

THE WALL STREET JOURNAL.

WEDNESDAY, MAY 9, 2018

© 2018 Dow Jones & Company, Inc. All Rights Reserved.

CIO JOURNAL.

Seeking Insights into Rare Diseases, Pfizer Scales AI Analytics Platform

Data science shouldn't be confined to mathematicians, says CIO Jeff Keisling

By Sara Castellanos

Pfizer Inc.'s recently-built analytics platform is helping employees from different divisions within the company collaborate in an effort to identify patients with rare diseases that might previously have gone undiagnosed, company executives said.

With the help of machine-learning algorithms, the analytics platform is now yielding new insights that were previously difficult to identify in a short amount of time because there too many disparate datasets to sift through.

The cloud-based Virtual Analytics Workbench tool officially launched in 2017 and more than 350 employees are using it including mainly data analysts and research and development teams. The goal is to scale it more broadly across different departments.

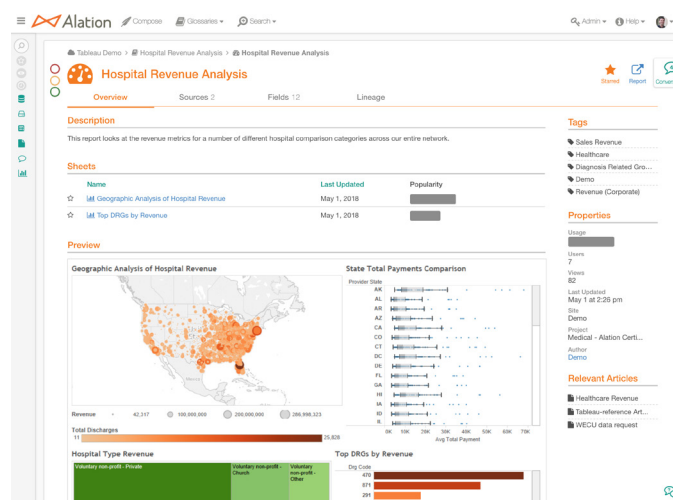
"While data science is a critical skill set in our company, you don't need a mathematician or data science expert to use (it)," said Jeff Keisling, Pfizer's chief information officer, in an email. It also gives analysts "more time to pursue insights versus finding them," he said.

The idea for the analytics tool came about in 2014 when Pfizer was beginning to explore different capabilities around health care and technology at a time when data was proliferating, said Julie Schiffman, Pfizer's vice president of business analytics.

"Data has become an extremely powerful currency for any company, and what we were



Julie Schiffman, Pfizer's vice president of business analytics
PHOTO: PFIZER INC



A screenshot of the Virtual Analytics Workbench tool, developed by Pfizer and vendors including Alation Inc. PHOTO: ALATION INC.

finding is the data was very fragmented," Ms. Schiffman said.

Back then, the company did not have one platform that could support searching, analyzing and visualizing different types of structured and unstructured

data, including physician notes, lab reports, demographics and comorbidities — which refers to the presence of more than one disease occurring simultaneously in a patient.

Such data points can number

in the millions. In one example, the tool is being used to explore those types of data and, using machine learning algorithms, build models to potentially identify patients with a rare form of heart failure called transthyretin cardiomyopathy, Ms. Schiffman said. This type of disease often goes undiagnosed because the symptoms are similar to other forms of heart failure.

Ultimately, the findings could be used to educate physicians about how to best identify transthyretin cardiomyopathy in a patient or how to identify patients at risk of the disease, she said. The tool could also eventually be used to find patients that would be good candidates for clinical trials, she said.

The platform was built by Pfizer using collaborative data science software tools from Alation Inc. and Dataiku, as well as data visualization software from Tableau Software Inc.

It took about two years to get the tool up and running, in part because Pfizer and the vendors had to spend time locating the data, figuring out who had administrative privileges to the existing data sources, and deciding which data sources to integrate, said Satyen Sangani, CEO of Alation. A lot of time was also spent training Pfizer employees on how to use the tool, he said.

Alation crawled through more than one million different data sources to discover potentially useful information based on how Pfizer analysts had used and trusted data in the past, Mr. Sangani said.