

CARNEGIE MELLON UNIVERSITY

School of Architecture

College of Fine Arts

Thesis

Submitted in partial fulfillment of the requirements for the degree of

Master of Science in Computational Design

Rethinking PropTech:

Drawing insights about the real estate technology
industry through technical experimentation

Ian Friedman

Accepted by advisory committee:


Eddy Man Kim *Principal Advisor*

5/11/2020

May 11, 2020


Valentina Vavasis *Advisor*

5/11/2020

May 11, 2020


Daniel Cardoso Llach *Advisor*

5/11/2020

May 11, 2020


Josh Panknin *Advisor*

5/11/2020

May 11, 2020

Rethinking PropTech:

Drawing insights about the real estate technology industry
through technical experimentation

Ian Friedman

Master of Science in Computational Design

Carnegie Mellon University

2020

Eddy Man Kim

Carnegie Mellon University

Primary Advisor

Valentina Vavasis

Carnegie Mellon University

Daniel Cardoso Llach

Carnegie Mellon University

Josh Panknin

Columbia University

Abstract

Over the past ten years, the real estate technology industry (also known as PropTech) has undergone a massive expansion. This rapid expansion, fueled by a >140x growth in venture capital funding over the past decade, has propelled PropTech startups to actively shape the ways in which we engage with the built environment; physically, digitally, financially, and otherwise.

This rapid change calls for rigorous inquiries into the nature and mechanics of PropTech. Prior work that attempts to understand PropTech has focused on establishing broad definitions of the industry, occasionally scoping down to highlight a few issues at play at a technical level. While these approaches have been incredibly useful in defining an otherwise nebulous industry, they do not examine how the technical issues of PropTech relate to and inform the structures at play within industry, an important perspective to take on an industry that is fundamentally driven by technology.

This thesis draws insights about PropTech by taking a decidedly different approach; one which uses technical experimentation, data analysis, and industry immersion as a basis of examining the broader industry from the inside. Specifically, this thesis discusses the connection between business models and data distribution algorithms, highlights the necessity to account for macroeconomics in data and algorithms, presents an approach to addressing data scarcity in real estate by parsing unstructured data, investigates Zillow's automated valuation model, analyzes the geospatial specificity of its predictive accuracy, discusses the methods of communicating this accuracy, probes the issue of geographic specificity of value, and reflects upon hands-on experience within industry to illustrate the vastly different considerations at play within different PropTech companies.

In taking this sociotechnical approach, new insights emerge and a different mode of thought arises - one which is hierarchical, contextualizing technical issues within areas of industry and highlighting their relationship to the much broader phenomena of macroeconomics, regulation and politics, culture and society, and the physical environment.

Acknowledgements

To Eddy Man Kim, for working so closely with me over the years. Your guidance has been invaluable for my personal and academic development and words cannot fully express my gratitude.

To Valentina Vavasis, for holding me accountable, endless proofreading, keeping my head in the right place, and for never failing to make me laugh.

To Daniel Cardoso Llach, for always pushing me in the right direction.

To Josh Panknin, for the valuable feedback and PropTech perspective.

To my cohort, for our shared experience throughout this program and especially through the strange times that we currently find ourselves in. Thank you Yufei Cheng, Zhihao Fang, Yixiao Fu, Siyu Guo, Yun Hao, Yaxin Hu, Yichin Lee, Hongtao Ma, Vincent Mai, Weixin Qiu, and Erik Ulberg.

To the broader CMU computational design community that I have had the pleasure of interacting with over the past few years. Javier Argota Sánchez-Vaquerizo, Ardavan Bidgoli, Özgüç Bertuğ Çapunaman, Wei Wei Chi, Scott Donaldson, Emek Erdolu, Cecilia Ferrando, Runchang Kang, Zheng Luo, Atefeh Mahdavi, Jinmo Rhee, Pedro Veloso, Humphrey Yang, and many others.

To my coworkers at Cadre for giving me an experience that proved to be invaluable in producing this thesis. I specifically want to thank my team: Samantha Buttegieg, Luke Grymek, AJ Hallac, Ross Mechanic, Clement Miao, Sam Nguyen, and Ashwin Raghu. And a big thanks to everyone else at Cadre who made my time there delightful.

To Harsh Kedia, for being an excellent friend and for endlessly stimulating conversation.

And finally, to my family, because I would not be here without them.

Table of Contents

<i>1 - Introduction</i>	6
<i>Personal statement</i>	6
<i>Defining PropTech</i>	7
<i>Why PropTech</i>	8
<i>Background</i>	12
<i>Methods</i>	15
<i>2 - Framework</i>	17
<i>Externalities</i>	18
<i>Industry</i>	19
<i>Technical matters</i>	20
<i>Discussion</i>	21
<i>3 - Methods</i>	22
<i>3.1 - Facade analysis with Google Street View</i>	22
<i>Data collection</i>	22
<i>Analysis Pipeline</i>	25
<i>Results</i>	27
<i>Insights</i>	28
<i>3.2 - Zillow error rate analysis</i>	30
<i>Insights</i>	35
<i>3.3 - Context-specific notions of value</i>	37
<i>Data and setup</i>	37
<i>Results</i>	41
<i>Insights</i>	42
<i>3.4 - Cadre</i>	44
<i>Cadre as tech-enabled operator</i>	45
<i>Cadre as transaction innovation</i>	46
<i>4 - Conclusion</i>	49
<i>Contributions and future work</i>	49
<i>Concluding thoughts</i>	50
<i>Bibliography</i>	51

1 - Introduction

Personal statement

This thesis details the work that I have produced in relation to PropTech (defined later in this introduction) over the past two and a half years. In this time, my work has shown a clear progression of ideas and research ranging over a closely related set of driving questions. Though a thesis question is posed and an emphasis placed upon conceptual frameworks in PropTech, this represents the culmination of a large body of work that extends even before the scope of what is covered in this thesis.

My path started in the technical questions of automated valuation and ultimately progressed, out of necessity, into issues of understanding PropTech as an industry. At the lowest level of abstraction I found a necessity to understand much bigger and broader questions about industry and global context. Without the proper understanding of these issues, decisions can be made at a technical level which would compromise the validity of the technical solution. Unfortunately, there is very little academic research which probes the issues of PropTech as an industry. With this realization, my efforts shifted to try to understand and break down the complexities of PropTech.

This thesis compiles and documents the work that I have produced, provides some context for my thoughts, and perhaps starts a much needed conversation about how we research PropTech. However, I hope that those who are not interested in PropTech might also find some value in this work, the connections between the social and technical realms, the approach taken toward uncovering new questions, and the technical methods employed here.

For those in PropTech, I hope that this thesis provides a foundation for future discussion and research about the industry.

Defining PropTech

PropTech, an abbreviation of ‘Property Technology’, refers to the software, hardware, and digital products which relate to or augment the ways in which we use, understand, and transact upon the built environment. Some examples of PropTech that have reached a broad public awareness are cashierless stores (e.g. Amazon Go), online real estate databases (e.g. Zillow), and tech-enabled landlords / development (e.g. WeWork). However, many key technologies in PropTech have not quite made it to the same level of public perception. These will be discussed in more depth later in this chapter.

PropTech also refers more broadly to the collective of companies which compose an industry within which these technology solutions are produced and utilized. It is important to note that companies within PropTech do not need to exclusively operate within the realm of PropTech. For instance, Alphabet, Amazon, and Apple are all active within PropTech, but their PropTech products do not produce their primary sources of revenue. However, these players are making a significant impact in this space and excluding them from discussion would be inaccurate.

The definitions of PropTech above are my own, but the discourse around PropTech is still incredibly young and definitions are still somewhat in flux. While the definitions of PropTech listed above are mostly in line with other definitions, there is a lack of consensus about the bounds of PropTech. Notably absent in the definition presented here is a notion of how the built environment is created. Construction technology (ConTech), design technology, and engineering technology are considered part of PropTech in a few key works^{1,2,3}, while a few others discuss construction and design technology as an important and adjacent but ultimately separate industry^{4,5}. Coming from a background in architecture technology, I see a distinct difference between the types of issues at play within PropTech and ConTech, design technology, and engineering technology. As an example, robotics, generative machine learning practices, and 3d modeling are much more relevant within ConTech, design technology, and engineering technology than they are within PropTech. That is not to say that there is no overlap between these spaces (e.g. BIM can be a helpful representation of data for real estate), but unconditional

1 - Block, A., & Aarons, Z. (2019). PropTech 101: Turning Chaos Into Cash Through Real Estate Innovation (1st ed.). New York: Advantage Media Group.

2 - Rivaton, R., & Pavanello, V. (2018). Make Real Estate Great Again: Proptech, Real Estech, Construction Tech. RETech Publishing.

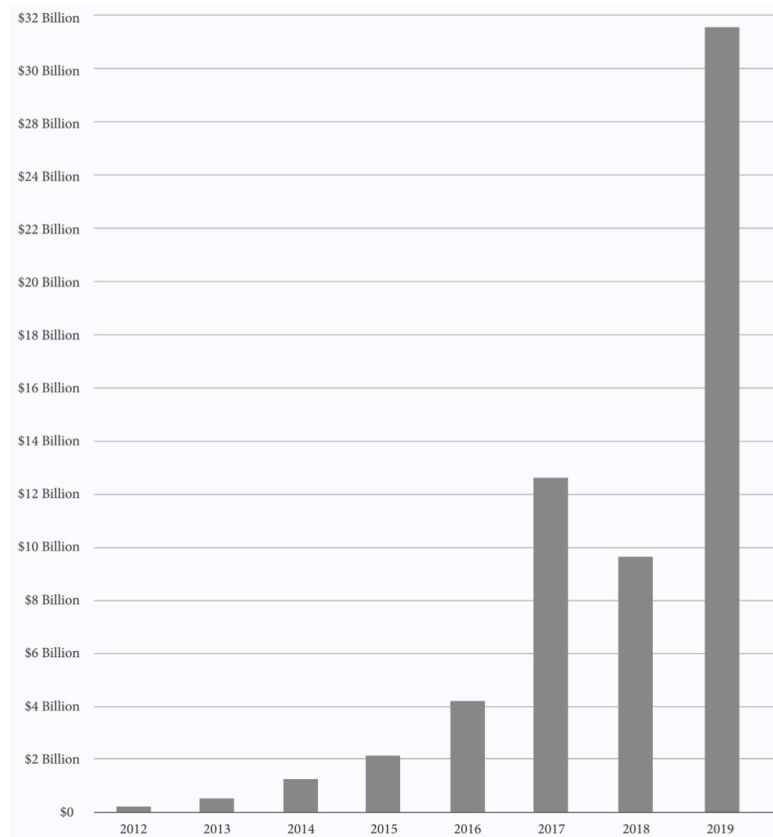
3 - Brown, R. W. J. (2018). #PropTech: A guide to how property technology is changing how we live, work and invest. Independently published.

4 - Baum, A. (2017). PropTech 3.0: The Future of Real Estate. Oxford, UK.

5 - Baum, A. (2020). PropTech 2020 : The Future of Real Estate. Oxford, UK.

inclusion of these other industries would make discussion here a bit unwieldy, as they produce entirely different classes of products that cater to a different demographic and utilize technologies which are otherwise uncommon within PropTech.

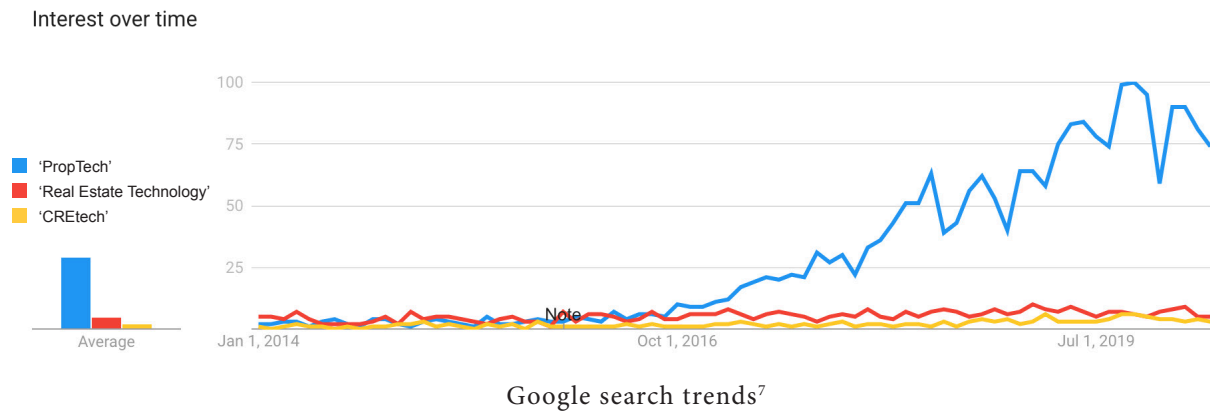
Why PropTech



PropTech Venture Capital Funding from 2012 to 2019⁶

Over the past several years, PropTech has expanded at an incredible rate. Venture capital funding in PropTech startups grew from \$218 million in 2012 to \$31.6 billion in 2019, representing a compound annual growth rate of over 103%. This massive cash injection has fueled a large number of innovative startups that have already changed the way we operate as a society. Companies like AirBnB, WeWork, Nest, and Zillow have already become household names, but many more less commonly known PropTech companies are quietly changing the world around us.

⁶ - Data from:
Wong, M. (2018). The rise of real estate tech. Retrieved April 24, 2020, from https://www.cbinsights.com/reports/CB-Insights_Rise-Real-Estate-Tech-Briefing.pdf
CRE Tech. (2020). 2019 End-Year Report. Retrieved April 24, 2020, from <https://www.cretech.com/product/cretech-2019-end-year-report/>
CBInsights. (2017). Real Estate Tech Funding Reaches New Highs In 2016. Retrieved April 24, 2020, from <https://www.cbinsights.com/research/real-estate-tech-startup-funding/>



Separately, but similarly, PropTech has been gaining traction in the public eye. The graph above shows the number of queries made to Google⁷ using the case-insensitive phrases 'PropTech' (in blue), 'real estate technology' (in red), and 'CREtech' (in yellow), a term which predates PropTech. The y-axis represents the percentage of queries in a given month in comparison to the peak number of queries. Using this data, we can calculate that the number of monthly queries for 'PropTech' made to Google has a compound annual growth rate of over 92% between 2016 and 2019. As startups in the space become more mature and widely recognized, this growth trend will likely become even more pronounced.

