

Data Privacy

Data Collection, Monetization, and Ramifications

Aaron Dawson, Shengjia He, Tom Lancaster, and Danniell Sotelo

Maseeh College of Engineering and Computer Science

Portland State University

Portland Oregon U.S.A.

dawson22@pdx.edu, shengjia@pdx.edu, thomas.m.lancaster@pdx.edu, dsotelo@pdx.edu

ABSTRACT

Data privacy is a ubiquitous concept in today's world. Sharing data provides insight for business, but too much of the user's identity could be revealed and lead to serious consequences. We discuss how internet browsers, phone apps, and retail stores collect data and use it to enhance customers' experiences and make money off it. The data collection industry has become profitable over the years thanks to targeted advertisements. It also has serious consequences for many people when their personal information is found out. Finally, we make recommendations on how to protect your data better.

CCS CONCEPTS

• Security and privacy~Access control • Security and privacy~Authorization • Security and privacy~Social aspects of security and privacy • Security and privacy~Data anonymization and sanitization

KEYWORDS

application, browser, data, mobile, personal, privacy, transparency

1 INTRODUCTION

The term “data” in data privacy is all personal information that is used to identify a customer or person. For example, this information can include name, address, email addresses, phone numbers, credit or debit card information, birthdates, or social security numbers. This information is typically stored in a database with different users of the database having different access privileges. Some users can see all of the information while others see only certain parts. How much of the database that should be shared leads to an important tension in the data business of gaining business insights versus invading someone's right to data privacy.

Being able to access data to notice trends and patterns makes businesses more competitive and better able to serve their customers. The aggregated data can improve day-to-day activities such as traffic congestion in major cities and give businesses a better customer profile. For example, if you are selling baby products, isn't it important to know if your customer had a baby recently? On the other end, too much information can identify individual customers and become an breach of privacy.

1.1 Identified Celebrity NYC Taxi Rides

For example, self-described data junkie Chris Whong used the Freedom of Information Act to receive over 50GB of taxi driver data from The New York City Taxi and Limousine Commission. Chris made all the encrypted data available for download online.[1] A Google software developer named Jason Hall later uploaded a deanonymized version of Mr. Whong's database for the whole internet to review.[2] A Northwestern graduate student then figured out which taxi rides carried specific celebrities thanks to paparazzi photos.[1] Through this exposed data, the world learned where the celebrity came from, where they were going, how much the ride cost, and most interestingly, how much they tipped.

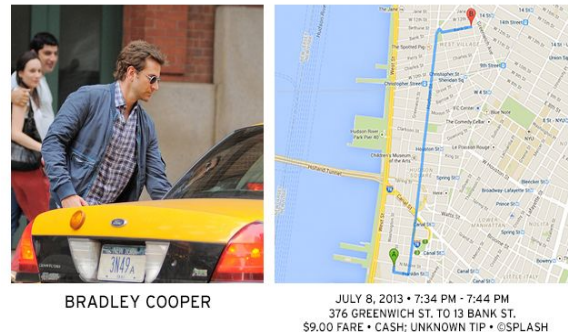


Figure 1: Through paparazzi photos, online users identified a NYC ride Bradley Cooper took in July 2013.

Is it fair that we now know that Jessica Alba, Jessica Biel, Amanda Bynes, Bradley Cooper, and Olivia Munn didn't tip their taxi drivers? Or is it fair that we can now calculate various taxi driver's income for 2013 fiscal year? Although this information is relatively harmless, what about other information that could have real consequences in people's lives? There is a reason why people put locks on cabinets and have safes in their house. There is certain information that you don't want others knowing about you and you have a right to protect it.

In the following sections, we will cover internet browser, phone app and in-store data; what is collected, how it is collected, how it is monetized, and the ramifications of this data.

2 BROWSER & MOBILE APP COLLECTION

According to a Pew Research Center survey, conducted between January 8 and February 7 of 2019, 81% of Americans report that they go online daily. Included in this figure are the 28% of Americans who report they are “constantly” online, up from 21% in 2015 when this survey was last conducted. With the increase in internet connectivity the lucrative industry of data collection is also increasing.

When we connect to the internet a plethora of data is sent out and collected but, what data is collected? This section of the report will focus on what type of data is sent out and collected and the section will be broken up into two parts, internet browser collection and application (apps) collection. First we will cover browsers.

2.1 Browser Collection

When you open up your favorite browser, type in your favorite url, and press enter what data is sent and collected? The first thing your browser shares when you go online is your Internet Protocol (IP) address. Your IP address is how you connect to the internet and will need to be shared in order to connect to the internet. However, your IP address can be used to approximate your location. Your browser, in addition to your IP address, reports its name, plugins, your desktop or mobile Operating System, Central Processing Unit (CPU) and Graphics Processing Unit models, display resolution, and device battery level. Even your mouse movements and clicks can be collected. Your browser of choice also sends out additional data which can be collected and used to create a unique signature. The browser also sends your time zone, system fonts, platform, touch support, and more. This unique signature can be used to identify you when you connect to a website.[8]

The information we have covered so far is only what your browser sends out. But, what about what the websites collect? Some websites do not collect much data while others try to collect as much as possible. But what are they collecting?

To collect information on users, websites place a HyperText Transfer Protocol (HTTP) cookie in your system. What is a HTTP cookie you may ask? According to Wikipedia,

“An HTTP cookie is a small piece of data sent from a website and stored on the user's computer by the user's web browser while the user is browsing. Cookies were designed to be a reliable mechanism for websites to remember stateful information (such as items added in the shopping cart in an online store) or to record the user's browsing activity (including clicking particular buttons, logging in, or recording which pages were visited in the past). They can also be used to remember arbitrary pieces of information that the user previously entered into form fields such as names, addresses, passwords, and credit card numbers.”[3]

Cookies are generally harmless and actually useful. However, there are cookies, such as third-party cookies. According to Mozilla,

“Third-party cookies are cookies that are set by a website other than the one you are currently on. For example, cnn.com might have a Facebook like button on their site. That like button will set a cookie that can be read by Facebook. That would be considered a third-party cookie.”[3]

Third-party cookies are often used by advertising companies to keep track of your browsing habits. When you search for a new pair of shoes online third-party cookies will remember your search and display advertisements of shoes everywhere else you go online.[4] So, if you see online advertisements on every site trying to sell you shoes that is because the advertisement company has collected data on your search history through third-party cookies. Many consider this data collection an invasion of privacy and I would agree but, this is only the tip of the iceberg. There is also a vast amount of data collected while we use applications (apps).

2.2 Application Collection

According to Statista, the number of applications available for download for the first quarter of 2019 was approximately 2.1 million for Google Play and 1.8 million for Apple App Store. Additionally, according to Statista, 95.6% of the applications are free for Android versus 90% free applications for iPhone users.[5] In order for the free applications to become profitable user's data is collected and sold which is also what the browsers do.

