**7 Myths About Data Science**

**By Arnaud Davy M.M**

Data Science (DS) may be defined as a new powerful data-driven approach and a collection of disciplines which combines aspects from applied mathematics to statistics, programming, and visualization techniques to turn structured and unstructured data into information in semi or fully automated fashion.

Data scientists provide invaluable insights that enable stakeholders to better manage strategic, tactical and operational performance to gain a competitive advantage on competitors.

* Do Data Scientists Predict the Future?
* Is Data Science Limited to the Tech Industry?
* Are Data Scientists Programmers?
* Can Data Scientists Do Everything?
* Is Data Science (DS) Solely About Predictions?
* Is Data Science (DS) Solely About Data Analysis?
* Is Data Science (DS) a Single Discipline?

Before you hire KAMA-Tech Solutions, it's important to learn the truths behind the myths that surround this industry. Once you know what to expect, you'll approach KAMA-Tech Solutions with confidence.

**1. Data Scientists Predict the Future**

While Data Scientists (practitioners of DS) use data to predict trends, potential issues and user behavior, they're not right 100% of the time. Likewise, there is no algorithm that predicts with 100% accuracy score. This doesn't mean that you hired the wrong team--it simply means that no one can predict the future or predict with 100% accuracy in prediction, either. For example, an unexpected catastrophe could disrupt the supply chain, forcing your team to rethink their predictions.

In the same vein, powerful machine learning (ML) includes deep neural networks and statistical imputations can't solve the business. It needs the combination of strong business acumen, a deep intellect, machine learning methods and inference to solve the problem.

When you hire Data Scientists, don't expect them to be right 100% of the time. However, data scientists can take virtually everything into consideration, including sudden accidents or disruptions, to give you the most accurate possible predictions. Most of the time, you can rely on their data for accurate conclusions.

**2. Data Science is Limited to the Tech Industry**

Most people immediately think of the tech industry when they hear the term "Data Science." While Data Scientists often gather and analyze information for tech companies, Data Scientists work with a variety of industries including medical, healthcare, manufacturing, customer service, sales, financial or banking, marketing, travel, transportation, telecommunication, retail, postal, oil-field, FMCG industry, agricultural industries et al.

If you need data gathered, cleansed, analyzed, visualized and interpreted, never assume that Data Scientists can't help you. Instead, reach out to KAMA-Tech Solutions to discuss your situation. Talking to us is free, and we'll determine whether we can help your business. If we can, we'll discuss your next steps so you can start gathering predictions and stay ahead of your competitors.

**3. Data Scientists are Programmers**

While Data Scientists use programming tools and write code sometimes, they're not programmers themselves. Programmers write code, develop software, create tools, build apps and update existing tools, products and software. Conversely, Data Scientists use these tools such as Python but with different libraries such as Numpy, Pandas, Seaborn, Matplotlib, Scikit-Learn, and statistical software like R, SAS and SPSS to analyze data and turn it into actionable, intelligent insights. Essentially, they use the products that programmers create.

Likewise, not all Data Scientists come from math, science or engineering backgrounds. Many Data Scientists started off in other industries or had a different college major. This doesn't mean that they’re less knowledgeable--it just means that they've had diverse experiences, which bring new expertise to the team.

**4. Data Scientists Can Do Everything**

One Data Scientist can't do everything. In fact, a whole team of data scientists can't do everything. Since Data Scientists have a specific skill set, you might need to hire additional professionals. When each team member specializes in a different field, you'll have a comprehensive knowledge base that's not tied to one specialty.

Since Data Scientists have a specific skill set, you might need to hire additional professionals such as engineers, researchers, project managers, software engineers, product designers and managers, and people from several other fields. Everyone contributes a unique background, perspective and job history to the group. The Data Scientists work with the rest of your team to complete the project according to your specifications.

Depending on the project, Kama-Tech Solutions could assist you in locating and hiring engineers, designers, project managers, researchers, et al.

**5. Data Science is Solely About Predictions**

Though prediction is at the heart of DS, depending on certain business purposes, Data Scientists make either predictions or inferences, but hiring a data scientist isn't a straight shot to learning what your customers need or to making predictions. Even to make predictions, data scientists follow a systematic process such as understanding a business problem, the nature of the industry, general functionality, bottlenecks, and specific challenges to the industry.

