

URBAN TECH ON THE RISE: MACHINE LEARNING DISRUPTS THE REAL ESTATE INDUSTRY

Featuring interviews of:
**Marc Rutzen and Jasjeet Thind by Stanislas Chaillou, Daniel Fink
and Pamella Gonçalves**



Stanislas Chaillou, is an architecture master candidate at Harvard University as well as Director of the CitiX Initiative at The Future Society. He has previously worked for international renowned offices such as Shigeru Ban Architect, and Adrian Smith and Gordon Gill. He is also the recent winner of the American Architecture Prize (student category). Stanislas is a Fulbright scholar, an Arthur Sachs Fellow and Jean Gaillard Fellow at Harvard University.

Daniel Fink is a SMarchS candidate at MIT. He is director at Urban Agency and worked previously at Grimshaw Architects.

Pamella Gonçalves is a MBA candidate at MIT. She previously worked at Endeavor Brasil, as Head of Research and Public Policy.

KEYWORDS

- PREDICTION
- DATA ANALYTICS
- REAL ESTATE
- MACHINE LEARNING

The practice of AI-powered Urban Analytics is taking off within the real estate industry. Data science and algorithmic logic are close to the forefront of new urban development practices.

How close? is the question—experts predict that digitization will go far beyond intelligent building management systems.

New analytical tools with predictive capabilities will dramatically affect the future of urban development, reshaping the real estate industry in the process.

INTRODUCTION

In his introduction to ‘Smart Cities,’ Anthony Townsend raises the issue clearly: “Today more people live in cities than in the countryside, mobile broadband connections outnumber fixed ones and machines outnumber people on a new Internet of Things.¹” Yet neither the glossy marketing of major IT players such as IBM and Cisco nor the dystopian theories of critical scientists like Adam Greenfield admit that the digital revolution washing over cities has yet to be fully evidenced. Instead, over the last decade we have witnessed the slow emergence followed by strong growth of the computational paradigm applied to urban planning and real estate.

As the travel and tourism sectors demonstrate, Big Data and machine learning can radically transform entire industries. Widespread disintermediation combined with newfound efficiencies have massively empowered consumers while destroying the traditional roles played by brokers and agencies. Similar trends are occurring in the real estate industry. Although the disruption is still elusive, transformation is underway. In fact, the fragmentation of the sector and the inertia of the profession maintain market opacity and mask reality.

The urban analytics realm has been growing steadily since the 1980s and the inception of personal computers. From a simple aggregation of data on the web to database scraping and filtering, the 90s saw a refinement of real estate-tech tools. With cloud computing and high-speed internet, online platforms are now bringing the potential of machine learning, neural networks,

¹ Anthony Townsend, “Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia.”

and artificial intelligence² to real estate market forecasting. Such tools have already disrupted traditional practices in other industries. Today, they are starting to challenge real estate, letting us wonder how close we are to a structural disruption of the whole real estate industry.

² Machine Learning explores the study and construction of algorithms that can learn from and make predictions on data, without being explicitly programmed. Deep Learning is part of a broader family of machine learning methods based on learning representations of data. It attempts to make better representations from large-scale unlabeled data. Neural Network is a statistical technique that is inspired by the way biological nervous systems, such as the brain, process information. It is used mostly for pattern recognition (reading images for instance) or data classification, through a learning process.

URBAN DEVELOPMENT AND THE URBAN REAL ESTATE LANDSCAPE

THE DEVELOPMENT OF AMERICAN CITIES

In the United States, the first cities were concentrated on the east coast, bringing together merchants trading with Europe. Two centuries of industrialization and the development of the service industry enabled cities to emerge and grow throughout the country, expanding the American domestic market. For most of the twentieth century, a clear majority of Americans pursuing the American Dream left crowded urban centers behind to build homes with backyards on the outskirts of cities. The most recent stage of the development of American cities observes a resurgence of urban life in parallel with the rise of the “knowledge economy,” with innovation fueling an optimistic perspective on the prosperous future of cities³.

Economics drives urban change. Access to public service goods (water, gas, electricity) and reduction of transaction costs (transport and communication) have influenced the location and pace of urban development. As a result, the role of government in city making has shifted over time from being passive and reactive to proactive, even pre-emptive.

Demographic changes helped. As shown in Figure 1, the number of US cities with more than 50,000 people has soared since the 1940s. Only the rise of social and environmental problems (traffic, pollution, and crime, amongst others) has forced the government to reconsider its role—it had to develop not only planning but deal-making capabilities whenever necessary. Its progressive involvement in urban development helped to reshape its goal of making cities attractive and healthy places for people to live.