However, startups are not the only ones getting into this space. They are joined by large tech companies like Amazon, Alphabet (parent company of Google), Facebook, and Apple, which have entered into the space with technology like cashierless stores, online marketplaces, built environment mapping / modeling, and smart home / smart city technology. Real estate incumbents like JLL, CBRE, and Blackstone are launching internal technology initiatives and are also directly investing in more nimble startups, forming strategic relationships to avoid being left behind in the wake of technological transformation. Meanwhile, more established players in PropTech, like Yardi and CoStar, have expanded their product and customer bases through acquisitions.

7 - Google. (2020). Google Trends. Retrieved April 24, 2020, from [https://trends.google.com/trends/explore?date=2014-01-01 2020-04-24&q=proptech,%22real estate technology%22,CREtech](https://trends.google.com/trends/explore?date=2014-01-01%2020-04-24&q=proptech,%22real estate technology%22,CREtech)



Cashierless Amazon Go store⁸

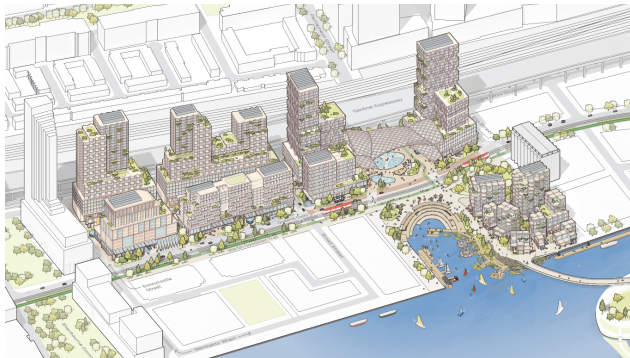


Link NYC Kiosk⁹



Latch Keyless Entry¹⁰

Products which blend physical and digital worlds, such as wifi-producing digital billboards, cashierless stores, and smartphone-controlled door locks, have already made their way into the public eye, but much is yet to come. As machine learning / AI gets more sophisticated and the internet of things (IOT) becomes more pervasive, digital technologies will continue to augment the ways in which we have traditionally interfaced with our physical environments.



Sidewalk Labs Proposed Quayside Development¹¹



WeWork Location in Nashville¹²

Beyond ubiquitous computing, PropTech companies are also changing our skylines. Sidewalk Labs, a subsidiary of Alphabet, has been partially approved to construct a 2.65m square foot development in Toronto, CA, to the tune of \$1.3 billion¹³. In 2018, WeWork became the largest office tenant in NYC, after just eight years since its inception¹⁴.

8 - Levy, N. (2018, December 12). New compact Amazon Go store opens the door for locations in office lobbies, hospitals and more. Retrieved April 24, 2020, from GeekWire website: <https://www.geekwire.com/2018/new-compact-amazon-go-store-opens-door-locations-office-lobbies-hospitals/>

9 - Barone, V. (2018, April 4). MTA bus times will display on LinkNYC kiosks. Retrieved April 24, 2020, from Newsday website: <https://www.newsday.com/mta-bus-times-linknyc-1.17863056>

10 - Higgins, M. (2016, January 1). New York Discovers Keyless Entry Systems. Retrieved April 24, 2020, from New York Times website: <https://www.nytimes.com/2016/01/03/realestate/new-york-discovers-keyless-entry-systems.html>

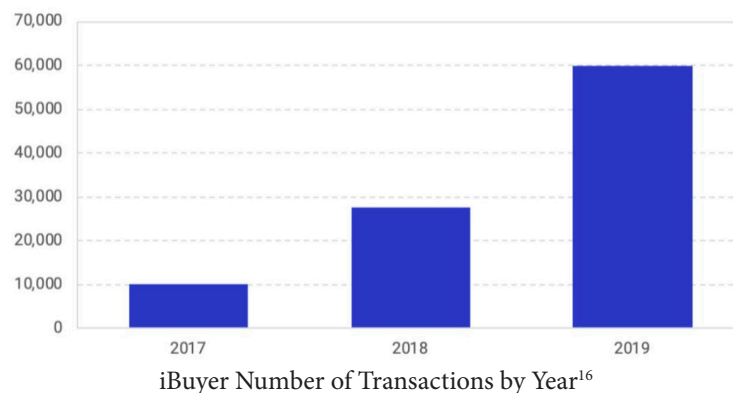
11 - Sidewalk Labs. (2020). Quayside: A Complete Community and a Proving Ground for Innovation. Retrieved April 24, 2020, from <https://www.sidewalktoronto.ca/plans/quayside/>

12 - Garner, M. (2018, August). Battle over downtown signs heats back up. Retrieved April 24, 2020, from Nashville Business Journal website: <https://www.bizjournals.com/nashville/news/2018/08/31/battle-over-downtown-signs-heats-back-up.html>

13 - Sidewalk Labs. (2020). Quayside: A Complete Community and a Proving Ground for Innovation. Retrieved April 24, 2020, from <https://www.sidewalktoronto.ca/plans/quayside/>

14 - Lash, H. (2018, September 18). WeWork tops JPMorgan as No. 1 New York tenant as coworking booms. Retrieved April 24, 2020, from Reuters website: <https://www.reuters.com/article/us-usa-property-wework-milestone/wework-tops-jpmorgan-as-no-1-new-york-tenant-as-coworking-booms-idUSKCN1LY34R>

PropTech has also produced significant changes to the ways in which we engage the built environment digitally. From 3d scans and online virtual tours produced by companies like Matterport¹⁵, to smart home technology which connects our HVAC, electrical, and plumbing systems to our phones, our built world is becoming even more connected to our digital world. Even our activity in the built environment is creating digital footprints, as companies move to track the way consumers move, shop, and respond to stimulus throughout our cities. Each of these products and businesses generates more data that can be used to drive insights for real estate. It is no secret that data has become one of the most important commodities in our society, and this trend is likely to accelerate in the future.



From a financial perspective, PropTech is also changing the ways in which we transact upon real estate. New business models have emerged, such as iBuyers, a business model in which homeowners can get an instant offer to buy their house online. iBuyers accounted for 60,000 home transactions, and \$8.7 billion of homes purchased in 2019 alone, representing a more than 2x growth from the year before¹⁶. Online real estate investment platforms have allowed individuals to trade relatively small stakes in \$100+ million real estate deals, making real estate investment more similar to trading stocks. Compass, a tech-enabled brokerage founded in 2012, now boasts that it is the largest independent real estate brokerage in the United States¹⁷. Similar disruption is being made to the way in which individuals can secure real estate insurance and loans.

15 - Matterport. (2020). Retrieved April 24, 2020, from <https://matterport.com/>

16 - DelPrete, Mike. "The 2019 iBuyer Scorecard." Received by Ian Friedman, The 2019 iBuyer Scorecard, 4 Feb. 2020.

17 - Compass. (2020). About Us - Compass Real Estate. Retrieved April 24, 2020, from <https://www.compass.com/about/>

Despite the massive disruptive potential of this industry, it remains largely unexamined by academia. A better understanding of this growing phenomenon will allow for greater insight into the opportunities and risks within the industry, as well as the ways in which PropTech will affect the adjacent industries (AEC, finance, government, etc.) and perhaps even how it will affect society as a whole.

Background



Real estate itself is relatively new as an academic area of study. The first real-estate-specific graduate program was founded in 1967 at NYU, with other academic institutions slow to follow. Despite a growing academic presence, critical perspectives on technology did not play a large role in shaping the field. Given this, and the meteoric rise of PropTech, it is not surprising that academic discourse in this area is just emerging.



A few PropTech-focused academic groups have begun to emerge, such as MIT's Real Estate Innovation Lab (REIL), Columbia's Real Estate Technology Initiative, Oxford's Future of Real Estate initiative, University of Colorado Boulder's Real Estate Technology program, and Stanford's Disruptive Technology and Digital Cities program. A few institutions have also

started hosting classes on PropTech but don't have formalized programs or initiatives yet, such as University of Pennsylvania, NYU, and Cornell. The oldest of these groups was founded in 2015, and very little research in the space existed before then, so literature on PropTech is incredibly limited.



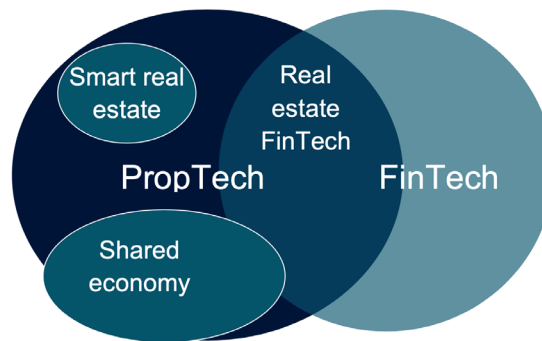
PropTech Market Map (Oct. 2015) ¹⁸

In some ways, the PropTech industry was quicker to self-reflect than academia was to scrutinize it. Real estate technology-specific venture capital firms date back to as early as 2009¹⁹, and real estate technology-specific media companies, which aggregate and publish news specific to PropTech, date back to 2013. As funding ramped up and more startups entered the ecosystem, other industries took notice and latched on. Market intelligence companies, those who collect and sell information about industries, started to create 'market maps' for real estate technology. These maps are intended to show all the companies within a space, and they served as the first representation of PropTech as an industry.

18 - Wong, M. (2015). Disrupting Real Estate: 58 Startups Reshaping How Property Is Bought, Sold, And Managed. Retrieved from <https://www.cbinsights.com/research/real-estate-tech-startups-market-map/>

19 - Crunchbase. (n.d.). Camber Creek. Retrieved April 24, 2020, from <https://www.crunchbase.com/organization/camber-creek>

20 - Crunchbase. (n.d.). Navitas Capital. Retrieved April 24, 2020, from <https://www.crunchbase.com/organization/navitas>



Taxonomy proposed in PropTech 3.0 ²¹

In 2017, Andrew Baum, then a visiting professor at Oxford, released *PropTech 3.0: The future of real estate* and announced the Oxford Future of Real Estate Initiative. One of the primary goals of this 95 page report was to define PropTech as its own independent industry, as opposed to a subset of FinTech (financial technology). By analyzing market maps of the industry and pointing to key trends, he established a taxonomy of PropTech as defined by three subsectors; ‘Real Estate FinTech’, ‘Shared Economy’, and ‘Smart Real Estate’ ²². This document and taxonomy have become PropTech canon and are widely referenced (often without attribution) in writing and discussions throughout the industry.

Analysis & Financing	Space Identification + Listing	Site Selection & Negotiation	Diligence	Development + Construction	Process Automation	Space Usage & Management	Payments + Services
• Zoning • Appraisal • Budgeting • Modeling	• Marketplace • O2O • Marketing • Review	• Brokerage • Tour • Comps	• Prospecting • Title • Econometrics • Risk pricing	• Project oversight • Architecture • Design • Engineering • Compliance • Construction	• Sales • Leasing • Underwriting • Broker tools	• "Basement Tech" • HVAC • Facilities management • Security	• Rental • Utilities • Service providers • Trade • Credit

Taxonomy proposed in *PropTech 101* ²³

In 2019, Aaron Block and Zach Aarons published *PropTech 101*, in which they examined PropTech through the lens of venture capital. In this book, Block and Aarons discuss Baum’s taxonomy, saying that it is a ‘good organizational starting point’, but ultimately define and present their own taxonomy, shown above. The taxonomy they present as the ‘most important, useful, and unique’ is a taxonomy by ‘real estate value chain’, which breaks down

21 - Baum, A. (2017). *PropTech 3.0: The Future of Real Estate*. Oxford, UK.

22 - In 2020, Baum updated this taxonomy to include a fourth subsector; ‘Data Digitalization / Analytics’

23 - Block, A., & Aarons, Z. (2019). *PropTech 101: Turning Chaos Into Cash Through Real Estate Innovation* (1st ed.). New York: Advantage Media Group.

the space by the lifecycle of a building, ‘from dirt to disposition’. This taxonomy provides a much higher-fidelity view into PropTech than Baum’s taxonomy but functions very differently. Block and Aaron say that their taxonomy ‘emphasizes both exactly where a startup is adding value... and where it’s earning revenue’ by organizing according to ‘the function of a start-up or product, not what’s under the hood’. Like PropTech 3.0, PropTech 101 has been widely accepted in the PropTech community and has even been used as a course textbook to teach PropTech²⁴.

Methods

Existing literature on PropTech takes a very top-down approach to discussing this industry. This thesis approaches PropTech from a bottom-up perspective, using automated valuation and industry immersion to drive insights.

The methods employed in this thesis trace a progression of ideas about automated valuation of real estate, the application of this technology within PropTech, and fundamental limitations and constraints of this technology. Each project answers a few questions but ultimately raises many more that motivated the next project. Along the way, these projects generate insights about the broader context of real estate technology, which will be discussed in more depth within the methods section. Here I will provide a brief overview of the work, to illustrate the broader trajectory of the research.

The first project, ‘Facade Analysis with Google Street View Images’, employs machine learning and computer vision to determine whether a building’s facade can be analyzed to predict the value of that building. The intent of this project was to illuminate the value of design and utilize real estate value as an objective metric with which to validate design analysis. Among other issues, this project raised questions about data ownership and analysis, as well as the applications of automated valuation in industry.

24 - Duranton, G., & Mack, R. (2019). Real Estate Disruptions Syllabus. Retrieved April 24, 2020, from Duranton, G., & Mack, R. (2019). Real Estate Disruptions Syllabus (p. 11). p. 11.

This led to the next project, which analyzed Zillow's application of automated valuation through data visualization, analysis of their algorithm's accuracy, and an investigation into the motivations behind their application of this technology. Ultimately, this study showed that their algorithm performs very well in some parts of the country and very poorly in others. It also showed that they utilize this algorithm very differently based on the business context within which they are operating. Among other things, this analysis raised questions about why performance would vary so drastically across geographic regions.

Trying to address this question of geographic variance of accuracy, I probed issues of data availability and context-specific notions of value. To do this, I analyzed the data available in three different cities, created models to predict real estate value using the shared representation of data, and compared the relative importance of building features in predicting value in different cities.

