

# Data Analysis and Machine Learning to Predict Property Risk in Buffalo

L by Lokesh Bollini

# Goal

This case study examines a comprehensive real estate analytics initiative for Buffalo, New York.

The project combines historical landmark analysis, code violation tracking, housing court data, and zombie home identification to provide actionable insights for local authorities, developers, and community organizations.

This serves as a strategic tool for revitalizing Buffalo's historical properties through data-driven investment decisions, preserving the city's rich architectural heritage while addressing contemporary housing challenges.

# Project Workflow



# Buffalo's Housing Challenges

## Historic Landmark Risks

- Lack of proper care and funding is damaging Buffalo's historic buildings.
- Quick action is needed to protect these important cultural treasures.

## Code Violations

- Many properties suffer from poor maintenance and code violations.
- These issues create unsafe living conditions and hurt neighborhoods.

## Zombie Homes

- Foreclosed homes left empty lower property values and destabilize communities.
- They can attract vandalism and criminal activities.



# What kind of Data?



**Demographic info**

- Country: United States
- City: Buffalo
- Address: 3 Fargo Ave
- Parcel ID: 99.84-3-14
- Neighborhood: West Side
- Latitude: 42.89804822
- Longitude: -78.08515232

**Image section**



**Historic District: West Village**

The West Village Historic District comprises a residential neighborhood developed around Johnson Park in the mid to late nineteenth centuries, and encompassing styles ranging from Italianate, Second Empire, Greek Revival, to French Gothic.

**Drivers and motivators**

- Opportunity to raise the overall market value by ~200K
- Improved overall block value
- Preserve an important historic property

**Code Violations**

- Exterior Walls
- Guards
- Overhanging Extension
- Rods and Drainage
- Stairways deck, porch & balcony
- Woods

**Cases**

- Housing
- Police
- Street/Sanitation

**Property Valuation**

Property Value

Value	Count
0 - 100,000	1
100,000 - 170,000	1
170,000 - 240,000	1
240,000 - 460,000+	1

Land Value

Value	Count
0 - 10,000	1
10,000 - 20,000	1
20,000 - 30,000	1
30,000 - 40,000+	1

Area

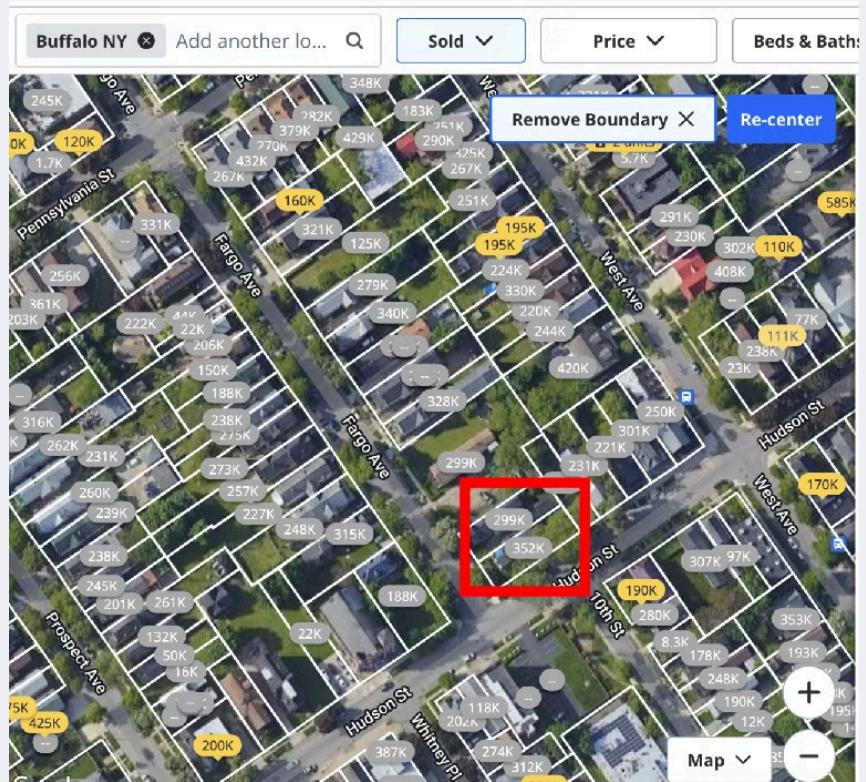
Value	Count
0 - 1,500	1
1,500 - 3,000	1
3,000 - 4,500	1
4,500 - 6,000+	1

I analyzed properties with significant code violations and legal cases to establish a benchmark for my "Risk" metric.

