications across industries.

Don't like to shop? Don't worry. This report encompasses the essence of how fashion brands and startups are using data to drive big sales—and how you can, too. It will also become clear that there is an overlap between fashion and other, more technical industries—relating to everything from using algorithms to relying on natural language processing. In addition, we can learn lessons from the most innovative fashion start-ups that apply well beyond the fashion industry.

One of the things that fashion has always done very well is to have two-way conversations with customers. "Most companies—Google, Yahoo!, Netflix—use what they call inferred attributes: they guess. We don't guess, we ask," says Eric Colson, who spent six years at Netflix before becoming the Chief Algorithms Officer at Stitch Fix, a personalized online shopping and styling service for women. This is an attitude that most other industries would do well to incorporate.

Trends in Fashion Data

> In difficult times,
> fashion is always outrageous.
>
> —Elsa Schiaparelli

Part of the challenge in fashion is something that every industry faces: "fitting the world into rectangles," a phrase Columbia University professor Chris Wiggins once used to describe the process of translating everyday experiences into spreadsheets.

However, fashion's rectangles are the kinds that change color, shape, and size every season—and even much more frequently these days. That rapidly shifting landscape offers big opportunities, but also comes with a unique set of challenges.

We'll start by looking at some of what makes fashion's relationship to big data unique: the emotional and unpredictable aspects of the industry; the lifecycle of the industry and how data is playing into every part of that cycle; the entire crop of new startups addressing big data in myriad ways; and the unique kinds of inputs and outputs that are making the relationship work.

Irrational Fashion

> Fashion is instant language.
>
> —Miuccia Prada

As we mentioned, fashion is not just about clothing; it's also about identity and expression. Even the most rational among us make decisions about how to clothe and accessorize ourselves based on irrational factors. We don't "need" a new pair of pants because the old ones don't function as pants anymore; we "need" them because

we don't like what the old pair is saying about us (that we're dirty, or careless, or behind the times).

Still, it's possible—and necessary—to find ways to correlate data with that emotion. Shawn Davis is currently Senior Director of Advanced Retail Analytics at Nike, and previously served as VP of Analytics at ModCloth, an online retailer for indie clothing, accessories, and decor. Shawn told us about his experience at ModCloth: "We'd be regularly sitting in meetings with our merchandising team, and listening to them describe why they think something is hot, or why they think the customer is going to love a particular product, and then trying to translate that into something that we could surface analytically in the data."

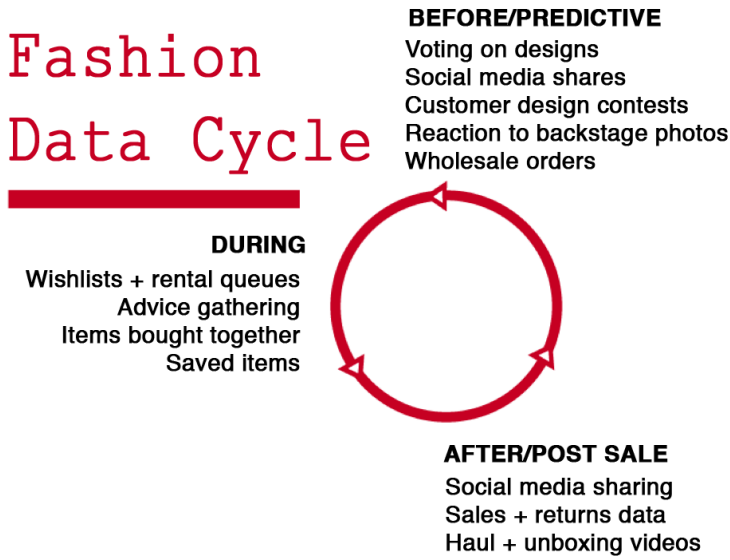
Lorraine Sanders, a San Francisco-based journalist who's written extensively about the intersection of fashion and tech, and is the host of the "Spirit of 608" podcast, puts it this way: "We're in the middle of a time when big data is becoming an important factor in just about every industry that deals with human behavior, in order to generate revenue. It's happening because, frankly, we're just hitting that time in history where the ability to collect data is becoming widespread and, in many ways, democratized."

Lorraine goes on to add that, "in a lot of ways, everyone everywhere can collect big data. It's the question of what to do with it that's the interesting part. With fashion, the data collected from consumer interactions, engagement, and reactions to products has the potential to add a ton of value in helping brands hone in on what's going to sell and become more efficient at getting the products they decide to invest in and produce to the exact right people, at the exact right time."

As Lucie Greene, Worldwide Director of The Innovation Group at J. Walter Thompson put it, "Fashion is about newness and novelty. We see something and feel compelled to buy it." Studies claim that 90% of all purchasing decisions are made subconsciously, and that those decisions are completed within 2.5 seconds. We buy products, especially fashion goods, based on having our emotions evoked in one way or another. The challenge of data is to find ways to understand, quantify, and use that emotion in a way that both serves customers' needs and drives sales.

Fashion's Data Lifecycle

Another unique aspect of the fashion industry is what is colloquially referred to as the "fashion cycle"—the time it takes to get a garment from idea, to runway, to factory, to store. What's happened recently in fashion is that consumers are an integral part of the full fashion cycle—before the fashion is even made, we see a range of consumer engagement—from designers asking for consumer votes on sleeve lengths, to brands holding contests for user-generated designs, and high-fashion brands taking consumer orders based only on samples shown on the runway. This engagement continues through the sales cycle all the way through post-sales data opportunities, such as the proliferation of online "haul" videos (consumer-recorded videos of recently purchased items) and outfit-based social media posts.



In addition to companies that are finding clever ways to use big data throughout the fashion cycle, a growing number are starting to use data to circumvent the traditional data cycle entirely.

Large companies like IBM and SAP and startups like fashion data analysis company Trendalytics are starting to tap social sentiment analysis (often correlated with historic demand) to more accurately predict trends—and specifically to identify when those trends are likely to begin and—also crucially—when they might end. Compa-

nies such as New York-based Moda Operandi, London-based Wowcracy, and Hong Kong-based LuxTNT offer customers the ability to pre-order fashions directly from the runways, instead of waiting up to six months to buy goods when they hit stores.

Regardless of whether they are trying to supplement or circumvent traditional cycles, fashion brands make copious use of different types of data during the design, manufacture, and sale of goods.