THE REAL ESTATE INVESTMENT LANDSCAPE

For developers starting a project, the greatest challenge is often securing initial equity partners. Since those equity investors typically become the property owners, they bear much of the project risk. The market is densely concentrated. Today, as shown

Number of cities with more than 50,000 people in the US



Source: KIM, 1999

Figure 1

in Figure 2, equity funds constitute the main group of investors at around 40% of all properties. Blackstone leads this group with about 20% of the market. They are four times the size of Lone Star, the second largest player.

Although investors commit huge sums, their decision processes remain based on limited financial considerations. Our interviews with Real Estate Fund Managers validated this bias. The golden rule to date is to combine (1) the cheap purchase of land, (2) immediate signing of leasing contracts, and (3) optimal capital structure for the deal. The approach is then to mitigate development risk through basic asset portfolio diversification.

The former CEO of a leading traditional real estate development firm admitted that even when dealing with large, high-risk investment programs they would, at best, commission specialized market research to validate their assumed demand. Nevertheless, no structured forecast was conducted and urban analytics were ignored beyond elementary demographic data.

In fact, three main players—the Tenant, the Lender, and the Developer (as shown in Figure 3)—control the investment process. In a typical real estate investment, these three players manage the different time horizons of the transaction. As their financial logic prevails, they prevent the system from fully exploiting market research data and improving the overall economic performance of development projects.

³ Sukkoo Kim, “Urban Development in the United States, 1690–1990,” Southern Economic Journal, Vol. 66, No.4.

While it is demonstrated that the access to public service goods and the reduction in transaction costs facilitate urban development, the quality and accessibility of these resources have influenced the decisions of individuals and companies to locate in certain places. Apartments located close to subway stations, for example, command a higher price than those further away. If such an observation is not new to the real estate industry, urban analytics are offering the opportunity to quantify and weigh the impact of proximity on the end price of any given property.

Outstanding questions remain: How can real estate data be used to improve the investment decision process and optimize returns? Can descriptive and predictive analytics result in greater efficiency and less uncertainty for the community at large?

