ALEX MORRISE

# DATA SCIENCE


## ABOUT THE AUTHOR

Alex Morrise is Chief Data Scientist at Pythian, a global data services company. He holds a PhD in Theoretical Physics from the University of California, Santa Cruz and is passionate about incorporating transformative intelligence into applications. Prior to joining Pythian, Alex founded TensorML, a machine learning company that provides intelligent applications using the full tensor structure of data. Alex built the intelligence powering Beats Music's streaming service (later acquired by Apple) and has served as Senior Data Scientist at Idle Games, Quid.com and advises several technology companies. When he isn't working, Alex can be found climbing trees with his daughter or training martial arts.


*While data science itself is not new, until recently it was the domain of an elite few—usually large enterprises that could afford teams of data scientists to explore and analyze large quantities of data. Now, data science is not only available to smaller enterprises, it's a mandatory competitive differentiator.*

*Data science lets you take the terabytes of machine-readable data that you already have, play with it, and test hypotheses to better understand your customers in completely new ways. It's a scientific approach to uncovering new information in your data.*

## UNLOCKING THE VALUE OF YOUR DATA

Business is becoming much more scientific and metrics driven. The volume of data and power of analytics are bringing new, previously unavailable insights that help companies truly transform the way they do business.

The key to unlocking the real value of your data is first understanding the data you have. At Pythian, we combine advanced algorithms with the best minds in the business and automation tools to explore your data and help you uncover insights in your existing data. From there, we can help you incorporate and analyze new, unstructured data for deeper user insights. Those insights become real features that change as your users' preferences evolve. The result is real-time, predictive analytics and a 360-degree view of your customers.

Retailers in particular have terabytes of behavioral data—click-stream data, page impressions, purchase history, navigational patterns, and more. Previously, retailers relied primarily on segmentation analysis. On the surface, this is a reasonable split because women tend to buy women's clothes and men tend to buy men's clothes. They also grouped people by spending or geographic clusters. In our experience, however, these splits don't allow for any granularity or in-depth analysis.

## CASE STUDY – GAMING

I recently did some work for a gaming company in California that was struggling to sell more games and to upsell within its existing user base. I looked at their user-generated data—how much time gamers spent on their own island versus exploring other islands; how much money they spent on their island versus another island; who worshipped which gods, etc.

I analyzed 40+ variables. Many of these variables might co-occur with each other, hence, to reduce this number, I looked for themes within the data—latent, behavior-driven characteristics—that are closely related to user personas.

This reduced space allowed me to see seven distinct behavioral clusters. People who spent time on other islands also did x, y, and z. Gamers who kept to themselves tended to do a, b, and c.

These behavioral clusters really come to life, however, when you regress demographic information onto them—female, male, age, location, etc.—and then ask questions. We discovered that one of the seven clusters was made up of 90% females, which was an enormous shock to the gaming company, who had assumed that most of their users were young men. Even more interesting was that these women were primarily between the ages of 34-45, were stay-at-home moms who played every day for a couple of hours and stopped around 2 pm—wherever they were in the world. We found that women in the Philippines, Singapore, and Europe, for example, behaved in the same way irrespective of where they lived. We also learned that this cluster also spent 3x more money in the game than other clusters.

Another cluster with approximately 22,000 gamers had a completely different set of behaviors. They tended to go to other islands and 'hate' on other players. When we regressed demographic information on this cluster, it turned out that these players were on average 18-26 years old, logged in sporadically throughout the day, including long periods during the night, and the gender was split evenly between male and female.

The gaming company also had a business intelligence (BI) team that did their own analysis using traditional BI tools. They looked at players between certain ages, number of log ins per day, average time per session, average money spent in a session—no behavioral data. To test and compare each group's findings, the gaming company split their ad buy money between the data science team and BI teams. We used our money to target the two behavioral clusters we identified as the spenders. The return on investment from our ads was 42% above baseline, while the BI team lost money.

At Pythian, we have the best and the brightest who will help you do things with your data that you couldn't do before. We reduce the complexity and cost of finding deep, multi-dimensional user insights. The result is real-time, predictive analytics and a 360-degree view of your users. Speak with one of our data scientists today to find out how we can help you.