**Data science: Expectation Vs Reality**

**By Arnaud Davy M.M**

In the last couple of years, the largest companies (Google, Amazon, Microsoft, Facebook, etc) and others in the entire world used data to their advantage. Data Science (DS) has consistently ranked as one of the best and sexiest jobs in America, both from a company’s perspective and from an employee’s perspective. Demand across the world for Data Scientists is high, and in no way of slowing down while supply is low. The lack of competition and talent for these jobs makes DS a very lucrative option for a career path. It has gained a lot of curiosity from learners, and many want to get into the field. Everyone gets more curious to learn about DS with better paychecks in mind.

Some people define DS as a new powerful approach to make discoveries from data, others refer to it as an automated way to analyze enormous amounts of data and extract information, still others term it as a new discipline that combines aspects from applied mathematics to statistics, programming, and visualization to turn data into meaningful information.

* What exactly does a data scientist do?
* Does the expectation meet reality in this?
* Do people understand Data Science?
* Is Data Science about modeling alone?
* Can managers calculate an explicit monetary value to a Data Scientist’s contributions within the company?
* Can Data and Analytics lead certain firms in an unethical direction?

In answering these legitimate questions, let's find out what answers Kama-Tech Solutions suggests.

1. **Most People Do Not Understand Data Science**

DS plays a critical role in data-driven decision-making. Depending on the project and the broad nature of DS, Data Scientists describe, diagnose, analyze or predict, prescribe, Interpret or explain, and then recommend data-driven solutions that guide business strategy and operations. For example, such solution may be aimed at cost-reduction, improvement in operational efficiency, better marketing (e.g., marketing promotion campaign costs, and analyzing the ROI of all marketing campaigns), improvement in sales efforts (e.g., refining target audiences and, customer segmentation, and measuring the likely lifetime value of customers), find business opportunities and new markets, all of which are designed to create an edge over competitors.

Although experts may disagree over some details and approaches, everyone agrees that DS combines domain expertise and scientific methods to provide insights into underlying data.

Many people have ventured into data science without a proper understanding of what it entails and what you can expect to do as a Data Scientist. Learners hope that DS is all about handling problems using ML (Machine Learning) and AI (Artificial Intelligence) software. After you acquire that job, the reality clicks in that you have to understand the nature of the industry in which your DS project is being performed including the specific metrics and challenges to the industry.

Simply put, you need broad industry knowledge combined with an understanding of how one particular metric or solution could impact the wider business.

More surprisingly, many firms look for Data Scientists with a job specification I ironically call a “ full stack Data Scientist” or a “ Data Scientist unicorn”, by this I mean a combination of data engineer, data analyst, machine learning engineer and cloud computing . They highlight the following skills: Python, R, SAS, JavaScript, SQL, PowerBI, Tableau, Hadoop, Apache Spark, Apache Kafka, Apache Airflow, Scala, Tensorflow, Pytorch, Computer vision, NLP, AWS, communication and interpersonal skills. It seems like a Data Scientist master and knows everything related to data.

According to their job specs, it's clear that some firms have no idea what DS is and what it can do. They want to hire anyone with the label of Data Scientist thinking the ideal candidate is going to be able to solve anything related to data.

1. **You May Be the Only Data Scientist in the Company**

Most firms with over 100 employees only have one data scientist, others hire Data Scientists without a suitable structure in place. They expect their Data Scientist to be all-seeing, all-knowing, and ready to solve any problem within the company, whether sales, customer support, engineering – to analyze data and respond to every request.

The reality clocks in when you can't solve all their problems or spend most of the day doing the fun things you think of as using ML. You spend the whole day organizing and scrubbing data. Even surprisingly, trying to troubleshoot computer issues or working with an outdated version of the software.

Four years ago, if my memory serves me right, I remembered working in a big project as a consultant with other fellow Data Scientists, two of them were fired and one talented fellow Data Scientist resigned. I ended up being the only Data Scientist in the company.

At the beginning, we were working with big data using *MS Excel 2010*. Yes, *MS Excel 2010 and a computer of 8 GB RAM, 250 GB storage* before we switched to cloud. The computer was crashing more than 10 times a day. Even when we created a lot of subset data, it always took 5 hours to execute the nested IF(COUNTIF()) to flag a unique observation as 1 and a duplicate one as 0.