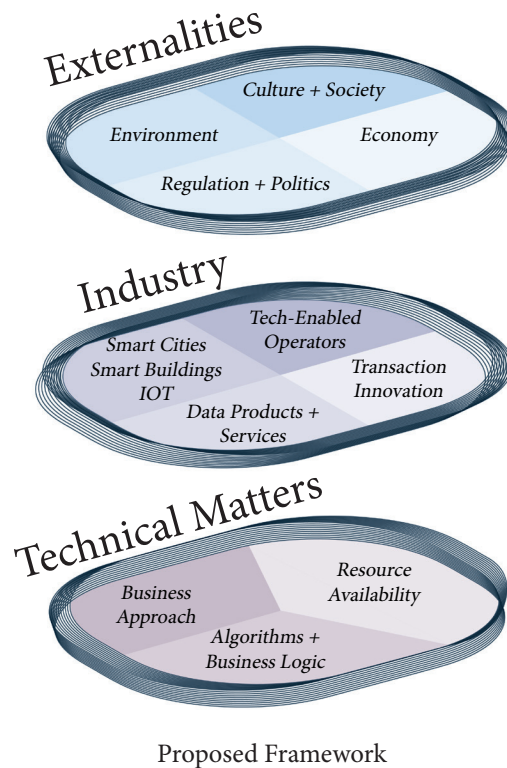
These three academic projects were followed up by a six month immersion within Cadre, a PropTech startup, which provided many additional insights into how PropTech functions. In this thesis, I use my deep knowledge of the business model and issues at play within Cadre to reflect upon its position within industry. In doing so, I also make observations about the broader industry and the factors which shape businesses within it.

These projects typically highlighted insights which related to but did not ultimately fit into existing models of PropTech. Specifically, each experience highlighted connections between very technical issues, such as data availability and algorithms; the industry context within which these technical issues are encountered; and much broader issues, such as macroeconomics, regulation, and cultural issues. Without explicit representation of this hierarchy of issues, existing taxonomies did not prove to be useful representations for discussing the insights gained from these projects. To frame these issues properly, I devised my own framework that gives explicit representation to technical issues, industry, and externalities.

2 - Framework

A conceptual framework is a tool for and an artifact of a researcher's learning.

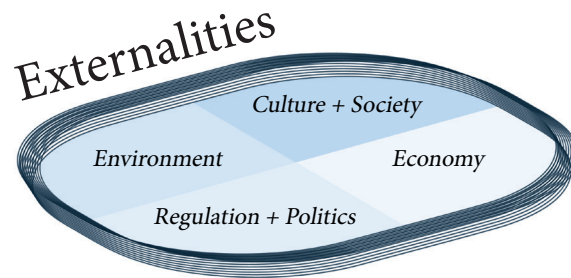
Ravitch and Riggan, *Reason and Rigor*¹



The primary intent of this framework is to provide a representation that can be used to view PropTech through a sociotechnical lens which enables connections to be drawn between technical issues, their context within industry, and much broader issues such as macroeconomics, regulation, and cultural issues. This framework is both graphic and conceptual, meaning that it is both a graphic container for ideas and also a way of structuring thoughts about PropTech.

The framework contains three key elements: conceptual groups, taxonomies, and relationships. The conceptual groups, 'technical matters', 'industry', and 'externalities', represent the most broad distinction within this framework and provide clear separation between the issues uncovered in my methods. Taxonomies provide a higher fidelity representation of each conceptual group, allowing for more specificity in discussion of each. Relationships are what allow you to create connections across the framework.

Externalities



Externalities conceptual group and taxonomy

The ‘externalities’ conceptual group relates to things which exist outside the scope of PropTech. Some examples include federal interest rates, the civil rights movement, demographic changes like aging population, regulatory change, climate change, and pandemics. Each of these can have profound effects upon PropTech, and PropTech may also affect these externalities as well. However, these externalities are characterized by scopes which are much more global in nature.

The taxonomy for the ‘externalities’ conceptual group is divided according to the ways in which external phenomena affect PropTech. This taxonomy consists of four segments; ‘culture and society’, ‘economy’, ‘regulation and politics’, and ‘environment’. These should be somewhat self-explanatory, but further definitions are provided below.

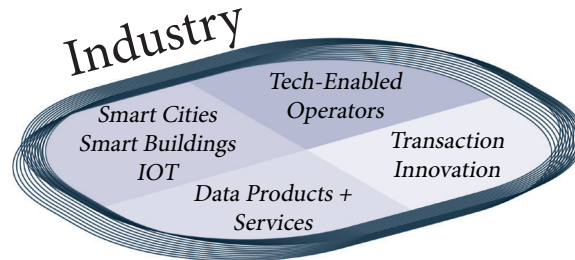
Culture and society relates to humans, their perception and reasoning, and their broader social movements. This segment might contain things such as a cultural aversion to mass surveillance, the civil rights movement, or demographic changes such as an aging population.

Economy relates to issues of capital, market forces, and the exchange of goods. Inflation rates, supply and demand, interest rates, and capital market performance are some examples of economic factors.

Regulation and politics refers to any legal or bureaucratic processes. Zoning and planning, reporting requirements, and financial restrictions are a few examples.

Environment relates to physical context (local or global), climate, and ecology. Climate change, pandemics, biodiversity, and urban fabric are examples of environmental factors.

Industry



Industry conceptual group and taxonomy

The ‘industry’ conceptual group relates to the PropTech industry itself. The four subsectors within the taxonomy for industry defined here are ‘smart cities, smart buildings, and IOT’, ‘tech-enabled operators’, ‘transaction innovation’, and ‘data products and services’. This is similar to Baum’s most recent taxonomy, except for the inclusion of ‘tech-enabled operators’ and exclusion of ‘shared economy’¹. As this taxonomy of PropTech serves to discuss the connections between industry, externalities, and technical matters, the subsectors should partition the industry in such a way that separates companies with very different technical considerations and external factors at play.

Smart cities, smart buildings, and IOT relates to businesses that enrich the built environment with sensors or actuators in order to collect data, drive insights, and/or augment the functionality of physical space with digital technology. Cashierless retail, digital billboards, digital HVAC management, smartphone-enabled locks, and smart appliances are just a few products that companies within this space are producing.

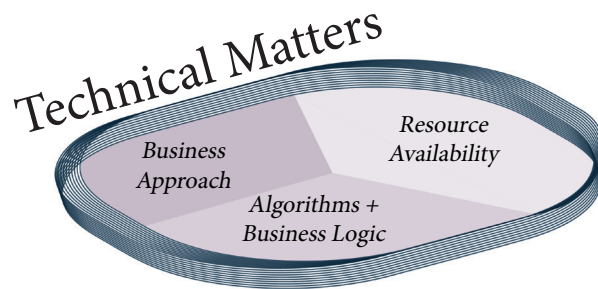
Tech-enabled operators refers to companies that function in traditional real estate roles (e.g. broker, assessor, investor, property manager, etc.) but use technology to make their operations more efficient and effective.

1 - Baum, A. (2020). PropTech 2020 : the future of real estate. Retrieved from <https://www.sbs.ox.ac.uk/sites/default/files/2020-02/proptech2020.pdf>

Companies within transaction innovation are primarily concerned with disrupting the ways in which we exchange money in real estate. Companies within this subsector may focus on things like mortgages, investment in properties, facilitating digital real estate transactions with blockchain, and crowdfunding.

Data products and services contains companies that focus on the collection, processing, management, analysis, and/or the distribution of data.

Technical matters



Technical Matters conceptual group and taxonomy

The ‘technical matters’ conceptual group relates to things which can explicitly be controlled by PropTech companies. These factors can be technological, like the choice in automated valuation algorithm, processes for data collection, and monitoring of key performance indicators. They can also be more traditional business considerations, such as talent availability and recruitment practices, marketing strategies, revenue model, and corporate culture.

The taxonomy for technical matters is divided into ‘algorithms and business logic’, ‘resource availability’, and ‘business approach’.

Algorithms and business logic refers to processes or technologies which drive key business activities. Artificial intelligence, machine learning, search algorithms, database structures, and operating logic are some examples of these. These can be controlled by PropTech companies in that they have the ability to make decisions to utilize or develop these

technologies, but that is not to say that these technologies are exclusively under the purview of PropTech.

Resource availability is self-explanatory, but refers to the ability for a company to access the resources that it needs to operate. Resources can be human capital, financial capital, data, raw materials, computational power, low network latency, or any other thing which helps a company run.

Business approach refers to the decisions, structures, and business processes put into place by a company. Corporate hierarchy, corporate culture, marketing strategy, and target revenue streams are some examples of issues that fall under business approach.

Discussion

A conceptual framework should represent the key factor, variable, construct, and the relationships between them⁴. If a conceptual framework is an argument, as Sharon Ravitch and Matthew Riggan suggest in their book about conceptual frameworks³, then the argument of this framework is that PropTech is fundamentally connected to issues outside of PropTech. Although this framework is novel within the context of PropTech, similar ways of framing industry have been accepted elsewhere. For instance, the Technology-Organization-Environment (TOE) framework models technological innovation and decision making as a process of an organization, technology, and the external environment⁴. This has been widely used and adopted as a way to model and understand technological adoption and diffusion within various industries. While the TOE framework focuses on decision making processes, as opposed to modeling an industry itself, it reflects a very similar thought process about the impact of externalities and technology upon an industry.

In the methods section to follow, this framework will be used to indicate the arguments and insights being discussed within each project. This should illustrate how the framework can be used as a tool to guide thoughts in PropTech.

2 - Miles, M., & Huberman, M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.). SAGE Publications, Inc.

3 - Ravitch, S., & Riggan, M. (2012). *Reason and Rigor: How Conceptual Frameworks Guide Research* (1st ed.). SAGE Publications, Inc.

4 - DePietro, R., Wiarda, E., & Fleischer, M. (1990). The Context for Change: Organization, Technology, and Environment. In L. Tornatzky & M. Fleischer (Eds.), *The Processes of Technological Innovation* (pp. 151–175). Lexington Books.

3 - Methods

This chapter presents the research, experiments, and findings that I used to investigate PropTech. The conceptual framework presented in the prior chapter provides a lens through which this work is analyzed and will be used to illustrate the insights generated from each project.

3.1 - Facade analysis with Google Street View

In an attempt to quantify the monetary value that architecture creates, Maria Rodriguez (a CMU undergrad in statistics and data science) and I designed a project around the computational analysis of building facades. The experiment was structured such that Google street view images of buildings would be fed into a machine learning + computer vision pipeline and used (along with other data) to predict real estate value.

Data collection

The first step in doing this was to collect images of facades, data about property values, data that could be used to predict value. It became immediately clear that this task was not just about data - much broader questions of value, macroeconomics, and real estate came into play. These questions were often addressed by constraining the problem via the data collected.

Data on the features of properties was collected from Boston property assessment data². This dataset contains information about properties that is used by the city to assess property value, which is used to determine municipal property taxes. Features of the dataset include number of floors, square footage, number of rooms, and Boston's valuation of the property, among other things.

1 - Ravitch, S., & Riggan, M. (2012). Reason and Rigor: How Conceptual Frameworks Guide Research (1st ed.). SAGE Publications, Inc.

2 - Boston Assessing Department. (2020). Property Assessment - Datasets - Analyze Boston. Retrieved April 24, 2020, from Analyze Boston website: <https://data.boston.gov/dataset/property-assessment>

Notably, the Boston property assessment dataset contains data for properties across all property types (commercial, residential, agricultural, industrial, etc.). Each property type is valued differently and some property types have much more complexity factoring into the value of a building. For example, valuing undeveloped agricultural land would likely utilize the comparable sales approach, which would entail examining what other agricultural land has sold for recently and using that to derive the value. Valuing a commercial property, which typically uses an income-based approach, is much more complex. Typically, this would require creating a discounted cash flow analysis, projecting the market out several years into the future, accounting for risk of vacancy, operating expenses, costs to improve the property, etc. In order to simplify, Rodriguez and I opted to operate only on residential properties, which has less complexity than other property types and has more reliable sale data.

It should be pointed out that the city's assessed value is for tax purposes and is not indicative of the market value, which is typically much higher. I later revisited this discrepancy between assessed value and found that there was a 26.1% median absolute percent difference and a 52.6% mean absolute percent difference between Boston's assessments of value and the actual market value, as determined from sales. This means that the assessed values for these properties were not reliable as an indicator of value, and other data would need to be collected in order to provide an accurate representation of this vital information. Since the intent of the project was to use real estate value as a proxy to a notion of public perception of architecture, market value (the dollar amount that a building can be sold on the market) would need to be the target variable.

Market value for real estate is difficult to get exactly right, as buildings are incredibly heterogeneous, value is closely aligned with the state of the market and macroeconomic trends, and sales happen very infrequently. Sales within a specific geographic area are subject to the same geographic market and macroeconomic trends, so the Boston dataset provided a helpful natural geographic constraint. However, just as value is geographically specific, it is also temporally specific. Value is affected by ever-changing factors such as inflation, changing supply and demand, federal interest rates, demographic trends in an area, and many more. In order to maintain consistency in valuation, these factors need to be controlled against in some way. For this project, we decided to define market value from the sale prices of homes sold within the past 6 months, under the premise that there would be minimal change in that period of time.

To gather sales data, we scraped the real estate listing website Zillow³. This proved to be no small task, as Zillow works very hard to maintain control over their data. The details of their scraping-defense infrastructure aren't relevant here, only that it was a significant obstacle in the process and that it is indicative of a wider industry trend to capitalize on data.



A typical Google Street View image, viewed through Google's online portal⁴

For images of building facades, we used images from Google Street View. Street View has an application programming interface (API), which allowed us to easily gather images of the properties that we had in our dataset. However, these thousands of images had to be manually checked by a human to ensure that the image was clear. Many building facades were obstructed by trees or cars, while others were partially or completely out of frame. Other buildings simply could not be found on Street View (likely due to an address mismatch).