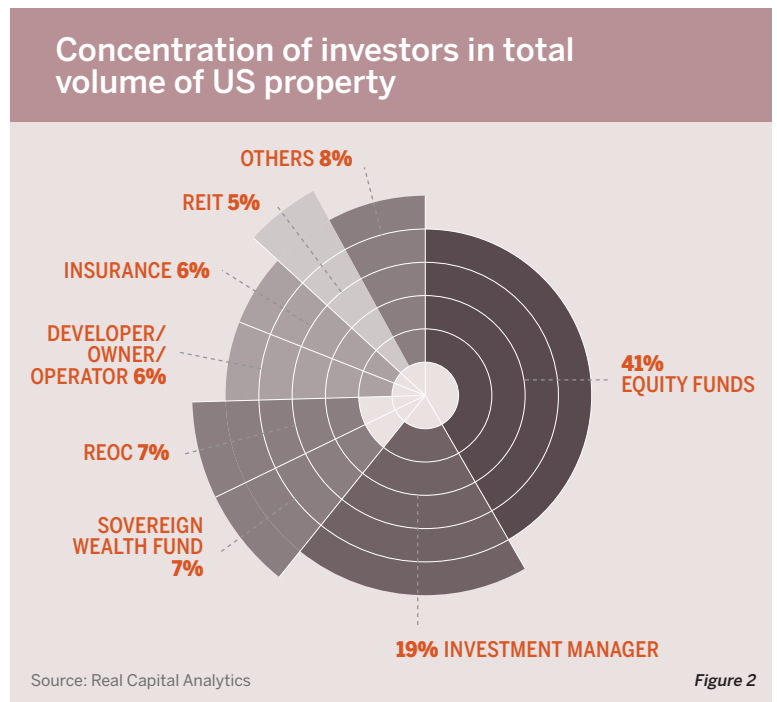


Figure 2

The Real Estate Investment Landscape

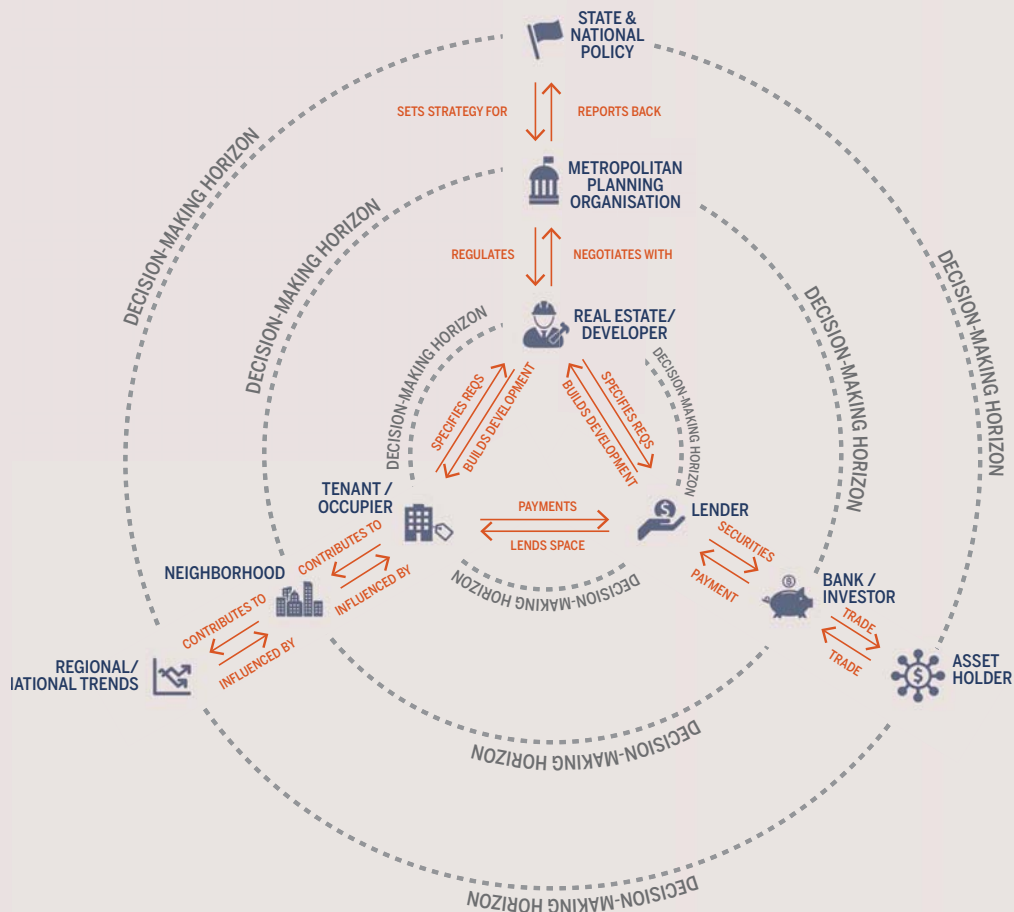


Diagram by authors

Figure 3

THE GROWTH OF DATA PLATFORMS

TOWARD ANALYTICS FOR REAL ESTATE AND URBAN DEVELOPMENT

New aspirations toward urban lifestyle and data science are driving the current boom in urban-centric technology companies. On one side, the renaissance of 'aspirational urban life' was initially described by journalist Alan Ehrenhalt in *The Great Inversion and the Future of the American City*. The concept was then supported by the foundational work of contemporary urban economists Ed Glaeser and Paul Krugman⁴ who identified agglomeration effects in urban areas. On the other side, the current boom in data and analytics is both produced by, and descriptive of, cities. Various referred to as the *Smart City movement*, *Urban Big Data*, or the *New Science of Cities*, this paradigm draws its pedigree from technocratic interest in computer-assisted social analysis promoted in the 1970s by the cybernetics and control systems movements. Current computational resources have enabled massive amounts of data on the urban realm to be recorded and analyzed.

These two phenomena — people moving to the city, and our newfound fidelity in recording and analyzing the city — coupled with a massive increase in the investment of global capital, have contributed to significant investment momentum in real estate tech firms: 2015 marked a record \$1.5 Billion in venture capital to real estate startups, according to VC analyst firm CB Insights⁵.

As Figure 4 suggests, the momentum in real estate tech has been growing, sometimes blurring the lines between its different facets.

