Analytics mindset

Alternative data

Overview

Background

Accounting professionals have long relied on traditional data — the stream of information traditionally provided by companies (e.g., earnings releases, SEC filings, management presentations), governments (e.g., GDP, trade data, unemployment, surveys) and analysts (e.g., company and industry reports). In recent years, a variety of stakeholders like investors, analysts, corporate executives, auditors and consultants are increasingly turning to alternative data — data that flows outside the stream of traditional data — in order to improve the speed and accuracy of their decision-making. In this case, you will learn about factors that have fueled the growth of alternative data, some common types of alternative data, how the data is being used by professionals to improve decision-making and challenges that often emerge when using alternative data.

What is alternative data?

Imagine that Mother's Day is approaching, and you want to buy a gift for your mother, or a mother figure in your life. So, you hop online and search. You can't find the right gift online, so you drive to the mall, browse a few stores and finally make a purchase. You drive home and share with your friends on social media your excitement for Mother's Day.

This simple purchase has created several types of alternative data. What alternative data did your Mother's Day purchase create? When you searched online, websites captured all of the item pages you looked at, how long you looked at each item, whether you added those items to a purchasing cart, your IP address and other browsing behavior through web cookies.

As you parked your car at the mall, a satellite in orbit captured a picture of the mall parking lot to see how full it was. Then, as you browsed stores at the mall, the weather app in your cell phone tracked your location. The weather app sold this data — along with your demographics — to a company that compiled a record of how long you paused and shopped at each store throughout the mall.

Your credit card company also recorded this transaction and combined it with your other credit card transactions to compile information about your purchasing behavior. Finally, several analysts and companies captured your social media post in sentiment analyses that they are conducting to forecast industry-wide sales and earnings growth.

Growth in alternative data

Alternative data is growing at a rapid pace. Worldwide production of data has more than tripled to several dozen zettabytes per year over the last five years and is expected to increase exponentially over the coming decades (see Figure 1 following). To put this in perspective, you would need approximately 660 *billion* Blu-ray discs to store all of the data that was created in 2018 alone! Much of this growth is driven by the supply and demand for alternative data. For example, investors have rapidly increased spending on alternative data.

SCORE no. 16691-221US_4

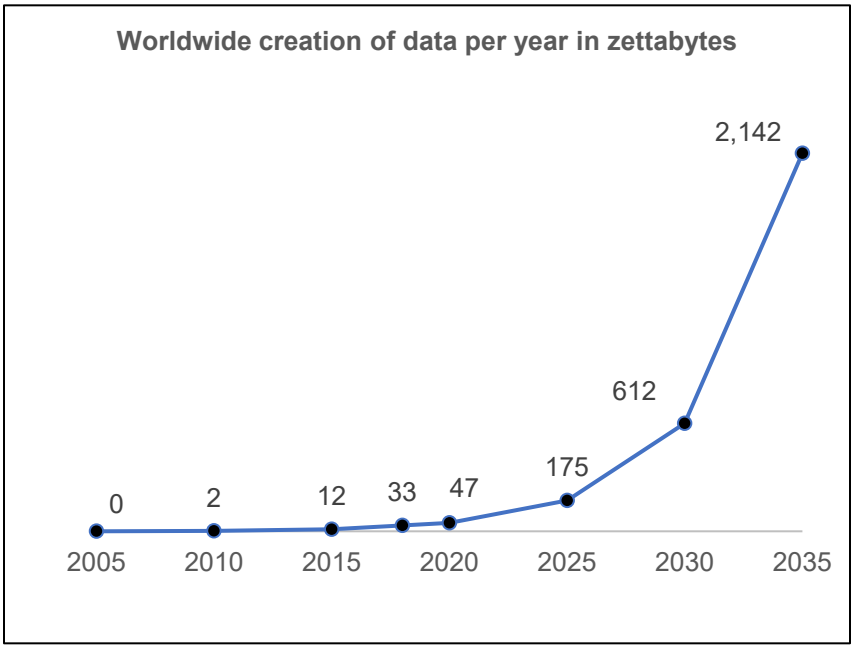

Figure 1: Growth in alternative data
Source: Statista Digital Economy Compass 2021