Fashion's Data Startups

When we published the first edition of this report in Fall 2014, we noted nine different fashion-tech startups that focused on big data. In the time since, the space has expanded rapidly, further proof of big data's importance to the fashion industry. Here are the different types of companies that are populating fashion's big data world.

Social Media and Influencer Analytics

Influencer marketing is big business in fashion, driving millions of hits for brands that partner with top bloggers. On top of that, social media, especially visually focused platforms, has been absolutely explosive for fashion—an industry built on knowing the "right" person and wearing the "right" thing.

CURALATE
: An analytics platform that is a darling of the fashion world.

TRIBE DYNAMICS
: A platform for fashion and lifestyle brands to track and analyze complex influencer programs.

D'MARIE
: Connects talent agencies and designers with fashion bloggers that fit their needs.

FOHR CARD
: A platform that gives brands access to verified social media stats of influencers.

Pre-Order

As the relationship between brands and shoppers becomes more symbiotic, a slew of pre-ordering platforms has opened up, circum-

venting the traditional fashion cycle in a way that can benefit everyone.

WOWCRACY
: A platform for independent designers to buy and sell goods as pre-orders.

MODA OPERANDI
: Sells luxury fashions on pre-order straight from the runway.

LUXTNT
: Asia's first luxury pre-order platform.

NINETEENTH AMENDMENT
: A platform for connecting emerging designers with customers that also manufactures the goods on their behalf.

Buying Platforms

It truly wasn't that long ago when buyers "wrote" wholesale orders, they really *wrote* them—on paper. A new class of startups is providing not only digital ordering capabilities, but all of the big-data tools needed for buyers and merchandisers to make informed product and assortment decisions.

JOOR ACCESS
: A global wholesale marketplace for hundreds of major brands.

FASHION GPS
: Tracks and analyzes samples, images, digital assets, and event attendance. Closely aligned with NY Fashion Weeks.

MODALYST
: A buying platform for connecting independent designers with small retailers.

Buying Tools

This new crop of high-tech, big-data companies is providing brands with in-depth analysis of the massively differentiated product assortments, constantly shifting trends, and rapidly shifting social sentiment that they are grappling with.

TRENDALYTICS
: A visual analytics platform for predicting consumer demand.

EDITD
: A big data tool for fashion designers, merchandisers, and buyers that quantifies trends in real time by analyzing data from retail, social, and product metrics.

WGSN INSTOCK
: A retail analytics platform from the well-respected global trend-forecasting company that uses the same taxonomy cross-platform.

Consumer Facing

Here's one of the things that fashion companies do well: if they want to know how consumers are thinking or feeling, they just ask. The direct dialogue is one of the smartest things that fashion does; these startups make it easier by providing this data collection and analysis as a service.

POSHLY
: Beauty analytics company that utilizes quizzes and contests to gather in-depth data for brands.

RANK & STYLE
: Algorithm-driven Top 10 lists for fashion and beauty, harnessing user reviews, editorial recommendations, bestsellers lists, and other buzz.

CLOSETSPACE
: A closet-data-tracking platform that gathers data and insights for brands.

Customer Marketing and Management

When consumers make emotional decisions and have extremely nuanced fashion needs—due to weather or occasion or self-expression—highly targeted and segmented information can provide the best service possible, and the highest chance for a sale. These are just a couple of the startups who are focusing on highly-targeted marketing.

CUSTORA
: Predictive analytics platform for ecommerce customer acquisition, retention, and segmentation.

DATAPOP
: An advertising analytics platform that focuses on highly targeted messages delivered at scale.

Sales Data

Going beyond the typical points of data collection, and diving deeper into things like trending products and location-based preferences, these companies are using sales data to drive deeper engagement and higher purchase rates.

42
: Tracks POS information, product, location, and customer data to provide insights.

INSPARQ
: Trending products feeds, ads, and modules; social sharing tools and tracking.

Preferences In, Fashion Out

Many fashion brands use the same software and tools as other large companies—especially other large retail companies. However, there are some ways that fashion companies gather and use data that are unique, and they have some unique inputs and outputs as well.

Many fashion brands and companies have mastered the idea of give-and-take conversations with customers. Lorraine Sanders, the fashion tech journalist, told us that, "Fashion does a really good job of engaging its audience in a two-way conversation, and that the two-way conversation that takes place can only make the big data collected from it richer and more meaningful."

One popular data-collection technique in fashion, for example, is the use of "style quizzes" that give consumers fashion advice or a curated selection of products in exchange for answering questions about their preferences (for example, see Refinery 29). In fact, it's become almost par for the course that fashion brands offer some kind of way for customers to filter products based on their style of product preferences.

"Styles" are particularly hard to quantify, as we'll outline in the section on natural language processing. While machines don't necessarily know the nuanced differences between "Boho-chic" and "Editor-

off-duty" styles, the consumer taking the quiz will have very specific ideas about whether or not they want to see a fringed bag, for example, in the search results.

Therefore, a variety of types of data collection are imperative in fashion.

Types of Input	Types of Output
Q&A/Style quizzes	Style types
Social media "shares" and "likes"	Color and silhouette preferences
Private clubs and loyalty cards	Aversion or attraction trends
Pre-ordering and ordering directly off the runway	Brand loyalty
In-store sensors; beacons; RFID	Purchase intent

