DATA

It's not how big it is,
but what you do with it.

The long-lived nature of real estate, and real assets more broadly, massively exposes real estate investors to the emergence of new technologies, such as rideshare and homeshare platforms, and increased transparency through "big data."

Dr. Nils Kok, *Chief Economist at GeoPhy (NY) and*
Associate Professor at Maastricht University (NL)

For investors in buildings that stay around for decades, sometimes even centuries, it is tempting to dismiss technological innovations as irrelevant. But as the real estate sector is poised to become more data driven, this will 1) lead to (partial) automation of now-manual tasks, such as site selection and due diligence for acquisitions, valuations, and mortgage origination, 2) enhance efficiency and liquidity in the commercial real estate market, leading to an increased capital flow, and 3) affect traditional drivers of real estate demand, with significant consequences for the existing capital stock, including its pricing.

"Big data" has quickly emerged from obscurity into a buzzword, rising from non-existent in 2011 Google Trends numbers, to peak popularity in 2016. Indeed, peak popularity of "big data" was last year. In 2017, the world no longer cares so much about how big data is, but rather, what you can do with it. Because without application and analysis, data is like raw oil without the refinery.

For the commercial real estate sector, the emergence of "big data" is part of the broader rise of "proptech," another catch-all term that includes the many sides of the real estate market. As a short digression, it is worth looking at "proptech" in a bit more detail. It is easiest to decompose the hundreds, or perhaps thousands of tech firms in the real estate space by using a well-known framework for real estate market analysis: the four-quadrant model. As academic as that framework is, the workings of the real estate market can be boiled down to a combination of the construction market, the market for space (where landlords and tenants interact), and the market for assets (the capital market, where sellers, buyers, and providers of capital interact). As illustrated by the graph above, each part of the market has its own set of disruptive "tech" firms – I listed just a few names, but the list is virtually endless, and growing every day.

Asset/Capital Market
GeoPhy
Opendoor.com
RealtyShares
Blockchain

Space/Tenant Market
VTS
CompStak
Reonomy
AirBnB
WeWork
AppearHere

Construction Market
OnTarget
Honest Buildings
Ravti
BIM
3D printing

Other
Ridesharing (Uber/Lyft)
Self-driving cars (Tesla)
Telehealth (Teledoctor)
Etc.

In the construction market, some big trends include BIM modeling (increasingly commonplace), 3D printing (the first commercial building using 3D printing has been constructed, in China), and the Internet of Things (IoT) creating "smart" buildings (OVG's "Edge" building in Amsterdam and Oxford Properties' "RBC Waterpark Place" in Toronto are early adopters of IoT technology).

In the space market, there are many example of disruptive technologies. From the obvious, such as the offline-to-online retail shift, more advanced brokerage platforms aimed to obviate expensive and unneeded brokers, to rideshare platforms changing the value of location. And don't underestimate the impact that platforms such as WeWork and AirBnB may have – with sticky supply (it's hard to move buildings) small changes in demand create large fluctuations in value.

In the real estate capital market, there are multiple disruptive technologies that have direct relevance for fund managers, REITs, developers, lenders, and other players in the capital market. As a source of capital, crowdfunding and arbitrage platforms are emerging, albeit mostly focused on the single-family residential sector, where liquidity is typically higher and ticket size smaller (see, for example, Opendoor.com). Two other fields in the real estate capital market that are ripe for disruption include the role of traditional research and analytics, primarily focused on site and asset selection, and the archaic role of appraisers.

Big data for site selection and due diligence purposes

For real estate developers and investors, finding "the next Williamsburg" or "the next Soho" is often considered the holy grail. But, how to identify what areas (may) gentrify? At the macro level, site selection is often done using data on demographic trends, including population, employment, and income growth. But once a city has been selected for capital allocation, Census data is neither granular, nor frequent enough to be useful.

Enter big data. Because at the level of the city, there are a multitude of data layers that can be used for further differentiation, as detailed as an inch, as frequent as every second. Think about data on bars, restaurants, coffee shops, art galleries, (music) events, crime, the location and frequency of pictures uploaded onto social media, homesharing (e.g. listings on AirBnB), WiFi hotspots, public transport stops, rideshare pick-ups and drop-offs. And all other data layers that Richard Florida wrote about in his "Rise of the Creative Class."

