Forecasting Atlanta Gentrification with Transformers

Gaurav Sett
Georgia Institute of Technology
gauravsett@gatech.edu

December 2022

Undergraduate Thesis

Approved by:

Jacob Abernethy (Advisor)
College of Computing
Georgia Institute of Technology

DocuSigned by:
D47008B2F846482...
12/11/2022 | 11:39 PM EST

David Muchlinski
Ivan Allen College of Liberal Arts
Georgia Institute of Technology

DocuSigned by:
David Muchlinski
494609921AE1446...
12/12/2022 | 10:05 AM EST

Abstract

Gentrification is an impactful trend in American cities, yet our ability to measure and predict this process remains weak. This thesis explores the use of machine learning to predict gentrification in Atlanta, Georgia. We use a dataset of land parcels collected by Fulton County Tax Assessors and set out to forecast changes in local land value. Inspired by progress in natural language processing, we apply a machine learning model called a transformer to forecast gentrification. We find our model outperforms typical methods for time-series forecasting of gentrification, and we discuss the implications of our findings for future research.

1 Introduction

As America's housing affordability crisis persists, contentious debates about gentrification have taken over local discourse. This process of neighborhood change is often associated with the displacement of lower-income tenants. Indeed, the nation has a long history of excluding marginalized groups from the prosperity generated by cities. Residential displacement is highly destabilizing for those affected, and a myriad of policies have been developed in response. Unfortunately, the area is difficult to study, leaving researchers with an incomplete picture of both the causes, effects, and efficacy of responses to gentrification.

The primary challenge is that gentrification is difficult to measure. An ideal dataset would include granular information on highly localized areas over long periods of time. This data is rarely available, forcing researchers to make the most of what they can collect.

We start with yearly tax assessor data from Fulton County, Georgia which comprises the vast majority of the city of Atlanta, detailing the location, size, and value of each land parcel. This data is typically collected annually by nearly every municipality in the United States. We use the key metric of land value to measure gentrification.

Inspired by advances in natural language processing, we use a machine learning model known as a transformer tasked with multivariate forecasting of land value for multiple time-series. Specifically, we leverage a variant of this model, the Temporal Fusion Transformer (Lim et al. 2021) which has been modified to produce interpretable forecasts. This model is trained on a dataset composed of 3.8 million observations from 2011 to 2021. We implement simple spatial analysis by creating features representing averages of surrounding areas.

We find this model outperforms a variety of typical forecasting methods, including a baseline model that uses only the previous year's value. We demonstrate the capability of the transformer for ordinary data science tasks, and illustrate the promise of this model for forecasting urban development.

2 Literature Review

The trend of gentrification, a process of change in which historically low-income neighborhoods experience economic ascent, investment inflow, and demographic change, has steadily increased over the past few decades in the United States. A reductive but pragmatic metric of gentrification is housing prices. In Atlanta, nearly half of neighborhoods from 2000 to 2017 have seen the cost of housing exceed the regional median, sparking a contentious discussion about gentrification in the city (Thomas, Driscoll, and Aguilar 2020).

The history of this story is important, bringing us back over a century. From the 1930s to the late 1960s, banks, following guidelines from the federal government, classified neighborhoods predominantly home to people of color as risky. Cutting off credit from these communities, the policies triggered continual disinvestment. Around the same time, "white flight" began to take off. As urban cores decayed, higher-income white residents moved to the suburbs (Landis 2016). Through explicit segregation ordinances, denial of mortgages, and other policies, these suburban neighborhoods became highly exclusive. Within cities, urban renewal projects further enforced these trends. The development of the Interstate highway system was often used to destroy black neighborhoods and separate them from the rest of the city (Zuk et al. 2018). In Atlanta, Interstate 20 divides the poorer, predominantly black south side of the from the central business district and suburbs. As stated by then Mayer Bill Hartsfield, the highway defines "the boundary between the white and Negro communities" (Kruse 2019).