You guessed it, the overall computer’s performance was way too poor including speed and computing power. I was challenged and frustrated by system issues. In addition, I was challenged to write desk level procedures for our data analysis process.  Anyone can imagine how happy I was to see the clock indicating that it was time to go home and how frustrated it was to start the day knowing what issues and pressure I was going to face at work.

I made recommendations to upgrade the CPU including system RAM, Disk storage, and video card, and I recommended suitable software for the actual DS work on the project. One year later , the company’s IT department finally decided to switch to cloud.

1. **You Keep Learning New Technologies**

Many learners who are taking DS courses for a career change or are new graduates think they will save the world by using ML to predict a target variable with high accuracy. Reality quickly clicks in, and they are left with realizing that they will need much more learning in order to somehow keep up with the ever-changing technology.

DS technologies and frameworks evolve so fast that it’s unwise and stupid trying to get stuck on any previous single skills we mastered 2-4 years ago. It is wise to constantly learn new concepts to update your tool kit of current state-of-the-art (predictive analytics) to respond to the emerging and broad nature of the DS ecosystem.

Amongst various kinds of knowledge gained over the course of my journey at Harvard Business Analytics Program, precisely in Innovation Technologies and Change (taught by Professor ***Karim R. Lakhani***) and Leadership, Innovation and change (taught by Professor ***Michael Tushman)***. I always remember:

***“In a fast-changing technology, fast-moving world and market, aspiring leaders and news generation of Data Scientists must keep up with change to stay competitive. In the same vein, firms tend to disappear if they don't either respond to the emerging digital technology or change the business and the operating model”***

***Likewise, exploring (failures, experiments, variants) and exploiting (consistency, incremental change and continuous improvements) both in business and personal life is key to success.”***

To illustrate, there was no demand for DS 18 years ago. The field didn't even exist due to slow computers, slow internet, low computational power, primitive programming languages, etc. It’s important to note that DS was driven by fast-changing technology. Today Data Scientists must be versatile or possess di-technical expertise.

1. **You May Be Lacking in Some Most Important Vast Knowledge**

Employers expect data scientists to have vast knowledge in computer science, mathematics, and statistical models and overturn business applications and domain expertise.

A Data Scientist should know how to present and communicate insights to others. An underemphasis on soft skills has therefore led to a communication gap. In essence, business knowledge, exhibition of great teamwork and flexibility in meeting aggressive deadlines will allow you to provide timely, accurate, high-value, and actionable information to improve strategic, operational and tactical performance.

To your tool kit of current state-of-the-art skills, you need to add soft skills such as storytelling skills, communication skills, and collaboration skills. Without any interpretation, prediction results or ML results are useless. A Data Scientist needs to articulate findings and communicate key hidden trends and patterns. Analytical results need to be interpreted and communicated to stakeholders for any given model or diverse audiences (e.g., executive decision makers, operation managers, marketing managers, sales managers, product designers and managers, and software engineers) in a clear, concise or precise, to-the-point manner.

In sum, it’s worth noting that DS itself can be viewed as a new common basis for communication, inquiry, and understanding in both science and industry.

1. **Is Data Science about modeling alone?**

Many people venturing into DS think it is about modeling all day and coming up with key insights that drive business change. But DS involves lots of cleaning and transforming data. Reading a lot about the business ensures you have a proper business understanding before building your model. There is also a lot of data collection if you don't have any data. Later you will realize that you spend more time learning and not just sitting around solving problems. For instance, Kama-Tech Solutions has worked on a research project without any data at hand. We add Natural Language Processing in the LinkedIn of the company to get data and turn it into meaningful information.

Not everything can be solved with machine learning, and inference may be more than you thought by building a model to predict the target variable. Very often, business managers just need a comprehensive snapshot of the company’s performance.

How could modeling interpret the result, provide the snapshot?

Arguably, a dashboard (whether it is Strategic, Analytical, and Operational Dashboards) is what's needed to help managers make informed decisions that dramatically impact business performance.

In the same light, a dashboard reliably provides the Key Performance Indicators (KPIs) regardless of various core departments within the company (e.g., sales performance dashboard, marketing dashboard, HR dashboard, etc.).