Three example images from the Architectural Style Dataset⁵

3 - Zillow Inc. (2020). Zillow: Real Estate, Apartments, Mortgages, and Home Values. Retrieved April 24, 2020, from Zillow website: <https://www.zillow.com/>

4 - Google. (2020). Google Maps. Retrieved April 24, 2020, from Google website: <https://www.google.com/maps>

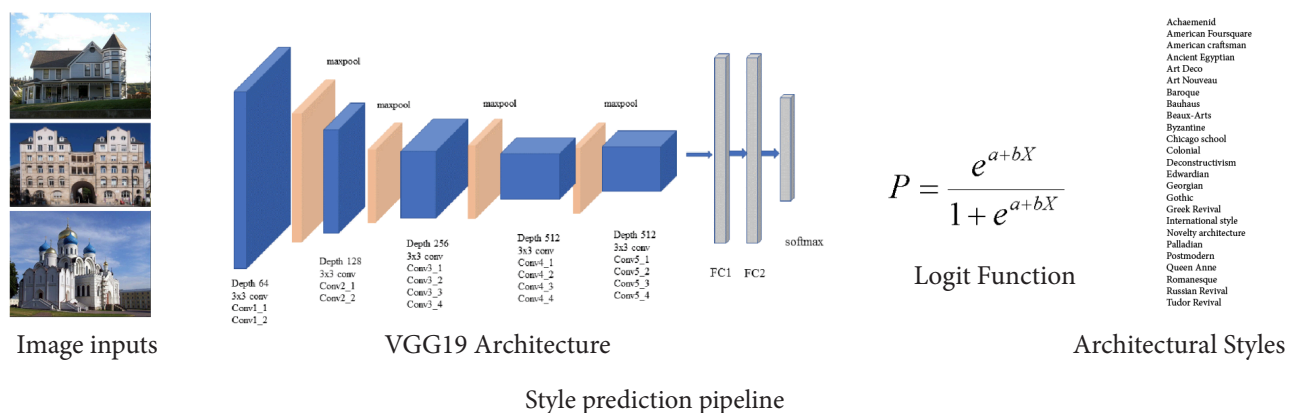
5 - Xu, Z. (2014). Architectural Style Classification using MLLR. Retrieved April 24, 2020, from <https://sites.google.com/site/zhexuutssjtu/projects/arch>

The final piece of the data puzzle was the architectural style dataset⁶. This dataset contains more than 5,000 images of building facades, labeled by their architectural styles. This was by far the easiest data to work with and would provide the backbone for the facade analysis detailed in the next section. However, it was not without its limitations. The dataset contains 25 architectural styles, ranging from neoclassical to deconstructivism. This was not particularly helpful within a context of Boston housing, where style differences are much more nuanced. Some processing would be required to utilize this data.

It is worth mentioning that all of the data discussed above needed some form of preprocessing in order to take advantage of the statistical properties of machine learning models. This is a standard part of any data science experiment and nothing novel was introduced in this project, so I will not go into depth on the methods employed. With this preprocessing complete, the analysis could begin.

Analysis Pipeline

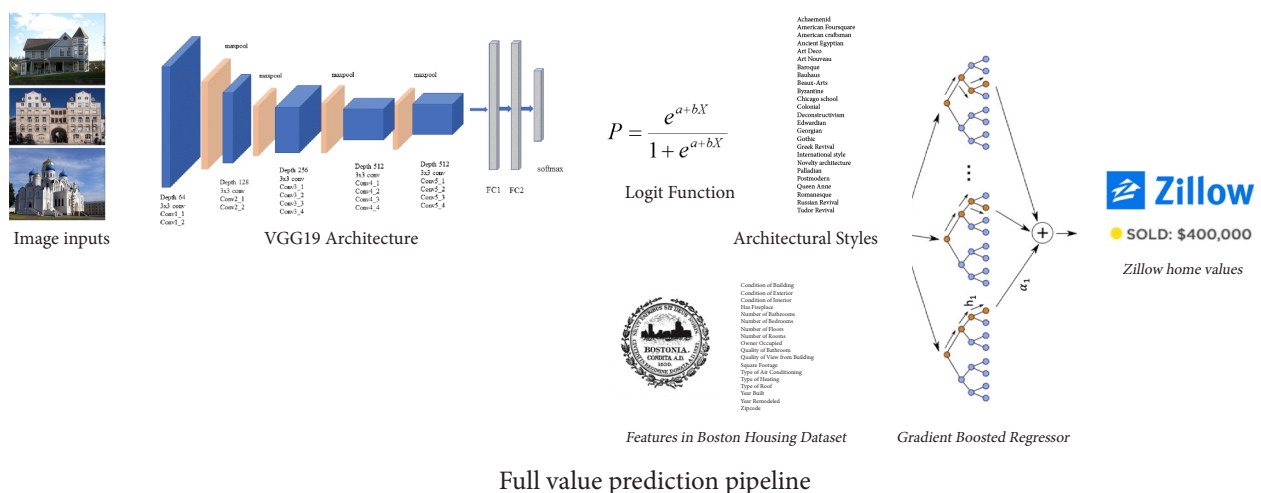
The analysis pipeline consisted of a few major components. The task at hand was essentially to predict market value using just real estate data, predict market value using real estate data and information about building facades, then compare the results between the two.



6 - Xu, Z. (2014). Architectural Style Classification using MLLR. Retrieved April 24, 2020, from <https://sites.google.com/site/zhxuutssjtu/projects/arch>

One big hurdle was the mismatch between the very broad architectural styles in the architectural style dataset and the more nuanced styles of the houses in Boston. While the architectural style dataset was not a good representation of Boston housing, it did provide some means of capturing information about a building facade. In order to get any information out of the Street View images, we would need some way to analyze the images. To do this, we used a VGG19⁷ model pretrained on imagenet, with the softmax layers removed, to break each 800x800x3 image (effectively 1.9m features) into a 230,000 feature vector which captures the low-level features of each image. In non-technical terms, we used a machine learning model to reduce our data into a more information-rich and condensed format which enables other non-image specific models to make predictions on the images.

We then trained a logistic regression model to predict architectural style from the output of the model mentioned above. This model correctly predicted architectural style for around 70% of the images in the architectural style dataset, which was enough for us to be convinced that this pipeline (VGG19 -> logistic regression) was capturing relevant information about the building facades. Since the actual styles themselves were not relevant within the Boston housing context, we decided to take the model's predicted probabilities of each of the 25 styles as a representation for future analysis. The idea was that even if no house was classified as neoclassical, the degree to which the facade resembles a neoclassical building can contain a lot of latent information about the appearance of the facade.



7 - Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR. Retrieved from <https://arxiv.org/pdf/1409.1556.pdf>

To test whether the facade images could be used to predict real estate value, we needed to compare the accuracy of a model using the facade images to the accuracy of a baseline model which predicted value without the images. To create the baseline model, we trained a gradient boosting regressor to predict the sales price using just the Boston assessment data. We also trained a gradient boosting regressor to predict sales price using the Boston assessment data and the raw predicted probabilities of architectural style from the street view images. The results are discussed in the section below.

Results

This baseline model achieved an R^2 score (using Friedman MSE) of approximately 37%. This means that the model explained 37% of the variance in sale price based on the inputs. After training the same model to predict sales price using both the Boston assessment data and the street view images, we were able to attain an R^2 score of approximately 55%. This was a significant improvement upon the baseline model, showing that building facades can in fact be used to predict real estate value.

It is important to note that R^2 is a standard statistic used in machine learning, but is not a typical metric for evaluating AVM accuracy. In reality, neither of these models performed well enough to use commercially. The median absolute percent error (MdApe), the typical KPI for AVM, was approximately 42% for the baseline model and approximately 39% for the model with street view images. This is likely due to a relatively small dataset (just a few thousand properties), not enough information captured, lack of complexity in the model, or too wide of a temporal window for ‘recent’ sales. However, this does not negate the findings of the experiment.

These results reinforced the idea that architectural design is a strong factor in the value of real estate. This work was produced in May 2018 but was never submitted for any publication. To my knowledge, no previous published research had utilized street view images to predict real estate value. Since then, a few papers⁷ have been released which employ similar strategies of Google Street View image analysis for real estate valuation, but none focus on building facades. They have each shown similar findings.

⁷ -Law, S., Paige, B., & Russell, C. (2019). Take a Look Around: Using Street View and Satellite Images to Estimate House Prices. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(5). <https://doi.org/10.1145/3342240>
Hamilton, T.L. (2018). Using Machine Learning and Google Street View to Estimate Visual Amenity Values.

Insights

In showing that images of facades can be used to help predict the value of real estate, this project examines the connection between business models and data distribution algorithms, highlights the necessity to account for macroeconomics in data and algorithms, and presents an approach to addressing data scarcity by parsing unstructured data.

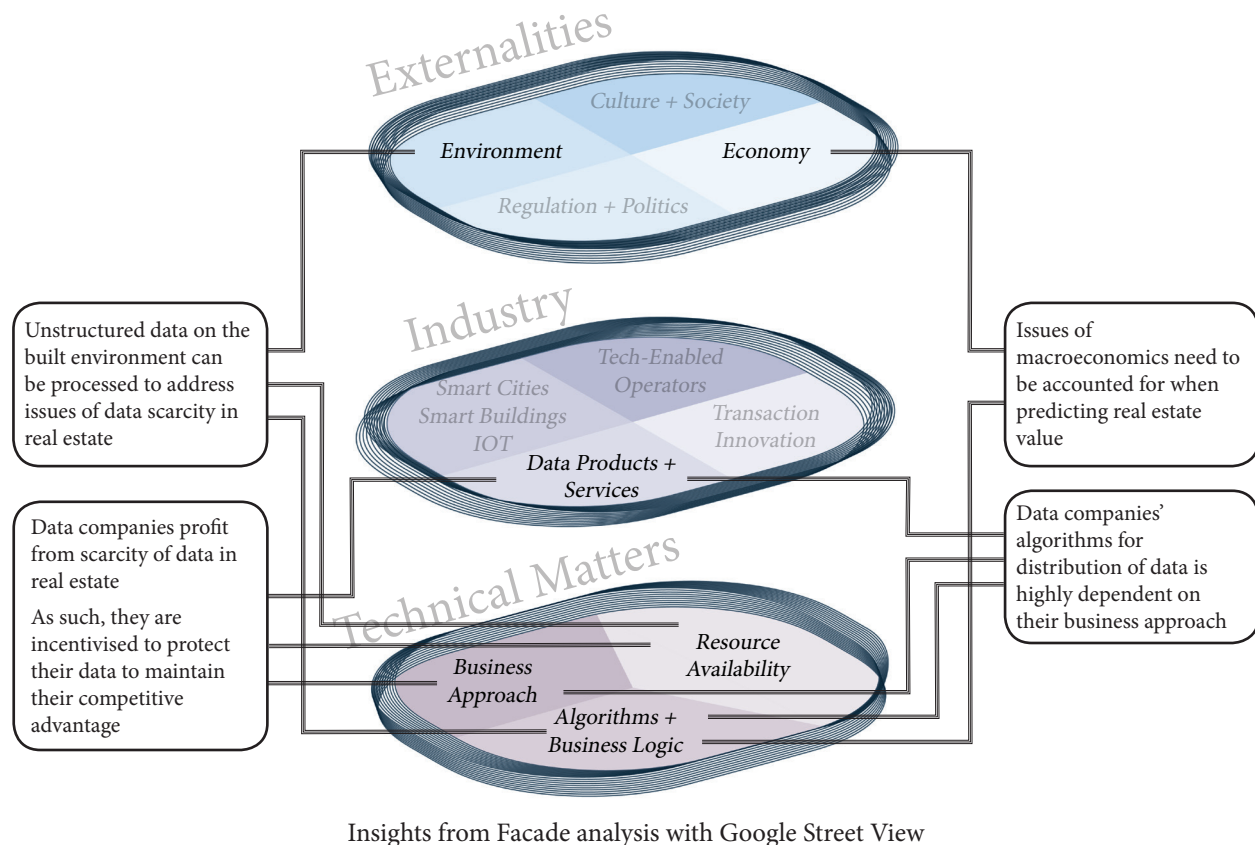
Before any technical experimentation could begin, the issues of procuring data showed a lot about the industry. Bypassing Zillow's data scraping defenses proved to be quite a challenge, but the reasoning behind why they have implemented these defense systems are much more interesting for the purposes of this thesis than the defenses themselves. Much of Zillow's value comes from their proprietary data products. Zillow leverages their massive real estate dataset to attract users that are looking for information on the real estate market, a resource which they would otherwise find difficult to obtain. As such, Zillow is incentivized to protect this resource and prevent others from collecting this data en masse.

Google's approach to monetizing their Street View data is entirely different. Rather than closely guarding their data, Google charges a small fee for each call to their API. Having an open API means that relatively little infrastructure is required for this data to be collected, though there are stipulations about this data can be used and how long it can be stored. The Street View images are being used as real estate data in this context, but the product is typically targeted more toward mobile applications and websites that want to give some context about a location. In other words, the user base is not primarily located within real estate and the business model dictates that data should be distributed as much as possible, as opposed to the restrictive data practices employed by Zillow. This highlights the role of the business model in determining the algorithms which surround the distribution of data, and the various ways in which data products approach the issue of the data availability within real estate.

Aside from the availability of data, other questions came in to play when determining what data should be used within the experiment. Since real estate value is affected by temporally-specific issues of inflation, supply and demand, interest rates, etc., the data had to be constrained to recent sales only to avoid having to adjust for these issues of macroeconomics. Without this constraint upon the data, the predictive accuracy would be reduced to the point of project failure. This is a critical connection between the technical considerations of automated

valuation and issues external to PropTech.

Finally, this project shows one way in which unstructured data can be analyzed and utilized within PropTech. A general rule of thumb is that 80-90% of all data is unstructured, where relevant information cannot be directly queried (e.g. images). In showing how images can be used to predict real estate value, this project highlights the ways in which PropTech can address scarcity of real estate data by digging into alternative data sources. Further, it suggests that aesthetics and design of buildings may have quantifiable value from a real estate perspective. This builds on the conclusions of prior work which shows that real estate value is higher in starchitect-designed buildings⁸ and in new urbanist environments⁹.



The diagram above shows simplified summaries of the insights gained from this project and the areas within PropTech to which they relate, highlighting the relationships between externalities, industry, and technical matters.

8 - Fuerst, F., McAllister, P., & Murray, C. B. (2011). Designer buildings: Estimating the economic value of “signature” architecture. *Environment and Planning A*, 43(1), 166–184. <https://doi.org/10.1068/a43270>

9 - Knaap, G. J., & Song, Y. (2003). New urbanism and housing values: a disaggregate assessment. In *Journal of Urban Economics* (Vol. 54).