⁴ Edward L. Glaeser, "Is There a New Urbanism? The Growth of U.S. Cities in the 1990s" http://scholar.harvard.edu/files/glaeser/files/is_there_a_new_urbanism_the_growth_of_u.s._cities_in_the_1990s.pdf
Paul Krugman, Masahisa Fujita, Anthony J. Venables, "The Spatial Economy, Cities, Regions, and International Trade"

⁵ CB-Insights, "Where Are the Top Smart Money VCs Investing in Real Estate Tech?", <https://www.cbinsights.com/blog/smart-money-vcs-real-estate-tech/>

Most new urban tech platforms have succeeded by either offering data at much finer granularity (*Compstack*, which provides details of specific real estate transactions to brokers) or aggregating large and diverse datasets in the same platform (*CoStar*, which aggregates large data sets at the zip code level for the entire country). Some firms leverage few datasets but aim for maximum accuracy (*Compstack*), while others touch many different sources of data to depict a holistic view of the market (*CoStar*, *Reonomy*, *NCREIF*, etc).

We can categorize the data being used into four main clusters: People, Place, Infrastructure, and Wealth.

THE THREE DISRUPTION WAVES

AGGREGATION, ANALYTICS, AND PREDICTION

When data is collected, organizing and analyzing it is a critical step in unlocking its knowledge. Software and cloud-based platforms are now implemented for this purpose. By visualizing, filtering, analyzing, or even simulating future scenarios, the industry can assess market trends, financial assets, and design decisions. It can even predict potential future outcomes.

There is fierce competition among technologies to address the market: traditional software is challenged by new cloud-based platforms which enable automated data aggregation into large valuable databases. These platforms — even cheaper to maintain — are disrupting common

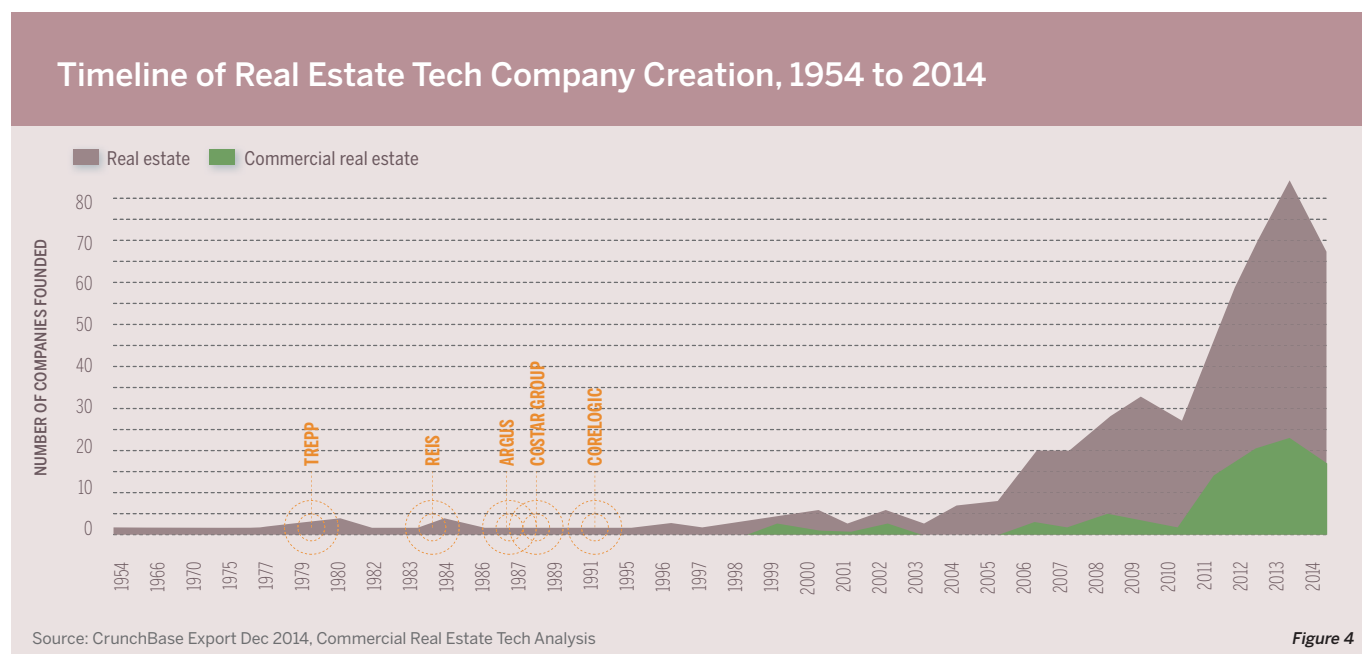


Figure 4

practices. They are expected to disintermediate certain stakeholders, to enable crowdsourcing of a new kind of data, and to eventually offer brand new insights to the industry.

