Every day, a small army of observers fans out through the markets of Mumbai and other Indian cities, snapping photos of onions and sundry other food items with their mobile phones.

Welcome to the era of advanced data correlation. The vegetable spotters are part-time employees of San Francisco, CA-based startup Premise, which uses data intensive, on-the-ground surveys of key indicators to pick up inflationary trends, often uncovering them long ahead of government sources.

Big Data is opening up a world in which millions of disparate data points can be mapped and correlated, illuminating economic patterns and providing business insights that weren’t previously available; until recently, Big Data was focused mainly on amassing large amounts of data and analyzing it. The emerging new trend is to cross-correlate large data sets in unconventional ways.

Premise, for example, addresses multiple data sets, including price data from shops in emerging markets such as India, China, Nigeria and
PRIMER: BASIC DATA STRUCTURES

Unstructured Data
Unstructured data encompasses a great deal of what occurs online and via mobile devices: emails, phone calls, text of all kinds, data streaming in from social media, internet searches, location data generated by GPS, Instagram photos, reviews, comments, video content, blog posts and much more. These masses of data do not fit naturally into normalized structures. This is quite different from the kinds of structured data most company IT systems were originally built to handle.

Structured Data
Structured data is data that neatly fits into preset rows and columns, and is readily retrieved through SQL (Structured Query Language) or other data retrieval methods.

Metadata
Metadata is data about data. For example: a photo is a piece of data. But that data source also contains a great deal of data within it. The location, date and time the photo was taken, specific elements within the photo such as individuals or objects, and so on. All of those pieces of data are metadata.

A great example of a company that makes use of metadata is movie rental and streaming service Netflix. It uses Big Data analysis to offer its customers viewing experiences that reflect—as closely as possible—their individual tastes. To keep them hooked, rather than relying on structured data like film ratings, Netflix looks at the millions of pieces of interactive data that users themselves generate. Thus, it knows, even before a viewer, that they’ll be watching the newest series.

To do this, Netflix uses language and speech models to gauge the impact of previews and promotional graphics on customers individually. The firepower required to process all this data exceeds the capabilities of conventional processors, so Netflix uses hundreds of linked processors to analyze a universe of data points. As this shows, Big Data is not only here to stay, but can be unlocked for myriad business advantages.

Summary
Structured data fits into rows and columns, unstructured data is everything that lives outside those boundaries, and metadata is data about data.

The emerging new trend is to cross-correlate large data sets in unconventional ways.
in combination with other information, such as map and economic data, to yield surprising insights and ideas. This is now possible due to the ubiquity and interoperability of the mobile devices and the innovative thinking of the company itself.

The correlations and insights gleaned are highly valuable to Premise’s customer base of hedge funds, banks, consumer goods manufacturers, retailers, national policy makers, and NGOs. Its platform also runs on Bloomberg terminals. The entire process represents a new frontier for Big Data, which is increasingly being referred to as “hyperdata.”

Premise is earning its stripes as part of a current crop of advanced analytics firms that are applying new thinking to the potential for correlation. This is possible due to the burgeoning supply of data sources in a digitally connected world. Backed by Google Ventures, Premise counts among its advisors Google’s chief economist, Hal Varian, and former White House economic advisor Alan Kreuger.

GETTING REAL VALUE OUT OF BIG DATA

Premise isn’t the only player in the emerging space. Others such as Kensho, Skybox, Relate IQ, and DataHug are joining the fray, each with its own take on unconventional data sets and the science of business intelligence.

As momentum builds, it becomes clear just how many opportunities there are hidden in this revolutionary new approach. Whether accessing external data or utilizing a client’s own data, correlative outputs can enable hedge funds to make more nimble investment strategies, banks to better serve corporate customers, sales professionals to generate better leads, and packaged food giants like Procter & Gamble to better understand their economics, to cite just a few examples.

Of course, correlation should not be confused with causation. As the startups themselves are quick to point out, smart people are still required to interpret the results. However, historical correlation patterns can help inform future projections.

Silicon Valley venture capital (VC) firms have already spotted the trend and see big opportunity for returns in a highly lucrative emerging space. Top tier Palo Alto, CA VC firm Accel Partners, for example, has raised two $100M Big
Data funds in the past three years in order to jump into that space, the second of the two focusing on software and analytics.

Hedge funds, unsurprisingly, were some of the first to take notice of these emerging startups, because of the edge they provide in showing potential market movements. In many other fields, the Big Data window into trend and event data should also enable companies that sign on to gain an edge in business intelligence initiatives.

**THE SKY’S THE LIMIT**

Mountain View, CA-based Skybox (acquired by Google in 2014) is a data analytics startup that uses high resolution satellite imagery to provide access to information about everything from activity in Walmart parking lots to ship movements in ports. Skybox is hard at work building a global network of small, nimble, and inexpensive “microsatellites.” Scanning the received satellite images and extracting metadata, the network is able to provide client companies with fine grained detail and deep insight into their business practices.

**PARTNERING WITH PREMISE**

As a direct result of the network of startups curated by Standard Chartered’s Silicon Valley Technology office, SC Studios, the bank forged a partnership with Premise to gather and analyze real-time, crowdsourced pricing data in Nigeria, a key emerging market country. The Standard Chartered-Premise Consumer Price Tracker (“The Tracker”) was launched in June 2014, providing the bank and its customers a significant competitive edge. The partnership also helped build goodwill in Nigeria through job creation and better economic insights.

“Our clients value highly our unique on-the-ground insights coming from our footprint research and we continue to work at the frontier of data sourcing, analysis and presentation with our innovative technology partners, such as Premise Data Corporation,” said Adrian Walkling, Head of Global Research, Standard Chartered Bank. “We recognise that differentiating our content through innovation keeps our research insights ahead of the pack.”