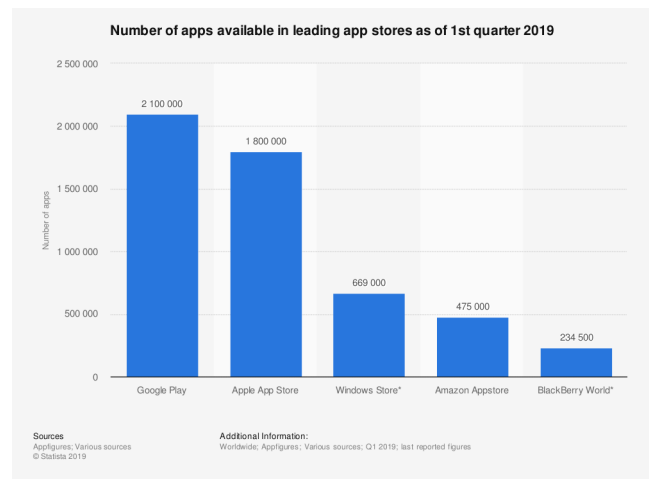


Figure 2: Number of Applications Per Application Store [5]

So, what is collected by applications? To figure this out I read *Mobile Privacy: What Do Your Apps Know About You?*[3] Written by Gillian Cleary, Senior Software Engineer at Symantec and member of their Security Technology and Response (STAR) Team. On May 3, 2018, Ms. Cleary and her team analyzed the top 100 apps in the Apple App Store and Google Play Store to figure the types of Personally Identifiable Information (PII) shared by these applications. Important to note that Ms. Cleary and her team performed their research and analysis for a blog news post for their company's website. The team did not include the tools they used or how they collected their data as a research paper would.

The data that qualified as PII in the STAR team's findings are phone number, email address, username, and address. [3] The team found that email address was the most shared PII with 44% of Android and 48% iOS applications sharing email

addresses. Username was next, 30% Android and 33% iOS, then phone number, 9% Android and 12% iOS, and followed by address, 5% Android and 4% iOS. [6]

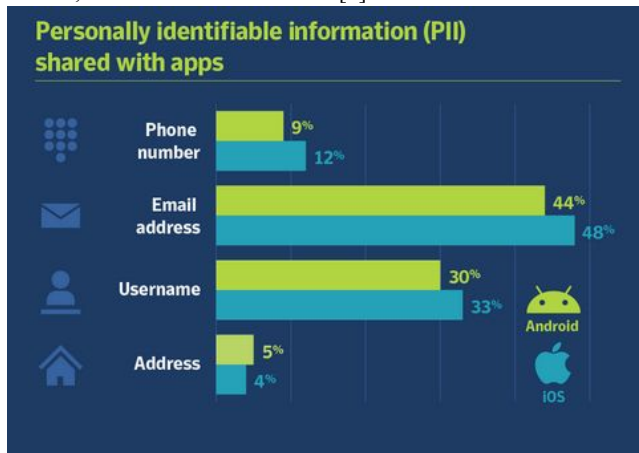


Figure 3: PII Shared with Applications [6]

The team stated that these statistics do not show the whole picture. Several applications utilized social media to login. This method allows users to login with their social media account without the need to register a new username and password, but there are drawbacks. As Ms. Cleary states, “*But this symbiotic relationship also allows the app to collect user data from the social media account, while also allowing the social media service to collect data from the app.*”[6] The team also reports they were able to see the types of PII shared by iOS applications that employed social media integration. However, they were not able to see the types of PII shared by Android applications because of the Android applications used Graph application programming interface (API). According to Facebook, “*The Graph API is the primary way for apps to read and write to the Facebook social graph. All of our SDKs and products interact with the Graph API in some way, and our other APIs are extensions of the Graph API.*”[6] The Android version of Graph also uses certificate pinning which prevented the team from seeing the types of PII shared. What is certificate pinning? In a nutshell certificate pinning is, “*It’s a security precaution that helps prevent attackers intercepting supposedly secure communications. It does this by ensuring the app only communicates with a server using the correct security certificate.*” [6]. Because social media integration, Android applications used Facebook’s Graph API the statistic stating 44% of Android applications share email addresses, and other PII, is most likely higher.

The Graph API may sound familiar because it has been in the news within the last few years. Quoting the team’s findings, “*Facebook Graph may be familiar to some people because it was used by Cambridge Analytica to compile personal information relating to 87 million Facebook users. This information was reportedly then used in targeted social media campaigns directed at voters during the 2016 U.S. presidential election campaign. Facebook responded to this incident by significantly tightening up its API and restricting the amount of personal information that can be shared through it.*”[6]

The Graph API may be the most famous integration service, it was not the only integration service nor was it the most used. The team found that 29% of iOS and 47% of Android applications used Google’s integration service, compared to 26% of iOS and 41% of Android applications that used Graph.[6]

Some applications also request permission to access other features on your device such as camera, locations, microphone, and so on. An application can request numerous permissions but permissions are not all the same. Because not all permissions are the same the team defined some permissions as “risky permissions”. Risky permissions are “*permissions that could provide access to data or resources that involve the user’s private information or could potentially affect the user’s stored data or the operation of other apps. Examples of risky permissions include access to the user’s location, contacts, SMS messages, phone logs, camera, or calendar.*” [6]

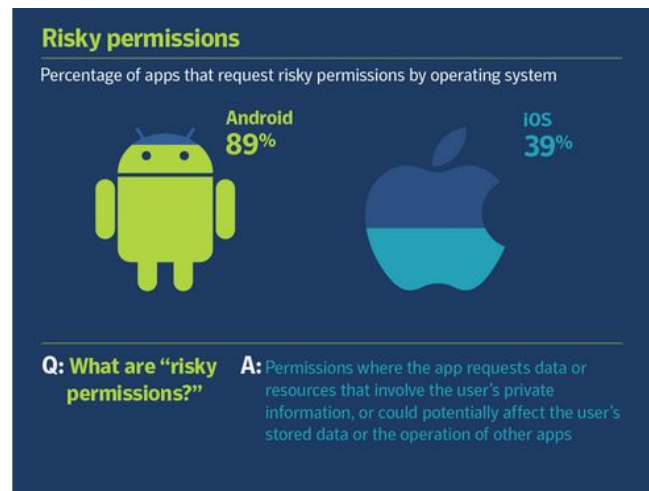


Figure 4: Applications with Risky Permissions for Android and iOS [6]

The team found the mobile device’s camera was the most request risky permission with 25% of iOS and 46% of Android applications requesting access. The statistics for location tracking was very similar to the camera statistics with 25% of iOS and 45% of Android applications requesting permission for location tracking. Next was microphone, 9% iOS and 25% Android, followed by Short Message Service (SMS), 15% Android, and lastly phone call logs, 10% Android. Reading phone call log and SMS messages is not available on Apple iOS.

