



Group Assignment

---

# **TDS 2101**

## Introduction to Data Science

### Data Analytics Implementation in Retail

Prepared by

Name	Khalid Ibrahim Adem Hamed	Yue XiangRong
ID	1161303501	1141327188
Tutorial Section	TT01	TT01
Email	FXTRyio@gmail.com	russellyxr@gmail.com
Contact no.	011 3903 9632	018 319 0203

## Table of Contents

1.0	INTRODUCTION .....	1
2.0	THE RETAIL INDUSTRY .....	2
3.0	BUSINESS USE CASES.....	4
4.0	CASE STUDY 1 .....	6
5.0	CASE STUDY 2 .....	9
6.0	COMMENTARY.....	11
7.0	FUTURE WORK.....	12

## 1.0 INTRODUCTION

Data Science has seen tremendous growth and incredible popularity in the last two decades. This growth is no coincidence, since Data Science provides us with tools that transform businesses and even industries. When the internet was young and technology was limited, data was scarce. But now we live in an exciting time. There are unprecedented amounts of data being generated, 2.5 trillion gigabytes of data every day. Within this mind-boggling amount of data lies the potential for powerful mechanisms of knowledge-extraction, and new avenues of technological advancement. Couple this with the fast growing development in hardware that gives us more computational power to explore this data for information, and you get a set of tools capable of transforming the landscape of modern industry.

This is exactly what we have seen in these last two decades – industry after industry leveraging data science and data analytics to increase productivity and provide better customer experience. This is apparent not only in business, but also in healthcare, sports, government sector, and so on.

In this project, we explore this transformative power of data. We select an industry that leveraged data science for its growth, and look at study cases of companies that employed data analytics techniques to further its development. We will look at the companies in question from the perspective of an outsider, who assesses how data science changed the company and in what ways it improved it. We pick a set of problems that these companies were facing before. We examine the scope of the problem and the effect it had on productivity. Then, we look at the data analytics tools that were used to solve these problems and examine the effectiveness of the solution and its transformative aspect on that industry.

For our project, we chose to focus on the retail industry. Retail is quite interesting because it has arguably benefitted the most from this surge in data and the power of data analytics. This is perhaps because the industry is also one of the biggest generators of data, and because retail is an integral part of our livelihoods. We contribute data to the industry every time we purchase, every time we view items, or even when we are unaware of it.

There were huge challenges that were overcome in this industry, and it is one of the ones that changed the most over the last few years. This growth would have not been possible if it was not for data analytics, and this is precisely what we want to explore.

## **2.0 THE RETAIL INDUSTRY**

Data Science can be used in a variety of ways in the domain of retail. To explore this, let us first take a look at what retail is and what types of retail can we find in our everyday lives. Then, we will look at a small history of the industry and how it changed so far. From this scope, we will be able to better understand the effectiveness of these business use cases of data science in this industry.

### **What is Retail?**

Retail is the process of selling goods to consumers through different channels of distribution. It basically includes every facet of purchase of goods, be it physical goods, digital, food, services, and so on. This takes form in all our day to day purchases, from our groceries, to our food, clothes, entertainment, accommodation services, and many more.

The industry that comprises all these purchasing activities is known as the Retail Industry. It is apparent now how huge and far-reaching this industry is, and how integral it is to our day to day life. This is why it has a special position in that it is a haven for data analytics expertise. The industry generates so much data that it is only natural that it benefits from data science the most.

### **The Story of Retail**

Retail dates back to ancient history. From the start of civilization, when humans changed from hunter-gatherers and switched to agriculture, there was the chance for everyone to specialize in a certain domain and for people to exchange goods. A potter would make pots and exchange them with food from a farmer, and such were the old ways.