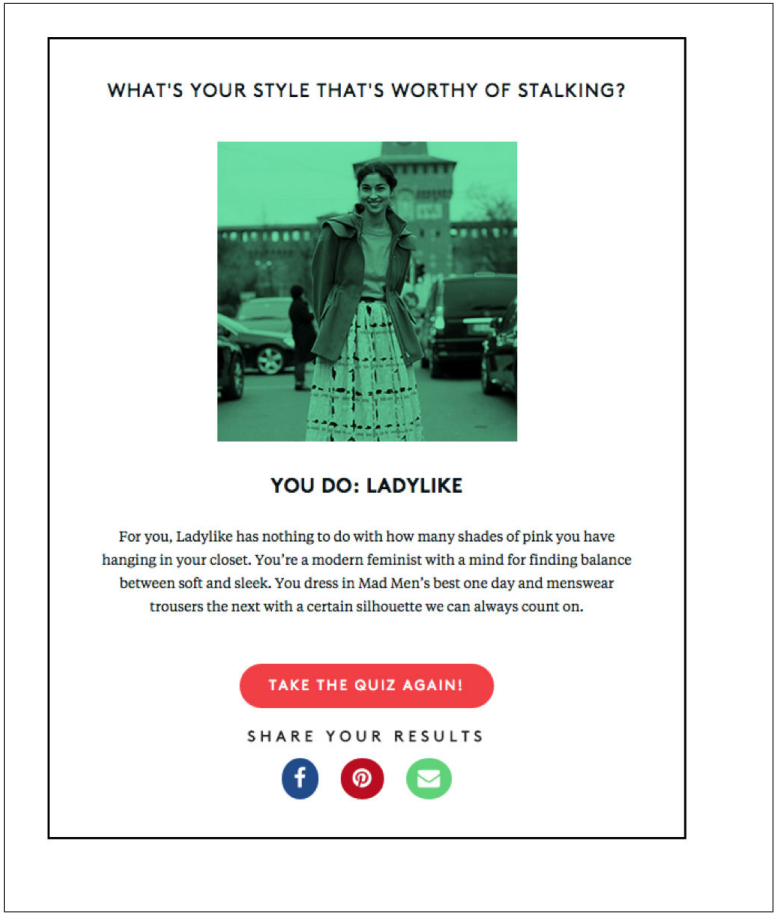


Figure 2-1. A results page from a fashion quiz on the website Refinery 29

Addressing the Challenges

> If you aren't in over your head,
> how do you know how tall you are?
>
> —T.S. Eliot

The fashion industry has some unique challenges, which we'll address in this section. At the same time, some of the challenges the fashion industry faces will be very familiar to those in any industry. Topics we'll explore in this section include the unique pace of supply and demand, the use of algorithms and natural language processing, the potential (and importance) of visual search in a highly visual market, and what's new with using data in menswear.

First, we'll explore the rapid pace of change in the fashion industry—demand for new products happens 8–20 times faster than in consumer electronics. For instance, consumers these days replace their mobile phones on average once every 30 months, while they shop for clothing as often as twice per month. Then, we'll look at geography, and how it does—and does not—affect demand.

Even when working with hard data, it's important to let the softer side of humanity shine through. We'll look at how companies are finding ways for humans and machines to work together successfully, as well as explore the challenges and benefits of focusing on natural language processing.

Algorithms are an integral part of many fashion businesses today, and we'll look at companies that are implementing algorithms in some new and interesting ways. Fashion is emotional, though, and so it's also worth comparing the use of algorithms to the continued use of curation and discovery tools.

Visual search is a huge opportunity in fashion—one that many are trying to tackle but where few are making progress. Still, it's a massive untapped opportunity, so we'll explore the various approaches and levels of success with visual search.

We'll also look at why analyzing big data has become a popular method for tackling the quickly growing menswear industry.

Let's dive in.

The Only Constant Is Change

It may be cliché, but fashion really does change constantly. The hottest colors, the newest silhouettes, and the latest "must have"—they are all in a continuous state of upheaval, by design.

For instance, the color-forecasting company Pantone announces their "Color of the Year" annually (this year it's Marsala—and yes, it is roughly the color of red wine), but they also release dozens of "official colors" for many industries each season. One key way to inspire consumers to buy new jeans, for example, is to manufacture them in the "newest" colors.

Like consumer electronics, fashion trends are designed to become obsolete and turn over very quickly. But whereas even the latest mobile phone or new laptop is usually good for at least a couple of years, clothing can go out of style in a season. Trying to do market analysis, design, prototyping, manufacturing, customer engagement, and returns/feedback at that kind of breakneck pace is dizzying.

Until recently, fashion trends were conceived by designers (with the help of trend forecasters), shown on a runway, brought to stores six months later, and then knocked off by cheaper brands, moving down the sales cycle from high-end brands to "mass-tige" (mass prestige) brands, all the way to discount store clearance bins. That cycle has been completely flipped on its head.

Now, unlike the truly exclusive runway shows of the past, everyone with an Internet connection can watch the shows live from their computers—this means that from the very moment designs are shown, fast-fashion brands such as H&M and Forever 21 set to work knocking them off. In a feat of infrastructure, these "inspired by" designs often hit the stores many months before the originals. For example, Spain-based clothing chain Zara can get goods into stores

within two weeks—at a fraction of the price that the original designs will sell for. The supply-chain logistics for this new process are incredibly complex—and supported in myriad ways by big data.

Geography as a Shorthand for Style

Fashion also changes by geography. Igor Elbert, Distinguished Data Scientist at the membership-based designer discount site Gilt, says: "Region makes a huge difference with brand recognition and so on. There are some marquee brands that are universally recognized, but still—if you plot it by country, you will see that popularity varies a lot by purchases and views."

Geography is so predictive, says Elbert, that when a new member comes to the Gilt site, the best way to create a good first impression and show the customer something she will be interested in is to *use her IP address to determine her location*, and show her things that members in the same location have liked. "Some parameters are more predictive than geography, but often geography is the only thing that we have," he says.

Of course, no one wants to wear exactly what everyone else in their neighborhood is wearing. At the end of the day, this is an industry based on self-expression—about telling the story of who each individual is, through their clothing. Therefore, many of us want to own garments that are different in some way from what our peers are wearing—but not too different, as it turns out. Despite the important influence of data and trend forecasting, designers will always have a central place in fashion.

"People do want both," says Stitch Fix's Eric Colson. "They want things that are popular: they want to look like everyone else, there are social pressures. But they also want stuff unique to them."

The business challenge here is one of scale. If you have many cohorts, and you're contending with geographic variability even on the popular trends, as well as the desire to own unique items on top of popular items, then you're talking about an inventory that is *broad* instead of deep. Not everyone is interested in segmenting by geographical locations, though. Ricardo Cuervo, Founder of Genostyle, a startup that is quantifying style data into "style genomes" for shoppers and brands, is taking a more global approach. "Rather than focusing on typical segmentation variables (geography being

one of them), we look into the style traits and characteristics that define a brand and a potential buyer (i.e., the genostyles)."

He's not ruling it out, though: "Having said that, we have the ability to 'dissect our data' both for brands and buyers, across typical segmentation variables (such as geography or other demographics) should it be of interest to any particular client." Geography can be extremely helpful for the decoding process—but it's style types that are the true goal.