TRENDS AND FUTURE PERSPECTIVES

Mapping the inception of specific platforms and software is quite telling: it reveals the shift from 1990s-era software to the cloud-based platform wave of the early 2000s. In addition, the untapped market of predictive analytics is soaring. Real estate tech companies such as *SpaceQuant*, *Mashvisor*, *SMartZip*, *Enodo Score*, and *Zillow* all rely on large datasets for simulating future investment outcomes or assessing potential market trends.

However, as databases grow exponentially, existing software and platforms must adapt to handle such massive amounts of information. On the demand side, according to market research conducted by *Syntheticity*, companies are not prepared to pay for and adopt those new tools. The “spreadsheet mentality” seems to be the default practice.

Regardless, the trend toward urban analytics applications seems irresistible. If data filtering, standardization, and privacy might hinder the growth of such platforms, the needs of larger entities such as cities or governments, which are already pressuring platform and software providers for their services, will counterbalance current practices in the industry.

PREDICTIVE ANALYTICS

THE EARLY STAGE OF DISRUPTION

The third and last “wave” of the real estate industry disruption began five years ago with the advent of machine learning. “Predictive analytics” is the name of this last phase. Companies such as *SpaceQuant*, *Enodo Score*, and *Zillow* have tried to tap into the potential of advanced statistical technics. New algorithmic logic enables predictions based on datasets aggregated over the past 20 years by government websites or large real estate data platforms such as *CoStar*. It is less a revolution of scale (volume of data, computational power, etc.) than a disruption of intelligence. The once-shortsighted real estate market can now use forecasting for a wide range of topics: from rent price forecasting (*Enodo Score*), to tenant turnover in commercial real estate (*SpaceQuant*), and mortgage default rate forecasting. The disruptive potential of predictive analytics relies on the growing time span of the predictions and

their increasing granularity. As Marc Rutzen⁶, CEO of *Enodo Score*, explains, the endgame is better accuracy in less time in real estate deal-making. In other words, a feasibility study that used to take, on average, 4 hours and 15 minutes for a cost of \$10,000, is now automated, taking 5 minutes and functioning with greater accuracy.

THE UNDERLYING TECHNOLOGY

To understand the emerging real estate predictive analytics market, it is important to get a sense, albeit superficial, of the kind of technology at stake. Rutzen explains that the tools used are primarily based on statistical methods applied in a more sophisticated way. By using **machine learning**, **deep learning**, or **neural networks**, startups like *Enodo Score* take traditional statistical tools to a new level.

All of these advanced statistical methods follow the same process. First, a phase of “training” in which the machine is fed a dataset and “learns.” In other words, it assimilates the dataset’s complexity and tries to weigh the impact of each factor (house characteristics for instance) on the output value (house price). Following that is a “testing” phase in which the machine previously trained is tested against a set of data where the output is known. We observe here how accurately the algorithm predicts the output value. Once the machine is calibrated (after a certain number of iterations), it is ready to predict. This is the prediction phase, in which we use the algorithm to guess the output value of a dataset with unknown output value.

Rutzen adds:

“Once the machine is trained, and estimates of property values are generated, the UI [User Interface] allows for a feedback from our clients. Our model might not be able to take into account certain types of very granular information, such as wood-floor finish, but the feedback of our clients will ultimately be used to further train the machine, and adjust the model. The goal here is to increase the accuracy of the tool by giving it some flexibility. Actually, it is the true value of our platform!”

There is a great deal of effort put into the fine-tuning of the tool, in which user feedback, geolocation data, and other types of information are incorporated to increase the accuracy of the model. Overall, the statistical techniques described above are at the core of the current real estate tech momentum. They are methods developed over the past 20 years in data science that have proven efficient and reliable and are at the doorstep of the real estate industry today.

TIMESPAN AND GRANULARITY

The real estate industry has invested a great deal of effort into using simple regression models to complete short term analysis. The predictive precision of machine learning is pushing the boundaries of forecasting. From a few months to a five-year span, the ability to predict the future of deals is radically changing the perspectives of investors. Rutzen explains, *“Prediction is taking decision making to a whole new stage. By looking at the statistical meaning of the data,*

⁶ Marc Rutzen is the Co-Founder and Chief Technology Officer of *Enodo Score*, a predictive analytics platform for the commercial real estate industry that measures the institutional investment grade of multifamily properties. He directs the development and implementation of the platform, including front-end design and development, development of data sharing partnerships, beta testing, customer feedback and business development. Marc is a Licensed Managing Broker in the State of Illinois, and earned his Master of Science in Real Estate Development from Columbia University.

the real estate industry is now invited to look at the probability of deals success.”