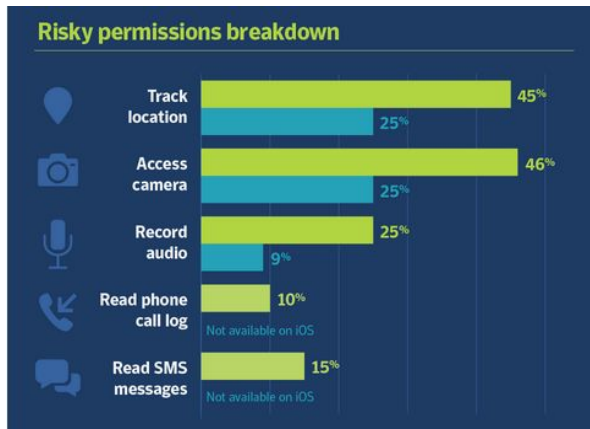


Figure 5: Risky Permissions Breakdown [6]

The team did term these permissions as risky permission however, this does not mean a user should not grant access to these requests. A majority of the time there are valid reasons an application requests a risky permission. For example, if you are using Google Maps to find directions to a particular location the application may request the user's location. This is a valid request and the application notifies the user when the application needs this feature. These permissions should make the user follow discretion when allowing applications permission to access certain features on the device. If you do not think the application needs the requested permission then refuse the permission.

The team also found an interesting case. When they analyzed the same application on both iOS and Android they found seven Android applications requested SMS messages and one Android App requesting phone call logs.[6] The iOS applications did not in both cases because this type of permission is not allowed on iOS, but they found in interesting that Android applications requested the permission when clearly the permission are not required. If the iOS application works fine without the permission, why do Android applications request these risky permissions?

A pessimistic view is that developers are purposely adding certain features to request more permissions so more data can be collected. This certainly can be a possibility for some applications but not for all. There are many players in data collection such as application developers, the application stores, the companies that own and run the stores, the companies that make the phones, telecommunications companies, advertising companies, and the user. Trying to change how data is collected is nearly impossible for user. A user does not have billions of dollars nor does the user have an army of lawyers and lobbyists. A user can control what applications to install on their mobile device. So, prior to installing and running applications on their phone the STAR team recommends a user ask themselves three questions, "Do I know what information and permissions an app is requesting? Am I comfortable sharing personal information with this app developer? Does the app really need the device permissions it is requesting?"[6]



Figure 6: Privacy Checklist [6]

These three questions are a good starting point to protecting a user's privacy. Additionally, a user should read applications' privacy policy. This in itself can be tricky because privacy policies can be difficult to read. Kevin Litman-Navaro of the New York Times analyzed the privacy policies of 150 "popular" applications and websites. He found the majority of privacy policies surpass college reading levels. Then there applications with no privacy policies. The STAR team found four percent of Android and three percent of iOS applications had no privacy policy whatsoever even though these applications requested risky permissions. Yet, as Ms. Cleary wrote, "Even when apps do have privacy policies, users can still find it difficult to keep track of what they are consenting to. While each app has its own set of permissions and privacy policies, there are several complicating factors." Many applications may need additional applications or are linked to third-party websites to run properly, linking applications and third-party websites may have their own privacy policy, and first-party applications claim no responsibility of the data used by those third-parties. [6]

The necessity of additional applications and third-party websites is troublesome because Ms. Cleary states, "A significant number of apps that request risky permissions are tied to third-party apps. Of the Android apps that require risky permissions, 40 percent have links to third-party apps. Either normal app functionality is interrupted with advertisements or there were links to third-party apps for normal functionality (for example purchase links to seller sites). Meanwhile, 16 percent of the iOS apps that require risky permissions have links to third-party apps."[6] If you are keeping track there appears to be more risk with Android applications then iOS applications. There are some reasons for this. The Apple App Store takes more time to review scrutinize a developers application before uploading the application to the App Store. The Google Play Store takes less time which is why the total number of applications is higher but there are also more poorly developed applications for the Google Play Store. Again, in the end the user will have to decide which applications and their permissions are worth utilizing.

After covering browser and application data collection, which is better? Fortunately, Northeastern University wrote a research paper on that topic in 2016. The paper is titled *Should You Use An App For That?* and their data can be accessed at <https://recon.meddle.mobi/appvswweb/> and the site is easy to use and fun to interact with. The research team include Christopher

Leung, Jingjing Ren, Christo Wilson and was led by David Choffnes. The research team sought to find out, “*given that the Web-based and mobile-app-based ecosystems evolve independently, an important open question is how these platforms compare with respect to user privacy. In this paper, we conduct the first head-to-head study of 50 popular, free online services to understand which is better for privacy—Web or app?*” [7] Their conclusion was, “well it depends.” [7]

Choffness and his team did conduct a thorough analysis of web vs application. They used free applications available in the Google Play Store and Apple App Store. The application had to be popular, free, and provide the same functionality as its web counterpart. An application that did not make the cut was Instagram because the website does not provide the same functionality as the application. They also used the default web browser installed on a mobile device, either Chrome or Safari. This is because they used a Nexus 4, Nexus 5, both using Android 4.4, and an iPhone 5 running on iOS 9.3.1. [7] They excluded applications that could not work on all devices such as Pandora. Back then Pandora did not stream via Chrome on Android. [7]. All phones were factory reset.

They found that websites leaked more names and locations than applications, but applications leaked unique identifiers such as MAC address and IMEI. [7] And there were even cases of passwords being shared over HTTPS on both web browser and application. The culprits were Grubhub, JetBlue, The Food Network, and NCAA Sports. Grubhub was sending passwords to taplytics.com, their analytics provider. Grubhub stated this was a bug and has since been fixed. The passwords were said to have been deleted from taplytics.com. Jetblue sent passwords to usablenet.com for authentication services. When contacted by the team Jetblue said, “that in addition to using encryption to send the password over the network, it is also encrypted before storing.” The Food Network and NCAA Sports send passwords to Gigya. Gigya is a third-party identity management service. [7]

They found that applications leak more PII than websites which makes sense. Applications on mobile devices can directly access more types of PII than a web browser can. Additionally, Education and Weather services contacted more third-party domains than other categories of services. The majority of third-party domains on both applications and websites were to advertising agencies. They also found Facebook as the most contacted domain by the applications they tested. [7].

At the conclusion of their study there was no clear winner. Websites contacted more domains than applications which accessed their cookies and installed third-party cookies. The applications however leaked more device identifiers which can easily be used to identify the individual using the device. Similarly to the Symantec analysis done by Ms. Clearly, the team for Northeastern University stated, “In short, there is no single answer to the seminal question in this work; rather, the answer depends on user preferences and priorities for controlling access to their PII.” In lieu of legislation or mass customer revolt, the responsibility of protecting user’s data privacy currently falls on the user. You must ask yourself if using a website or application is worth sharing your data. For me, the answer is sometimes yes and sometimes no.