.

1. **Not Everything Can Be Solved Through Machine Learning or Statistical Modeling?**

A company may think it can solve different problems by throwing them to MI and getting results. However, this is not always the case because human beings can solve some of these problems. Many problems that company throw at MI are not well thought out. For instance, people would have done that by now if you could toss a punch of variables and predict the stock market. But that hasn't happened yet.

1. **Are Data Scientists Incredibly valued within the Company?**

Because data is breaking down frontiers, creating new industries and fields no one could have thought of, enabling everything from autonomous cars to Internet of Things, from operation efficiencies to uncovering new markets, the value and need of Data Scientists will keep growing.

Likewise, leading organizations in every industry are wielding data and analytics as competitive weapons, operational accelerants and innovation catalysts while data roles continue to grow and expand within industries. Evidence suggests that all dimensions of DS competency significantly improve decision quality while significantly increasing decision efficiency.

Sadly, nearly all managers rarely calculate an explicit monetary value to measure a Data Scientist’s contributions. It’s way too difficult to calculate a business value or put an explicit price on analyses that help a company derive significant value.

To illustrate my point, imagine a Data Scientist working on a targeted investment project in which through analysis he/she achieved payment metrics including $96.5M in payments.

How can anyone measure the Data Scientist’s impact or put a Data Scientist’s price to this project, especially knowing that the Data Scientist worked with managers of diverse departments.

Similarly, a Kama-Tech Solutions’ Data Scientist worked on a project in which he reduced logistic and marketing promotion campaign costs and increased the ROI from these campaigns through marketing analysis. The Data Scientist empowered management to make smart data-driven decisions which led the company to a cost reduction of $8.73 million. Some say he saved the company from major risk and losses of $8.73 million, others say the company didn’t spend $8.73 M. A passionate debate on matters such as this can eventually erupt. Feel free to join the debate and draw your own conclusion.

It's not unreasonable to presume that the connection between the performance metrics, the prediction model Data Scientist builds, and the explicit monetary value is less clear.

1. **Can Data and Analytics Lead firms in an Unethical Direction?**

Data is a new and exciting discovery like oil was a century and half ago, and to be a successful company in the 21st century you have to use data to your advantage. The largest companies in the entire world are data science fueled enterprises. Each uses DS to build and leverage algorithms that improve customer satisfaction and maximize profits. The power of data can be used for nefarious purposes as Facebook- Cambridge Analytica scandal (Facebook allowed Cambridge Analytica to gather data under false pretenses about 50 million users and their friends), the Google nightingale data controversy, and the photo-scraping scandal of IBM to name just a few. Given that, we can say that the power of Data and Analytics can lead certain firms in an unethical direction.

In what follows, In my sincere opinion there is no correlation whatsoever between a data-driven culture and crossing ethical frontiers in the use of data and analytics. In the same light, even pure human decisions that aren’t relied on data can create lots of ethical problems. It's human nature or decision that creates an ethical problem. Likewise, countless firms exist that rely heavily on analytics and data, but have a clean record regarding ethical infractions.

Furthermore it's not unreasonable to proclaim with certainty that the collection and use of consumers’ personal information is always unethical and the use of data and analytics leads to ethical problems. In the same light, even pure human decisions that aren’t relied on data can create lots of ethical problems. For example, few will argue that some individuals take a stand and even resign from their job to avoid perpetuating misleading facts.

In sum, it is worth noting that when thinking about investing in a DS career, you need to understand that DS is not about custom algorithms alone. It is not about complicated models to be solved by ML and AI.

DS involves several steps and processes including collecting and understanding data (data exploration), cleaning, analyzing and visualizing, interpreting to extract value from data that guide business strategy and operations.

As you venture into DS, Kama-Tech Solutions advises you to connect with communities of DS practitioners (Data Scientists), within either large firms or externally or find a mentor.

Data scientists need to participate in new conferences, webinars and belong to the formal or informal associations that are emerging to support collaboration and technology sharing.

In the same vein, companies should encourage Data Scientists to become involved in these activities.

Data Scientists will also need to constantly update their skills and sharpness to keep up with the fast-changing technology and a fast-evolving field of DS.

.