Humans, Meet Machines

It's clear to see that fashion is all about people. Yes, there are supply chains and databases and sales figures in the mix, but given that fashion is a self-expression engine, it should come as no surprise that even the most data-driven startups are ultimately seeking human-scale processes and solutions.

What machines still can't do at all, for instance, is invent new popular trends from scratch. Camille Fournier is the CTO for Rent the Runway, a site that rents out designer fashion (especially dresses.) She says, "there has to be this creative element. There has to be the person who puts two things together that you never expected, and people see it and they're like, 'wow.' That 'wow' factor comes from the right mix of surprise and delight—two very human emotions that we have yet to quantify."

But that doesn't mean no one is trying. Shawn Davis from ModCloth adds: "one of the challenges for me and our analytics team, broadly, is to try to translate some of that creativity or intuition into more of a data-driven type of a structure."

When it comes to putting products in front of customers, the hybrid approach has many benefits. Many online fashion sites are using a mix of algorithmic recommendation engines and human stylists and/or buyers.

At Stitch Fix, "we use both machines and expert humans, because they're just good at different things," says Eric Colson, Chief Algorithms Officer. Machines are indefatigable and typically work much faster than people, but people are capable of understanding unstructured data much more quickly and successfully than machines. Having a stylist read a note from a customer can often tell them everything they need to know about the occasion or the customer's needs

in just a few words—such as "Bahamas cruise," for example. "This is doable with machines—just not with the confidence that we need," says Colson.

Stitch Fix also uses human experts to capture certain data that will be used by both humans and machines. "How does it hang on the shoulders, how would it look on a tall person, how edgy is it—we record all that stuff. It is expert knowledge, but it is capable of being persisted down into data, which makes it available to our stylists and machines," notes Colson.

And the machines give an analytical supercharge to stylists' instinct. "A human can use their intuition to know that a higher price means less [product] likely to be sold, but they can never tell you how much," says Colson. "They can't make exact trade-offs, but machines are great at that." Machines, as always, are great at the logic piece. But in fashion, especially, humans will always be needed for human-level, editorial filtering.

Experiences that *Feel* More Human

InSparq is a company that uses big data to drive experiences that *feel more human*. Founder Veronika Sonsev told us how InSparq is using data to help retailers market and merchandise products through features like trending product feeds, modules, ads, and through tracking social sharing.

Sonsev explained that a major challenge is that "consumers don't trust [trending] results, because some retailers (not our clients) will take things like trending products or bestsellers—and they'll just pick the things that they want to sell." This can breed distrust. She adds: "And so, customers don't trust those results—they feel like they're fake, and it's just retailers trying to hawk more stuff."

One way that InSparq approaches this challenge is to do things like show consumers the share counts of certain items. Sonsev adds, "those things create some social proof around why something is trending—and even though that doesn't completely humanize the experience, it does give some real validation, instead of being a kind of forced merchandising." Finding ways to let that humanity shine through is imperative. Shoppers care about brands, but they care about people a whole lot more.

> Style is when they're running you out of town
> and you make it look like you're leading the parade.
>
> —William Battie

Natural Language Processing

Colorways, silhouettes, style-types, seasonality, trends… all this complex information has to be understood—and the average fashion buyer or merchandiser cannot be expected to have a data science background. Natural language processing is an important part of making data useful for fashion businesses. After all: one shopper's midnight blue is another's navy—and what a buyer might consider an overcoat, a shopper might call a parka.

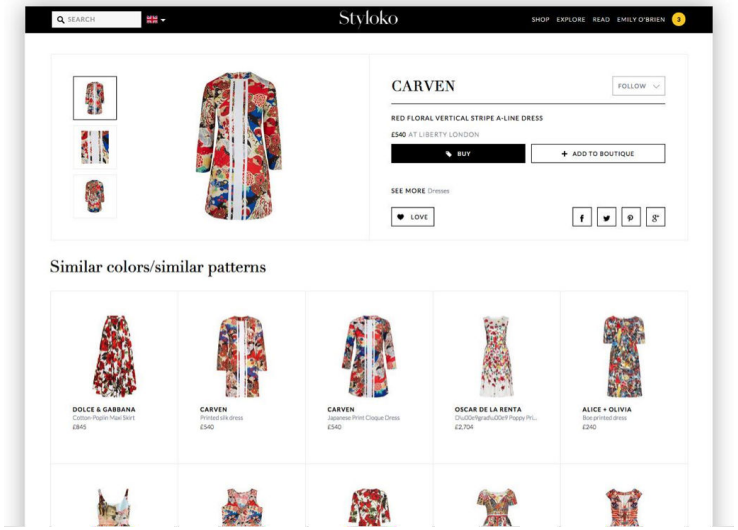
This speaks to a huge problem, in fashion as well as in other product industries: the lack of federated ontology on the web. This is the "black blazer problem": try searching the web to find a black blazer to wear to work, and more often than not, sports utility vehicles will come up in search results. The truth is that fashion is complicated, and descriptive words are often subjective. Exactly what color, for instance, is "macaroon," or "parchment"?

Figure 3-1. A common example of the issues facing the fashion industry with regard to natural language processing: is this a necklace, a choker, or strands of pearls? Are those pumps, high-heeled shoes, or stilettos?

Fashion brands to-date have had to solve this issue by building proprietary taxonomies, which at best work only for that one company. Many "fashion search engines" have tried to build federated ontologies (and so have some of the bigger search engines), but in order for these languages to work, they must be widely adopted—this is a simple definition of what federation means. Keep in mind though: this is an industry that trademarks color and considers the fit of a garment to be proprietary information. Fashion brands don't often play well with each other. This is one place where machine learning can make an enormous difference. These nuanced silhouettes and obscure names for colors (marsala!) all have to translate into something searchable, that makes sense to merchandisers as well as to shoppers. In essence, we have to use the data to make sense to (and of) humans.

A number of fashion startups of various sizes have tackled this issue head-on: by offering search engines that aggregate across multiple sites, to use human language to search a vast variety of different cross-site taxonomies. ShopStyle has made enormous headway in tackling the problem—and is monetizing their efforts through affiliate links. ShopItToMe offers daily sale alert emails, customized by consumer style and brand preferences.

Styloko bills itself as a "fashion discovery engine" and uses a variety of methods to surface results—from visually similar items to editorial content that can be "shoppable."