Three factors have enabled growth in alternative data:[1]

- ***Increase in the prevalence of devices and sensors that capture data.*** Today, 5 billion people interact with devices and sensors like cell phones, computers and smart watches that capture their data every day. This figure is expected to grow to 6 billion by 2025, with each person creating new data every 18 seconds.[2]
- ***Advances in cloud computing.*** Advances in cloud computing have simultaneously increased data storage and computing power while reducing costs for both, allowing stakeholders to store and analyze large quantities of alternative data in an economical manner.
- ***Advances in machine learning and artificial intelligence.*** Scientists have developed new capabilities in machine learning and artificial intelligence that permit users to gain insight from large unstructured and varied data sets.

Types of alternative data and its application

Numerous sources of alternative data are already being leveraged by companies, but current sources will soon be outnumbered by new emerging sources. Eagle Alpha, a pioneer in alternative data, currently identifies 1,200 alternative data sets that span 24 different categories. They forecast that this number will more than quadruple in just the next year alone.

---

[1] Kolanovic, Marko and Krishnamachari, Rajesh T., "Big Data and AI Strategies," *J.P. Morgan*, https://www.cfasociety.org/cleveland/Lists/Events%20Calendar/Attachments/1045/BIG-Data_AI-JPMmay2017.pdf, accessed July 2022.
[2] Reinsel, David, Gantz, John, et al., "The Digitization of the World From Edge to Core," *IDC*, https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf, accessed July 2022.

A brief overview of some of the most common types of alternative data is provided below. This overview focuses on three overarching categories:

- Geospatial and location data
- Data captured through open sources
- Data captured through proprietary sources

Within each category, the examples provided focus on how companies are leveraging alternative data.

Geospatial and location data

A variety of alternative data sets identify the location, movement and status of customers or products. Investors, auditors and analysts use this information to increase the speed and accuracy of their sales and production estimates.

Geolocation data. Eighty-one percent of Americans own smartphones, which typically have several apps that track the smartphone's location.[3] Alternative data companies purchase this location data from the apps and use it to estimate customer traffic, customer dwell times and ultimately customer sales. They can also use this data to estimate manufacturing production.

Example 1: Using cell phone location data to estimate Tesla car production[4]

Advan is a leading provider of geolocation data. They created geofences — virtual boundaries for real-world physical locations — around Tesla's new car parking lots (see Figure 2 below) and use cell phone location data to capture how much employee foot traffic occurs within the geofences.

The company argues that employee foot traffic within new car parking lots is a reliable indicator of Tesla's car production. In fact, Advan has been able to produce highly accurate forecasts of Tesla's car production over the past several years. They report a very high correlation (0.94) between their forecasts and Tesla's company-reported production.

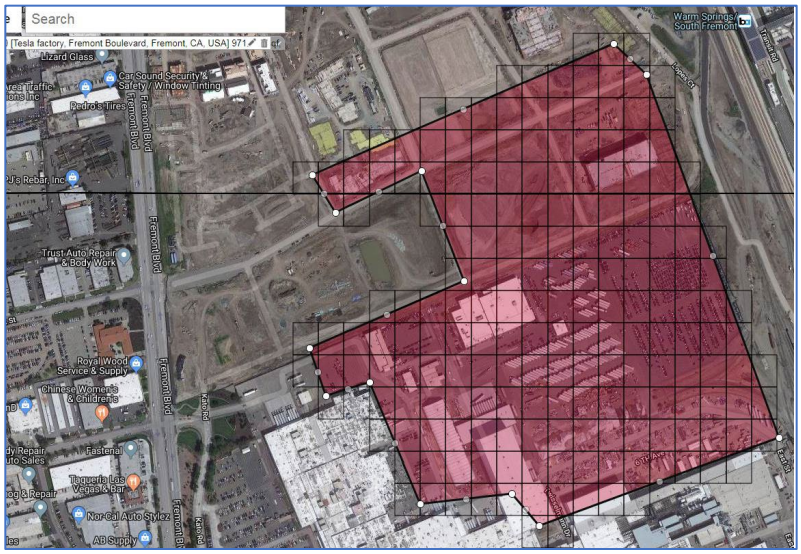

Figure 2: Advan geofence of Tesla new car lot [4]