3 RETAIL STORES

In daily life, we go shopping when we need to purchase something. If we enter stores and begin to do some shopping, a business relationship between stores and customers will be built up. This relationship must be based on the information related to both sides. The information of the stores certainly refers to the goods they sold in the store while the customers’ information is partly about their privacy. Therefore, what information about customers do the retailers want to know and how do they get it? I would like to give some examples to answer these questions.

3.1 Fred Meyer

On the website, you can see customer service. [9]. It includes Contact Us, FAQs, Privacy Policy, etc. From these items, win-win result can be seen. On one hand, to save customers’ time and money, and to make their shopping experience better, stores collect their customers information. On the other hand, to create a better, safer experience, develop new products and services, and better understand the use of products, services and websites, stores need to collect customers’ information. Stores collect customers’ information including name, address, phone and email, payment information, shopping preferences, even your password, and your driver’s license number.

There are many methods they can use to collect all kinds of information they need, for example, when customers want to make a call to the stores, track their packages and use the stores’ website to do shopping. Some of the technologies can be used, such as the cookies. I will talk about them.

In the Fred Meyer’s website, it has “Customer Service” web page for helping customers. There are six sub-web pages in the “Customer Service” web page, such as “Contact Us”, “Track Order”, “FAQs”, “Returns”, “Privacy Policy” and “Terms & Conditions”. This time I will talk about the “Contact Us” and “Track order” how to collect customer’s data.

The picture below is a screenshot of the “Contact Us” from the web page. [10]. In this web page, there are many functions for customers to make a call, such as “My Prescriptions”, “Gift Cards”, “Floral”, “i-wireless”, etc. If customers want to make a call to “Prepaid Debit Cards”, customers need to provide their personal information to the stores, because the stores can help customers with the help of the information. Based on the FAQs webpage, we can know what information the stores can collect according to the customers’ questions. [11]. For example, “how do I pay using a credit or debit card?” The stores can get your payment information by typing your card number. The stores welcome customers’ comments, questions and suggestions on their goods.

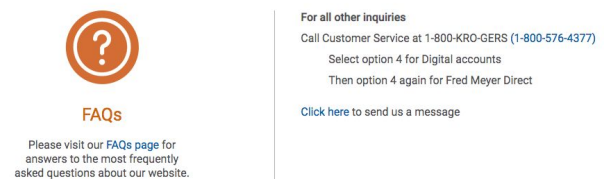


Figure 7: Contact Us

Besides the stores can also use “Track Order” program to collect customers’ personal information. [12]. I made a screenshot of “Track Order” from the web page. It is clear that it is very easy to get the customer’s information by filling this form including your zip code, order ID, Billing Last Name and email. When customers provide their zip code, the stores can know where you live.

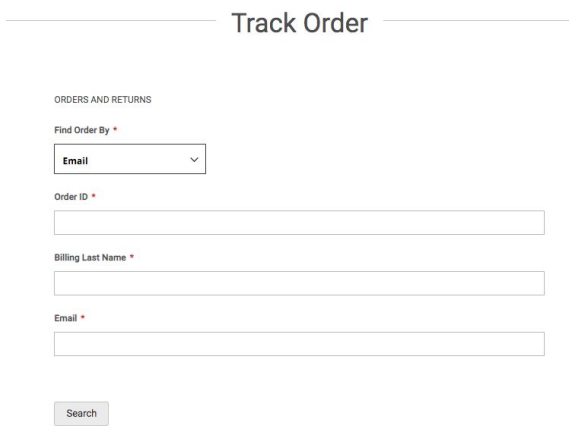


Figure 8: Track Order

The Fred Meyer wants to give more benefits to customers, the stores launch the “Prepaid Debit Card” program. This program can help customers benefit a lot including earn meaningful rewards, no credit check, no overdraft fees, international usage and the card security. If customers know these benefits of this card, customers would apply for it. I made a screenshot of the application form of this card. [13]. In the form below, customers need to provide their personal information, such as SSN, the full name, physical address, email address, etc. Through the physical address, the stores can know where you live.

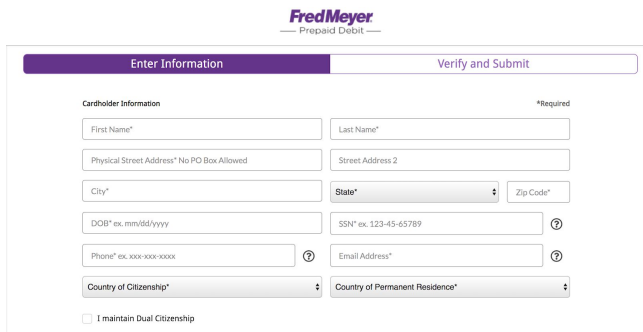


Figure 9: The application form of Prepaid Debit Card

Furthermore, the Fred Meyer also use technology methods to collect customers’ personal information. The stores use cookies to do it.[14]. Cookies is a small piece of data. It comes from a website. It can be stored on the user’s computer when the user is browsing. Cookies contains information which can be read through the web browser. That is to say, when customers visit the store websites, use the application or read one of their emails, their information is collected by tracking

technology, such as customers’ Internet Protocol address, system and browser type, etc.

3.2 Walmart

On the website, you can see customer service. [15]. It includes Order Status, Returns, Gift Cards, Pharmacy, etc. From these items, both the stores and customers can receive benefits. On one hand, while doing shopping, customers can save their time and money, and have a good shopping experience. On the other hand, in order to enrich a better and safer experience of selling, develop new products and services, and better understand the use of products, services and websites, stores need to collect customers’ information.

When you interact with the stores, websites, and mobile application, customers’ name, email address, physical or postal address, phone number, date of birth, and payment information. In addition, customer’s device information, browsing information and location information are included. [16].

There are some methods to do so. Walmart collects data through volunteered data, Walmart MoneyCard [17], and facial recognition system. I will talk about them separately.

With the development of technology, customers can do shopping online or in stores. When shopping online, customers have to register an account on the store website or mobile services. Participating in Saving Catcher, and requesting customer service, customers offer different types of personal information. Besides, in order to improve the quality or personalization of service, consumer reporting agencies provide customers’ personal information.

Walmart launches a Walmart MoneyCard program. [18]. The function of Walmart MoneyCard is that it pays to save, that is, the more customers pay, the more they save. If you fill this form to apply Walmart MoneyCard, your personal information will be offered to Walmart, including your social security number, and account balances. Even when customers are no longer Walmart MoneyCard holders, Walmart continue to share their information.