Keep is an app-based platform that is highly editorially focused, and surfaces products primarily through brand and influencer recommendations. Stylight tackles the issue by aggregating merchandise from over a hundred different online shops and presenting it in a way that looks like a unified store. Clearly, this is a problem that is recognized—and is beginning to be addressed in new and interesting ways.

Algorithms to Create and Decode Style Genomes

Ricardo Cuervo, of Genostyle, has set his sights on developing algorithms that break style down into quantifiable data, which they are using to build out a "style genome." It's the same fundamental idea that lets Pandora create radio stations from the "DNA" of a song or artist. Once the genome is built, Cuervo says, "That genome then powers the relationship between the retailer and the consumer, creating a common, quantifiable language of style that guides the consumer towards the products they love, and the retailers towards selling those products. In short, we are developing a platform where a 'next-gen product recommendation engine' meets 'your personal (virtual) stylist.'"

Genostyle uses natural language processing to clean and prepare the data that they use for their style prediction and matching algo-

rithms. "We are tracking close to 10,000 brands, and have decoded the genostyles of over 4,400 brands." He describes some of the company's challenges with natural language processing (NLP) this way: "We faced two challenges worthy of mentioning: applying NLP to the world of fashion, and determining the 'words that matter' to create meaningful '*fashion text quora*', by brand. We solve the first one with our proprietary fashion style taxonomy, which provides us with a vector representation of the styles in the same n-dimensional space where we model the brand styles. The second challenge requires a constant evaluation of the *text quora* representative of the brand, with an emphasis on significant style variations for any given brand at any point in time."

Today's sweet spot is often found when machine learning, specifically natural language processing, is combined with human intellect. This machine-meets-human matchup extends deeper into the general industry culture, as well, particularly in the new generation of fashion startups—many of which were built with analytics in mind from the start.

Shawn Davis of ModCloth greatly appreciates that mindset. "One of the things that brought me here was the fact that everyone in our company tends to think about how they can use information to improve their decision making, whereas at other points in my career I've found myself campaigning for why this is even something to consider. I think that mental shift is maybe one of the most notable elements."

All About that Algorithm

> Simplicity is the ultimate sophistication.
>
> —Leonardo da Vinci

Many retailers have found the importance of using hard data to drive the sales of soft goods. For fashion in particular, the move to embrace algorithms as an important part of the business model has been swift.

InSparq has a central focus on algorithms. Founder Veronika Sonsev shared with us: "InSparq helps retailers market and merchandise their trending products. We have been able to demonstrate that trends help drive sales, by showcasing what's hot." InSparq watches social media and what customers buy on site, and offers "tools to

help retailers feature those trending products on their ecommerce site, in email marketing, and also in advertising," Sonsev adds. InSparq has access to retailers' product catalogs, so they understand all of the variations of the inventory and track all of the actions against it.

"Once we capture all of that information, we run it through a proprietary trending algorithm that we've developed, that essentially puts more emphasis on recent activity than activity that's older." She explains it further: "When an action first happens, it has the most impact, and over time that impact decays (so to speak). Based on that, we rank the products, and then what our widgets do is update that ranking every 30 minutes." Retailers can choose to filter by gender, category, or price—and can exclude specific products if they like.

Social Media Is Aspirational

One discovery they've made at InSparq is that "social media is very aspirational. People share what they're passionate about—and they buy what they need." From one client, they found that people were sharing sneakers and jackets, but buying socks and baseball caps. On a baby site, people were sharing nursery items like cribs and rockers, but buying clothes and toys. Sonsev adds, "The benefit is that by featuring these aspirational products more heavily, we're basically inspiring their customers to buy more of them. And by doing that, we're able to drive higher conversion rates and higher average order value."

There are numbers to back it up: "By featuring these trending products prominently in their sites or in their digital marketing activities, we're able to demonstrate higher click-through rates, higher engagement—ultimately higher conversion rate and higher average order value." InSparq has found that people who engage with trending products are 50% more likely to make a purchase, and spend 30% more on average.

The Impact of Different Signals

But not all data is created equally. According to Sonsev, "Different signals have different impact on different retailers…. Understanding which signals really drive results has been one of the challenges that we've been trying to figure out—and how to use machine learning to

figure that out and automatically optimize. Right now we can look at the results and tweak the results for each client, but we've been trying to figure out how to automate that." There's still quite a bit to learn.

Startups are tackling the issue from different directions, too. FashionMetric is one such company—they use algorithms to offer SaaS services to retailers; their focus is on reducing return rates by matching customers to clothes that will fit well. True & Co focuses on lingerie, using algorithms to tackle the wild and wooly world of intimate apparel fit. StyleSeek uses a "Style Game" quiz to map a customer's "StyleDNA," which shoppers can then use to identify influencers they'd like to follow, or items they might like to browse and buy.

At Genostyle, algorithms for finding style types are at the front and center of the business model. Founder Ricardo Cuervo asks, "How many times do you search for a fashion item online, just to feel frustrated with the recommendations that come back? And worse, what about 'cookies' that follow you constantly, providing recommendations that, frankly speaking, miss the mark?" He cites a blog post in Marketplace that includes comments from Brian Boland, VP of advertising technology at Facebook, lamenting that big data isn't smart enough, specifically when it comes to those pesky retargeting ads. Boland bemoaned: "I would agree and you would agree that that's not a great experience for you as a person—it's just frankly annoying."

Improving Style Algorithms

The hope is that algorithms can continue to improve to the point where that experience can be improved. Cuervo is working on just that. "At Genostyle we are solving that 'annoyance' problem. While retailers have invested in predictive data tools and 'linear-regression-based' recommender engines and digital marketing models, none have sought to map their customers' style. We believe that doing so is the key to unlocking a constructive relationship. Our tools offer the means to build productive relationships between brands and consumers that are premised on information that is actually relevant to the shopping experience. And because we track that information as it evolves, both with the customer and the brands, we can empower companies to better understand their customers, and customers to better understand themselves."

Algorithm Challenges across Industries

There are, of course, challenges to working with algorithms. At Genostyle, Cuervo has faced three key challenges that will be familiar to anyone working with algorithms: "Our never-ending challenge will be to continuously improve the accuracy of the algorithms (in terms of predicting people's styles and matching them to the brands that matter). Secondly, our algorithms need to be flexible enough to adapt or reflect changes, for example in people's taste, or in changes in brand style direction, for example when a collection is launched."