[3] "Mobile Fact Sheet," *Pew Research Center website*, https://www.pewresearch.org/internet/fact-sheet/mobile/, accessed July 2022.
[4] "Tesla Motors Analysis Q1 2019," *Advan,* https://www.advan.us/downloads/Advan%20Tesla%20Analysis%20Q1%202019.pdf, accessed July 2022.

Satellite and drone imagery. Several alternative data vendors have launched a collection of nanosatellites into space in order to capture daily pictures of specific locations. For example, some satellites take daily pictures of retail parking lots in order to estimate the number of customers within the store. Other satellites take daily pictures of oil reserves, mineral stores, crop fields or shipping routes to estimate production and trade. More recently, companies have started using drones to monitor specific locations. Drones can be equipped with specialized sensors (e.g., thermal sensors and chemical sensors) that allow users to extract additional information, such as the health and maturity of agricultural production.

Corporate aviation data. Many investors have started tracking the movement of corporate jets in an attempt to anticipate major corporate transactions — such as mergers and acquisitions.

Example 2: Using corporate aviation data to predict Johnson & Johnson's $30 billion acquisition of Actelion[5]

On January 26, 2017, Johnson & Johnson announced a $30 billion acquisition of Actelion, a biotechnology company. Days before this official announcement, three hedge funds tracked a Johnson & Johnson corporate jet as it parked for several days near Actelion's headquarters near Basel, Switzerland. The fund managers used this information to predict and trade on Johnson & Johnson's acquisition of Actelion before the acquisition was publicly announced. The fund managers reported that they "grew increasingly confident of a big score" every day the jet was parked near Actelion's headquarters.

Aviation data is available online from various public sources and aggregated by private vendors (see Figure 3 below).

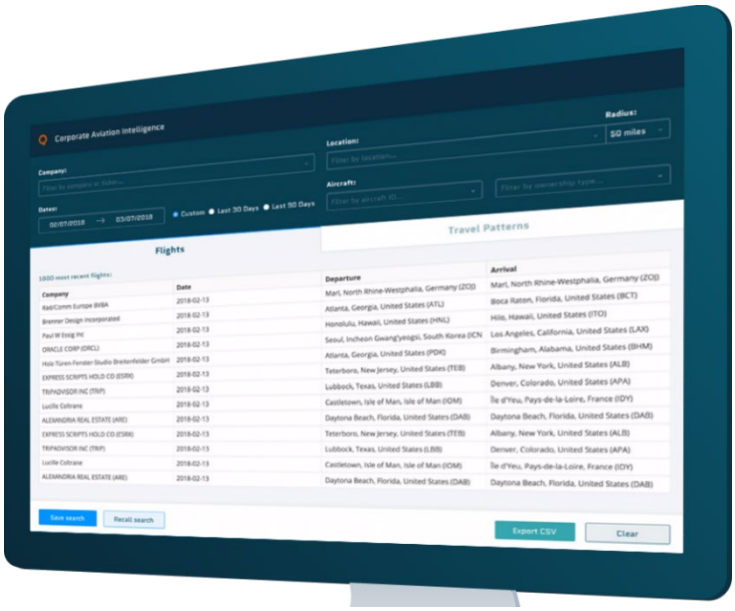

Figure 3: Corporate aviation data
Source: Nasdaq Corporate Aviation Intelligence (https://data.nasdaq.com/jets-aviation-data-corporate-intelligence)

[5] Kumar, Nishant, Nair, Dinesh, et al., "Hedge Funds Strike Paydirt on Actelion Deal After Tracking J&J's Jet," *Bloomberg website*, https://www.bloomberg.com/news/articles/2017-01-27/hedge-funds-track-j-j-private-jet-for-an-edge-on-actelion-score, accessed June 2022.