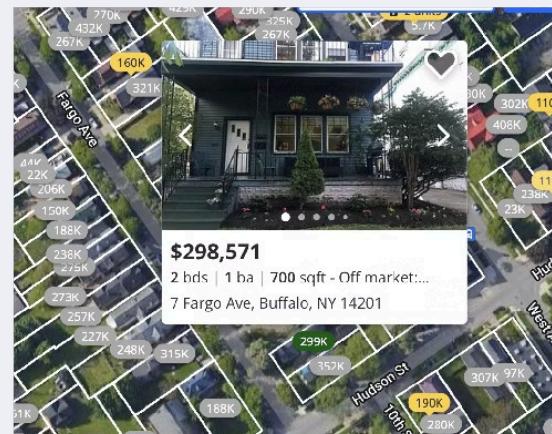
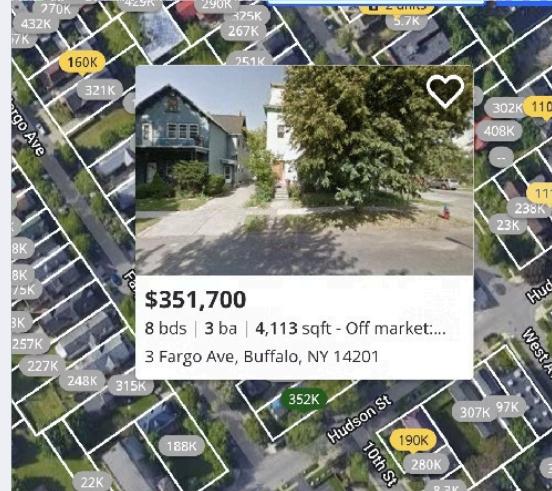
The risk level is determined by the severity of these code violations and other related factors. To ensure data integrity, I cross-verified the code violation information through visual checks on Google Maps.

Additionally, property valuations were calculated using data sourced from Zillow.

# Why?



Target Property - 8beds with less valuation



There's a notable price drop in 8-bedroom houses versus 2-bedroom houses—mainly due to a lack of renovation. This project targets these properties by categorizing them as High, Medium, or Low risk, with the goal of spurring neighborhood development and generating revenue.

# Data Sources and Collection

## Property Records

Assessment Rolls: SBL, address, owner, characteristics, location data.

Permits: Construction and property-related activities.

## Historical Data

Historic Landmarks: Location, features, and significance.

Historic Districts: Areas of architectural importance.

## City Services

311 Service Requests: Issues and code violations reported to the city.

Code Violations: Property and building code violations.

Housing Court Cases: Legal cases related to housing.

Conducted field visits and collected data from open source: <https://data.buffalony.gov/>

# Analysis Workflow



# Property Risk System

## Risk Factor Identification

The system calculates property risk based on multiple factors including mortgage status, code violation history, property values, and neighborhood characteristics.

## Feature Engineering

Combines BERT embeddings with quantitative features like 'Year Built', 'Case Reference' counts, '#ofBaths', '#ofBeds', 'GEOID20 blockgroup', 'Number of Units', 'Total Value', and 'CostPerSQFT'.