Genostyles' third challenge is to isolate noise. "When a celebrity wears a particular outfit and it becomes trendy for a short period of time, we need to ensure that effect on our algorithms is properly isolated and properly accounted for in determining the true style traits of the designer or the potential buyers." Blips on the data radar, for fashion, can be as short as a walk down the red carpet.

Recommendation engines

The rapid lifecycle in fashion causes challenges as well. Sonsev from InSparq describes how they have seen the effects: "If you look at competitive recommendation engines, the main approach that they use is that they're correlating products—'people who viewed this, also viewed this.' So when your products are going in and out of stock really quickly, it's hard to get enough critical mass to make that relevant."

"The other challenge is that the product they might correlate with might not be in stock any more. If you're a smaller retailer, you also take a while to accumulate data, but by that time you're changing inventory. So what we do, on the merchandising side, is to show trending products, but to show them in the context of the page. If you're looking at shoes, we're going to show you trending shoes," add Sonsev.

"We've done a lot of testing, and what we've seen is that we know that on high-level discovery pages (like home pages and landing pages), we outperform traditional recommendations. What we're starting to test is on the product detail pages where intent is known, we're starting to test our contextually relevant recommendations, and what we're seeing right now is that we're at par with them on those pages. So, better on discovery pages, at par on product detail

pages, overall we're better. And that's really exciting, especially for small-to-mid-sized businesses."

No matter the challenges though, the use of algorithms in fashion will continue to expand, as the applications become increasingly diverse and sophisticated.

Curation, Discovery, and Inspiration—versus Algorithms

> Fashions fade, style is eternal.
>
> —Yves Saint Laurent

Curation has come to be a bit of a dreaded word in publishing circles, and that is starting to become true in fashion also. While most online retailers (in any industry) are using recommendation engines these days, the combination of such engines with human experts and stylists qualifies as something different, although whether or not that actually adds value is still a matter of debate.

Stitch Fix subscribes heavily to the value of curation. Eric Colson says that many customers write to say that they discovered a new item or style that they wouldn't have previously considered on their own. "Because they're going to be trying it on, it opens up their eyes to things they were not previously open to...there is an aspect of discovery," says Colson.

"Very few [customers] have the exact same preferences for style, but then you add in their preferences for spending, and you add in their size, their height, their weight, whether or not they're a mom, their preferences for taking a risk—all those things make a customer unique, and very quickly at that," says Colson. As a matter of fact, every "Fix" has been unique, too. "We're counting at the style level, so even if there were different colors or different sizes, we're counting that as the same. And yet, even at that definition we've never sent the same five items or styles to two different customers. Which is pretty amazing. I had to check it five times." Colson attributes this in part to the variability of customers, in part to the variability of stylists, and in large part to the breadth of the inventory. In a standard (n; k) or "n choose k" problem, he explains, the probability that two stylists are going to choose the same five items from an inventory as large as they offer is pretty small.

Companies like ModCloth also believe strongly in the aspect of discovery. "As we have more and more [products], there's a level of curation that comes into play so that we're able to help the customer find the things she's most interested in," says Davis, "without having to wade through all of the products that we have available."

On the other hand, companies like Gilt, whose inventories are smaller and revolve more quickly, are doing personalization with machine learning algorithms alone. "For example, we can say: 'Those brands often are bought together. But one of them is known internationally, and another is not,'" says Elbert. "So in terms of personalization, we can offer an unknown brand as a substitute for a known brand. It's especially important on sites like ours, where we don't sell unlimited inventory all the time." Knowing that two brands are often correlated in terms of purchases or page views allows Gilt to offer one as a substitute for the other, depending on available inventory.

Another way to facilitate discovery using only algorithms is through the surfacing of trending products. This is something that InSparq has worked on a lot—they offer trending products pages, modules, and ads. Showcasing trending products can be a great way to add elements of social buzz and add activity to a site—but it can also be a way to offer up new items to customers in an exciting way.

Fashion is, as we know, a very emotional purchase transaction, one that is often begun well before the point of sale. Discovery and inspiration sites and apps, such as Pinterest, The Fancy, Instagram, and VSCO, are shown to directly lead to sales. Unlike commodities-based businesses, fashion cannot exist without inspiration and discovery—the thrill of the hunt, if you will. While algorithms can be exceptional at surfacing items that a shopper is likely to buy, they work at their best in tandem with a strong editorial point of view. Frankly, most items of fashion can be bought at a lower price on Amazon. However, fashion shoppers often don't want to search for fashion, they want to discover it. Fashion consumers need to be inspired.

Visual Search: Oh No, You Didn't

If there's one thing on everybody's wish list in the fashion industry, it's better image processing and an increased ability to capture structured data from photographs. There are a huge variety of startups

and large companies trying to crack this valuable nut, but the technology remains wildly imperfect, often requiring enough text input that a user may as well have just searched for it in the traditional way.

Rosario Martinez is a fashion lover and engineer who specializes in machine learning. She has created an automated process for analyzing images in fashion blogs, but has encountered several challenges along the way—some relating to issues as basic as color. She reports that because the colors are interpreted by computers, they are not always accurate and the difference between orange and red, for example, can be an issue. It's worth noting, though, that humans aren't so great at this either hence the infamous gold vs. blue internet spectacle.

To mitigate this issue, she has invested time in hand-tagging a training set of about a thousand images with information about color, style, and retailer. "I used these images for building the recommendations and then, with them, I can try to relate the ones that I don't have tagged, and try to classify them," she says.

Multiple fashion tech startups claim to have solved this issue, but with each new iteration, the problems have proven to be too myriad to solve with existing technology. "It's really complicated, but very interesting...having the correct information for images would be awesome," Martinez says.

Advancements in Image Processing

Although there are many problems, there have also been a number of recent advancements.

WhereToGetIt uses humans: users post a picture and other members offer suggestions on where to find goods. Snap Fashion offers an app, an in-browser bookmark button, and a web interface to let customers search pictures for fashion items in their database. ASAP4, Trendabl, and StyleThief offer apps that use various methods to provide visual search engines. SnapUp searches through screenshots (along with price notifications on the items searched for.) In addition, there are companies starting to offer visual search as a service, including Slyce, Wide Eyes, and Cortexica.