Data captured through open sources

Many alternative data sets are compiled from information publicly available on the web. In order to create these data sets, providers identify useful information and implement techniques (e.g., web scraping) to repetitively capture and store that information over time.

Social media and news sentiment. Social media and traditional news media can be used to evaluate the financial prospects of products, companies, industries or the economy as a whole. Decision-makers often extract information from social media and news media through a technique called sentiment analysis. Sentiment analysis involves gathering large quantities of data (e.g., tweets on Twitter), identifying the source and topic of each instance of data, using natural language processing to extract meaning from each instance of data and aggregating that meaning across the data.

Example 3: Company use of sentiment analysis[6]

Companies can use sentiment analysis to monitor positive and negative tweets, such as those shown below, in real time in order to better understand how consumers are reacting to their products and brands.

Figure 4: Examples of positive and negative tweets in sentiment analysis
Source: Fictional tweets

The output of sentiment analysis can also be used to improve quantitative models that incorporate other sources of information. For example, academic research provides evidence that the aggregate opinion from individual tweets on Twitter can predict the earnings that companies report in their quarterly earnings announcements, along with stock price movements around those announcements.[7]

[6] "Twitter Sentiment Analysis in Real-Time," *MonkeyLearn website*, https://monkeylearn.com/blog/sentiment-analysis-of-twitter/, accessed June 2022.
[7] Bartov, Eli, Faurel, Lucile, et al., "Can Twitter Help Predict Firm-Level Earnings and Stock Returns?" *American Accounting Association website*, https://doi.org/10.2308/accr-51865, accessed June 2022.

SCORE no. 16691-221US_4

Internet Reviews. Customers often post product reviews online, and employees often post reviews about various aspects of their employers online. Companies, investors and analysts can use this information to evaluate products, employee morale and overall company performance.

Example 4: Using Glassdoor employee ratings to predict stock price movements for Netflix[8]

Glassdoor.com provides a platform where employees can leave reviews about their employers. Accern Corporation, a vendor of analytics for institutional investors, used Glassdoor employee ratings to predict share price movements for Netflix. As seen in Figure 5 below, a rise in the average employee rating for Netflix preceded a similar increase in the stock price.

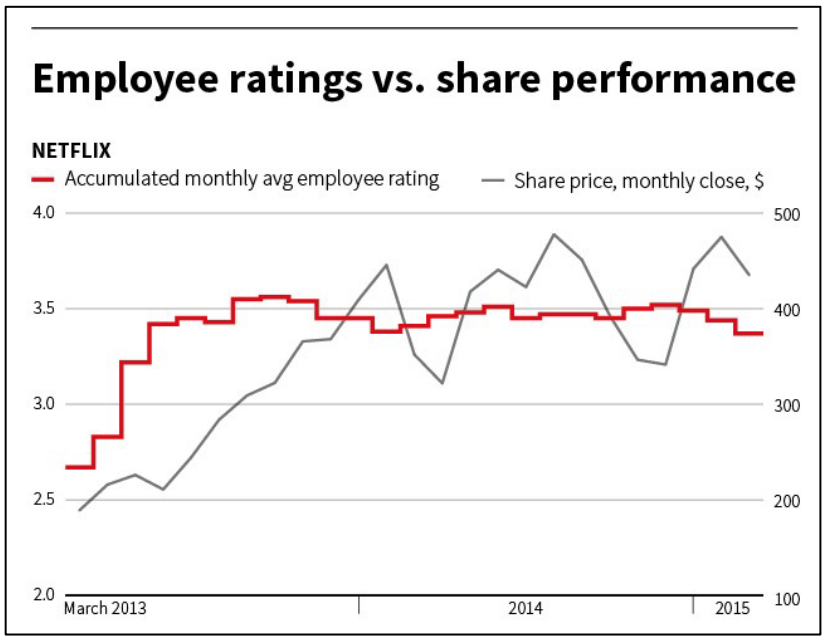

Figure 5: Netflix employee ratings and share performance

Company nontraditional data. Companies post a great deal of information online that can be used in nontraditional ways to evaluate those companies. For example, companies post product prices and new job postings online. In turn, this information can be used by investors and analysts to predict sales, profitability, growth and stock price movements. Eagle Alpha has conducted analyses showing that new job postings can be used to predict positive stock returns, resulting in profitable trading strategies.[9]