3.2 - Zillow error rate analysis

Building off of my work with street view images, I decided to look more into automated valuation models (AVM). I looked at the increasing relevance of AVM within society, which I attributed to increased data accessibility, an increased public confidence in computational technologies, and relaxation of regulation. In doing so, I also started my first investigation of PropTech as an industry, primarily through the lens of automated valuation employed in commercial contexts.

As perhaps the most widely-known example of an AVM, Zillow's 'Zestimate' became an immediate source of interest. Following previous research¹, I was tipped off to some questionable reporting practices of commercial AVM, and I was keen on investigating further.

In the United States, there are no reporting requirements for commercial AVM methodologies or accuracy rates. Despite the lack of reporting standards, companies like Zillow are incentivized to release some information about the performance of their models as a means of establishing credibility. However, this data is often curated and presented in a misleading way.

For instance, Zillow in 2018 boasted a median absolute percent error (MdAPE) of 4.3% between sale prices and Zillow's predictions of values. At first glance, this appears to indicate that the Zestimate provides a reliable valuation, as this metric indicates that more than half of the properties in the United States are within a 4.3% margin of error of the sales price of that property. This would suggest that one could assume a property whose market value is \$100,000, would have a 'Zestimate' valuation between \$95,700 and \$104,300 at least half of the time. However, this assumption is quite dangerous.

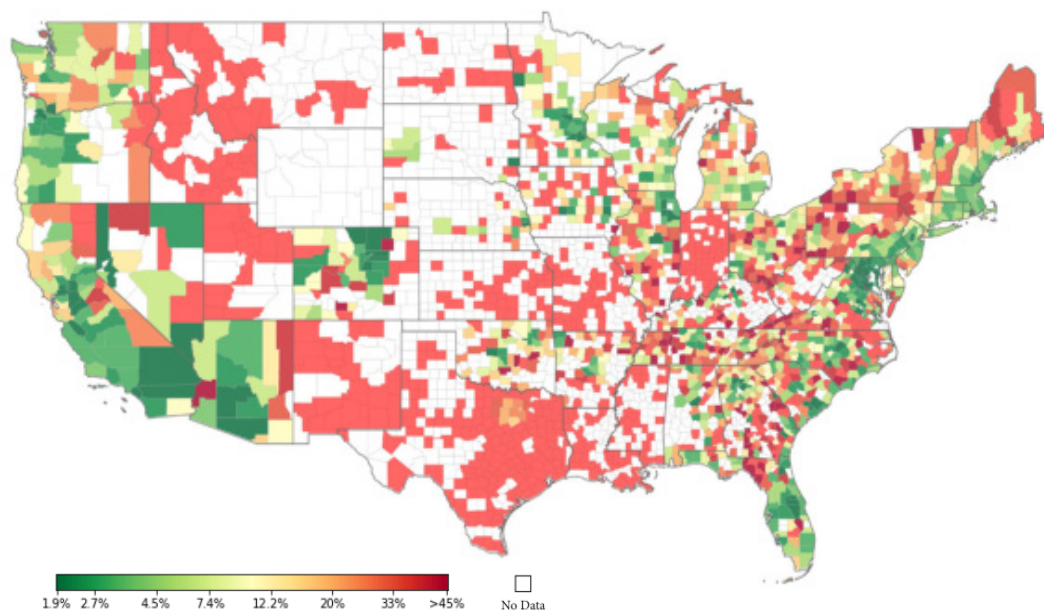
The first issue is that the use of the median performance of the AVM can obscure important statistical properties of the distribution of errors produced by the model. In particular, the median masks a long-tailed distribution (one where much of the distribution falls far away from the median). Since the 'Zestimate' has an MdAPE of 4.3% nationally, we

¹ - Matysiak, G. (2017). Automated Valuation Models (AVMs): a brave new world? Wroclaw Conference in Finance, (August). Retrieved from <https://www.researchgate.net/publication/319355261>

know that half of the distribution of errors fall between 0% and 4.3% (since an absolute value cannot be negative), but the other half of error rates has no upper bound. In other words, an AVM with an MdAPE of 4.3% may produce a wildly inaccurate valuation half of the time.

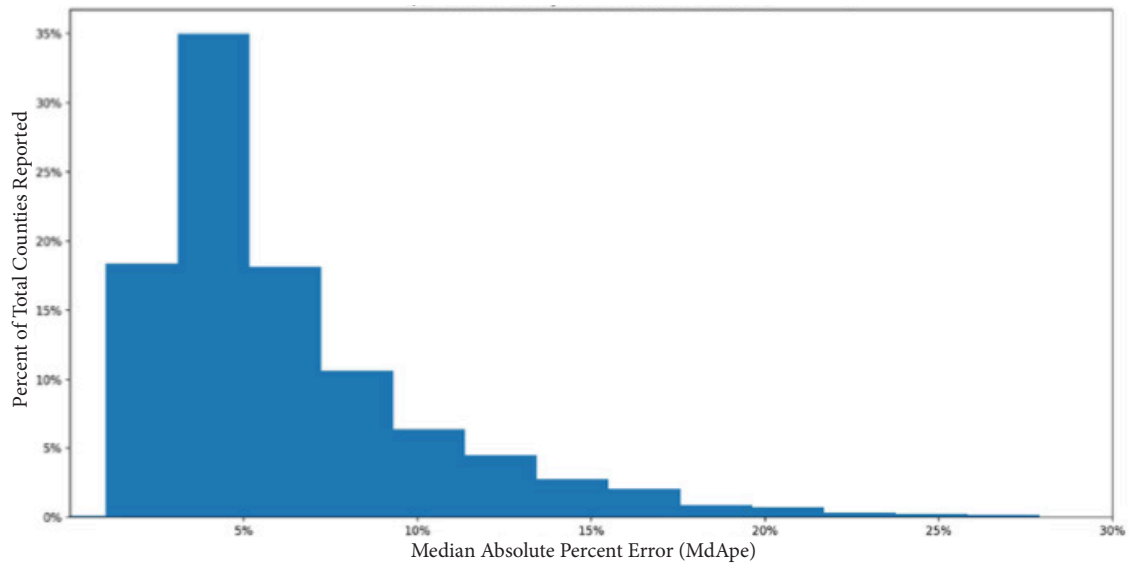
The second issue is that the national error rate of an AVM is not necessarily uniformly distributed across the country. Zillow releases error rates² by county, curated in such a way which gives Zillow the most credibility. This dataset contains data for 2,145 counties out of the 3,007 counties in the United States. Of the 2,145 counties, only 1,280 contain information about the accuracy of value predictions within the county. Zillow provides a four point ranking system for each county, based on the MdAPE of the Zestimate within those counties. Only counties with two points or greater have accuracy data provided, even though counties with a single star represent more than 18% of the properties with Zestimates provided. Since counties with two points have between 5% and 45.9% MdAPE, counties with one point presumably each have over 45.9% MdAPE.

I dug into these two issues with a series of data visualizations and basic statistical analysis.

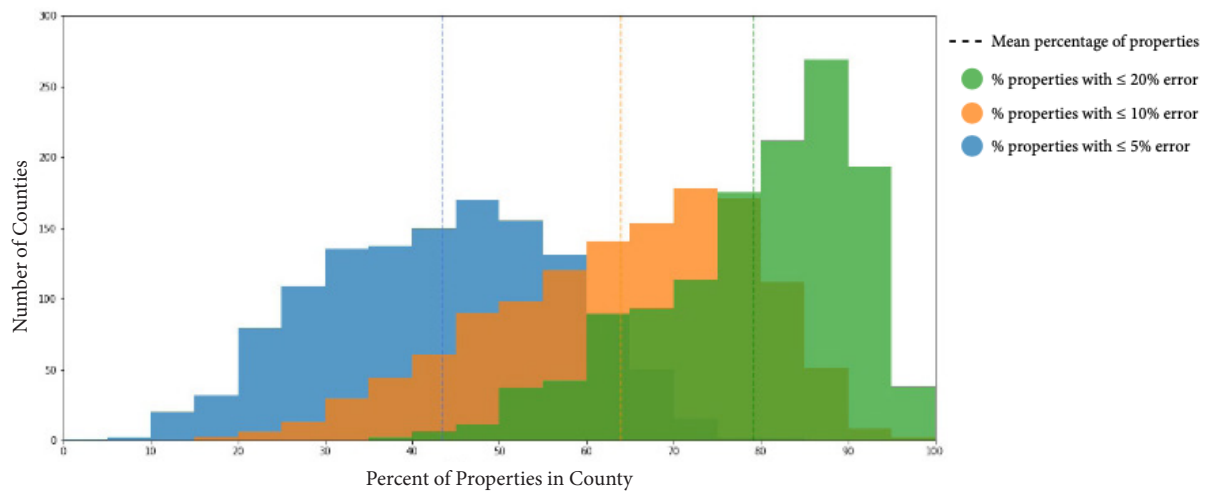


Zillow Median Absolute Percent Error (MdAPE) by County

2 - Zillow Inc. (2018). Zillow Data Coverage and Zestimate Accuracy. Retrieved from https://s.zillowstatic.com/cms/20180724032913427104:DataCoverageAndZestAccuracy_Recent3Months_Ending_2018-06-30.xls



Zillow Median Absolute Percent Error (MdAPE) by County



Number of counties by number of properties in county below x% error

Geographically, ‘Zestimates’ tend to have higher accuracy around the coastlines and much lower accuracy toward the center of the country. This trend does not seem to be caused by population density, as there are many rural areas in the northeast with high accuracy and many Midwestern cities with low accuracy. Other’s work has shown that further disparity occurs across different value ranges of real estate.

The implication is that there are many places in the country where the 4.2% national MdAPE of the Zestimate is far from the localized performance of the Zestimate. Zillow is often one of the first places where consumers look when searching for a home, and the Zestimate for a property anchors perceptions of value. Zillow does not advertise the county-level performance of the Zestimate, but they often boast their national MdAPE. As consumers are unlikely to dig into the dataset as I did, it is likely that many people are relying upon vastly inaccurate information.

This brought a line of questioning about why Zillow would mislead users about the Zestimate’s accuracy. An intuitive response is that Zillow is incentivized to drive traffic to their website through whatever means possible. The Zestimate is a large resource for users as they go about house hunting. Advertising a low error rate makes the Zestimate seem trustworthy, while advertising sub par accuracy rates would make the Zestimate a much less attractive resource to users and would likely reduce the traffic (and thus revenue) to Zillow’s site. However, this seemed to be specific to Zillow’s particular use-case and raised questions about the peculiarities of other use-cases of AVM within industry.

Given another use-case, Zillow takes a very different stance on how its AVM should be employed. Zillow Offers, Zillow’s online home buying program, utilizes Zillow’s AVM to automatically make offers to buy the homes of interested users. In doing so, homeowners receive an expedited selling process and Zillow receives a scalable way to invest in single family homes. As of April 2020, Zillow only offers this service in 24 cities across the United States. When investigating these cities on the map of their error rates, these largely coincide with the areas which have the lowest ‘Zestimate’ error rates. This is understandable, as Zillow does not want to sink money into investments where they don’t have a high confidence in their model. However, this highlights a different set of business considerations which lead to a different

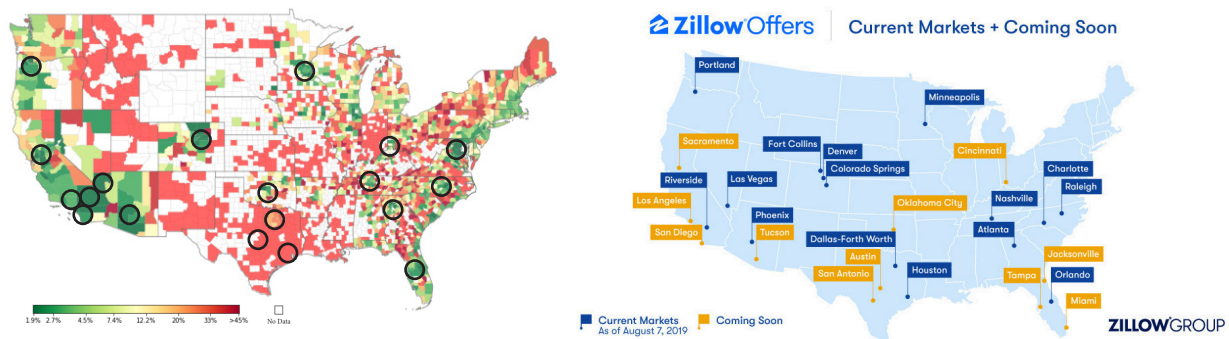
3 - Matysiak, G. (2017). Automated Valuation Models (AVMs): a brave new world? Wroclaw Conference in Finance, (August). Retrieved from <https://www.researchgate.net/publication/319355261>

Skip the hassle. Sell to Zillow.

Get a no-obligation offer directly from Zillow — no staging,
no hidden fees, no surprises.

Get started

Zillow Offers web portal⁴



Zillow Offers Operating Markets⁵

application of the model. It also raises potential ethical questions - if Zillow is not willing to operate on their own AVM, should consumers?

Beyond the application of the AVM and incentives for obscuring high error rates, there remained the question of why AVM performance was so bad in the first place. In the original paper, I made the argument for greater use of alternative data (like street view images and building designs) and deeper analysis of the built environment with computation. However, future work would probe the difficulties of valuation across different geographic contexts.

