

Why Khan Academy Chose Algebraix's Data Science as a Service

The Benefits of a Co-Sourced Data Science Model



CASE STUDY / EDUCATION

Introduction

Today, with a wealth of information at their fingertips, companies are capitalizing on valuable data sets like never before. However, how can businesses leverage this mountain of data and turn it into something meaningful, insightful, and actionable?

A recent Bain & Company survey, *The Value of Big Data: How Analytics Differentiates Winners*, highlights the fact that most companies have not yet leveraged their data to their advantage. It concluded that only 4% of businesses are really good at advanced analytics, and only 50% of mid-market companies properly use business intelligence (BI) to create accurate key performance indicator (KPI) reports. BI typically provides information on *WHAT* has already happened in the past. BI does not provide root cause correlative understanding of *WHY* things happen or their drivers. Additionally, advanced analytics and data science can probe even further not only into why things happen and their root causes, but also, through the use of predictive analytics, what is likely to happen in the future.

The ROI generated from predictive analytics and advanced data science, documented in multiple sources, is typically in the 250% range. The bottom line is that companies that implement advanced analytics outperform their competitors.

What is Khan Academy?

Khan Academy is a non-profit educational organization created in 2006. The organization produces micro lectures, practice exercises, and tools for educators and learners – available for free to anyone around the world. In 2009, Khan

Academy received the Microsoft Tech Award for education from Bill Gates. As of January 2016, the Khan Academy channel on YouTube has over 2.4 million subscribers, and the Khan Academy videos have been viewed more than 700 million times (Wikipedia).

"We are changing education for the better by providing a free world-class education to anyone anywhere." - Khan Academy

Khan Academy Business Challenge

Khan Academy was interested in understanding its learner population, which is equivalent to a customer base for the non-profit education company. To understand these learners, Algebraix Data segmented the user base into different categories and types, with specific definitions belonging to each learner segment.

Khan Academy receives roughly 20 million unique users per month, a large portion of which are non-registered. Having such a large amount of anonymous web traffic makes it challenging to ask questions regarding what varieties of content perform better than others, how users prefer to learn, and what specific demographics the user base consists of. Algebraix worked hand-in-hand with the Khan Academy's analytics team to produce user segments for deep analysis of usage, user demographics, engagement per technology platform, and product uptake per geographical region.

However, before the analysis, Algebraix Data had to go through Khan Academy's large historical database in the cloud to pull massive, multi-terabytes of data.

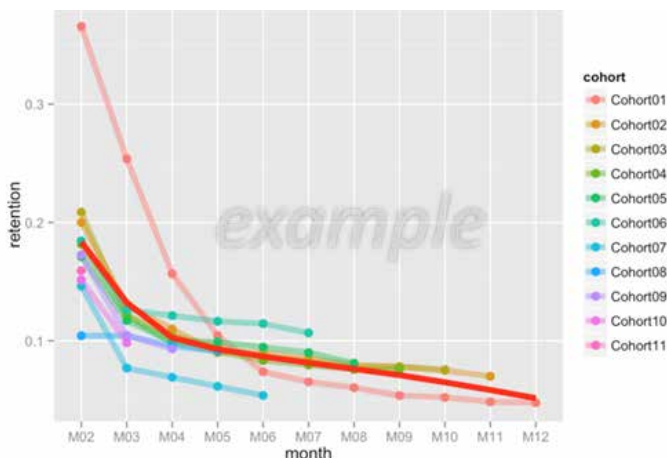
Algebraix Data Solution Deployed

With the help from Khan Academy's existing analytics team – and using Google BigQuery – Algebraix was able to query massive amounts of data in a performant fashion and find interesting nuggets of insights about Khan's learner segments. To simplify this large amount of learner segment information – and to summarize all results across the various user segments, device categories, geographic regions, subjects, and more – Algebraix created a “grid of insights” document. It allowed Khan Academy's senior team to answer complex questions about the users across many different slices of the data without needing any prior technical knowledge.

Algebraix Data continued to work hand-in-hand with Khan Academy, providing cohort and retention analysis, which involved understanding and calculating retention rates across subjects, segments, and domains.