[8] Toonkel, Jessica, "Insight: Does a happy employee make for a healthy stock price?" *Reuters website*, https://www.reuters.com/article/us-investors-workers-sentiment-insight-idUSKBN0MZ01V20150409, accessed June 2022.

[9] "The Rise of Data – Leveraging Data in Your Investment Process" *Eagle Alpha* https://www.eaglealpha.com/2020/04/01/the-rise-of-data-leveraging-data-in-your-investment-process/, accessed July 2022.

Data captured through proprietary sources

Many alternative data sets are compiled from information not publicly available on the web. Some of these data sets are generated as a by-product of other business processes, or "exhaust" data. For example, the primary purpose of supermarket scanners is to identify the price and quantity of products purchased in order to compute a transaction total, record the sale and generate a customer receipt. However, the information produced by supermarket scanners can be used for a variety of other purposes, such as identifying product and customer trends or measuring employee productivity. Other alternative data sets are created for the express purpose of improving decision-making.

Web traffic and search. Web traffic and search have been a long-standing source of alternative data for companies, investors and analysts. Trends in web traffic and search can be used for a variety of purposes, including measuring demand for products, evaluating the effectiveness of marketing campaigns, measuring company growth, etc. Traditionally, data vendors have tracked the web traffic and search of individuals. However, data vendors have also started to track the web traffic and search of companies by identifying the IP addresses associated with specific companies. Company-specific web traffic and search trends can be used for a variety of purposes, such as identifying businesses that might be interested in specific B2B products (see example 5 below), identifying challenges that specific companies are facing, inferring the R&D activities of companies, etc.

Example 5: Salesforce uses corporate web traffic to increase ROI on digital advertising[10]

Bombora is an alternative data provider that has partnerships with thousands of websites that share anonymized web traffic and search data with Bombora. Using IP addresses and natural language processing, Bombora identifies the company associated with the web traffic and the topic that is being researched. They then aggregate this data to identify topics that different companies are researching.

One of Bombora's customers, Salesforce, turned to Bombora to help the company identify potential corporate customers. Using Bombora's data, Salesforce was able to target corporate customers who were already researching topics related to Salesforce's products. This allowed Salesforce to increase its ROI for digital advertising by 271%.

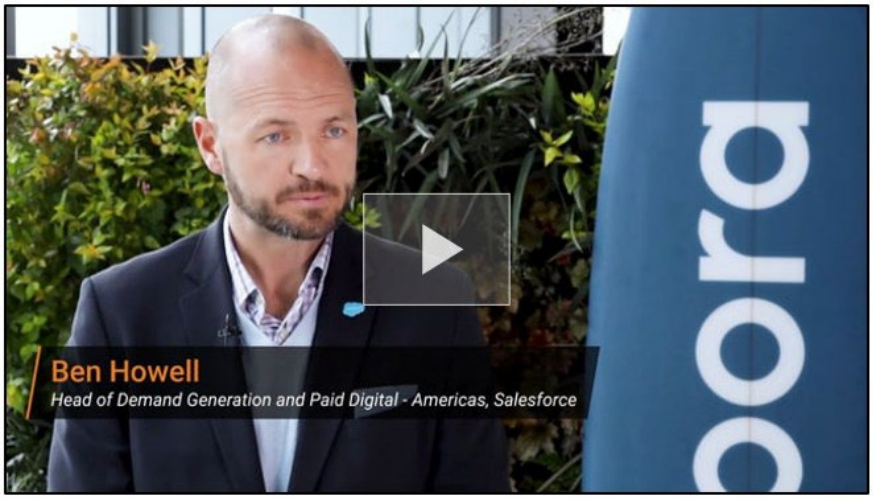

Figure 6: Video on Salesforce and Bombora[10]

[10] "Intent data powers display and paid social," *Bombora website*, https://bombora.com/blog/case_study/salesforce/, accessed July 2022.