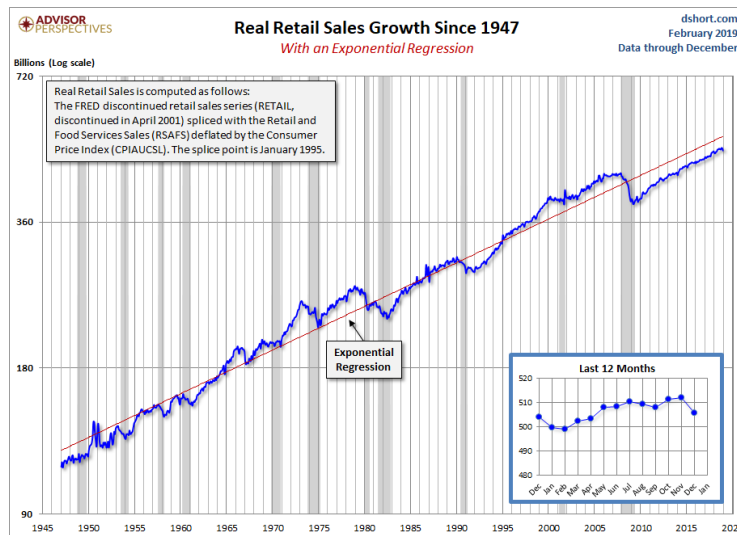
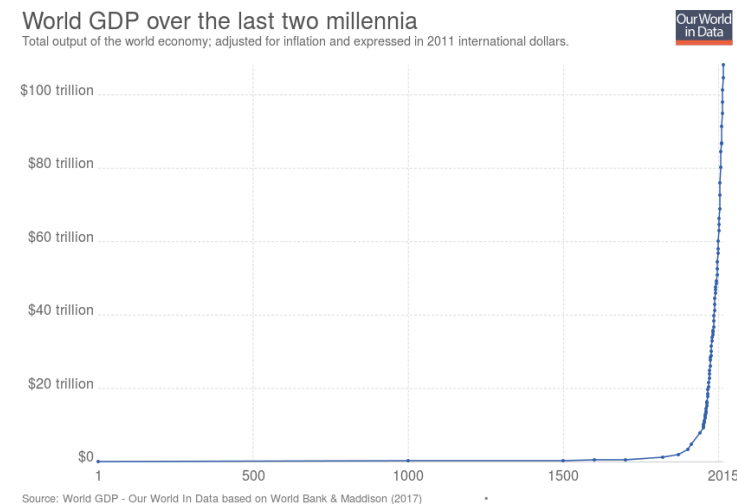
As time progresses, the act of making goods and selling them (now with the use of currency instead of bartering) became more sophisticated. There was starting a separation between the maker of goods and the seller. A person would make goods and another would collect multiple goods from different manufacturers and would sell them all in one place. From here was born the concept of the “Store”. Admittedly, it is something that we all take for granted, but this was revolutionary, as it enabled people to conveniently get whatever they needed from one place instead of having to go to multiple people for different items. This new efficiency only grew and stores become more and more convenient (to the point where they are even called “convenience stores”).

Stores are now drastically different from what they were centuries ago. They are now giant conglomerates that adjust and tweak every aspect of the sale process to maximize efficiency and convenience for the customer, and increase sales and profit for the retailer.

## Modern Retail

Modern retail has taken many shapes and forms. In order to increase efficiency, the idea of the store had to change. Now, we don't only have physical stores, we also have online stores and virtual stores. We buy from the comfort of our own houses. This enabled a new way for retailers to collect more data and employ data in new ways to increase sales. In addition to that, data science was used to optimize everything, from the way the websites look and the recommendations given to the overall shopping experience.

We can see a tremendous growth in products throughout human history:



This recent sharp growth is mainly due to technological advancement, and a big part of that is the implementation of data science in businesses.

## **3.0 BUSINESS USE CASES**

As we have seen in the previous chapter, data science can dramatically change businesses. But what does data do exactly that makes the business so much better? What are implementations of data science and how do they work? Here, we will look at some of the most effective and widely used set of applications of data analytics in retail.

### **3.1 Recommendation Systems**

A Recommendation System or Engine is a tool that predicts how much an item would be liked by a person when newly introduced. The purpose of this system is to figure out, based on different mechanisms, what customers might be interested in purchasing and recommend it to them. The idea is quite simple, yet the implementation of recommender systems is a powerful way of increasing customer satisfaction in a platform or an online store.