An example of Khan Academy's cohort and retention analysis was to look at all “algebra” subject learners to see how many learners they retained per week. As a hypothetical example, if, in Week 1, there were 100 algebra learners, and in Week 2, there were 50, this was a retention rate of 50%. This cascading retention rate from week to week can be visualized as a retention curve (see Figure 1). A retention curve has a normal drop-off, but it levels off to a natural retention rate for particular subjects.

Figure 1 : Retention Curve



Algebraix calculated and created retention curves for many subjects, topics, and domains across different learner segments.

By extending this analysis even further – detecting and understanding the common patterns of the retained cohorts – Algebraix Data and Khan Academy will be able to use common factors to predict churn. This predictive analytics model helps guide Khan's product development to maximize engagement and site interactions.

In the end, Algebraix Data was able to work with Khan Academy to query and analyze massive amounts of data. This allowed Algebraix Data to provide valuable insights on learner segmentation and retention. But most important, by co-sourcing with Algebraix Data, Khan Academy made specific insights into its own massive database without “hiring, firing or training” a single person.

Why Did Khan Academy Choose Algebraix Data?

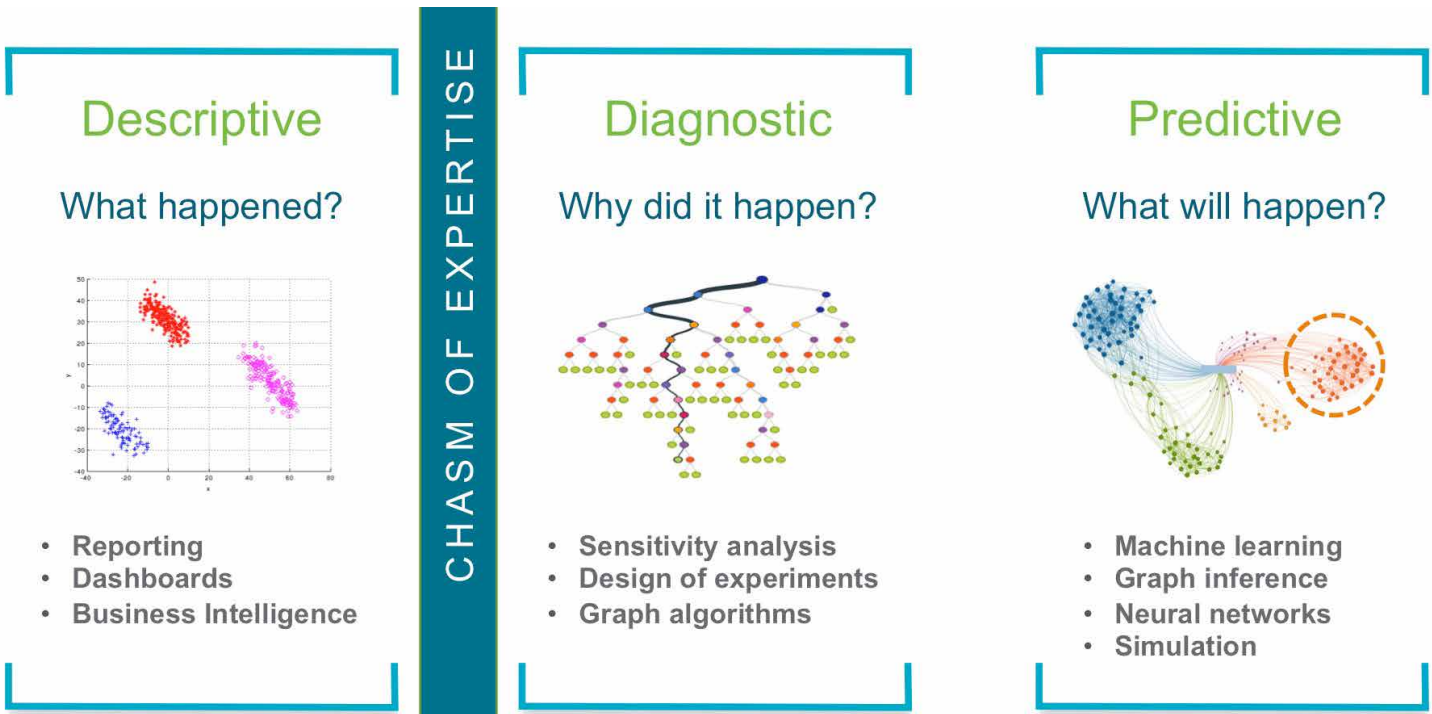
- **Data Engineering and Data Science Expertise**
- **Flexible On-Demand Delivery Model**
- **Low Risk & Affordability**



The Spectrum of Analytics and the Chasm of Expertise

Many companies today have a handle on KPI dashboards and BI reporting. However, very few are doing high-impact predictive analytics and data science because of the *chasm of expertise* (See Figure 2).