4 - Zillow Inc. (2020). Sell Your Home with Zillow Offers. Retrieved April 24, 2020, from Zillow website: <https://www.zillow.com/offers/>

5 - Falcon, J. (2019, August 8). Zillow Offers to buy and sell homes in 4 more cities. Retrieved April 24, 2020, from HousingWire website: <https://www.housingwire.com/articles/49807-zillow-offers-to-buy-and-sell-homes-in-4-more-cities/>

Insights

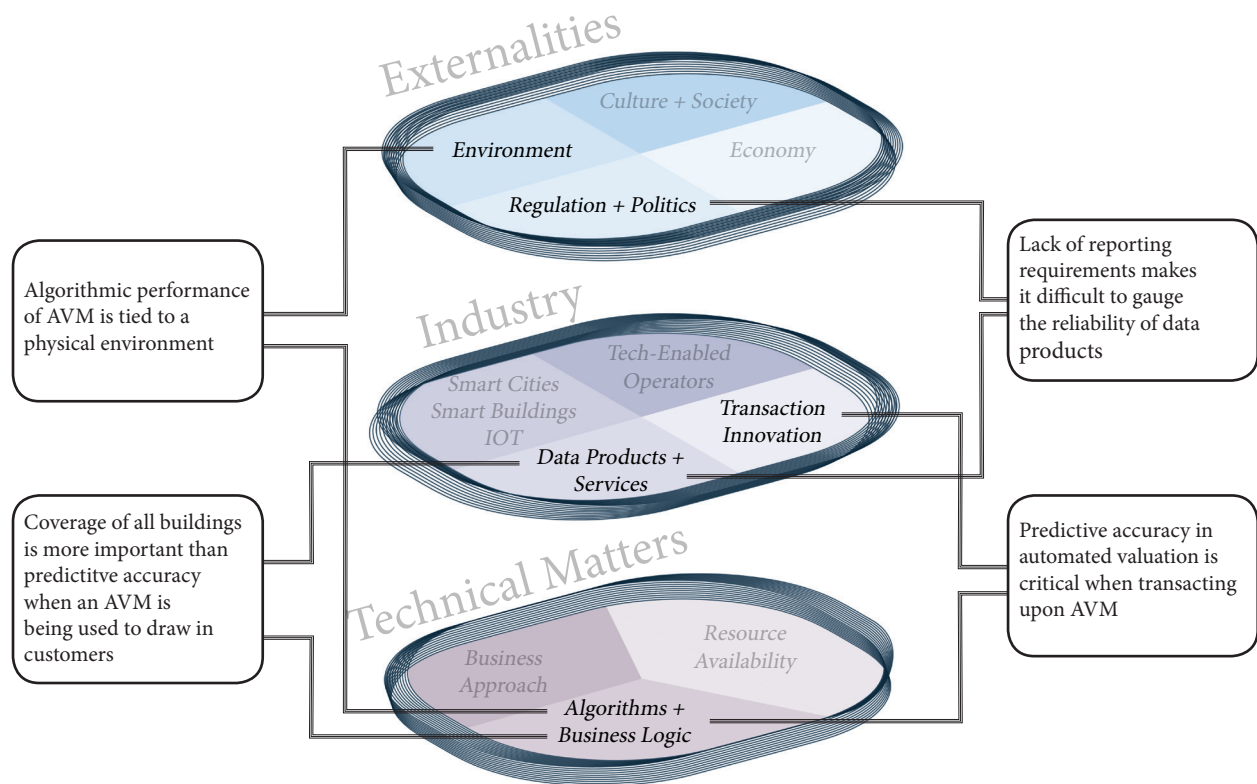
The application of Zillow's AVM shows that different technical considerations arise from the application of distinct use cases. In a data products and services context, where Zillow gives away its AVM data in order to draw users in, Zillow is incentivized to provide valuations for as many homes as possible, even if those valuations are not accurate. However, when they operate as a transaction innovator, providing new ways to buy and sell homes through Zillow Offers, they are incentivized to restrict their operations to only operate upon accurate valuations.

This study also builds on other work⁶ showing that just because an AVM performs well in one context does not mean that it will perform well in all contexts. In this case, the work highlights that algorithmic performance of AVM is tied to a physical environment.

Without regulation on reporting requirements for AVM, it can be difficult to know how accurate and reliable an AVM's output is. In the case of Zillow, their published accuracy metrics are incomplete and the advertised national error rate could mislead users into thinking that the national metric is indicative of localized error rates, which are not broadcasted as widely. When this study was conducted, county-level error rates were being published monthly, but as of April 2020, the county-level error rate data was last updated in June of 2019⁷. The webpage where data is available only shows data for cities where the error rates are very low. The data used in this study is now only available after downloading the raw data. This supports the argument that Zillow is incentivized not to advertise the error rates in areas where the AVM does not perform well.

6 - Matysiak, G. (2017). Automated Valuation Models (AVMs): a brave new world? Wroclaw Conference in Finance, (August). Retrieved from <https://www.researchgate.net/publication/319355261>

7 -Zillow. (2019). What is a Zestimate? Zillow's Zestimate Accuracy. Retrieved April 24, 2020, from Zillow website: <https://www.zillow.com/zestimate/>



Insights from Zillow error rate analysis

3.3 - Context-specific notions of value

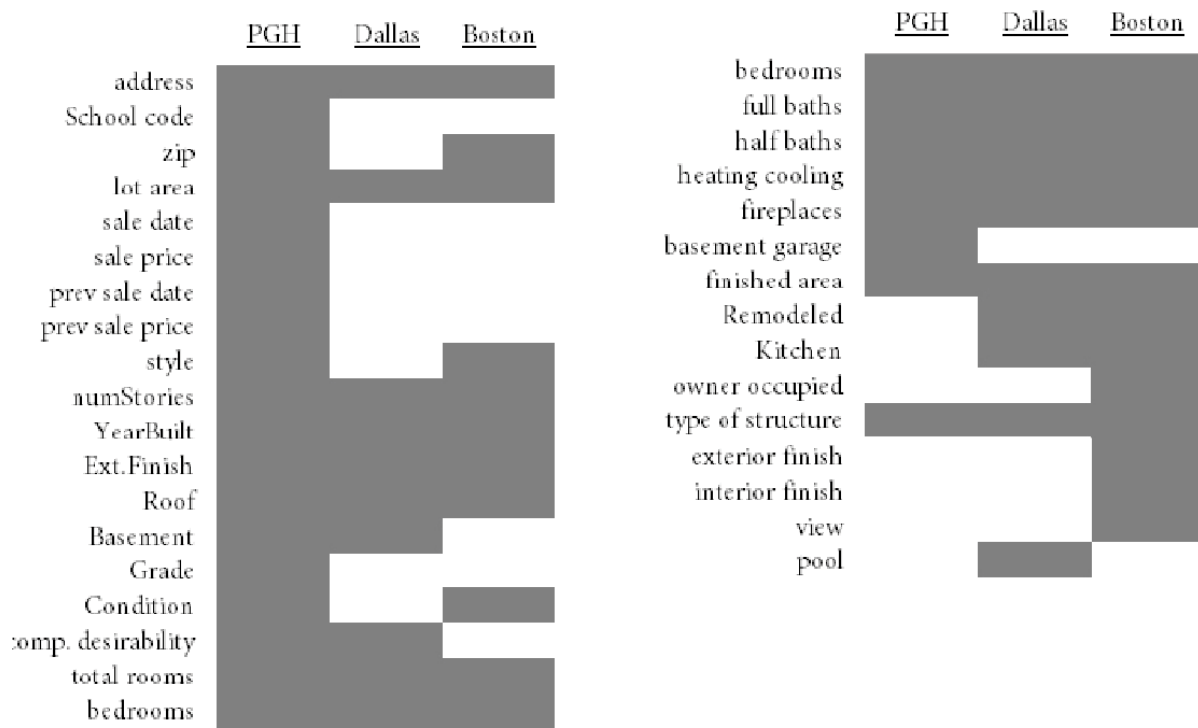
With newfound questions about geospatial variance in value, I began to dig into the context-specificity of value. Influenced by object-oriented ontology and notions of hyperobjects, I started to understand buildings beyond their physical manifestation. From this perspective, a building significantly impacts and is impacted by its economic, cultural, temporal, and social contexts. As such, a building's value cannot be understood without understanding the contexts within which it is situated. With this understanding, the phrase 'location, location, location' took on an entirely new character.

Intuitively, location is the reason why a 2000 ft² home in Tulsa, Oklahoma is a fraction of the price of a physically identical home in Los Angeles, California. The conditions which define a location are complex and operate on many different scales - from the climate of the area, to the surrounding geography, the number of trees on the streets, the urban fabric of the city, the culture of the people within it, and the municipal, regional, and national economic trends. Places have a plethora of defining characteristics, but the value of each of these characteristics is neither constant nor geographically uniform.

With these ideas in mind, I devised an experiment to use automated valuation to illustrate the context-specificity of value. By training separate machine learning models to predict value in different cities, these models would learn mappings between features of buildings and the value that those features add. Comparing the relative weights of these models should show that these models found value in different things.

Data and setup

The first thing I did was collect property data from Pittsburgh, Boston, and Dallas. These datasets contained different information about the properties, so the first step was to figure out what could actually be used. This turned out to be quite revealing.



Property-specific data available for Pittsburgh, Dallas, and Boston

There wasn't much overlap between the data available, but this was telling in and of itself. In the diagrams above, a grey box indicates that the city (listed at the top) contains data for the feature listed on the left. Since these datasets are used by the municipalities to assess the value of properties, they paint a picture about what is considered valuable in each area. For instance, the data shows that Dallas assesses properties using the presence of a pool, hot tub, deck, and lawn sprinklers. Pittsburgh uses the presence of a basement garage, and Boston uses information about the view from condos. Each of these sheds light on the culture and character in the respective cities; Dallas is comprised mainly of single family homes with ample space for lawns and pools, Pittsburgh is a more population-dense city^{1,2} built on a hilly terrain, leading to more low-rise developments and garages in basements. Boston is much more of a vertical city, where views take on much more relevance than in Pittsburgh or Boston and multi-family developments and a walkable urban context renders garages all but irrelevant.

1 - Pittsburgh Population. (2020-02-17). Retrieved 2020-03-22, from <http://worldpopulationreview.com/us-cities/pittsburgh-population/>

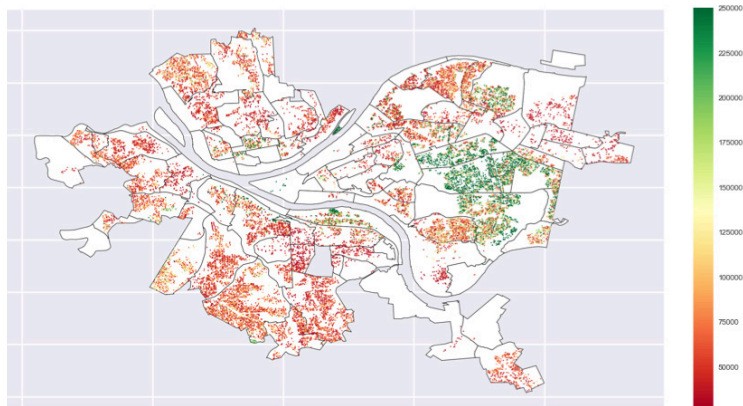
2 - Dallas Population. (2020-02-17). Retrieved 2020-03-22, from <http://worldpopulationreview.com/us-cities/dallas-population/>



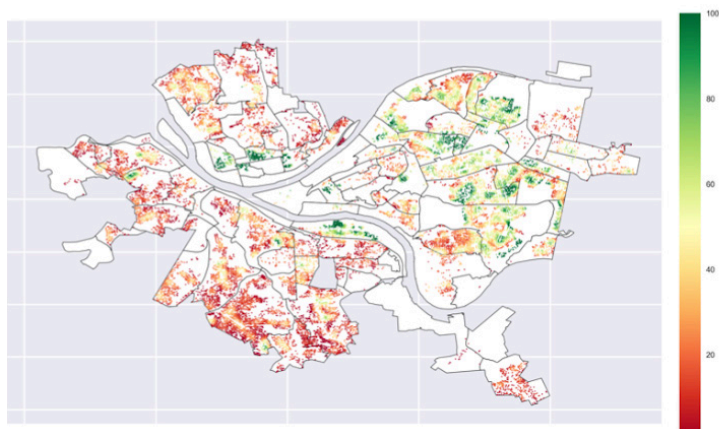
Citywide data available for Pittsburgh, Dallas, and Boston

Though context is latent in the data on individual properties, some notion of the surrounding area would push this point even further. After surveying the available data for each city, it was clear that there weren't many common datasets between these cities. After careful consideration, street tree datasets were chosen to be included in the AVM as a way to capture some spatial element of the urban context. This did mean that Dallas would be excluded from the rest of the analysis, as they do not publish data on their street trees, but their unique notion of value was adequately conveyed in the data availability at the individual property level.

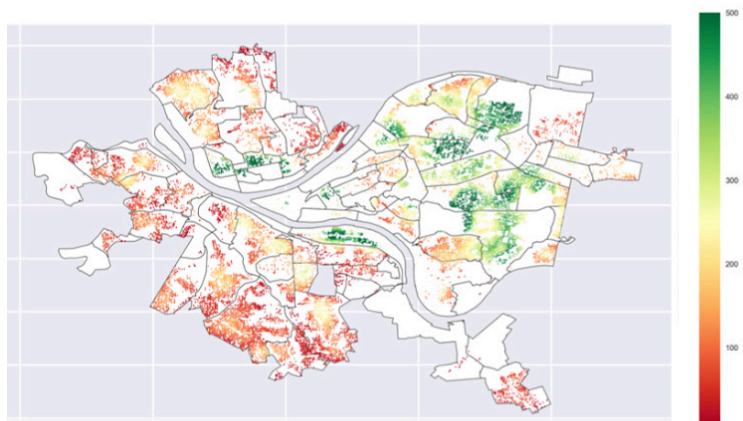
Using the tree datasets from Boston and Pittsburgh, I created two new features that describe the street tree density around each property. These two features, number of trees within a 500ft radius and number of trees within a quarter mile radius, were created by comparing location information from the tree datasets and the property assessment datasets.



Pittsburgh Home Prices



Pittsburgh Trees within 500 Feet



Pittsburgh Trees within 0.25 Miles

With the data in place, training the models could commence. The model of choice here was the random forest model³. Without getting into too much detail, this machine learning algorithm works by constructing thousands of models to predict value, then combining all of those models to produce one final output. This type of model is able to capture a significant amount of complexity while still remaining legible. One important metric produced by this model is a notion of ‘feature importance’, which is an indication of the relative weight that any feature of the data has across the thousands of models within the random forest. In other words, a high feature importance indicates that there was some consensus between the thousands of models that this feature has a large impact upon the value of a property.