The racial and economic segregation of cities like Atlanta set the stage for today's patterns of gentrification. Nationwide, supply-restraints on housing are driving a housing affordability crisis. Predominantly, restrictive zoning that limits density has led to skyrocketing rents and home prices (Uhler and Sisney 2016). Control over community development has been largely handed over to localities which are resistant to neighborhood change. Thus, developers often face fierce opposition to new projects. Again, the housing landscape in Atlanta is largely divided between neighborhoods that are mostly lower-income and black, and those that are higher-income and white. In recent decades, real estate investors have come to understand they stand the best chance targeting low-income neighborhoods with low political capital and economic vulnerability exacerbated by the 2008 housing crisis (Dougherty 2020). In contrast to the ongoing discourse in many cities, gentrification cannot be simplistically labeled as a force for good or bad. Gentrified neighborhoods typically bring more jobs, nicer amenities, higher quality housing, reduced crime, and better schools. However, this change is often exclusionary. As the neighborhood gentrifies, vulnerable low-income residents may be forced out in a process called displacement, either by high prices, evictions, or deteriorating living conditions. As old residents leave, landlords are able to rent to new, higher-income residents at higher prices.

Where displaced residents end up depends on local geography, but one thing that is certain is that displacement is destabilizing for tenants. Some experience homelessness, and the process can be highly stressful and traumatic, especially for children. Most who must find a new home not only lose out to improvements to their old neighborhood, but also have to deal with longer commutes, reduced economic opportunity, substandard housing, and the anxiety associated with the disruption (Marcus and Zuk 2017). Policymakers responded with a wide variety of policies. Rent control, subsidies, tenant's rights, stronger zoning, and weaker zoning have all been applied, but our understanding of the results and the underlying process remains weak (Zuk et al. 2018).

Data on gentrification is hard to collect. The first issue is the locality of data. Gentrification happens at the neighborhood level, and driven by specific developments. Thus, we are limited to data providers deeply interested in the Atlanta community. Second is the temporal aspect. Neighborhood change is a long process that also may progress in rapid intervals, so we are limited to data providers who are committed to collecting information at most yearly, indefinitely (Easton et al. 2020). The third issue is data confidentiality. For a neighborhood's residents, we are interested in not only their socio-economic profile, but also their residential history. There are few providers able to access such personal details (Cohen and Pettit 2019).

In sum, these constraints are quite difficult to overcome. Thus, there is sparse quantitative literature for such a hot topic. Indeed, much of the advancements in the literature come from unique datasets. For example, the Federal Reserve has used their Consumer Credit Panel/Equifax dataset of individual credit records including census geography identifiers (Ding and Divringi 2015). Other researchers have secured data from online services like Yelp, detailing neighborhood amenities and businesses (Glaeser, Luca, and Moszkowski 2020). Data scraping methods have also been employed, such as using web scraping to track eviction filings in court databases (Raymond, Stein, and Haley 2021). There is promise in the growing availability of data, especially as the concept of high-tech smart cities progresses. However, the necessary data remains difficult to collect, requiring effort, expense, or connections.

Furthermore, unique datasets will not give us a broad understanding of the topic. Researchers typically note in their discussion that their results cannot be generalized due to differences in cities' housing policies (Dragan, Ellen, and Glied 2019). Thus, these studies lacking descriptive power will certainly lack predictive power when applied to different areas. While these results are informative, there is a clear need for stronger methodology for gentrification forecasting.

The growing use of geographic information system (GIS) technology was vital for field of urban informatics. Large datasets composed of small geographical units require complex methods for spatial analysis. Paired with machine learning, GIS has empowered many researchers to take advantage of such data. Many researchers and policymakers have collected census, tax, and other readily available data to create predictive models for different cities (Reades, De Souza, and Hubbard 2019). Employing spatial analysis, these models look at neighborhoods in whole, rather than individual phenomena. This approach captures neighborhood character in a way that was previously impossible. Machine learning has further advanced this analysis, bringing forecasting and classification models to our tasks of identifying gentrification trends (Steif 2020; Graff 2020). Rather than simply describing the data or fitting curves to time-series, these models can identify predictors of key outcomes and forecast future trends.

In the field of natural language process, the transformer has demonstrated a impressive ability to model sequential data (Vaswani et al. 2017). Beyond language, the transformer architecture has been applied to a wide variety of domains, including numerical time-series forecasting. This model implements the concept of attention. Attention is a mechanism that allows the model to focus on specific parts of the input data. This is especially useful for time-series forecasting, where the model must learn to predict the future based on the past. Key examples include the forecasting of seasonal influenza, city traffic, and weather (Wu et al. 2020; Zhou et al. 2021). These models have been shown to outperform traditional time-series models, such as ARIMA, and have been applied to a wide variety of tasks.