Figure 10: FIGURE DESCRIPTION

A news report comes from BBC.[19]. It says that Walmart has confirmed that it uses image recognition cameras at checkouts to detect theft. It also says the cameras track items rather than people. If an item is spotted being put in a shopping

bag before it has been scanned at the checkout, the system can call an employee to “help”. As the retailer said, the loss of products due to theft or error had decreased since the technology had been deployed. But the cameras also capture the customers’ personal information which may enter the stores’ database.



Figure 11: FIGURE DESCRIPTION

3.3 Starbucks

Starbucks is a retail coffee and snacks store. When customers receive from Starbucks online services or visit stores, they should offer personal information including full name, birthday, gender, username, password, email address, postal address, phone number, financial account information. Besides, customers’ purchasing information, device and usage information, location information are included. [20].

Starbucks uses a variety of technologies to collect information about customer’s device and use of store’s websites and mobile applications. This section I will talk about some technique skills which collect customer data. Starbucks uses these technologies to collect information about customers’ browser or device, including the type of device customers are using, operating system, browser, internet service provider, etc.

There are many sections about customer services in the webpage. [21]. I will use one of them to talk about how to collect customers’ information reflectively. From the screenshot below, customers need to fill up their full name, email address and card number. [22]. It can also be known that the personal information is collected by Starbucks after filling the rewards card application.

Starbucks Cards

Choose a general topic... *

Message *

Card Number

First Name *

Last Name *

Email *

Figure 12: Starbucks Card services and general Card questions.

Starbucks has a rewards loyalty program. This program can let the new Starbucks Rewards members to earn Starts toward free Rewards from the day they join, bringing immediate value to customers. When customers earn 25 starts, customers can get extra espresso shot and dairy substitute or additional flavor. Similarly, 50 stars can earn brewed hot coffee, hot tea or select bakery items. 200 Stars can let customers have lunch sandwich, protein box or salad. Customers can hardly resist these temptations. If customer wants to register into this reward loyalty program, they need to provide their personal information, such as full name, email address and their password. [23].

PERSONAL INFORMATION

First name

Last name

ACCOUNT SECURITY

Email address

Password

Figure 13: The application form of account

3.4 Safeway

Safeway is an American supermarket chain. I will talk about what customer’s data is collected and how Safeway collect customers’ data.

To enjoy newsletters, articles, product or service alerts, new product or service announcements, saving awards from Safeway, customers need to download and register application, register at store websites, or register for a Safeway Club Card.

The personal information will be collected by Safeway, including name, age, address, telephone number, bank card number, transaction related information associated with a Safeway Club Card, etc. [24].

This part I will talk about how Safeway collect customers' information. I only talk about web technologies and club card. In order to help store to recognize customers, customize or personalize their shopping experience, Safeway uses cookies, web beacons, Uniform Resource Locators (URL) and similar technologies to collect and track information about customers and customers' activities online at the store website, such as customers' computer IP address and operating system, browser type and information about customers' preferences.

In order to prompt many customers to enter the store to do shopping, Safeway launches the Club Card.[25]. The club card is a membership card that can be used to save money on grocery shopping. The Club Card application requires that the individual provide their full name and current address. Customers need to provide their driver's license number and birthday.

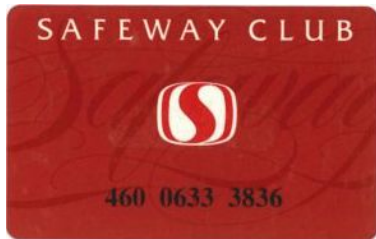


Figure 14: Safeway Club Card

4 DATA MONETIZATION

Why is all of this data being collected in the first place? The answer to this question is simple: because collecting all this data is profitable. This fits with the definition of what it means to monetize something. At its core, monetization is the act of finding ways to gain value from something, and companies are doing just that; finding ways to gain value from data collected from users.

4.1 Monetization in Stores

Let us start by looking at how monetization works in retail stores like Walmart, Fred Myer, and Safeway. Like any business, their goal is to increase net profits. Simply total up all the money you made as a business and subtract from it all the expenses you had to make to get that total, and what is left is your net profits. From this equation, it is easy to see the two ways a business can increase their net profits. They can reduce their expenses, and they can increase their total revenue. In both cases, customer data can be used to increase this net profit.

4.2 Reducing Money Spent

In the first case, the store will try and reduce the amount of money they spend. There are two main ways that a company can use collected customer data to reduce spending. They can Minimize Churn, and they can improve internal processes. [26]

Churn refers to losing a customer, and reducing this is beneficial for the company. Reducing churn helps reduce the amount of money a store spends because it is cheaper for a

company to keep its customers than it is to find new ones. [27] These companies will spend time and money on building prediction models that are able to predict if a customer is likely to churn, the likely reason why, and what can be done to retain the customer.

One reason to churn might be because of a health and safety concern. When faced with a recall, any store will want to warn their customers as soon as they possibly can in order to tackle the problem head-on. In turn showing their customers that they care about their safety. This will hopefully decrease their likelihood of churning. However, instead of informing absolutely every shopper on file, what if a company could personally inform only those customers who purchased the recalled item? With all this data being collected on what each customer buys and when that is exactly what happens. After a fruit recall, "Costco took just one day to use the data it collects on its members to create a list of all the customers who could have purchased the potentially dangerous fruit." [28]

Internal processes are all the small tasks/job activities that a company has that allow it to function smoothly. There are security-related processes (like checking bags and having theft protection), safety-related processes (like cleaning up spills, and incident reporting), and merchandise-related processes (like inspecting stock, replenishing product, and maintaining a nice atmosphere for the customers). [29] All of these processes are things that Walmart, Fred Myer, etc. will have, and some of these internal processes can be reviewed and changed because of the data being collected on their customers.

To understand how these internal processes can be improved, let's look at a few examples. Over the last few months Store "A" has experienced increased pressure by a lot of shoplifters. Although, these aren't your typical "customers," data is still being collected on them. How they act, how they steal, and what is stolen can all be seen on the surveillance footage. Over the months of data collecting, a pattern is found; almost all of the shoplifters come in with backpacks, and the most frequently stolen objects are batteries. From this, Store "A" creates a new rule for customers: "All backpacks should be left at the front of the store or be subject to potential searches by the loss prevention team." In addition to this, Store "A" moves almost all of its stock of batteries to be next to the registers. A final change is made to the product stocking team's process. They are instructed to place a maximum of two packages per item type on the shelves that are not close to the registers.

In this example, we see that the company notices that there is something less than ideal happening, and uses the data being collected to change one or more of their internal processes. Shoplifting was less than ideal, so they observed months of data, changed some loss prevention processes as well as stocking team processes, thereby reducing the amount of money they spend rebuying/restocking those items which were stolen.