### **3.2 Personalised Marketing**

It is a method that uses data analytics to provide individualised marketing that is tailored to a person's interest and offer them advertisement messages or product deals based on their interests. This method saves a lot of time and money for companies because a lot of a company's budget is spent on marketing its products. But if one can know how effective certain marketing is on people and whether they will be interested in the product in the first place, a lot of unnecessary marketing can be saved.

### **3.3 Price Optimisation**

Price Optimization is a technique used to determine how customers will react to different prices. This is important because it can be a way by which to set prices so as to maximise profit. It uses data analytics to select the most reasonable price for different items and to best beat competitors in terms of low prices and high profitability. It can also be used to set prices in real time based on changing factors, like airline fares based on booking dates or on increased demand.

### **3.4 Market Basket Analysis**

This is a metric that analyses how products are related to each other. It uses Frequent Pattern Mining to create associations between different products so that we can know what products to group together, how to group them together, and on what bases. For example, many physical retail stores use Basket Analysis to figure out what products are

purchased together, and then places these products in the same shelves or near each other in order to maximize the likelihood that more of these products will be bought.

### **3.5 Loyalty Analysis**

This is a system that identifies the most loyal customers and rewards them for their loyalty. It is an inference system from which businesses can tell which customers can be the most frequent and profitable and focus its resources on those people. It also identifies factors that affect loyalty in order to incentivise more customers to be loyal.

As one can imagine, there are many possibilities of data science application in the field of retail, and the examples provided here only touch the surface of what is being used in the industry and what is possible in the long term. To get a better picture of the impact of data science, we will closely examine two case studies of its usage in the industry. We will analyse the problems that some companies were facing, how they implemented data-driven solutions, and how these solutions impacted the companies.

## 4.0 CASE STUDY 1

In our first case study, we will look at the biggest online retail company in the world: Amazon. Amazon is a company focusing on e-commerce, cloud computing, and artificial intelligence. It was started in 1994 as an online bookstore, but soon diversified into selling all kinds of media, electronics, clothes, and many more. It is the highest valued company in the world, with its heavy disruptive potential and its sophisticated usage of technology, including data analytics, in the operation of the business.

Amazon was not always like that. When it first started, it barely had the capacity to meet the small demand of the online book market. Soon, as the business grew and it became steadier, the sales were increasing and so were the variety of items. This caused an issue to the company since it had to figure out how to best capitalise on this to maximise sales. From the beginning, Amazon employed data analytics to its platform, to the point where it is now an integral part of its operation. We will examine the problems that might've faced the company had they not applied data analytics, and what impact did this have on the company as a whole.

### 4.1 The Problem

Human beings are very easily overwhelmed. Studies show that when people are presented with so many choices and little information about the choices, they tend not to make a decision, and choose nothing. This was a problem that Amazon had to deal with from the beginning. There were so many choices, but customers left with only one item even though they might have benefitted from getting a similar item that goes hand in hand with the one that was purchased.

### 4.2 Questions Asked

In order to solve the problem, it was translated into a series of questions that needed to be answered. These questions were:

1. ***'How many options should the user be given when looking for a product?'***  
Since the central issue was the limitations of many options, it was imperative to understand then how many options were too many to the point that it was limiting the customer's decision process.
2. ***'What options are likely to be useful to the customer and which ones will only add to the user's confusion?'***  
If we are providing many options to the customers, it stands to reason that we provide options that would likely be helpful to the customer.



3. *‘What list of items can be recommended to the customer, that he is likely to buy?’*

This is the central question that the problem boils down to. If one can answer this, we can know what options are the most useful to the customer (since the customer is more likely to buy it, it must be useful) and we can also know how many of them to present to the user as recommendation.

Once the questions were posed and the problem was formulated, it is time to use data science to implement a solution.

### 4.3 Solution

To solve this problem, the Amazon team used data analytics to their advantage. It therefore implemented a **recommendation system** as a solution for its problem.