However, like many machine learning approaches, transformers are difficult to interpret. This is a problem for forecasting, as we are often interested in the underlying factors that drive the model's predictions. Most methods for interpretability in machine learning do not work well for time-series models as they do not model the temporal relationships between features. In response, Lim et al. 2021 developed the Temporal Fusion Transformer which implements several improvements to the transformer for time-series forecasting that enable the explanation of how model features are used to make predictions.

3 Data

We accessed the Fulton County GIS Data Portal to collect tax assessors data for each year from 2011 to 2021 (*Fulton County GIS Portal* 2022). Each file comprises about 350,000 parcels for a total of about 3,850,000 parcels in the dataset. Land parcels are the units by which land is bought and sold. Each parcel is the basis for property taxes, which is why nearly all American local governments regularly collect this data. Each data point includes the appraisal value of the property, appraisal value of the land, appraisal value of the improvements, assessment value of the property, the assessment value of the land, assessment value of the improvements, land acreage, owner information, zoning code, and land use code of the land. This data is formatted as a shapefile allowing for processing using GIS software.

Land parcels often change shape, and thus we created a 100×100 mesh grid which we interpolated our parcel data onto. Each grid cell contains the average of the values of the parcels that intersect it, producing dataset that is uniform in size and shape. For each cell in our grid, we computed spatial features by averaging the features of the surrounding 9×9 grid, centered on the cell. For non-numerical features such as class code, we used the most common value in the surrounding cells.

Our data spans many years and thus intersects with many significant trends such as COVID-19 and the aftermath of the Great Recession. To account for such phenomena, we standardize our data by each year by substracting means and dividing by standard deviation. Thus, our features should be understood as relative to other land in Fulton County.

4 Methods

4.1 Forecasting Task

We are interested in forecasting land value across metro Atlanta as a measurement for urban gentrification. This is a spatiotemporal multivariate multiple time-series forecasting task. For input parcel data over n years $[\vec{x}_{t-n+1}, \ldots, \vec{x}_t]$, we predict future parcel land-value over an m year horizon $[y_{t+1}, \ldots, y_{t+m}]$. Note that input parcel data points $\vec{x}_i$ are vectors containing multiple features, including our target variable land value. Future target values are masked to prevent the model from using them as input.

We train our model on a 5 year input window from 2011 to 2015 and predict a 5 year horizon from 2016 to 2020. We also generate a 5 year forecast with input window from 2016 to 2020 and predict a 5 year horizon from 2021 to 2025.

4.2 Model Architecture

The model we use is based on the transformer architecture (Vaswani et al. 2017). The transformer is a sequence-to-sequence model that models the relationship between input and output using a concept called attention. Similar to how humans focus on specific stimuli, attention mechanisms allow models to automatically learn which parts of the input are most important for generating the output. The model is composed of two components, an encoder and decoder. The encoder takes in a sequence of input data and produces a sequence of hidden states. This hidden state is a fixed-length vector that represents the entire input sequence. The decoder takes in the hidden states and produces a sequence of output data. Because the transformer is an autoregressive model, the decoder is designed to iteratively predict the next element in the sequence.

We employ the Temporal Fusion Transformer (TFT) (Lim et al. 2021), a modification of the transformer that is designed for time-series forecasting through the implementation of several mechanisms throughout the architecture. The TFT accepts both static and time-varying inputs from past and future time steps. While neural networks provide the advantage of being able to learn non-linear relationships, we do not know with this task whether we need to model such interactions between inputs and outputs. To favor parsimony, the TFT employs gating mechanisms in the form of a Gated Residual Netowrk (GRN) to determine which features to apply non-linear processing to.

We will make reference to these gates throughout the model description. Our inputs are subject to a variable selection mechanism to remove features contributing noise. This paper provides an overview of the relevant components of the TFT, but we refer the reader to the original paper for a more detailed description.

4.2.1 Encoder

After variable selection, our time-series features are fed into Long Short-Term Memory (LSTM) encoders. LSTMs are a type of recurrent neural network (RNN) that process input sequences of any length, maintaining memory that captures dependencies between elements in the series. The outputs of the LSTMs are fed through GRN components before being concatenated and fed into the decoder as the hidden state.

4.2.2 Decoder

The TFT applies a self attention mechanism, allowing it to focus on specific parts of the series. This model uses multi-headed attention, where the self-attention is applied multiple times in parallel. To enable interpretation of the model's attention to each feature across attention heads, the TFT ensures that the value weights are shared for each head. This results in a single attention weight for each feature. This attention layer is also modified to handle temporal data, masking the target outputs in the input data. After, the data is fed through an additional feed-forward network making use of the weights from the earlier GRN which are shared across the decoder component. The output of the decoder is are 7-quantile forecasts, specifying the probability that the true value lies within a given range.