If predictive capabilities open new horizons to urban development and real estate, data accuracy is the heart of the matter. By constantly training algorithms, platforms such as *Enodo Score* will, over time, be able to predict prices with greater accuracy and detail. A good example is the gap between the local price average provided by *Zillow* and the average price per square foot for a given building forecasted over the next 5 years generated by *Enodo Score*. It predicts more precisely, and further into the future, than older platforms.

However, there is a trade-off between the time span and the range of predictions: *Enodo Score* is able to calculate a 5-year forecast, as they only focus on multifamily investment properties. If *Zillow* predicts only for the year to come, the range of property types they offer is much wider than their competitors. After interviewing Jasjeet Thind⁷, VP of Data Science at *Zillow*, the explanation is clear: the number of datasets is such that processing them and adapting the infrastructure to their ever-growing size makes it hard to achieve long time span prediction, accuracy, and a large set of property types. One has to focus on a property type to forecast far into the future, or to limit forecasting ability to embrace a wider range of property types. Either way, the accuracy of the model is paramount.

LIMITS

If forecasting should soon be game changing for the real estate industry, every company in this field is still struggling with the issue of reputation. The traditional approach of risk analysis—spending almost unlimited resources on feasibility studies—is still recognized as standard. Most investors and real estate players trust the old system and react with skepticism to *tech-evangelism*. Rutzen explains how hard it is to get the client to trust his platform's ratings. *Zillow*, which benefits from its primacy in this industry and around 70% of the market share, has been able to position its metrics as industry standards. The *Zestimate* Index, a metric that reflects the valuation of any given property in the country, is widely accepted among real estate professionals.

Another issue for predictive analytics is the integrity of the datasets. If large databases have been aggregated over time, it is often hard to judge the quality of the data. Also, users are sometimes invited by certain platforms, such as *Zillow*, to claim their own property and enter data themselves—data that is then factored into their predictions. The filtering of such data is a crucial and sometimes problematic step. In the age of data science, filtering methods are part of the art, and standard machine learning procedures allow firms like *Zillow* to “clean” the data before running their analyses. Thind affirms that after removing outliers and filtering the data, the quality of the prediction is reliable enough. As proof, *Zillow* regularly publishes their “scoring”—the accuracy of their predictions—on their website.

⁷ Jasjeet Thind is the senior director of data science and engineering at *Zillow*. His group focuses on machine-learned prediction models and big data systems that power use cases such as *Zestimates*, personalization, housing indices, search, content recommendations, and user segmentation. Prior to *Zillow*, Jasjeet served as director of engineering at Yahoo, where he architected a machine-learned real-time big data platform leveraging social signals for user interest signals and content prediction. The system powers personalized content on Yahoo, Yahoo Sports, and Yahoo News. Jasjeet holds a BS and master's degree in computer science from Cornell University. <https://www.crunchbase.com/person/jasjeet-thind>

Lastly, the precision of the prediction depends to a large extent on the amount of data processed. Developing the infrastructure, or “pipeline,” to catch up with the ever-growing size of data sets is more than challenging, as Thind explains. The open source tools integrate with difficulty; ensuring the robustness and scalability of the prediction tools is quite problematic.

CONCLUSION

The boom in urban data collection and its use in real estate technology shows that the traditional silo-ing of industry knowledge is fracturing. Key data links between data types and data providers means that more comprehensive analytics over the breadth of the urban development and real estate industry is possible. Furthermore, statistical predictive analysis and simulation, based both on trends in data platforms and the complexity of the phenomena being modeled, are gaining relevance. Initiatives such as SpaceQuant, Enodo Score, and Zillow are forging new ways of forecasting and prediction at lower costs.

With newly available analytic and simulation technologies combined with the opportunity of integrating datasets, the decision-making horizon for players is suddenly expanding. Conventionally finance- or urban planning-focused players are invited to begin considering wider inputs into their decisions.

As Rutzen asserts:

“I think predictive analytics have still a long way to go in real estate. I see, in fact, more forward-looking predictions. What we have today is—at maximum—predictions one year from now. I think our tool could predict a 5-year time span. ‘What will the multi-family housing market be 5 or 10 years from now? How should I invest today if I want such returns in 5 years?’ are questions that we should try to answer. A longer time span and greater granularity of the predictions—these are the big perspectives.”

As our ability to simulate and predict increases and the cost of these efforts decreases, forecasting the future of urban development may be common practice in the coming decades.