Figure 2 : **Chasm of Expertise**



For most companies, predictive analytics seems out of reach. However, Algebraix Data has recently developed a co-sourced data science model that can provide on-demand data science resources for any sized business. This on-demand data science resource is exactly why Khan Academy chose Algebraix's Data Science as a service to augment its existing analytics team.

"Data scientists are difficult and expensive to hire and – given the very competitive market for their services – difficult to retain. There simply aren't a lot of people with their combination of scientific background plus computational and analytical skills." - **Harvard Business Review**

Data science is on track to become the most sought after marketable skill set in 2016. However, the reality of advanced data science is that data scientists are expensive and hard to find.

Only 4% of Companies Are Really Good at Analytics

In the Bain survey, only a few companies – “the 4 percenters” – said they had not only the right people, tools, data, and intent to draw meaningful insights from their data but also the ability to act on them.

If the business benefits are clear and compelling, why are so few enterprises really good at competing on analytics? What is keeping the vast majority of companies from achieving similar results?

Historically there have been two huge barriers: 1) High-cost proprietary hardware/software, and 2) expensive, scarce analytical expertise.

Expensive and Scarce Analytical Expertise

Companies have been limited to two primary strategies: Either build an in-house team or hire specialized consulting firms to perform their analytics. Each approach has unique challenges.

In-House Analytics Teams: These groups consist of data engineers, data analysts, and data scientists. Three to five people are required to effectively perform all the tasks needed for advanced analytics. This quickly becomes an annual employee cost of at least \$500,000 in salaries and benefits, not including recruiting and training. These experts are also in high demand, especially data scientists, so retention is a challenge.

IT TAKES A TEAM 22 Skills of a Data Scientist

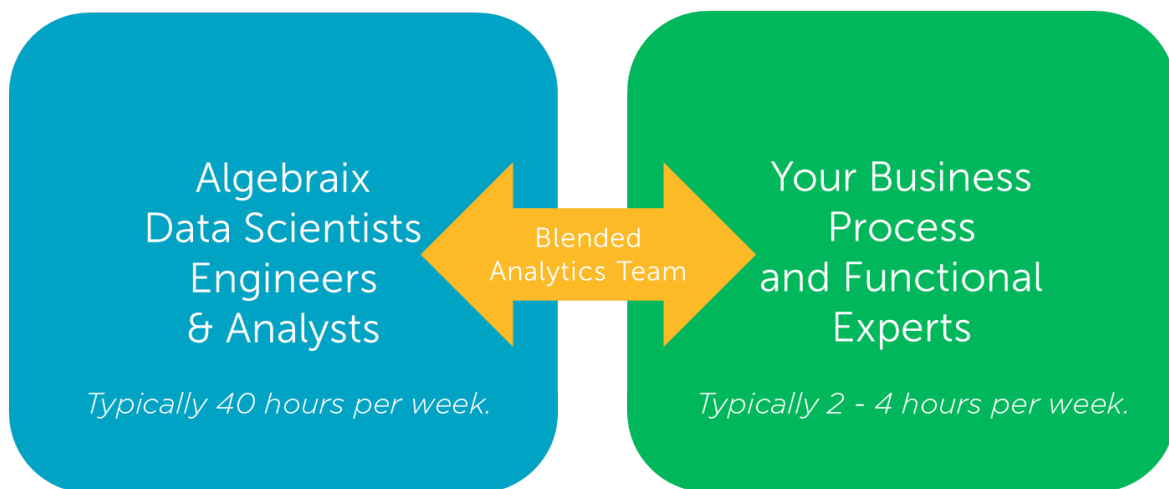
-2013 O'Reilly “Analyzing the Analyzers”