4.3 Training

Our model is trained on 80% of our data, predicting 5 time steps our from 5 years of input. The encoder LSTMs have a hidden size of 128 and a dropout rate of 0.2. We use 2 attention heads for the decoder. Following our outputs, we use quantile loss to train our network. We use a learning rate of $1e - 3$ with the Adam optimizer and train for 10 epochs. We use a batch size of 32.

4.4 Evaluation

To account for city wide trends in our data, we standardize our data by subtracting the mean and dividing by the standard deviation for each year. This helps account for wide scale economic trends like COVID-19. Our outputs are thus reported as differences in normalized values.

We evaluate our model using the remaining 20% of our data with the mean absolute error (MAE), the root mean squared error (RMSE), mean absolute percentage error (MAPE), and the median absolute percentage error (MdAPE).

We compare our transformer model to a variety of typical time series forecasting methods. Most previous approaches to the gentrification task leverage econometrics models, although new interest in machine learning has furthered the field.

Baseline We model a naive forecaster that returns the latest target value for the entire forecasting horizon.

VARMAX We use a multivariate VARMAX model with only our continuous data labels.

LSTM We built a simple LSTM forecasting model. This model has a embedding layer of size 128, 2 stacked bidirectional LSTM layers of size 128, and a linear layer for output. We use a learning rate of $1e-2$ with the Adam optimizer and train for 8 epochs.

5 Results

5.1 Performance

	MAE	$RMSE$	$MAPE$	$MdAPE$
TFT	**0.182**	**0.453**	**0.552**	**0.174**
Baseline	0.641	1.015	0.944	0.917
VARMAX	0.631	1.001	0.893	0.843
LSTM	0.334	0.601	0.652	0.435

Table 1: Forecasting evaluation across models

In Table 1, we can see that our transformer-based architecture performs well in our prediction task. Notably, its median absolute percent error is much lower than the other models. This is a good indicator that the transformer is a significant improvement over other methods when forecasting typical values. As expected our simple econometric models perform worse than our machine learning approaches. Furthermore, our transformer sees a significant improvement over our LSTM model. Across all metrics, our transformer outperforms the comparison models.

5.2 Analysis

Due to the limitation of the medium of this paper, we are unable to provide a full analysis of our model. We will focus on the model's performance and give a brief overview of its spatial forecasts. Because our outputs are normalized, we will emphasis the direction and relative intensity of our predictions rather than specific values.

5.2.1 Feature Importance

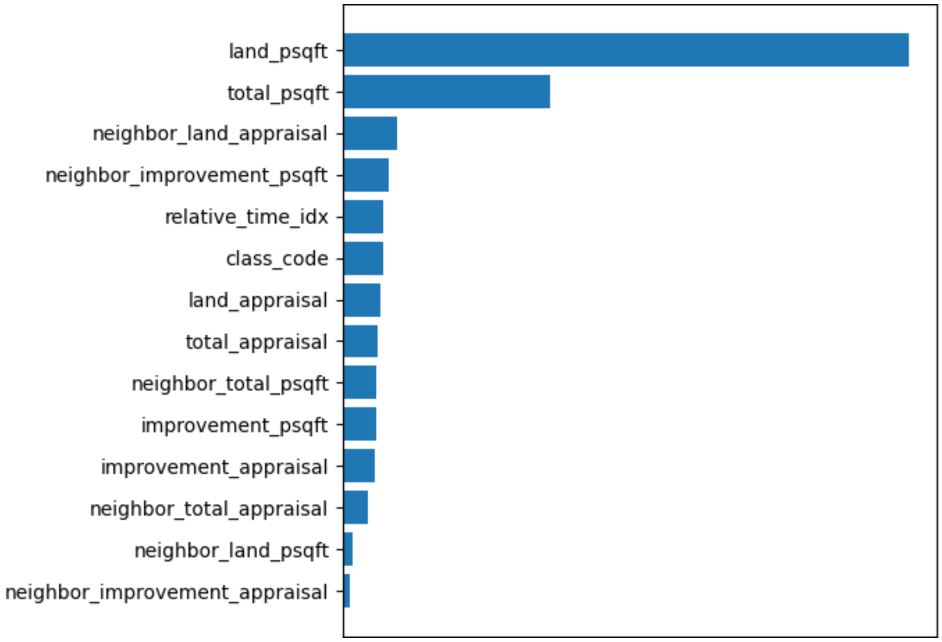


Figure 1: Relative importance of features used by the transformer model

In Figure 1, we can see the importance of features in our transformer model. These are the weights of the attention mechanism, which are used to determine which features are most important for the model to make predictions. As expected, the most relevant feature is past values for our target, land appraisal per square foot. This is followed by the total appraisal per square foot, the land value of neighboring cells, the time index within the forecasting window, and the zoning code of the property.