A second example showcases a less extreme scenario. Store "B" uses their rewards card program to collect data on what each customer buys. Using this information, a program has been made that tallies up the movement (number of items sold of a particular product) of perishable goods. The store sees a lot of perishables expiring on the shelves and wants to reduce this. The program reports any edge cases it detects. An edge case can either be a fast-moving product, or a slow-moving product. Store B gets

a report from this program every week and uses it to adjust how many cases of each product they buy. Over time, Store “B” gets better at predicting how many perishables they will sell, and have changed the way they order certain products in order to reduce the number of items that expire on the self.

Again, we see that the company notices that there is something less than ideal happening, and uses the data being collected to change their internal processes. Products expiring on the shelves were less than ideal, so they created a way of organizing the data, and changed the process of ordering some products, thereby reducing the amount of money they spend on products that expire in the store.

4.2 Increasing Total Revenue

Now we take a look at the second way a business can increase their net profits, by increasing their total revenue. Just like the first part of the equation, there are two main ways that a company can use collected customer data to do this, by providing new products or services and maximizing customer satisfaction. [26]

Providing new products or services can be greatly aided by the collection of data. Many organizations analyze data trends to inform their decision-making processes. On July 19, 2017, Wells Fargo produced 5 steps highlighting this process: First, identify what data is to be tracked. Second, plan how frequently the data will be polled. Third, find the variables that will be represented. Fourth, graph the collected data. Ending with step five, analyzing the results and using the data to inform decisions. [30] Using this, stores can determine things like which products customers typically buy together, and when certain items are more popular.

Increasing total revenue can also be achieved through good customer satisfaction. In a review by Peter Kriss, Published in the Harvard business review in 2014, it is shown that a good customer experience score of seven or better equates to a customer spending 50% more, and up to 140% more for customers scoring their experiences a 10/10. In addition to this, good customer satisfaction also fends off churn. [31]

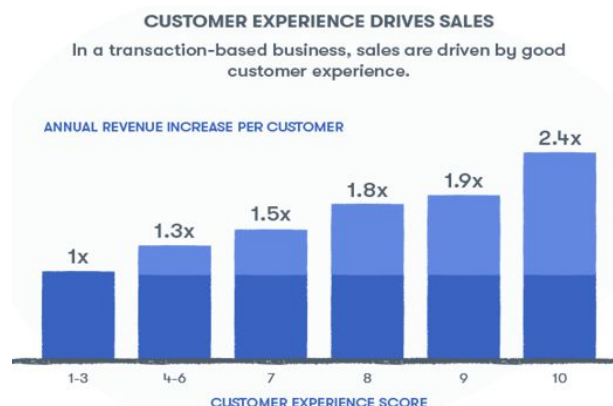


Figure 15: Sales Organized by Customer Rating

This is why so many companies put effort into collecting data on their customers’ satisfaction, measuring and monitoring what the

customer thinks about their experience helps stores figure out what they could do better.

4.1 Monetization Online

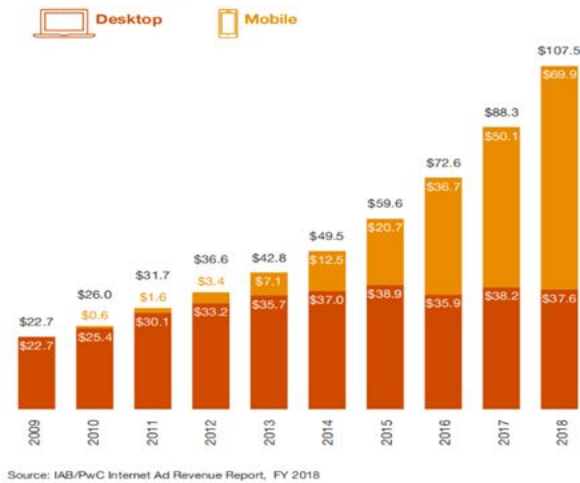
So far, we’ve focused on monetization in stores and shops, now let’s shift to talking about the monetization of user data in an online setting. First, we will talk about how much a user’s data is worth. Then we will go over how this data is monetized, and finish with some statistics on said monetization.

So what should the value of an individual’s data be? This is a rather personal question to answer as it almost entirely relies on what kind of data we are talking about. For example, “small” stuff like name, age, and location, net less than a penny at \$0.0005, whilst your SSN can be valued at over a hundred. [32] [33] The Financial Times built a calculator for calculating the approximate worth of an individual’s data. The average person is usually within the range of \$0.20 - \$0.40. If we take a look at Facebook’s current active user base as of the second quarter of 2019, 2.41 billion, and multiply that by our range, Facebook has around \$482 million to \$864 million worth of user data. \$0.20 – \$0.40 is nothing compared to an SSN however. As shown by the Equifax data breach of September in 2017, our SSN is apparently valued at a maximum of \$125 [33].

As concerning as a leaked SSN is, it is not the norm. As stated above, the data that companies have on you is usually worth far less than a dollar, but how does it make money for these online companies? From web browsers and apps to companies like Facebook, the main way this data is monetized is through advertising.

When it comes to advertising “Companies are capitalizing on their trove of customer data, including purchasing and browsing history, enabling advertisers to programmatically reach audiences in more effective ways.” [34] Much of the data that is collected on us will be stored with data brokers. Companies such as Acxiom, Experian, Epsilon, CoreLogic and many others, collect data from public records, networking and social media sites, and can even get access to past purchase history and viewing habits using the aforementioned strategies highlighted in section 2. From this, the data brokers create profiles that link all the data from all of these sources for each individual. These profiles are “often made up of thousands of individual pieces of information, such as a person’s age, race, gender, height, weight, marital status, religious affiliation, political affiliation, occupation, household income, net worth, home ownership status, investment habits, product preferences and health-related interests” [35]. These profiles are then sold to various other groups, with most of them using the data to target users with advertisements.

Figure 16 is a bar graph provided by The Internet Advertising Revenue Report which compares desktop and mobile advertising. In 2018, the combined internet advertising revenues in the United States totaled \$107.5 billion for the full year. That’s an increase of 22% from the previous year and a 48% increase in two years. Of this \$107.5 total, 70 billion of it was from ads on mobile devices. This means that as of 2018 advertising on mobile devices in the U.S. makes up over 65% of all reported internet advertising revenue. [34]

Desktop vs. mobile internet advertising revenue
(Full year results, \$ billions)**Figure 16: Yearly Total Revenue Comparison (Billions)**

4 RAMIFICATIONS

What are the ramifications of third parties obtaining our information either by consent or illegal means? It is a double-edged sword. There are times when this information enhances the customer's experience and times when it negatively affects the customer's wallet and Here are a few scenarios of how personal data can be used from benign to malicious.

4.1 The Good

First, I'll present a benign scenario. In the movie *Big Fat Liar*, Frankie Muniz's character Jason Shepherd steals the palm pilot of a dishonest movie executive named Marty Wolf who stole his high school assignment and made a big motion picture film out of it. Loaded with his daily itinerary from the palm pilot, Jason puts blue color dye in his pool before Mr. Wolf's daily swim, orange dye in his shampoo, and super glue on the earpiece of his frequently used headset.[36] Obviously, these are more like pranks that don't have lasting negative repercussions.

