DATA SCIENCE AT RISA

In the last decade, the number of devices generating and collecting data has expanded exponentially. Consequentially, the amount of data that organizations have to contend with has also expanded at an even greater rate. In 2016, IBM estimated that, on a daily basis, we generated 2.5 quintillion bytes of data. To bring that into some perspective, that means that 90% of all data in existence has been created in the last two years. (IBM "What is Big Data?" 2016). Adding to the feeling of drowning in a sea of data is the fact that most of this data (estimated at 80% of all data) is "unstructured". It is not organized in any pre-defined manner. This information is typically heavy on text, but can also contain other data, such as dates, numbers, other non-text data points. The popular solution for storing all of this unstructured data is to collect it all, along with structured data, into a large "data lake", with the intention to analyze it for actionable value at some future time. Data science is the means by which this seemingly overwhelming amount of data can be "wrangled" into a usable form and provide business intelligence to decision-makers inside your organization.

At RISA, our data scientists first determine the business questions that your organization is looking to answer with an analysis of your data. We want to make sure that any analyses we perform are aligned with your organizational goals and provide value. These can range from efficiencies in your business processes and purchasing, to undiscovered opportunities revealed by trend analysis of your organization's data.

Once the business value (and success criteria) is defined, our data specialists then proceed with data gathering. What are the sources of data within your organizations If it is like most organizations, relevant data will not be collected in a structured manner in one place. It will likely be in a structured and unstructured form, and distributed amongst multiple databases, servers, and other data storage locations. It may even be stored in both structured and unstructured form in a data lake. No matter the case, those data sources and the data they contain will need to be examined to get an idea of the best way to approach the data and "wrangle" it into a usable form. This is also the point at which our team will perform data profiling. Not only getting to know the source(s) of the relevant data, but also examining it for characteristics such as: the size of the data set we're working with, the column heads, the data type of each column, the relationships between columns, the range of values in each column, the frequency of missing or junk data, and the number of rows.

Once our team has examined the data and gotten an idea of the myriad forms it may be stored in, the next step is data preparation. This is usually one of the most time-consuming parts of the process. When dealing with large quantities of data (i.e. Big Data) it is rare that the data is complete, accurate and in a common format. Most commonly, there will be errors or incomplete/missing entries in numerous fields and records. Our team of data scientists is experienced with automated ways of cleaning data to correct errors or remove incomplete data from the analysis. In addition, this step also includes any necessary transformation of the data such that it is all in a usable format for analysis. For example, if the data is gathered from disparate data sources, the fields from one source may not match the fields of another source. This might require transformation such that all the data is in a common format. Once the data is all in a "clean" form, this is when our team of data specialists begin their analysis. This may consist of descriptive analysis (which describes the current data landscape and the current environment), diagnostic analysis (which attempts to determine why a current condition or trend is occurring), predictive analysis (given the current data and trends in that data over time, what is the likely behavior of the system in the future over a similar time period), or even prescriptive analysis (given trends in the current data over time, and the likelihood of certain behavior in the future, what actions are likely to provide the most benefit by taking advantage of the predicted behavior).

Sometimes for business decision-makers and stakeholders, it is difficult to see the benefit data science/data analysis/"Big Data" can provide for their organization. Their question is, "What can it do for me to justify the expenditure?" Too often, providers of data science services are technical experts that can solve a problem once it is defined, but they have difficulty answering that question initially. Our data science team is composed of excellent technical subject matter experts, but also business-oriented personnel who can show the potential benefit of the analysis and how it can provide value to your organization. For a detailed case study, please visit: https://www.risadirect.com/data_science/case_study.html

CONTRACT VEHICLES

GSA Schedule
Contract # GS-35F-482DA

SEAPORT-NxG
Contract# N0017821D9316

COMPANY CERTIFICATIONS

SBA
SERVICE-DISABLED VETERAN-OWNED CERTIFIED

CONTACT US

RISA

CAGE: 4ACX2 AND DUNS: 195190611

UEI: **MKA4F1KQSSB5**

WEBSITE : WWW.RISADIRECT.COM

STEPHEN ROLSTON, Principal
813.929.3898
STEPHEN.ROLSTON@RISADIRECT.COM