5.2.2 Error Clustering

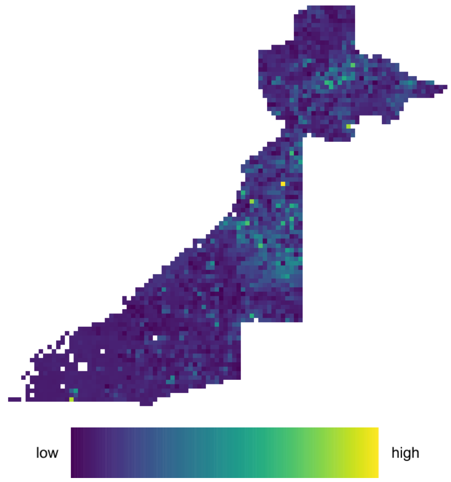
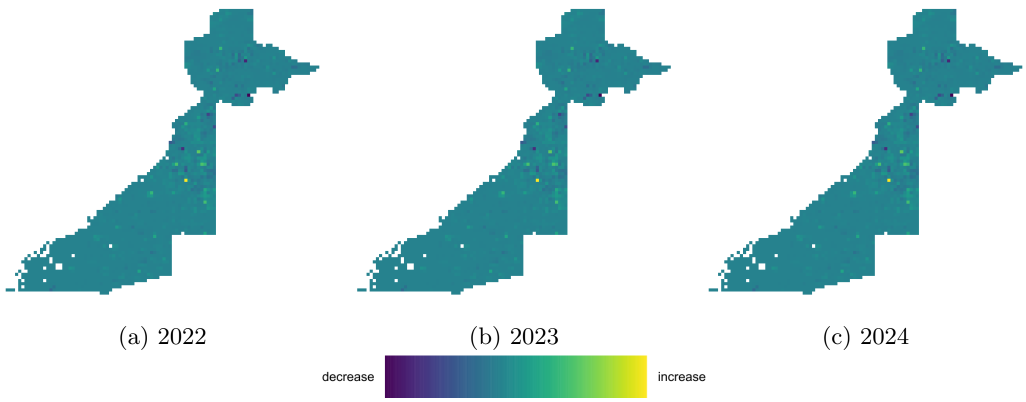


Figure 2: Clustering of errors from transformer model

In Figure 2, we can see the distribution of errors for our model. Note that these errors have been squared to make the distribution more uniform. We can see that the majority of errors are small, with a few large errors. This is a good sign that our model is able to predict the majority of values well, but struggles with a few outliers.

It appears our model is least accurate in forecasting in Alpharetta and the neighborhoods surrounding the urban core of Atlanta. This is likely due to the fact that these areas are rapidly developing and markets have yet to settle on prices. Unfortunately, this significantly intersects with areas experiencing gentrification, which is our model's primary forecasting goal.

5.3 Forecasts



(a) 2022 (b) 2023 (c) 2024

Figure 3: Forecasts for changes in land value from 2021

In Figure 3, we can see the forecasts of our model for the 3 years following our dataset. Here, we plot the expected changes in land value relative to 2021. We can see that our model predicts a significant increase in land value in the urban core of Atlanta, as well as in the neighborhoods surrounding it such as Forest Park. There are few isolated cells where our model predicts a decrease in land value.

6 Discussion

We set out to answer the question of whether machine learning can improve on existing methods for forecasting gentrification. Given the data's poor location specificity, infrequent collection intervals, and inconsistent formatting, the task of modeling urban development is a challenge for existing models (Zuk et al. 2018). As demonstrated by recent advances in Natural Language Processing, the transformers architecture shows promise in modeling sequential data. We found that our transformer model outperformed our econometric models and our LSTM model. While other work has approached gentrification prediction as a classification task (Chapple and Zuk 2016), creating taxonomies of urban development, our model provides a unique insight through its regression outputs. We also found that our model was able to capture the spatial and temporal patterns of gentrification in Atlanta. Whether this model can be generalized to other cities remains to be seen. Cities may differ due to economic growth, development restrictions, and socioeconomic characteristics. However, we believe that our model can be used as a baseline for future work in this area. We aim to provide an example of the usefulness of this data science tool which should be understood in the context of other work in this area.