Another common benign scenario is targeted ads. As previously mentioned, this is the most lucrative part of collecting data which is then used to provide ads that line up with the interest of the user. For example, when I was looking for engagement rings for my now wife, all of the sudden Facebook and other websites started showing me engagement ring ads. I'm sure you all have many similar situations. This is good for both the customers and businesses since it puts products and services in front of customers who are more likely to buy said product or service.

4.2 The Bad - Price Discrimination

On the flip side, companies can price discriminate against their customers. Price discrimination is when companies sell identical goods and services at different prices depending on a various contingencies. A classic example of this is the airline industry. A price for an airline flight can vary according to seat selection, time of day, day of the week, time of year, how close a

purchase is made to the date of travel, or even whether you cleared out the cookies, or data packet that tracks your online activity. For example, if you shop at higher-end priced stores, then this signals you have more disposable income and therefore, aren't as price-sensitive as other customers. This means the same exact seat on a flight could be higher for you than other potential buyers.

4.3 The Ugly - Catfished NBA player and Instagram Model

**Figure 17: China's surveillance camera recording their citizens every move which affects their social score**

There are many scenarios though where data can end up in the wrong person's hands. In 2012, then NBA player Chris Andersen's home was raided for potential child pornography. [37] He was prevented from entering his workplace and was eventually released from his team the Denver Nuggets. [38] His case wasn't resolved until six months later where an investigation revealed that he was "catfished" by a 31-year-old Canadian woman Shelly Cartier. She impersonated both Andersen and then 17-year-old aspiring model Paris Dunn who claimed she was 18 years old at the time. [37] Imagine your professional career being put on hold for six months. Imagine the mainstream media associating your name with child pornography as federal prosecutors investigated whether you were guilty of such a serious crime. All of this happened because a bored woman managed to hack the personal social media accounts of both Chris Andersen and Paris Dunn from the comforts of her own home. This shows you how easily your data can be used against you.

4.4 The Ugly - China Social Score



Figure 18: China's surveillance camera recording their citizens every move which affects their social score

In China, they are now keeping track of individuals' "social credit scores." Since China has security cameras everywhere, they are able to track every move of every citizen. The exact methodology of keeping score is secret, but an example of actions that lower your social credit score include smoking in non-smoking areas, bad driving, buying too many video games and posting fake news online. Negative consequences of a poor social score include preventing citizens from traveling, having higher internet speeds, putting your children in the school of your choice, being accepted to certain jobs, having your dog removed from home, and being publicly named a bad citizen. [39]

4.5 The Ugly - Future Job Applications

As computer science students aspiring for future tech jobs, imagine being rejected by a potential employer because you didn't exercise enough or frequently ate fast food. How about you vote with certain political party or affiliate with a certain religion that the company doesn't favor? Should job applicants be penalized for private habits that does not affect your work performance? Some data blackmail can hit too close to home.

4.6 The Ugly - Antifa Surrounds Home of News Anchor

In November 2018, Antifa protestors found Fox News anchor Tucker Carlson's personal address and mobbed his Washington D.C. home. They chanted, "Tucker Carlson, we will fight. We know where you sleep at night." They rang the doorbell repeatedly, broke a door and mentioned having a pipe bomb. Tucker and none of his four children were home at the time, but his wife was. She had to lock herself in the pantry and call the police. [40] This is a real-life consequence of people having your location information. They could potentially even get your real-time location activity from phone apps. This gives others the power to intimidate, stalk, or do something potentially even worse to you or your loved ones. This is another reason why it is so important to keep data private. It could mean your sense of safety or even your life.

5 SAFEGUARDING YOUR DATA

Access to personal data has real-life consequences and therefore, needs to be protected as best as possible. The following are industry recommended data protection measures:

- Use passwords on all devices (phones, tablets, computers, etc...)
 - No easy entry to your device
- Change your passwords regularly (2-3 months)
- Avoid free Wifi services such as airports or grocery stores
 - These wireless networks are usually not encrypted
- Update the software on your device
 - Fixes any vulnerabilities in your device's security
- Use two-factor authentication
 - Adds another layer of security that typically requires a phone or biomarker such as a fingerprint
- Freeze your credit when you are affected by a data breach
 - Minimizes any financial risk by keeping your credit history in that place and time which prevents fraudulent charges from affecting you
- Install a password manager
 - Creates a virtual vault of strong passwords that only requires one password to enter
 - Options include Dashlane, 1Password, KeePass, and LastPass [41][42]

User data collection is being implemented across all aspects of life which provides a sense of responsibility for both the businesses using the data and the customers themselves. This data can have benign to dangerous ramifications that could affect your finances, the ads you are presented or even the lives of you and your loved ones. Using the measures above helps you minimize your risk of data falling into the wrong hands and being able to use the great services that many of these businesses employ.

ACKNOWLEDGMENTS

We would like to thank Katherine Tufte, Professor at Portland State for sharing her pearls of wisdom and comments that greatly improved our paper. We would also like to thank Bruce Irvin, fellow professor at Portland State, for his assistance throughout the course.

REFERENCES

- [1] J.K. Trotter. Public NYC Taxicab Database Lets You See How Celebrities Tip. Retrieved August 8, 2019 from <https://gawker.com/the-public-nyc-taxicab-database-that-accidentally-track-164-6724546>
- [2] Chris Gayomali. 2014. NYC Taxi Data Blunder Reveals Which Celebs Don't Tip—And Who Frequents Strip Clubs. (October 2014). Retrieved August 8, 2019 from <https://www.fastcompany.com/3036573/nyc-taxi-data-blunder-reveals-which-celebs-dont-tip-and-who-frequents-strip-clubs>
- [3] Wikipedia contributors. 2019. HTTP cookie. (August 2019.) retrieved August 7 from https://en.wikipedia.org/wiki/HTTP_cookie
- [4] Andrew Perrin. 2019 About three-in-ten U.S. adults say they are 'almost constantly' online. (July 2019) Retrieved July 29, 2019 from