These "modern" data layers, which can often be accessed through APIs, lead to unparalleled market intelligence that enables real estate investors, developers and lenders to make more informed, data-driven decision. As opposed to, or perhaps in addition to, "heard on the street" or gut feeling. But the trick is often to access and make sense of all these different layers of information, as most real estate firms are not set up to ingest many different sources of primary data, but are rather used to consuming data in aggregated format (think: data on rent levels, price developments, etc.).

To illustrate how new, granular sources of data can be used by the real estate sector, my firm, GeoPhy, developed a simple site selection tool, the "Grapevine Dashboard" (after all, we often hear about new, upcoming, cool areas through the proverbial grapevine). The Grapevine Dashboard pulls a vast real-time dataset

Information on relative rental levels in a city in conjunction with the relative increase/ decrease in local economic vibrancy provides a "gentrification indicator" that can be used for site selection.
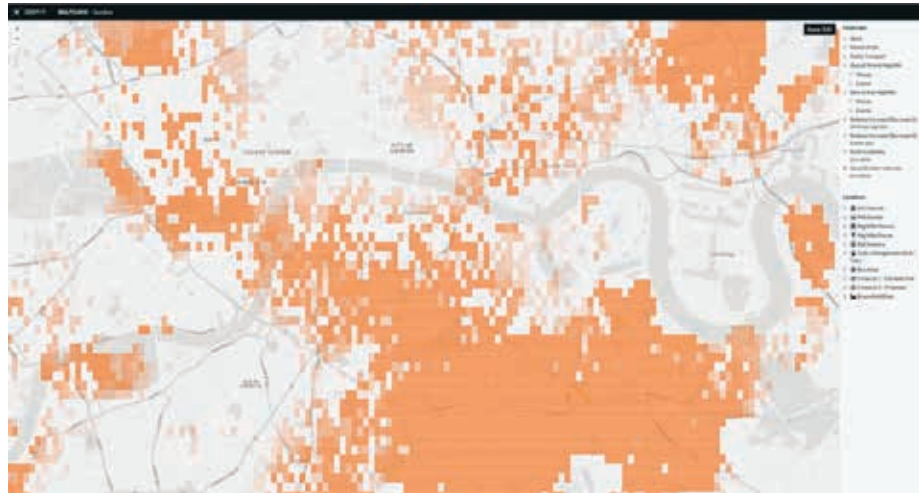
of physical venues and one-off events in a city to provide a detailed view of arts, nightlife and other leisure activities. This includes the exact location of amenities like bars, cafes, art galleries, music concerts, food festivals and exhibition openings. By utilizing proprietary methods to calculate locational proximity and distribution of such venues or events, we can spatially sort and rank this data across entire cities.

The picture above shows, for London, which areas have the greatest number of new bars, cafes and restaurants (in this case, those which have opened during the past two years). However, this view can also be filtered to look at a more specific subset of the data, such as restaurants or coffee shops that are not part of the major chains. The data also allow us to look at less spatially fixed behaviour, such as events that might not take place in a traditional venue; which could be pop-up markets, outdoor cinemas or all-night warehouse raves. In doing so, we can access a view of the city that is rarely seen; and understand which neighbourhoods certain groups of people are likely to go to in order to spend their money and enjoy themselves.

We then combine the increase and decrease in an area's popularity with rental price data. Information on relative rental levels in a city in conjunction with the relative increase/decrease in local economic vibrancy provides a "gentrification indicator" that can be used for site selection. Low rents and strong increases in economic activity, that's what developers and investors should be looking for, which is exactly what the Grapvine Dashboard points out.

Of course, macroeconomic data may shift some of the institutional focus away from London. But the concept of Grapevine specifically, and the use of big data more generally, can equally be applied in every market for every property type. As the availability of data grows, investment decision making will increasingly be informed through objective analysis. And if it makes you feel better, you can always add a dash of gut feeling.

BIG DATA FOR REAL ESTATE VALUATIONS

Determining the value of commercial real estate remains elusively hard, with a workforce of 74,000 appraisers in the U.S. alone still manually assessing the value of assets sometimes worth billions of dollars. In addition to the significant increase in the availability of data that the real estate industry has witnessed, there is the advent of machine learning techniques. Such techniques are now widely used in medical research, as well as in applications such as search algorithms and recommendation engines. (Examples of machine learning applications vary widely, from fraud detection by PayPal to the personalized online ads that we have grown used to.) In the single-family housing market, there are some nascent products that use this combination of data abundance and machine-learning modeling, including, for example, House Canary's automated valuation model, and Zillow's much-discussed "Zestimate."