We also believe that our model can be used to forecast other urban development patterns. For example, our model can be used to forecast the development of affordable housing, business development, and capital investment. However, each approach will require careful consideration of the features needed, the data available, and the model architecture. We hope that our work will serve as a starting point for future work in this area.

Our findings provide valuable insight for policymakers. Cities face significant resource constraints necessitating prioritization of many priorities. To identify which neighborhoods are seeing the most disruptive development trends, models like ours can provide a quantitative measure that serve as the basis for further investigation. We do not encourage the use of such models as the sole basis for evaluating gentrification. Rather, we hope that our model can be used as a tool to inform the decision-making process of policymakers.

7 References

Chapple, Karen and Miriam Zuk (2016). "Forewarned: The use of neighborhood early warning systems for gentrification and displacement". In: *Cityscape* 18.3, pp. 109–130.

Cohen, Mychal and Kathryn LS Pettit (2019). "Guide to measuring neighborhood change to understand and prevent displacement". In: *Washington, DC: National Neighborhood Indicators Partnership (NNIP). https://www. urban. org/research/publication/guide-measuring-neighborhoodchange-understand-and-prevent-displacement*.

Ding, Lei and Eileen Divringi (2015). "A Practitioner's Summary: Gentrification and Residential Mobility in Philadelphia". In: *Federal Reserve Bank of Philadelphia*.

Dougherty, Conor (2020). *Golden gates: fighting for housing in America*. Penguin Press.

Dragan, Kacie, Ingrid Ellen, and Sherry A Glied (2019). *Does gentrification displace poor children? New evidence from New York City Medicaid data*. Tech. rep. National Bureau of Economic Research.

Easton, Sue et al. (2020). "Measuring and mapping displacement: The problem of quantification in the battle against gentrification". In: *Urban studies* 57.2, pp. 286–306.

Fulton County GIS Portal (2022). URL: `https://gisdata.fultoncountyga.gov`.

Glaeser, Edward L, Michael Luca, and Erica Moszkowski (2020). *Gentrification and Neighborhood Change: Evidence from Yelp*. Tech. rep. National Bureau of Economic Research.

Graff, Alissa (2020). "Predicting Neighborhood Change in Detroit: A Data and Ethical Analysis of Data-Driven Policymaking". In:

Kruse, Kevin M (2019). "How Segregation Caused Your Traffic Jam". In: *The New York Times Magazine. August*.

Landis, John D (2016). "Tracking and explaining neighborhood socioeconomic change in US metropolitan areas between 1990 and 2010". In: *Housing Policy Debate* 26.1, pp. 2–52.

Lim, Bryan et al. (2021). "Temporal fusion transformers for interpretable multi-horizon time series forecasting". In: *International Journal of Forecasting* 37.4, pp. 1748–1764.

Marcus, Justine and Miriam Zuk (2017). "Displacement in San Mateo County, California: consequences for housing, neighborhoods, quality of life, and health". In:

Raymond, EL, S Stein, and V Haley (2021). *Atlanta Region Eviction Tracker*. Tech. rep. Georgia Institute of Technology.

Reades, Jonathan, Jordan De Souza, and Phil Hubbard (2019). "Understanding urban gentrification through machine learning". In: *Urban Studies* 56.5, pp. 922–942.

Steif, Ken (2020). "Predicting vacant land price for the Philadelphia Land Bank". In:

Thomas, Tim, Anna Driscoll, and Gabriela Picado Aguilar (2020). *Displacement and Gentrification Typologies*. Tech. rep. Urban Displacement Project.

Uhler, Brian and Jason Sisney (2016). *Perspectives on Helping Low-income Californians Afford Housing*. Legislative Analyst's Office.

Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in neural information processing systems* 30.

Wu, Neo et al. (2020). "Deep transformer models for time series forecasting: The influenza prevalence case". In: *arXiv preprint arXiv:2001.08317*.

Zhou, Haoyi et al. (2021). "Informer: Beyond efficient transformer for long sequence time-series forecasting". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 12, pp. 11106–11115.

Zuk, Miriam et al. (2018). "Gentrification, displacement, and the role of public investment". In: *Journal of Planning Literature* 33.1, pp. 31–44.