- **Algorithms**
(ex: computational complexity, CS theory)
- **Back-End Programming**
(ex: JAVA / Rails / Objective C)
- **Bayesian/Monte-Carlo Statistics**
(ex: MCMC, BUGS)
- **Big and Distributed Data**
(ex: Hadoop, Map/Reduce)
- **Business**
(ex: management, business development, budgeting)
- **Classical Statistics**
(ex: general linear model, ANOVA)
- **Data Manipulation**
(ex: regexes, R, SAS, web scraping)
- **Front-End Programming**
(ex: JavaScript, HTML, CSS)
- **Graphical Models**
(ex: social networks, Bayes networks)
- **Machine Learning**
(ex: decision trees, neural nets, SVM, clustering)
- **Math**
(ex: linear algebra, data algebra, calculus)
- **Optimization**
(ex: linear, integer, convex, global)
- **Product Development**
(ex: design, project management)
- **Science**
(ex: experimental design, technical writing/publishing)
- **Simulation**
(ex: discrete, agent-based, continuous)
- **Spatial Statistics**
(ex: geographic covariates, GIS)
- **Structured Data**
(ex: SQL, JSON, XML)
- **Surveys and Marketing**
(ex: multinomial modeling)
- **Systems Administration**
(ex: *nix, DBA, cloud tech.)
- **Temporal Statistics**
(ex: forecasting, time-series analysis)
- **Unstructured Data**
(ex: noSQL, text mining)
- **Visualization**
(ex: statistical graphics, mapping, web-based data-viz)

Eliminating the Expertise Barrier

Algebraix has built an expert team of data analysts, data engineers, and data scientists in order to deliver the most complex, high-impact analytics efficiently. Because Algebraix does the recruiting, hiring, training, and retaining of personnel, clients don't have to. The company has fine-tuned its delivery model to optimize its activities across all functional steps – delivering efficiency in data acquisition, rigorous data preparation, and data science.

Co-Sourcing Delivery Model



Algebraix Data and its co-sourcing delivery model allow a blended analytics team that includes Algebraix Data scientists, engineers, and analysts that co-exist with a company's process and functional experts. There is no need to replace anything or anyone.

The Benefits of a Co-Sourced Data Science Model

Algebraix has eliminated the traditional cost and expertise barriers to data science and advanced analytics, opening the doors to companies of every size that want to compete and win.

By co-sourcing data science, Algebraix Data's expert data team leverages clients' internal and external systems and tools. This wide, deep flexibility makes it possible to choose the best ways to rapidly deliver data science and advanced analytics with minimum cost and disruption.

Here are a few of the benefits:

- Much lower costs across the board, hiring to hardware
- Faster time to market because projects can be scaled quickly
- High-quality results from a balanced team of data scientists, analysts and engineers
- Flexibility to adjust resources and cloud computing to meet changing requirements
- Assistance with loading and preparing data
- Simple, clear pricing with a flat monthly, all-inclusive fee
- Low level of involvement from your team, typically only 2 - 4 hours per week

Algebraix's Data Science as a Service – How To Get Started

The reasons for sitting on the advanced data science sidelines are gone – no matter a company's size, it's the rapid adopters who will win. Mid-market companies powered by data science can now outperform and outmaneuver their larger competitors.

Three Steps to Advanced Data Science

Data-Driven Workshop: A free Algebraix Data workshop is the smart first step. The seasoned Algebraix team works with senior management to prioritize and plan a strategy based on business impact, existing data, technology, skills, and goals.

Initial Analytics Project: We will mutually agree to an initial project to attain early success in revealing hidden value in your company's data. The typical initial project combines use-case selection, basic data preparation, and highly concentrated data science. This high-impact project often produces strong results and generates momentum for future analytics.

On-going Data Science as a Service: By co-sourcing, Algebraix's data experts become part of your strategic team. They bring deep experience in statistical modeling, predictive tools, and new approaches to each environment and extract more and more valuable insights. Algebraix delivers lasting value by integrating modern data science into existing business processes and platforms.

"The competitive edge to be gained from advanced analytics is no longer limited to a few techy companies or data-intensive industries. It's here today in all sectors and, as our survey results demonstrate, companies that commit to making the most of their data and investing in analytics capabilities are already outperforming their peers financially. A wait-and-see attitude is a luxury that no competitive company can afford." – **Bain & Company**

**Contact Algebraix Data
to make an appointment to discuss your specific needs.**

858-381-4800 or Info@AlgebraixData.com