Visual search remains so imperfect that some retailers are trying to circumvent the issue entirely. Stitch Fix's customers often send its

stylists notes and links to photos online in order to indicate things they might be interested in. "Grasping context from an image is very tricky, especially Pinterest pages. Often, customers don't mean they want that exact thing; they mean something along those lines. It's aspirational," says Colson.

Still, there is a lot of potential. Fashion-tech journalist Lorraine Sanders says, "I think that some of the visual search companies out there like Slyce and Cortexica are doing interesting things with big data that's collected through monitoring users' image-based searches and habits, and then using that information to build a case for business conducted directly with brands." Ricardo Cuervo of Genostyle adds, "We are looking into visual search technologies as a channel to extract relevant big data from visual-rich social media channels (such as Instagram and Pinterest), to feed such data into our algorithms in determining buyer's genostyles."

Visual search just isn't "there yet" for fashion, but when it is, enormous potential will be unlocked.

Mining Menswear

> Simplicity is the keynote of all true elegance.
>
> —Coco Chanel

One of the great opportunities of data science in fashion is the ability to segment the market in new ways—not just by demographics and geography, but also by size and body type. Most startups are now using data about height, weight, and basic measurements, at the very least. Men's bodies have been shown to have less variability than women's, and so we've seen companies start to digitally tackle the fit of menswear.

One such company is R.F. Madison, a startup menswear company that is utilizing big data as they prepare to launch. Kevin Flammia, cofounder of the brand, described how the company is focusing on clothing for "tall, lean guys," and is using a 3D body scanner "to generate a database of touchpoint of anthropomorphic data on the back-end, to make it a real-time aspect of our business model going forward." They are working to collect enough data to make smart business decisions about what types of sizing tall (but not necessarily "large") men need.

Flammia says, "We went and purchased a 3D body scanner…and brought on an advisor to really help us get a better sense of how we can collect data and integrate it into an apparel company. The body scanning that has been done so far has really been done on the mass market—there's nothing on the tail-end." Interestingly, he notes that "most of the fit measurements we have today still come from body measurement data that was collected during WWII by the US military." Outdated information, to be sure. "To make clothes that actually fit these guys, we have to figure out what the population distribution would look like, and find out what the appropriate size would be."

Mapping Body Shape

Inventory management—and even more specifically cash flow—are the lifeblood of a retail business. R.F. Madison is using big data to help them optimize for that flow. As Flammia describes it, "On the one hand, we're collecting data, but what we're trying to do is understand—how can we make the right sizes and optimize our inventory with that information going forward. We know that dealing with a tail-end market is super capital-intensive, and logistically challenging, but if we can reduce those returns and create the right sizes, then it becomes more of a sustainable model down the road." Customers provide this highly personal information because of the value they gain in exchange: an improved ability to find clothes that fit well.

Fashion-tech journalist Lorraine Sanders says, "There are really fascinating things happening with mapping human bodies and trying to create predictable sizing through the use of big data that could aid a great many industries—and the fashion application just happens to look like one that could make money in the short term—but the long-term implications of the technology extend far beyond that." This is certainly something for all industries to pay attention to.

Lucie Greene, Worldwide Director of The Innovation Group, agrees: "New sizing and scanning technology is also becoming an important data pool. It's started out as a way to create personalized garments, but as more consumers upload data about their real size, shape, and proportions, companies will be able to create sizing and collections that are much more accurate—removing the guesswork from clothing sizing in ready to wear."

Body shapes aren't the only things that have been found to separate men and women's shopping habits. Jenny Griffiths, the founder of London-based Snap Fashion, a visual search engine, told Techcrunch: "We held a load of focus groups and found that men don't look to celebrities for their inspiration, but their peers." This has shaped the models of many different menswear brands.

And yes, men are shopping: a recent report from research firm IbisWorld showed that the sales of men's clothing is outpacing the sales of food, electronics, and even beer and wine. Menswear is a market that, unlike so many others, remains *unsaturated*. It's a market that is projected to expand by 14.2% in the next five years. Using big data as a tool to find ways to capture that growing market makes enormous sense.

Fashion Forward

> Style is knowing who you are,
> what you want to say,
> and not giving a damn.
>
> —Orson Welles

While each of the areas we've explored still have room for improvement, their basic advantages and limitations are already known. The companies we spoke with are more or less pushing toward common goals and best practices in those areas. However, there *are* territories that remain relatively untested—and are therefore quite exciting.

Here, we briefly explore the merging of online and offline shopping, and how that affects data collection. We take a short look at the implications of data collection in wearable tech—truly a merging of fashion with big data. And finally, we make a note about the important privacy implications of these new technologies.

Online Meets Offline

The proliferation of in-store devices that can collect and share data is a topic that could fill its own report—but here we'll simply mention it and encourage further exploration. As beacons of all types become less and less expensive—and as we start to understand the best ways to use them, they will become more and more widely adopted.

In-store data can go the other way as well: "smart fixtures," such as interactive tables, connected mannequins, and tablet signage, can push data to a customer in a way that can make product information part of a brand's story-telling experience.

Many retailers are turning to another hybrid model: temporary pop-up shops. Pop-up shops are a way for online retailers to experience some of the benefits of a physical presence, and the pop-up shops can provide data and analytic insights unavailable in an online-only experience.

Melissa Gonzalez, CEO of The Lionesque Group and an award-winning pop-up store producer, explains: "When a brand does a pop-up shop, they have a unique, isolated opportunity to collect both quantitative and qualitative data, data that can drive long-term ROI. It can affect manufacturing, merchandising, and marketing decisions in the future." However, you have to be prepared, Gonzalez says. "In order to really capitalize on that benefit, brands need to have elements in place to track relevant data. Some are basic, like getting Google Analytics set up, and understanding where your drivers of traffic are coming from. It could be utilizing sites like tagboard.com, and creating a hashtag for your program, and then going back and seeing all the people who interacted with your brand. Or, it might be as simple as literally going back to your own social channels and seeing what new engagement happened and how many times you were mentioned over the time period of your pop-up."

Whether it's data collection, using data for storytelling, or both, the way that we're using big data in stores is shifting quickly.

Wearables and Big Data

This is the place where fashion and big data *literally* overlap. As an increasing number of people wear an increasing number of data-collection devices—that are, in turn, collecting increasing types and amounts of data—we'll start to see more and more discussion happening around the appropriate ways to collect, parse, store, and use that data.