The data collection, which is the first step of the solution, comes directly from the platform itself. Since the platform already collects a lot of data about what customers like, their purchasing habits, their comments and reviews, and so on, it is at a privileged position. It has direct and useful data to use for model building.

The second part is to filter and preprocess this data. This would involve all the common transformations that data usually goes through, like normalisation, discretisation, etc. It will also include in this case feature engineering, which is used to determine which features of the data are useful for the task. It will also include some transformation, like changing text reviews of products to something machine-understandable like ‘positive’ or ‘negative’ review.

Next, the recommendation system is built. The purpose of the system is to predict the likelihood of purchase of a given item. There are broadly two ways of implementing recommendation systems:

1. Collaborative Filtering

This approach builds a model from a user's past behaviour (items previously purchased or selected or numerical ratings given to those items) as well as similar decisions made by other users. This model is then used to predict items (or ratings for items) that the user may have an interest in.

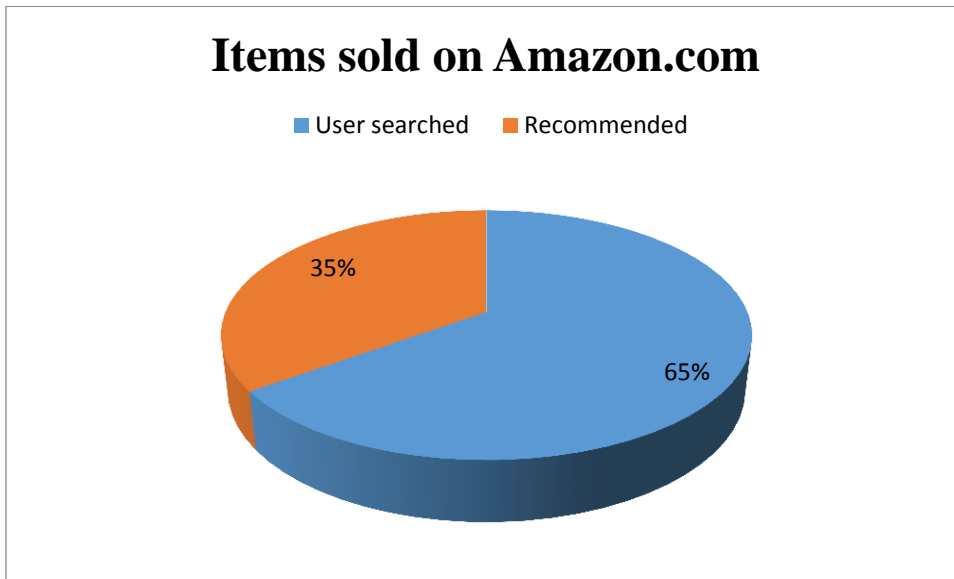
2. Content-based Filtering

This approach utilises a series of features or properties of a product and then recommends products that have the same properties.

Amazon’s solution combines these two to make a Hybrid Recommendation System. It predicts the “preference” or rating that a user might give to a product and recommends those products with high predicted preference rate.

#### 4.4 Impact

The impact of recommendation systems was huge. Amazon's core mechanism of driving sales relies primarily on data driven recommendation systems. It also contributed greatly to its increase in sales. McKinsey and Company estimates that 35% of Amazon's sales revenue comes from products that were recommended to customers. This is a total of \$52 billion annually.



In addition, it drastically increased user experience in an immeasurable amount. This could be discerned by user feedback regarding the website and the helpfulness of the recommendations provided.

## 5.0 CASE STUDY 2

In the second case study, we look at one of the biggest hotel chains in the world: Intercontinental Hotel group. It operates over 5000 hotels in more than a 100 countries all over the world. A company of this size and magnitude needs to carefully assess each business decision with utmost precision. This is because any inefficiency in operation can mean big losses, and therefore they need to strategize effectively.