From there, Kama-Tech Solutions Data Scientists implement the core methodology for solving a problem as follows: data collection or data acquisition, data cleansing or scrubbing, data analysis including model building, data visualization, interpreting the results and putting it into use or production.

Accordingly, analytics can broadly be classified under three categories: descriptive analytics, predictive analytics and prescriptive analytics. For instance, the data generation process itself goes beyond the selection of the sample data, it involves several steps, such as sampling theory, precision and estimation error, experimental design, and field testing to name just a few.

Ultimately, Data Scientists possess a keen sense of observation to ask important questions, find the answers to these questions, present what they learned, and make recommendations. They also stay on top of their field to use the latest software and provide the most accurate data. However, it's the business's responsibility to follow up on the project and use the data to their advantage.

**6. Data Science is Solely About data analysis**

Regardless of the DS project the data preparation involves several steps such as data cleansing or scrubbing and wrangling or munging (the process of converting and mapping data from one raw format into another). Often, raw data is not in a needed format to directly perform data analysis. Data may be available in structured, semi-structured and unstructured form including streaming data. Simply put, there are many types of data such as text, picture or image, audio, and video. Moreover, it's important to highlight different phases of the data life cycle ranging from data generation to data interpretation before putting the result into production.

Data discovery and data preparation are the most important steps for any organization to create a culture of quality data decision-making because it entails the process of fixing or removing incorrect, corrupted, incorrectly formatted, errors, missing values, duplicates or irrelevant observations within a dataset. Data scrubbing and wrangling have always been among the most time-intensive steps of any data science project knowing that a significant effort is needed to clean and wrangle data from a data set.

In the same light, data visualization, inferential statistics, Bayesian statistics, selection of probabilities, some areas of deep-learning such as computer vision, reinforcement learning (e.g., autonomous driving), digital assistants (Google Now, Amazon Alexa, et al), text-to-speech, explain the black boxes models are not about data analysis.

Considering these considerations, it is important to bear in mind that DS is a broad field with lots of ramifications and collection of disciplines, countless steps, sub steps and processes that can't be solely limited to data analysis. Simply put, DS encompasses a collection of disciplines, many steps, sub steps and processes including privacy and ethical concerns.

**7. Data Science is the single discipline**

How are Data Scientists different from Machine Learning (ML) Engineers, Financial Analysts, Statisticians Analysts, AML Analysts, Data Engineers, BI analysts?

Could they be interchangeably called Data Scientists?

Contrary to common misperceptions of DS, these roles can all arguably be considered as DS domain due to data's life cycle: data generation, data collection, data processing, data storage, data organization, data analysis, and data interpretation. DS is a fast-evolving field and has evolved in multiple dimensions. The ubiquity of data science led the field to an artificial ecosystem which I call "a new generation of data scientists''.

As stated by Harvard Data Science Review on page 10, I quote:

*" Like science, social science, or humanities, DS is best understood as a collection of disciplines with complementary functions, perspective, approaches, and aims, but with a shared grand mission. That is to use digital technologies and information of any kind to advance human society as a harmonious, responsible, and vital ecosystem".*

Even less surprisingly, many Data Scientists work today in business such as a data manager expert, who might be great at generating and organizing data in structured form but not at turning unstructured data into structured data nor at building machine learning models, nor at analyzing data to uncover hidden key patterns and trends to guide businesses strategy and operations.

Given the complexity of the field combined with increased digital dexterity among customers and fast-changing market conditions, uncertainties, and disruptions; many universities and bootcamps are adding big data coursework to their data science programs that provide students with exposure to actual big data challenges and frameworks.

To sum up, DS is not a single discipline by itself. It is a new powerful data-driven approach, a collection of disciplines which combines aspects from applied mathematics to statistics, programming, and visualization techniques to turn data into precise distinguishable metrics that enables stakeholders to better manage strategic, tactical and operational performance and gain a competitive edge on competitors. Also, there is no scientist who is an expert or possesses expert knowledge in all major scientific fields, such as physics, mathematics, chemistry, geology, biology, et al. In the same way, there is no data scientist who is an expert from all diverse fields and subfields of DS.

To gain the most benefit from a DS project, the business will want to implement practices that encourage early adoption of the Kama-Tech methodology and use of enhanced data.

To discuss Data Science solutions and learn more about the field, reach out to KAMA-Tech Solutions today (<https://kamatechsolutions.com> or [info@kamatechsolutions.com](mailto:info@kamatechsolutions.com) or call (281) 676-3571)