Results

Pittsburgh Data (81,148 data points)

With Trees:	Without Trees:	Assessment's Accuracy:	Zestimate Accuracy: (482,760 data points)
10.41% MdApe	10.47% MdApe	22.04% MdApe	4.90% MdApe
18.17% MApe	16.71% MApe	49.81% MApe	NA MApe
48.39% < 10% error	48.35% < 10% error	27.78% < 10% error	72.5% < 10% error

Boston Data (24,044 data points)

With Trees:	Without Trees:	Assessment's Accuracy:	Zestimate Accuracy: (221,472 data points)
31.16% MdApe	29.20% MdApe	26.31% MdApe	5.10% MdApe
100.09% MApe	91.90% MApe	52.60% MApe	NA MApe
19.06% < 10% error	19.60% < 10% error	20.33% < 10% error	74.6% < 10% error

The random forest models fit the data significantly, enough to justify that the models were accurately predicting value. Though the models did not beat Zillow's performance, they did outperform Pittsburgh's municipal assessments of value and came close to the performance of Boston's municipal assessments of value. However, achieving a high predictive accuracy was not the intent here. Rather, it was to train models well enough so that their algorithms could be investigated to derive insights about valuation in each physical context. This makes sense, as only common features between Boston and Pittsburgh were used. If using the full datasets available, we would likely see much higher accuracy.

In comparing the relative importances of features between Boston and Pittsburgh, we can see that each AVM has a unique mapping of features to value. For instance, we can

3 - Breiman, L. (2001). Random Forests. Machine Learning, 45, 5–32. <https://doi.org/https://doi.org/10.1023/A:1010933404324>

see that the east-west position of a home in Pittsburgh is much more important than it is in Boston. While interesting, this speaks more to the location of high value neighborhoods than it does about the value of physical attributes of the houses. To this end, note that the amount of living area is more important in the Pittsburgh model, whereas the age of the building is more important in the Boston model.

It should be noted that feature importance does not equate to a direct correlation between feature and value. The random forest model can find complex nonlinear relationships in data, so feature importance is more about the degree to which a feature was used to derive the value than it is about direct correlation.

Insights

This project highlights the geographic specificity of value. In different areas, cultures, climate, and demographics can drastically change the types of buildings that are required in a city, changing the nature of supply and demand of housing stock. This phenomena can be seen by simply evaluating the availability of data. Data, in this case, is supplied by municipalities which use the data for assessing the value of buildings. If a particular feature of a building is not prevalent or relevant for pricing, then the municipalities are likely not going to keep track of it.

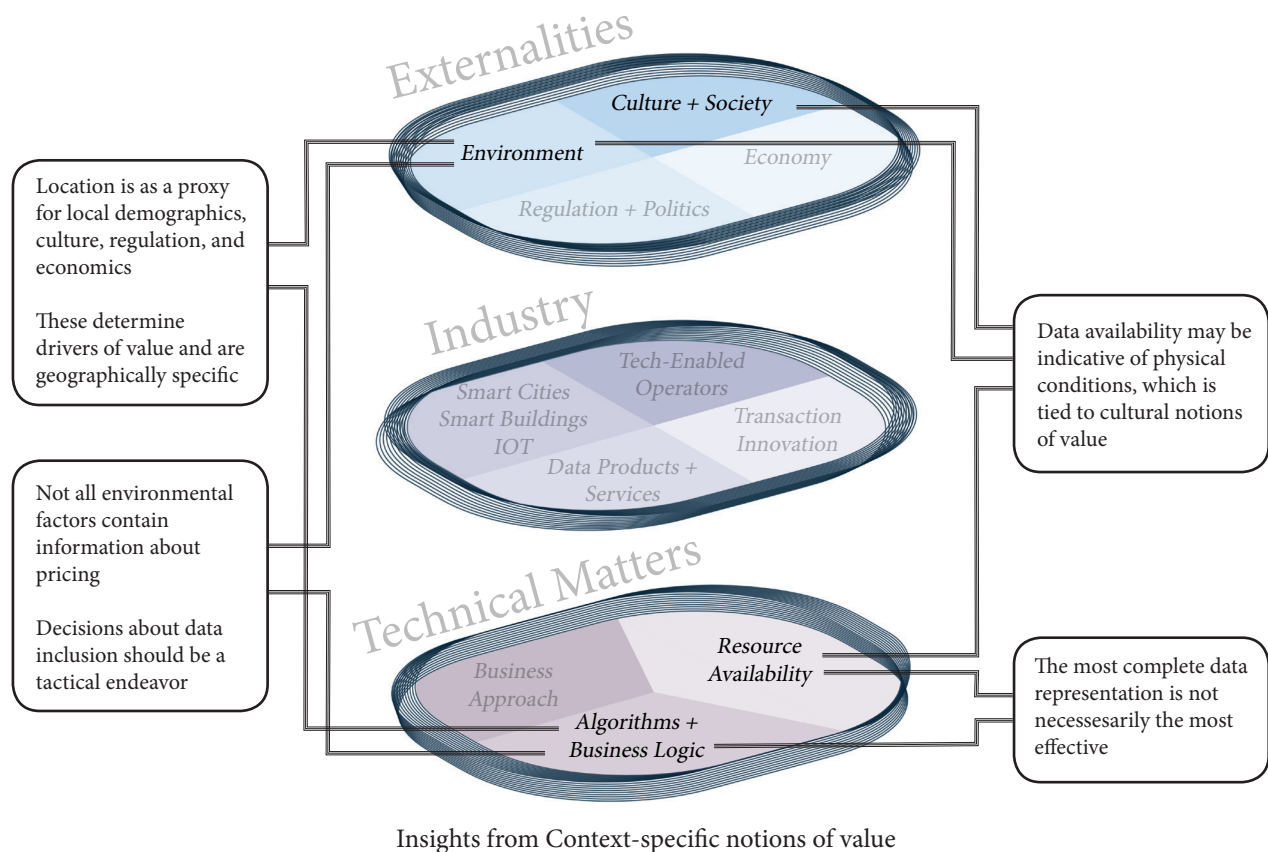
Even amongst the common representation of data, the features are not equally important in determining price. Thus, an effective AVM being applied in multiple distinct areas should be able to account for these geographic differences. This is not an easy task, but some AVM have been formulated that can account for these differences⁴. However, geography here is a proxy for cultural, environmental, and economic differences between areas. These factors play an important role in determining which of a building's attributes are important in determining value.

This project also showed that even though there is a lot of information that is latent in the data that we have on the built environment, not all of it has significance in determining real estate value. In this case, density of street trees did not improve the predictive accuracy

4 - Francke, M. K., & Vos, G. A. (2004). The Hierarchical Trend Model for Property Valuation and Local Price Indices. *The Journal of Real Estate Finance and Economics*, 28(2/3), 179–208. <https://doi.org/10.1023/B:REAL.0000011153.04496.42>

of the models. This is important because it shows that simply gathering the most extensive representation of the built environment will not necessarily lead to better predictive accuracy. In fact, statistical theory states that data with more features requires exponentially more data to accurately make predictions. So instead of gathering the most complete representation of the built environment, capturing the most effective representation for valuation is important.

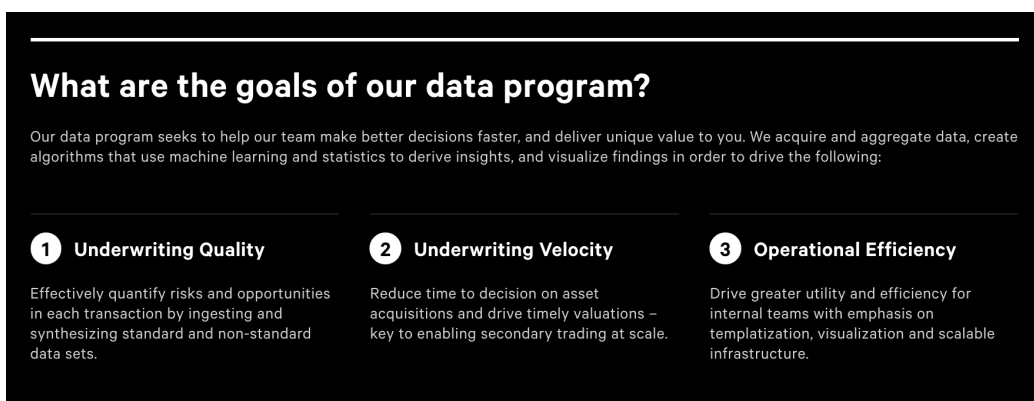
Another important point to make is that this study does not conclusively say that street trees are not valuable in any context, but just that the density of street trees is not a good predictor of value in Pittsburgh and Boston. However, this does not mean that density of tall street trees or wide street trees or deciduous street trees is not a good predictor of value. This study is suggestive of the ways in which geospatial data can be broken down for the purposes of automated valuation, although this particular implementation was not successful.



Though none of the insights here have a relationship drawn directly to industry, they apply to any application of AVM. Since any area of industry can apply AVM, these issues are not specific to any one subsector of industry but are relevant in all subsectors of industry.

3.4 - Cadre

Cadre is a PropTech company which invests in real estate as a limited partner, syndicates these investments to users on their online platform, and allows for users to buy and sell stakes in these investments in Cadre's secondary market. Cadre was founded in 2014 and last valued at \$800 million in 2017, and is financially backed by some of the most influential players in technology and real estate. Cadre's business model is a union between a tech-enabled operator model and a transaction innovation model and illustrates the specific considerations of these segments of industry.



The infographic is titled "What are the goals of our data program?" and explains the purpose of the data program: to help the team make better decisions faster and deliver unique value by acquiring and aggregating data, creating algorithms using machine learning and statistics to derive insights, and visualizing findings. It then lists three goals: 1. Underwriting Quality (effectively quantify risks and opportunities), 2. Underwriting Velocity (reduce time to decision on asset acquisitions and drive timely valuations), and 3. Operational Efficiency (drive greater utility and efficiency for internal teams with emphasis on templating, visualization, and scalable infrastructure).

1 Underwriting Quality	2 Underwriting Velocity	3 Operational Efficiency
Effectively quantify risks and opportunities in each transaction by ingesting and synthesizing standard and non-standard data sets.	Reduce time to decision on asset acquisitions and drive timely valuations – key to enabling secondary trading at scale.	Drive greater utility and efficiency for internal teams with emphasis on templating, visualization and scalable infrastructure.

Cadre's description of their 'data edge'¹

I spent six months, from June to December 2019, embedded within Cadre as a software engineer, where I observed and participated in tackling some of the problems that are specific to the business. Specifically, I was working on a team that engages with the collection, ingestion, and analysis of data and the production of tools for the acquisitions and asset management teams at Cadre. Among other initiatives, this team works to 'quantify risks and opportunities in each transaction' and 'drive timely valuations' through technological intervention, forming what Cadre refers to as its 'data edge'².

Through this hands-on experience, I gained valuable insights about PropTech. Where applicable, I will discuss how these insights might generalize or extend onto other areas of PropTech. The intent of this subchapter is to discuss Cadre in depth, with a particular focus on how the business fits into the framework presented in this thesis. In an effort to protect the intellectual property and trade secrets of Cadre, I will not dive into the highly technical aspects

1 - Cadre. (n.d.). Our Data Edge. Retrieved April 24, 2020, from Cadre website: <https://cadre.com/our-data-edge>

2 - Cadre. (n.d.). Our Data Edge. Retrieved April 24, 2020, from Cadre website: <https://cadre.com/our-data-edge>

of what I was doing. Instead, I will focus on the broader context and, where applicable, utilize content that is available online to reinforce my points.

Cadre as tech-enabled operator

As a tech-enabled investment firm, Cadre is infusing traditional real estate roles with technology. This brings about a unique set of circumstances which are prime for investigation.

As a company which aims to leverage its deep talent on both the technology and investment side, technological solutions take on their own set of considerations. Rather than fully automating the investment process (as iBuyers do), Cadre creates internal tools that make its investment professionals more efficient and effective. As investment professionals need to understand and have conviction in the technology-driven insights upon which they are operating, technical solutions (such as automated valuation) have an emphasis upon legibility of the algorithm. In these contexts, simpler and well-understood algorithms such as linear regression may be more amenable to black-box machine learning solutions that provide higher predictive accuracy.

This is not to say that the human requirements of a tech-enabled operator preclude sophisticated technological intervention, nor does it mean that automation is fully off the table. In Cadre's case, this means that solutions are geared more toward automating procedural tasks and driving insights that enable humans to make better decisions and unveil hidden realities of the current investment space. As an example³ of this, Cadre used high-dimensional Foursquare⁴ data on foot traffic, recent business formation, and business hours of operation to represent the vitality of a neighborhood. Encoding this information with a natural language processing technique called term frequency inverse document frequency (TFIDF) and using cosine similarity as a metric of distance, they identified the counties whose recent business formation today is most similar to Austin in 2014. The final result of this experiment was a list of cities which show promise as emerging markets for investment, which can be used to drive Cadre's investment thesis. The results, published online, are tempered by the statement that "this approach to understanding the trajectory of a market only provides part of the picture,

3 - Cadre. (2020). Finding the Next Austin. Retrieved April 24, 2020, from Cadre Insights website: <https://cadre.com/insights/finding-the-next-austin/>

4 - Foursquare - The Trusted Location Data & Intelligence Company. (2020). Retrieved April 24, 2020, from <https://foursquare.com/>

and should be paired with adequate fundamental analysis for a holistic view of the market”, highlighting this hybrid human-technology driven approach.

In a world where computational finance has provided technological solutions that operate almost entirely independently (e.g. high frequency trading), this more restrained application of technology may seem a bit counterintuitive. However, the world of commercial real estate is still very relationship-driven and dominated by traditional business practices. Specifically within an investment context, sourcing and evaluating deals is something that requires individuals with extensive networks and strong analytical capabilities. At this time, algorithms and computation simply cannot replace these human factors. This makes recruiting an incredibly important task, as strong human capital is a scarce resource and is in high demand. Cadre does exactly that and is known for recruiting highly talented individuals with extensive industry knowledge.