### 5.1 The Problem

The company employs a lot of service staff and invests a lot of resources to keep the business functional. However, the demand for the hotel services is not always the same at all times. Sometimes, you have many guests, but sometimes, the rooms are empty. This leads to a waste of resources and to staff being on standby for long periods of time. It would be commercially infeasible, or at least very costly, to keep operating at such times with the same price range as busy times. Therefore, there is a need to set prices in a way that makes it viable to operate at all times. It also needs to have competitive prices so that it is not unreasonable and would drive customers away.

### 5.2 Questions Asked

The problem manifests itself in the form of these questions:

1. *'What are the times of year or month or week that prices need to increase/decrease?'*
2. *'How much do we charge for a room in a given date and time?'*
3. *'What is the optimal price that is competitive enough, yet still profitable?'*

Answering these questions will require us to collect data about various products and services and to model a predictor of the best optimal prices for the given parameters.

### 5.3 Solution

In order to answer the questions posed here, the group resorted to data analytics. They used the data they collected from their daily operation and constructed a prediction model that predicts the best possible price depending on the parameters provided.

The model built was a regression model that takes in many factors, like date, time, location, type of room, size, among other factors, and gives you the best possible price to set. From the data provided, it learns which factors are the highest contributors to price and at what value do these factors produce the most profitable prices.

## **5.4 Impact**

To observe the effects of data analytics' integration in the business, the hotel group conducted an in-house study that analysed the techniques used and its impact on the company. The study concluded that among the 2000 hotels that implemented the analytical model, an incremental revenue of \$145 Million has been achieved due to the implementation. The study also anticipated that a full rollout of the data driven pricing would generate approximately \$400 Million per year in revenue.

## 6.0 COMMENTARY

The solutions implemented by the two companies are quite sophisticated. Amazon is essentially a data driven company. Its entire business strategy relies on data. This helps a lot in bringing out the best possible customer experience and nudging customers towards the best recommendations.

The issue, however, arises when Amazon taps into different sectors. It has dominated the online retail market thus far, but people don't only shop online. Despite the comfort and convenience, people still like to go out to physical stores. This is because online stores lack something essential to the shopping experience: the ability to touch the product. And by touch, we mean the ability to experience the product, pick it up, try it on, and be comfortable with the purchase. This is especially true when it comes to fashion, since people want to get the best fit, see how the clothes look on them and so on. It is also true with groceries to a certain extent. This is why Amazon is trying to penetrate the physical market domain, as it is evident by Amazon Go and its acquisition of Whole Foods.

The challenge for Amazon then becomes the integration of data analytics in these physical stores. How will the data be collected? What new models will have to be built to accommodate the new types of data? How will recommendations be tailored to specific customers?

One possible way to address this is to intertwine the online store with the physical one. This means having online memberships so that data profiles can be constructed for each member in order to best tailor the recommendations to these people.

Intercontinental group on the other hand does not rely on data as much as Amazon. It does leverage data, but not to its full potential. The company relies on data to optimise process. It even does that in a very good manner. But the factors used to determine price are quite limited. The prices need not only depend on the time or number of people. It can depend on factors that might seem quite far-fetched when thought about outside the context of data science. So, it is imperative to take into account more features and build an even more accurate and comprehensive optimisation tool.

In addition to price optimisation, data can be employed for many other things.

Recommendation is one of them, but it can also be used for other business cases like we saw in the first chapter. It can be used to market certain services to targeted individuals, or assess and reward loyalty, etc. There are no limits to what can be done to improve using data.

## **7.0 FUTURE WORK**

Amazon is a holistic, self-sufficient, data-driven company. There is very little that can be added to their work, since they are doing things very well and leading the industry. For Intercontinental group however, there is a lot of room for improvement. In addition to implementing data analytics on other domains, like recommendation and marketing, what the company needs to focus on as well is the implementation of data on optimising their prices with more features and taking into consideration a larger set of factors. This will help further improve the revenue gain, since it can give better prices that are a lot more profitable.

The ways in which data science can be implemented are only limited by the tenacity of the people in the organisation. It seems that there a lot more avenues of exploration to be done that could drastically improve the business.