SCORE no. 16691-221US_4

Mobile app usage. Several data vendors track detailed information about mobile apps, including downloads, usage and revenues. This information can be used to evaluate the success of products and companies that have a large digital presence (e.g., Netflix). It can also be used to identify larger industry and economic trends, such as the extent to which customers are moving from brick-and-mortar commerce to online commerce.

Advertising data. Corporate advertising can be tracked at the point of sale (tracking how much companies pay to advertise on different platforms) or at the point of advertisement (tracking advertising campaigns). Companies like MAGNA capture this data by rigorously monitoring television, internet and other media advertising and combining that data with traditional sources of data like advertiser revenues. This information can be used by outsiders to forecast product and company growth.

Challenges with alternative data

Alternative data creates many challenges and risks for the decision-makers who adopt it. We briefly discuss two of the biggest challenges and risks below.

Privacy and reputation risks

The public is increasingly concerned with companies acquiring and using data that could be considered private. Although most companies and alternative data vendors diligently comply with laws and regulations, many jurisdictions are pushing for more stringent privacy laws and regulations. Even when complying with laws and regulations, companies face reputational costs when violating social norms regarding personal privacy. Although there are concerns about privacy, much of the collection and use of alternative data is not transparently revealed, making it difficult to know what companies are actually tracking.

Blind alley data sets

Investing in and using alternative data can be costly, and users of alternative data often do not know the marginal benefit of alternative data before making the investment. This dynamic can lead to what J.P. Morgan refers to as blind alleys:[11]

> "Certain types of data may lead into blind alleys — datasets that don't contain alpha, signals that have too little investment capacity, decay quickly or are simply too expensive to purchase. Managers may invest too much unnecessary infrastructure (e.g., build complex models and architecture that don't justify the marginal performance improvements)."

Many data sets lead into blind alleys for three related reasons:

- ***Confusing correlation and causation.*** Many alternative data sets can demonstrate a strong historical correlation with a variable of interest without capturing a causal relation with the variable of interest. The predictive power of these data sets can change drastically over time, and in some cases become worthless over time.
- ***Data quality.*** Many alternative data sets produce a high volume of low-quality data. Although machine learning and artificial intelligence can sometimes produce valuable information from such data sets, low data quality often pushes costs above the marginal benefit of the data. In a survey, EY found that

[11] "Big Data and AI Strategies," *J.P. Morgan*, https://www.cfasociety.org/cleveland/Lists/Events%20Calendar/Attachments/1045/BIG-Data_AI-JPMmay2017.pdf, accessed June 2022.

50% of senior executives view poor data quality as a key challenge with big data, ranking highest among all the challenges cited by the executives.[12]

- ***Data paralysis.*** The most valuable alternative data sets enable companies to improve both the quality and speed of decision-making. However, alternative data can deteriorate decision-making when executives distrust the data, misunderstand the data, lack necessary skills to generate models or are paralyzed by data analysis.[13]

Required:

- Search online and find an example of how alternative data is being used to improve the quality or speed of decision-making.
- Identify an important decision that professionals often make in the area of business that you are most interested in.
  - Brainstorm how different types of alternative data could be used to improve decision-making in this area.
  - Brainstorm other risks associated with the use of alternative data in this area.

[12] "Global- Becoming an Analytics Driven Organization," *EY*, https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/digital/ey-global-becoming-an-analytics-driven-organization.pdf, accessed July 2022.
[13] Dykes, Brent, "Big Data Paralyzing Your Business? Avoid These 3 Common Traps," *Forbes website*, https://www.forbes.com/sites/brentdykes/2016/09/28/big-data-paralyzing-your-business-avoid-these-3-common-traps/?sh=c7e704128bc9, accessed July 2022.