Cadre as transaction innovation

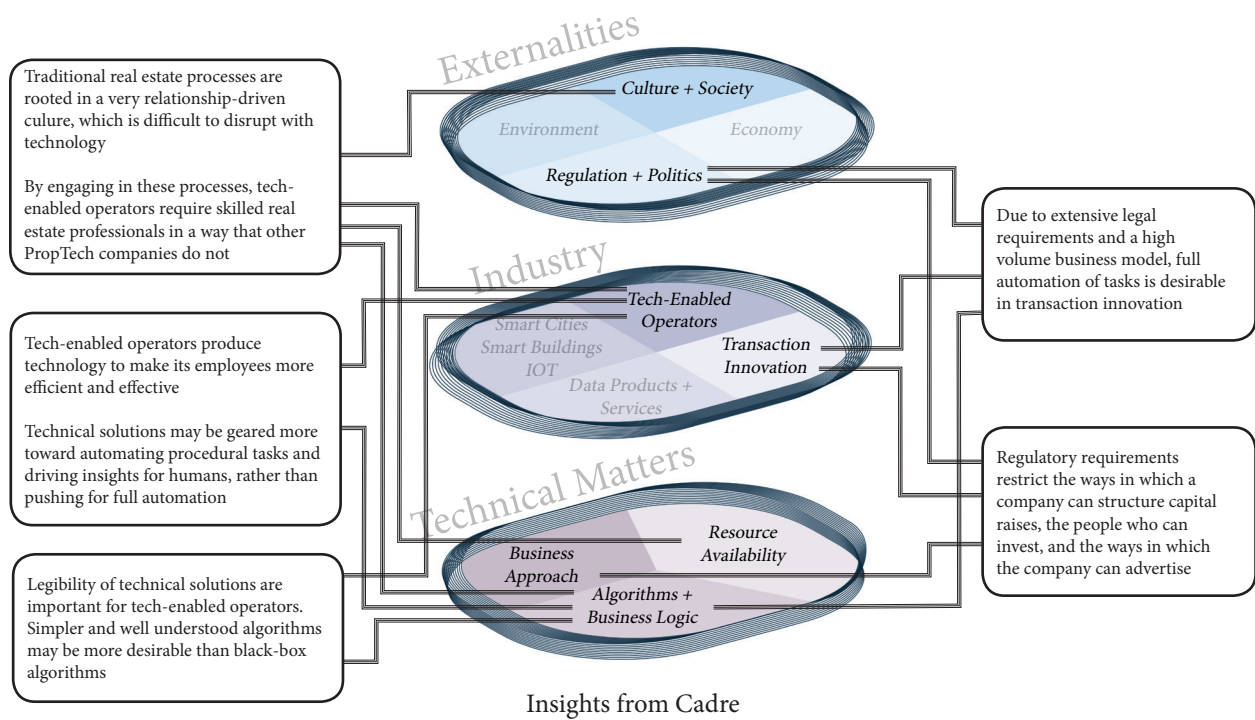
The issues at play in the production and management of Cadre’s online real estate marketplace are almost entirely distinct from those seen within Cadre’s role as a tech-enabled investment firm.

Whereas the technology in the tech-enabled investment side is geared toward driving insights, the technology on the transaction innovation is much more about automation. Cadre has had over \$3 billion invested on its platform. Each new deal can bring in tens of millions of dollars and users can invest as little as \$10,000 in each deal⁵. That means that on an individual deal, there may be thousands of investors involved. For each of these investors on each deal, paperwork needs to be drawn up, contracts signed, money moved, and a number of other logistical tasks need to be taken care of. In a typical setting, this would require an army of employees and a mess of paperwork. However, at the time that I left, Cadre only had around 100 employees, and only a handful of those actually handled the tasks around these transactions. This is only possible through an immense amount of automation.

5 - Cadre. (2020). Cadre | Real Estate Investment Platform. Retrieved April 24, 2020, from Cadre website: <https://cadre.com/>

Since real estate is a heavily regulated industry, raising funds for investment requires jumping through a lot of regulatory hoops. The SEC allows for a few different ways of structuring a real estate investment offering, each with their own implications about who can invest, how much can be invested, and how these offerings can be marketed. For instance, Cadre structures its offerings under SEC regulation D, which stipulates that only accredited investors (those who make over \$250k annually or have more than \$1m in assets) can invest. In order to comply with this regulation, Cadre must verify that each investor meets these requirements and cannot advertise an investment to someone who has not been verified. This effectively bars Cadre from openly advertising any deal on the internet, which makes marketing and growing the platform much more difficult.

After a deal is fully syndicated, Cadre has a secondary market which allows users to buy and sell their ownership in the investments on the platform. This allows users to enter and exit investments without the need for the building to be sold, creating a liquid investment vehicle for an otherwise illiquid asset. This paradigm of real estate transaction is much more akin to the way transactions work on the stock market than the process of deal sourcing and acquisition discussed earlier. As opposed to the tech-enabled investment side of the business, this transaction process is very amenable to computation and automated valuation, since the asset's performance is understood and some notion of 'market value' can be derived by trades on the secondary market, though this would not necessarily align with the value of the building. Though it was not the case while I was there, these transactions could theoretically be fully automated and an independent exchange could thrive with very little human intervention on Cadre's side.



4 - Conclusion

Contributions and future work

This thesis has two lines of contributions. The first comes from technical experimentation, and is the result of the questions being asked within each individual project. The second line of contributions comes from the more holistic approach and reflection across all of these projects, which speak to matters of PropTech more generally.

Facade analysis with Google Street View showed that images of building facades hold some information that can be used to value a building. In doing so, this project also suggests new ways of connecting design to monetary value and new ways of breaking down unstructured data for valuation. Future work could extend this project by doing higher-fidelity analysis of facades to pinpoint exactly what in the facade contains information that can be used for valuation. One possible approach might involve identifying individual elements within the facade (i.e. windows, doors, etc.) and representing some spatial relationships between them, then training an AVM to predict values based on this information and analyzing the results.

Zillow error rate analysis provided a geospatial visualization of ‘Zestimate’ error rates. It probed the motivations and drivers behind the largely unrestrained distribution of inaccurate data, juxtaposed with the motivations and drivers behind Zillow’s much more conservative utilization of the ‘Zestimate’ within their Zillow Offers platform. Future work might continue this investigation, looking at other PropTech companies and critically evaluating their business models and understanding the rationale behind their business practices¹.

Context-specific notions of value probed the conditions which produce such large geospatial variance in automated valuation models. It used AVM in a somewhat unusual way, as a tool to break down problems to drive insights, showing the potential for this technology to be used to highlight issues beyond property value. Future work might apply this methodology to understand how values of building elements change over time as a way to measure cultural perception of value. For instance, AVM applied to longitudinal real estate data may be able to show diminishing value of standalone dining rooms in residential real estate.

¹ -Mike DelPrete does similar investigations with data, but focuses on current performance rather than critiquing business context

Using my time at Cadre and my deep knowledge of its business model allowed for a discussion about how tech-enabled operators and transaction innovators differ. This reflected a perspective on the company which would be difficult to attain without immersion within the organization. The discussion here is limited by a non-disclosure agreement, but still provides a useful case-study.

As a group of projects, this thesis provides a bottom-up approach to investigating PropTech. In doing so, each project shows the connections between technical considerations, industry context, and issues external to PropTech. These connections also illustrate the utility and necessity of a framework which gives explicit representation to these key factors in PropTech. Admittedly, a top-down approach gives a much broader understanding of PropTech than the narrow but more precise view of a bottom-up approach. However, the more hands-on approach to investigating PropTech gives a much richer view into the processes and issues that are at play within industry. This approach can somewhat easily be grafted on to otherwise technical research with a bit of reflection about the broader insights about PropTech that are highlighted by the technical experimentation. Future work could also design technical experiments around probing very specific relationships between industry, technical matters, and externalities in PropTech. My work on context-specific notions of value may provide a good model for this type of work.

Concluding thoughts

PropTech academia is still in its infancy, and prior work has just started to define the industry. If arguments made in this thesis seem commonsensical, know that they are by no means trivial. The dearth of academic literature in the space calls for even simple insights to be backed up by data and experience. PropTech is experiencing a period of rapid change, so the specific conditions and companies described in literature may not hold for long. However, the hope is that this blended approach of technical experimentation and critical analysis illuminates issues and relationships and provides insights that have more longevity than the subjects upon which they are predicated.

Bibliography

Baum, A. (2017). *PropTech 3.0: The Future of Real Estate*. Oxford, UK.

Baum, A. (2020). *PropTech 2020 : the future of real estate*. Retrieved from <https://www.sbs.ox.ac.uk/sites/default/files/2020-02/proptech2020.pdf>

Bijlsma, E., Borst, R. A., D'Amato, M., Des Rosiers, F., Francke, M. K., Aurelio Stumpf Gonzalez, M., ... van de Hoef, P. (2008). *Mass Appraisal Methods* (T. Kauko & M. D'Amato, Eds.). <https://doi.org/10.1002/9781444301021.ch1>

Block, A., & Aarons, Z. (2019). *PropTech 101: Turning Chaos Into Cash Through Real Estate Innovation* (1st ed.). New York: Advantage Media Group.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/https://doi.org/10.1023/A:1010933404324>

Brown, R. W. J. (2018). *#PropTech: A guide to how property technology is changing how we live, work and invest*. Independently published.

Cannon, S. E., & Cole, R. A. (2011). How Accurate Are Commercial Real Estate Appraisals? Evidence from 25 Years of NCREIF Sales Data. *The Journal of Portfolio Management*, 37(5), 68–88.

Carlini, J. (2014). *Location, Location, Connectivity* (1st ed.). East Dundee, Illinois: CARLINI & ASSOCIATES.

CBInsights. (2017). *Real Estate Tech Funding Reaches New Highs In 2016*. Retrieved April 24, 2020, from <https://www.cbinsights.com/research/real-estate-tech-startup-funding/>

Chandra, L. (2018). *The PropTech Guide: Everything You Need to Know about the Future of Real Estate* (1st ed.). London, UK: PropTech Asset Management.

CRE Tech. (2020). 2019 End-Year Report. Retrieved from <https://www.cretech.com/product/cretech-2019-end-year-report/>

D'Amato, M., & Kauko, T. (2017). *Advances in Automated Valuation Modeling: AVM After the Non-Agency Mortgage Crisis* (1st ed.; M. D'Amato & T. Kauko, Eds.). Springer International Publishing.

DePietro, R., Wiarda, E., & Fleischer, M. (1990). The Context for Change: Organization, Technology, and Environment. In L. Tornatzky & M. Fleischer (Eds.), *The Processes of Technological Innovation* (pp. 151–175). Lexington Books.

Dubin, R., Pace, R. K., & Thibodeau, T. G. (1999). Spatial Autoregression Techniques for Real Estate Data. *Journal of Real Estate Literature*, 7(1), 79–95. <https://doi.org/10.1023/A:1008690521599>

Fink, D. (2017). *Linking Design to Finance*. MIT.

Francke, M. K., & Vos, G. A. (2004). The Hierarchical Trend Model for Property Valuation and Local Price Indices. *The Journal of Real Estate Finance and Economics*, 28(2/3), 179–208. <https://doi.org/10.1023/B:REAL.0000011153.04496.42>

Fuerst, F., McAllister, P., & Murray, C. B. (2011). Designer buildings: Estimating the economic value of “signature” architecture. *Environment and Planning A*, 43(1), 166–184. <https://doi.org/10.1068/a43270>

Knaap, G. J., & Song, Y. (2003). New urbanism and housing values: a disaggregate assessment. In *Journal of Urban Economics* (Vol. 54).

Kontrimas, V., & Verikas, A. (2011). The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing Journal*, 11(1), 443–448. <https://doi.org/10.1016/j.asoc.2009.12.003>

Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By* (1st ed.). Chicago: The University of Chicago Press.

Law, S., Paige, B., & Russell, C. (2019). Take a Look Around: Using Street View and Satellite Images to Estimate House Prices. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(5). <https://doi.org/10.1145/3342240>

Lockwood, T., & Rossini, P. (2012). Efficacy of the Geographically Weighted Model on the Mass Appraisal Process. *The 18th Annual PRRES Conference*, 15–18.

Loukissas, Y. (2018). All the Homes: Zillow and the Operational Context of Data. *10766*, 272–281. <https://doi.org/10.1007/978-3-319-78105-1>

Matysiak, G. (2017). Automated Valuation Models (AVMs): a brave new world? *Wroclaw Conference in Finance*, (August). Retrieved from <https://www.researchgate.net/publication/319355261>

Mccluskey, W. J., Mccord, M., Davis, P. T., Haran, M., & Mcilhatton, D. (2013). Prediction accuracy in mass appraisal : a comparison of modern approaches. *Journal of Property Research*, 30(4), 239–265. <https://doi.org/10.1080/09599916.2013.781204>

McCluskey, W. J., & Adair, A. (Eds.). (1997). *Computer Assisted Mass Appraisal: An International Review* (1st ed.). Ashgate Publishing Company.

Miles, M., & Huberman, M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.). SAGE Publications, Inc.

O'Connor, P. M., Wadsworth, W. M., Almy, R., Borst, R. A., Guerin, B. G., Kane, M. S., & Linne, M. R. (2003). Standard on Automated Valuation Models (AVMs). *AVM Standard Review*, pp. 1–36. International Association of Assessing Officers.

Pace, R. K., Barry, R., & Sirmans, C. F. (1998). Spatial Statistics and Real Estate. *Journal of Real Estate Finance and Economics*, 17(1), 5–13. <https://doi.org/10.1023/A:1007783811760>

Poleg, D. (2020). *Rethinking Real Estate: A Roadmap to Technology's Impact on the World's Largest Asset Class* (1st ed.). <https://doi.org/10.1007/978-3-030-13446-4>

Ravitch, S., & Riggan, M. (2012). Reason and Rigor: How Conceptual Frameworks Guide Research (1st ed.). SAGE Publications, Inc.

Rivaton, R., & Pavanello, V. (2018). Make Real Estate Great Again: PropTech, Real Estate, Construction Tech. RETech Publishing.

Rosen, S. (2018). Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR. Retrieved from <https://arxiv.org/pdf/1409.1556.pdf>

Wong, M. (2018). The rise of real estate tech. Retrieved from https://www.cbinsights.com/reports/CB-Insights_Rise-Real-Estate-Tech-Briefing.pdf

Wong, M. (2015). Disrupting Real Estate: 58 Startups Reshaping How Property Is Bought, Sold, And Managed. Retrieved from <https://www.cbinsights.com/research/real-estate-tech-startups-market-map/>

Xu, Z. (2014). Architectural Style Classification using MLLR. Retrieved April 24, 2020, from <https://sites.google.com/site/zhexuutssjtu/projects/arch>