We apply "big data" in combination with sophisticated modeling techniques, to develop an automated, machine-based valuation model for the commercial real estate sector. We first focused on the multifamily sector, enabled by access to a dataset of some 54,000 U.S. multifamily assets. This dataset is enriched by a wide set of both standard demographic and economic measures and more modern, "hyperlocal" metrics, such as proximity to music events, bars and restaurants, green space, and local crime incidence.

Rather than traditional hedonic models, which are limited both statistically and by a researcher's predisposition towards "standard" explanatory variables, we then apply assisted machine learning models that rely on (stochastic) decision trees. These models can sift through millions of combinations of thousands of variables, training and testing the model on randomly selected parts of the datasets, leading to precise out-of-sample tests of predictive performance.

We find strong evidence on the superiority of automated valuation models (AVM) over traditional appraisals – the median absolute error of the automated model we develop is below 9%, which compares favourably against the accuracy of traditional appraisals, while the model can produce an instant value at every moment in time, at a very low cost. We also find evidence on the importance of using "hyperlocal" information on the location of an asset. While the use of economic and demographic data at the Census tract and ZIP code level are standard practice in real estate modeling, new information layers gleaned from a wide variety of sources, including social media data, police records, and amenities related to economic vibrancy, add significant value to pricing models.

The implications of a well-functioning AVM are significant. First, timely estimations of property values are critical for real estate investors and lenders to make informed underwriting decisions, where systematic errors or biases in valuations may have adverse effects on the provision of equity or debt. Second, investors, regulators, and others rely upon appraised values to assess returns on the USD11 trillion U.S. commercial real estate market, and automated property valuations can provide a more accurate reflection of both the real estate stock (i.e. the value of real estate on the balance sheet) and flow (i.e. real estate returns from changes in capital values). Third, automated valuation models can be used for stress testing under adverse economic scenarios, which post-crisis remains a much-needed tool for regulators, banks, rating agencies and investors. Fourth, the availability of an instant, accurate property value may spur financial innovation in the real estate sector, such as automated loan origination by banks, defined-contribution products for private real estate investments, and arbitrage products for commercial real estate (comparable to emerging products for the single-family sector).

THESE BUILDINGS HAVE BEEN HERE FOR DECADES! (AND SO HAVE I)

For investors in capital goods that stay around for decades, sometimes even centuries, it is tempting to dismiss the emergence of "proptech" as irrelevant for the industry. After all, you may not sleep in an AirBnB, but rather in the Fairmont. And WeWork is probably not for you or your company, preferring a long-term lease with extensive fit-out package and dito incentives. And 3D printing is just for the Chinese, not for high-quality US real estate development.

But the long-lived nature of real estate, and real assets more broadly, massively exposes real estate investors to societal shifts and long-term trends. As an analogy: we all understand that small changes in demographics can have large impacts on real estate pricing. When Amsterdam lost (just) 20% of its population in the late 1700s, house prices declined by 80%. Similarly, just 6% of total retail expenditures moved from bricks-and-mortar shops to online channels, leading to the current bloodbath in retail real estate (although the retail landlords want you to believe differently).

Rather than denying change, the real estate sector should embrace it. Sure – some roles will change, and disintermediation may happen. Some will resist, comparable to the utility industry resisting renewables, and the proponents of the internal combustion engine resisting electric cars. But think about it differently: online retail means opportunity for real estate. Ridesharing means parking spaces can be repurposed to more effective use, both on and off-street. Big data creates transparency that will allow for further institutionalization of real estate as an asset class, upping the share that institutional investors will allocate. Uncertainty abound, but that's where companies led by innovative, smart executives thrive.

Note: The content of this article partially draws upon the GeoPhy Blog and a forthcoming article in the Journal of Portfolio Management, "Big Data in Real Estate? From Manual Appraisal to Automated Valuation.". For more on Google Trends as a leading indicator, I highly recommend the book "Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us Who We Really Are" by Seth Stephens-Davidowitz