- <https://www.pewresearch.org/fact-tank/2019/07/25/americans-going-online-almost-constantly/>
- [5] Appfigures, & VentureBeat. (August 6, 2019). Number of apps available in leading app stores as of 2nd quarter 2019 [Graph]. In Statista. Retrieved August 12, 2019, from <https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/>
 - [6] Gillian Cleary. 2018. Mobile Privacy: What Do your Apps Know About You? (Aug 2018) Retrieved July 28, 2019 from <https://www.symantec.com/blogs/threat-intelligence/mobile-privacy-apps>
 - [7] David Choffnes. 2016. Should You Use The App For That? Comparing the Privacy Implications of App- and Web-based Online Service (2016)
 - [8] David Nield. 2017. Here's All the Data Collected From You as You Browse the Web. (December 2018) Retrieved July 20, 2019 from <https://gizmodo.com/heres-all-the-data-collected-from-you-as-you-browse-the-1820779304>
 - [9] Anon. Customer Service. Retrieved August 13, 2019 from <https://direct.fredmeyer.com/customer-service>
 - [10] Anon. Contact Us. Retrieved August 14, 2019 from <https://direct.fredmeyer.com/contact-us>
 - [11] Anon. Frequently Asked Questions. Retrieved August 14, 2019 from <https://direct.fredmeyer.com/faqs>
 - [12] Anon. Track Order. Retrieved August 14, 2019 from <https://direct.fredmeyer.com/track-order>
 - [13] Anon. Retrieved August 14, 2019 from <https://fredmeyerenrollment.mmsselect.com/dtp>
 - [14] Anon. Privacy Policy. Retrieved August 14, 2019 from <https://direct.fredmeyer.com/privacy-policy>
 - [15] Anon. Hello, how can we help you today? Retrieved August 14, 2019 from <https://help.walmart.com/app/ask>
 - [16] Anon. 2019. Walmart Privacy Policy. (June 2019). Retrieved August 14, 2019 from <https://corporate.walmart.com/privacy-security/walmart-privacy-policy>
 - [17] Anon. You are about to leave walmartmoneycard.com. Retrieved August 14, 2019 from <https://www.walmartmoneycard.com/getacardnow>
 - [18] Anon. Privacy Policy. Retrieved August 14, 2019 from <https://www.walmartmoneycard.com/account/legal-info-page?doc=pp>
 - [19] Anon. 2019. Walmart uses AI cameras to spot thieves. (June 2019). Retrieved August 14, 2019 from <https://www.bbc.com/news/technology-48718198>
 - [20] Anon. Privacy Policy. Retrieved August 14, 2019 from <https://www.starbucks.com/about-us/company-information/online-policies/privacy-policy>
 - [21] Anon. Contact Us. Retrieved August 14, 2019 from <https://customerservice.starbucks.com/app/contact/ask/>
 - [22] Anon. Starbucks Cards. Retrieved August 14, 2019 from https://customerservice.starbucks.com/app/contact/ask_cards/
 - [23] Anon. Starbucks®. Retrieved August 14, 2019 from <https://app.starbucks.com/account/create?ReturnUrl=https://app.starbucks.com/>
 - [24] Anon. Privacy Policy. Retrieved August 14, 2019 from <https://www.safeway.com/emmd/Dominicks/v1/PrivacyPolicy.html>
 - [25] mgklous. UW Computer Security Research and Course Blog. Retrieved August 14, 2019 from <https://cubist.cs.washington.edu/Security/2008/02/10/security-review-the-safeway-club-card/>
 - [26] Vasundhara Bundela. 2018. Turning data into dollars: Here's how your company can master the art of data monetization. (November 2018), retrieved August 10, 2019 from <https://www.softwebsolutions.com/resources/data-monetization-strategy-for-enterprises.html>
 - [27] Jia Wertz. 2018. Don't Spend 5 Times More Attracting New Customers, Nurture The Existing Ones. (September 2018). Retrieved August 10, 2019 from <https://www.forbes.com/sites/jiawertz/2018/09/12/dont-spend-5-times-more-attracting-new-customers-nurture-the-existing-ones/#16cf9aa15a8e>
 - [28] Kim Bhasin. 2014. A Sort Of Creepy Reason To Love Costco. (July 2014) Retrieved August 10, 2019 from https://www.huffpost.com/entry/costco-recall_n_5618487?guccounter=1
 - [29] Brunot, Trudy. (n.d.). Internal Processes of a Retail Store. (n.d.) Retrieved August 10, 2019 from <http://smallbusiness.chron.com/internal-processes-retail-store-73736.html>
 - [30] Anon. 2017. Creating a sales analysis report. (July 2017). Retrieved August 10, 2019 from <https://wellsfargoworks.com/marketing-center/article/creating-a-sales-analysis-report>
 - [31] Peter Kriss, 2014. The Value of Customer Experience, Quantified. (august 2014). retrieved August 11, 2019 from <https://hbr.org/2014/08/the-value-of-customer-experience-quantified>
 - [32] Emily Steel, Callum Locke, Emily Cadman and Ben Freese. 2013. How much is your personal data worth? (June 2013). Retrieved August 11, 2019 from <https://ig.ft.com/how-much-is-your-personal-data-worth/#axzz2z2agBB6R>
 - [33] Anon. EQUIFAX DATA BREACH SETTLEMENT. Retrieved August 11, 2019 from <https://www.equifaxbreachsettlement.com/>
 - [34] Anon. 2018. IAB internet advertising revenue report (May 2019) Retrieved August 11, 2019 from <https://www.iab.com/wp-content/uploads/2019/05/Full-Year-2018-IAB-Internet-Advertising-Revenue-Report.pdf>
 - [35] Wikipedia contributors. 2019. Information broker (May 2019) Retrieved August 12, 2019 from https://en.wikipedia.org/wiki/Information_broker
 - [36] Anon. 2002. Big Fat Liar. (February 2002). Retrieved August 8, 2019 from <https://www.imdb.com/title/tt0265298/>
 - [37] John Ingold. 2016. Woman who catfished Chris "Birdman" Andersen online sentenced to jail. (April 2016). Retrieved August 8, 2019 from <https://www.denverpost.com/2015/10/28/woman-who-catfished-chris-birdman-andersen-online-sentenced-to-jail/>
 - [38] Anon. Chris 'Birdman' Andersen Investigation Targets Child Pornography. Retrieved August 8, 2019 from <https://www.christianpost.com/news/chris-birdman-andersen-investigation-targets-child-pornography--74777/print.html>
 - [39] Alexandra Ma. 2018. China has started ranking citizens with a creepy 'social credit' system - here's what you can do wrong, and the embarrassing, demeaning ways they can punish you. (October 2018). Retrieved August 8, 2019 from <https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4>
 - [40] Ashley May. 2018. Antifa protesters chant outside Fox's Tucker Carlson's home, break door. (November 2018). Retrieved August 8, 2019 from <https://www.usatoday.com/story/news/politics/2018/11/08/mob-tucker-carlsons-home-antifa-break-door-chant-fox-host/1927868002/>
 - [41] /@the_manifest. 2019. Data Privacy Concerns: An Overview for 2019. (March 2019). Retrieved August 8, 2019 from https://medium.com/@the_manifest/data-privacy-concerns-an-overview-for-2019-2ccea79aa6f8
 - [42] Allen St. John. 5 Easy Ways to Protect Your Digital Privacy in 2019. Retrieved August 8, 2019 from <https://www.consumerreports.org/privacy/ways-to-protect-digital-privacy/>