From physical- and mental-health data, to location and identity information, wearable tech is often a promise made based on a device's ability to collect data—and lots of it. All of the lessons that we've learned in big data (about helpfulness, privacy, and respect, to name just a few) will need to be continuously revisited in order to build a stable industry.

Privacy, Please

Of course, data usage is not all-or-nothing; there's also the question of how much (and which) data to collect, and how to use it.

While Stitch Fix asks explicitly for some very personal data, including physical characteristics and lifestyle factors (like whether a customer is a mother and how many times per week she goes out), the company is very conscientious about asking only for data that actually goes into improving the service. "If we don't use it, we'll probably take it away," says Colson. "If it doesn't add any value, we'll probably get rid of it so we don't needlessly collect information."

As another example—InSparq doesn't have privacy issues because they don't disclose any personal information, and the company's results trend in aggregate. As Sonsev explains, "Privacy issues come up when you start using personal information without people's permission, and disclose personally identifiable information (which we don't even use)."

Physical attributes can feel personal as well—not many people would want their exact measurements made public. It is critical to remember that transparency regarding how consumer data is collected, shared, and stored is not merely a wise choice, but a government-regulated factor across industries. Around the world, government regulators generally require that any website that collects consumer data include a privacy policy that describes the company's privacy and data security practices.

What's Next?

> The data was really our goldmine.
>
> —Veronika Sonsev, InSparq

The implications of what fashion is doing with big data has the potential to resonate among a huge variety of industries—potentially to anyone that makes or sells products.

Geoff Watts, CEO of the fashion analytics platform EDITD, told *Fortune*, "We help retailers have the right product at the right price and the right time. That's the kingmaking thing in retail. When you get that right, it unlocks a fortune."

Veronika Sonsev from InSparq put it this way: "The company's first enterprise product was a sharing and rewards product. One of the things that we saw with that product was that even though that product worked really well—we could really optimize sharing—it still was only affecting a small portion of business. We got some great advice that encouraged us to think about 'how can we take that small percentage of customers who share, and amplify what they do to benefit the rest of the customers.' That's when we started looking at the data, and seeing that the data was really our goldmine."

Journalist Lorraine Sanders agrees: "My sense is that the biggest dividends from the use of big data are going to come with reduced friction in both getting products to market and getting them in front of the right eyeballs once they are there. That, in turn, has the potential to affect margins and open the door to new types of business practices and internal structures that could change the way fashion is sourced, made, and sold around the world."

For so many industries, data can be a goldmine—and fashion has taught us some valuable lessons about how to make it happen.

The Big Wishes

Today, we're able to do staggering things with big data: parse enormous datasets, predict the future with some level of accuracy, and other truly fantastic things. But, we've also got a long way to go. When we asked people what they wished big data was able to do that it can't, the responses were enlightening.

Understanding Intent to Purchase

Lucie Greene, Worldwide Director of The Innovation Group, told us, "There is still an intangible emotive aspect to consuming, and digital (recordable) behavior only goes so far to explaining what inspires us to buy something. For example, if you're in a store, you have surveyed the whole terrain of the store and only try on what you really like. When we're online we'll put anything into a shopping basket to store them for later. We're much more fickle. Connecting the dots between the intent to purchase and the actual transaction will be the next phase. Already some exciting innovators are using facial recognition and neuroscience to understand our emotions on the screen and why we buy something." She adds, "It's difficult to predict what will have this effect based on previous behavior (digital or otherwise) because it's such a subjective, individual, and emotional thing."

When we asked Shawn Davis of ModCloth, which data parameter he would wish for, this was his answer: "One of the things that we struggle with at times is the context around why customers are coming to us. If I knew that [you were looking for a gift], I could help you connect with some of our great gift items, of which we have a ton. A lot of our customers are coming just for the entertainment of it for a few minutes. Being able to differentiate on some of those scenarios would be really helpful."

Connecting Various Online Personas

Veronika Sonsev from InSparq has a different wish: "Right now, it's hard to connect all of the versions of you that exist out there. Your version of you on your phone, and your version of you at your home computer, and maybe your work computer—those are still

pretty tricky to connect. There are companies that are doing that, but it's far from perfect. And really being able to see people across all of the different ways they connect online, and know that they're the same person, that would be really powerful." She adds, "It's not that it's completely impossible—it's just that it's not great."

What Drives Purchases

Kevin Flammia from R.F. Madison would love to use big data to get an "understanding of what products are being purchased that people are never wearing. You might buy something, and even if you don't return it, you just don't ever wear it and it just sits in your closet because it doesn't fit." He goes on to say that, "understanding what the distribution is, specifically for men, of purchases driven by style and brand association versus fit. We know that there's a tradeoff—some people will chose the same brand over and over again purely because they know it fits; other people have more of a brand attachment. Understanding that would be really helpful."

Conclusion

Data science is changing the fashion industry by allowing humans and machines to work together to solve problems that are, essentially, both emotional and data-driven. The hard problems, such as visual search and creating shared taxonomies, present major opportunities for investment and innovation. Other problems—such as optimizing the way that humans and machines work together—leave lots of room for exploration.

While it remains to be seen exactly what the landscape of fashion and big data will look like in the years to come, we know one thing for sure: there are myriad methods and a large variety of companies devoted to figuring it out. It will be interesting to watch it evolve.

About the Authors

Liza Kindred is the founder of Third Wave Fashion, a fashion tech think tank, and the editor of *Third Wave Magazine*, the world's first print magazine dedicated to fashion tech and wearables. She is the author of the upcoming book about the future of commerce, *How We Buy Now* (O'Reilly, 2015), and the co-author of an upcoming book about designing for wearable tech and connected devices.

Liza speaks and teaches frequently about wearable tech, the future of commerce, and the future of stores. Third Wave Fashion works with fashion brands like Bloomingdales and Bergdorf Goodman, and technology companies like Vodafone and Cisco. Liza lives with her family in Brooklyn.

Julie Steele is Director of Communications at Silicon Valley Data Science. She is coauthor of *Beautiful Visualization* (O'Reilly, 2010) and *Designing Data Visualizations* (O'Reilly, 2012). She finds beauty in exploring complex systems, and thinks in metaphors. She is particularly drawn to the visual medium as a way to understand and transmit information.