Through the partnership, the bank has been able to gain a much greater understanding of the Nigerian economy than was previously available. Government data on pricing is often inaccurate and out of sync with ground level realities. Premise deploys an army of employees using mobile phones in groceries and other shops in major metropolitan areas of the country, and builds a data intensive picture of the prices of staple goods such as grains, meats, and vegetables.

As early as the first two months that The Tracker was in operation, the bank found that its on-the-ground pricing data painted a very different picture of the rate of inflation than that published by the government. It was also able to develop a more finely grained picture, enabling it to see that certain staples—grains and flour—were driving the price downtrend.

During the Ebola outbreak, The Tracker enabled Standard Chartered to provide far greater insight into how it was affecting the economy as a whole. (See video on page 6.)

Clearly, this crowdsourced pricing data represents the wave of the future for financial institutions and others seeking new levels of insight and opportunity.
To accomplish this, Skybox combines metadata from their own satellites with image-based data from other sources: historical data from US government satellites, data extracted from aerial photography and images supplied by weather centers, for example. The goal is to create a web-based platform that enables an enterprise to monitor relevant aspects of its business, whether mining operations, retail activity or crop growth.

ARTIFICIAL INTELLIGENCE AND NATURAL LANGUAGE

Along another frontier, Big Data technology is becoming more accessible to non-technical staff by enabling users to interact with complex data engines in plain English. Cambridge, MA-based Kensho has developed a platform with the friendly name “Warren,” recalling IBM’s artificially intelligent computer “Watson” and its winning ways on the game show “Jeopardy!”

Warren, focused on finance, performs event-driven statistical analyses, and has built its own proprietary event database with, as the company described it, “a structured set of dates and events that move markets.” The platform can answer questions such as, “What happens to the share prices of energy companies when oil trades above $100 a barrel and political unrest has recently occurred in the Middle East?”

Warren can answer questions such as, “What happens to the share prices of energy companies when oil trades above $100 a barrel and political unrest has recently occurred in the Middle East?”

Wall Street investment banks and hedge funds —its chief users being portfolio managers, traders, and equity researchers. Because of the platform’s ability to map the correlation between market fluctuation and world events, there are myriad potential applications for a large swath of corporations including banks, especially as the technology develops and matures.

Currently, Kensho’s customer base includes leading startups in the space and acquisition of a third is an acknowledgment that advanced data correlation represents a massive market opportunity.

“If you look at what Apple is doing with ‘Siri,’ what IBM is doing with ‘Watson,’ and what Google is doing with ‘Now,’ it is clear that the most powerful technology companies are investing immense resources to make virtual intelligent assistants ubiquitous before the end of the decade,” said Kensho CEO Daniel Nadler in a news release.

The company is backed by Goldman Sachs — which led a $15M funding round in November 2014 — and top-tier VCs, including Google Ventures. Google’s backing of two similar
CUSTOMER RELATIONS IN THE BREACH

Some startups are taking aim specifically at CRM. Until now, such systems weren’t integrated with phone calls, emails or other key communications. This leads to time consuming manual data entry, errors and overlooked opportunities. As a result, sales relationships suffer, particularly now with the proliferation of tablets, smart phones, laptops, and desktops, devices that don’t allow for full syncing of cross-referenced and validated data.

Palo Alto, CA-based Relate IQ (acquired by salesforce.com in 2014) has stepped into the breach. Relate IQ ties together multiple data sources and gleans insights from sales employees’ interactions, acting as an intelligent guide to managing relationships, another way to use hyperdata. It isn’t a replacement for traditional cloud-based CRM management software, but rather a complementary and advantageous solution.

Relate IQ acts as an intelligent guide to managing sales relationships.

A similar startup, Datahug, based in Dublin, Ireland with an office in San Francisco, takes a different tack. Using a company’s own data, derived from corporate email and/or phone records, it creates a “dynamic business database” of connections within a company, and builds a map that shows every relationship in this ecosystem and connections to the outside world, tracked by volume and contact frequency.

Datahug can help turn a “cold” contact into a potentially “warm” one.

For example, if someone else in the company has a prospect in their network, the system tags the relationship so that the sales department can turn a “cold” contact into a potentially “warm” one. Datahug can also be deployed within an existing CRM system, providing a score or, as the company calls it, a “hugrank,” for each relationship’s value.

CONCLUSION

As the amount of data grows, the potential for better predictive analytics is also rising. Advanced data correlation has immense potential for competitive differentiation for the financial industry and beyond. By coming up with unusual and innovative combinations, a new wave of firms are pushing the boundaries of what can be discovered. This trend will no doubt gain momentum as its value becomes more widely known and understood.

Corporations can make use of the innovative startups in innovative ways. In addition to the insights they yield, there are additional internal advantages to advanced analytics. Some simplify processes and reduce errors by replacing manual entering of data with automation; most are cloud-based or rely on platforms that do
not require extensive IT resource allocations. And many of the new programs are designed for non-technical users and allow queries in natural language, or close to it, freeing up technical personnel for higher-level tasks.

In short, selectively and smartly applying this new type of advanced data correlation could open new opportunities.

Selective and smartly applying this new type of advanced data correlation could open new opportunities. Banks, their clients, and many others could benefit from predictions about how global events affect markets, and share prices could be modeled on multiple data sources. These new sophisticated analytics, and the unique insights generated by correlating different data sets, bring increased richness to data modeling.

This new set of tools can help drive business decisions, enable new customer relations, and provide insights that improve accuracy when it comes to identifying customer needs or serving up the most compelling product at the optimum time.

Economic impact of Ebola

Standard Chartered’s partnership with Premise provides insights into how food prices—a key economic indicator—are affected by the outbreak.