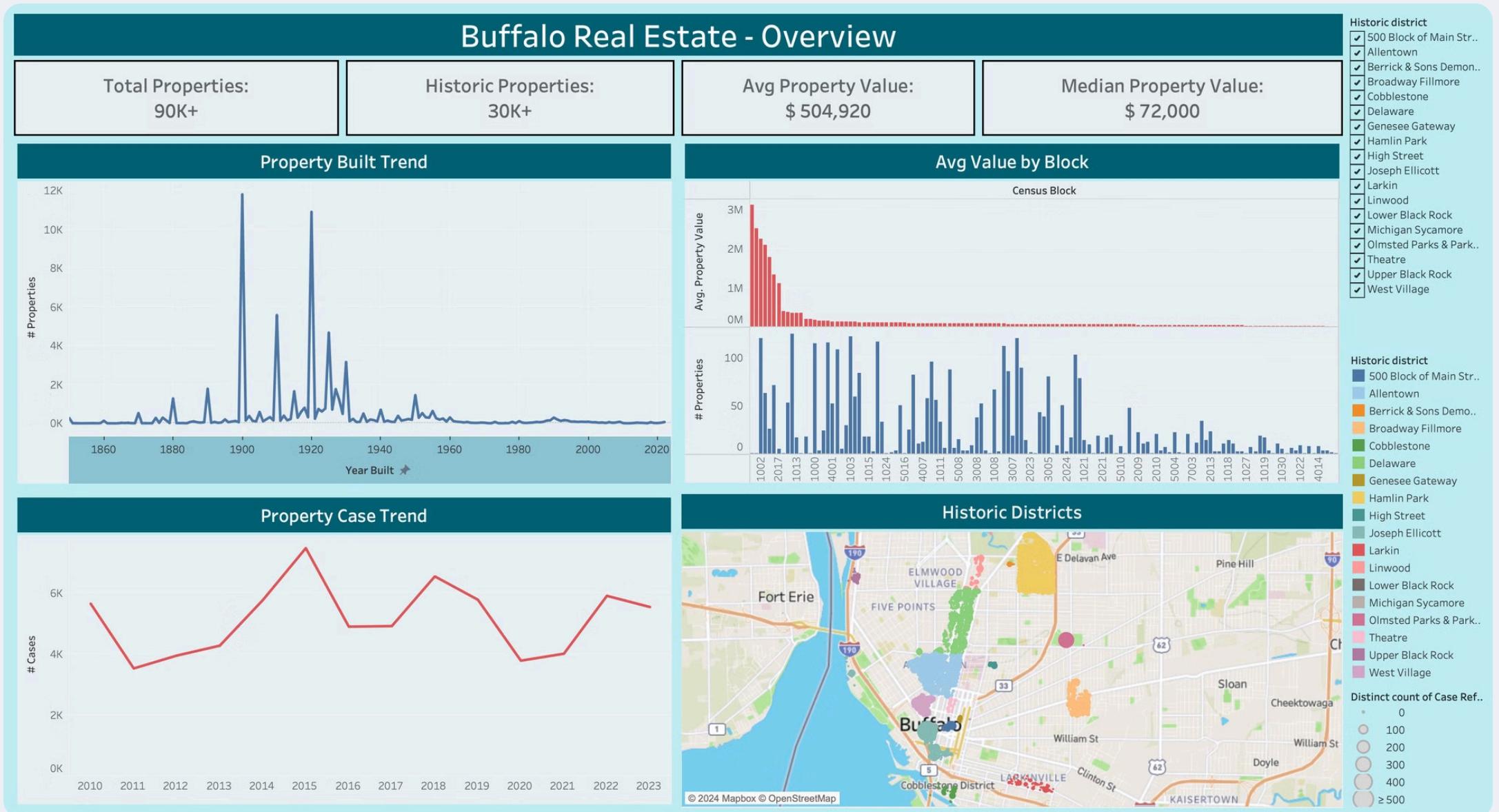
## Custom Weighting

Applies specialized weights to features such as 'Code\_Description', 'Comments', 'Prop Class Description', 'Neighborhood', and 'Dist Name' to reflect their relative importance.

## Machine Learning Model

Trains a Random Forest Regression model on scaled features to predict comprehensive risk scores for properties across Buffalo.

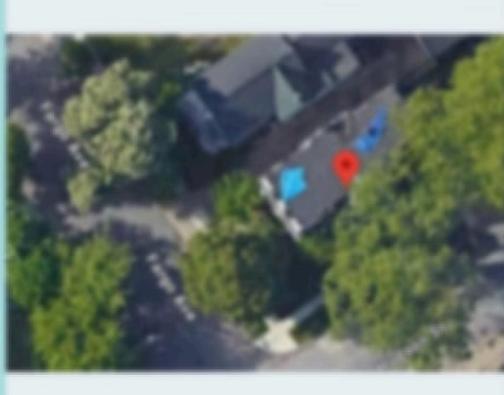
# Tableau Dashboards



# Business Use Case

## Property Profile - 3 Fargo

Total Value: \$230,000	Prospected Value: \$ 550,000	Year Built: 1890	Total Area: 4,113 sqft	Total Cases: 10	Total Violations: 8
---------------------------	---------------------------------	---------------------	---------------------------	--------------------	------------------------

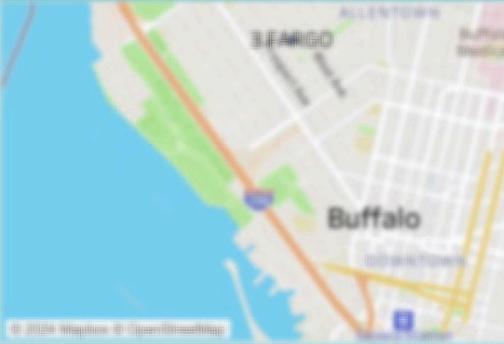


### Historic District

The West Village Historic District comprises a principally residential neighborhood developed around Johnson Park in the mid to late nineteenth centuries, and encompassing styles ranging from Italianate, Second Empire, Greek Revival, to French Gothic.

### Owners



Location	Violations	Case - Reasons
	<ul style="list-style-type: none"><li>Accumulation of rubbish or garbage</li><li>Exterior Walls</li><li>Guards</li><li>Overhang extensions</li><li>Protective treatment</li><li>Roofs and drainage</li><li>Stairways, decks, porches and balconies</li><li>Weeds</li></ul>	<ul style="list-style-type: none"><li>Animal Shelter</li><li>Forestry</li><li>Housing</li><li>Police</li><li>Streets</li><li>Streets/Sanitation</li></ul>

This property is identified as high risk, yet if renovated properly, it has significant potential for increased value.

Made with Gamma

# Tech Stack and Tools



## MySQL

For data mastering and complex joins on multiple datasets.



## Tableau

Used for creating custom maps and visualizations to communicate findings.



## Python

Used for data manipulation. Python libraries used include pandas, numpy



## BERT

For extracting contextual embeddings from text data to improve the property risk system.



## Random Forest

Machine learning algorithm used to predict risk scores.



## Confluence

For Project Management and Documentation



## Trello

Keeping track of project status and current pipeline



## Jupyter Notebook

For data exploration, analysis, and modeling.

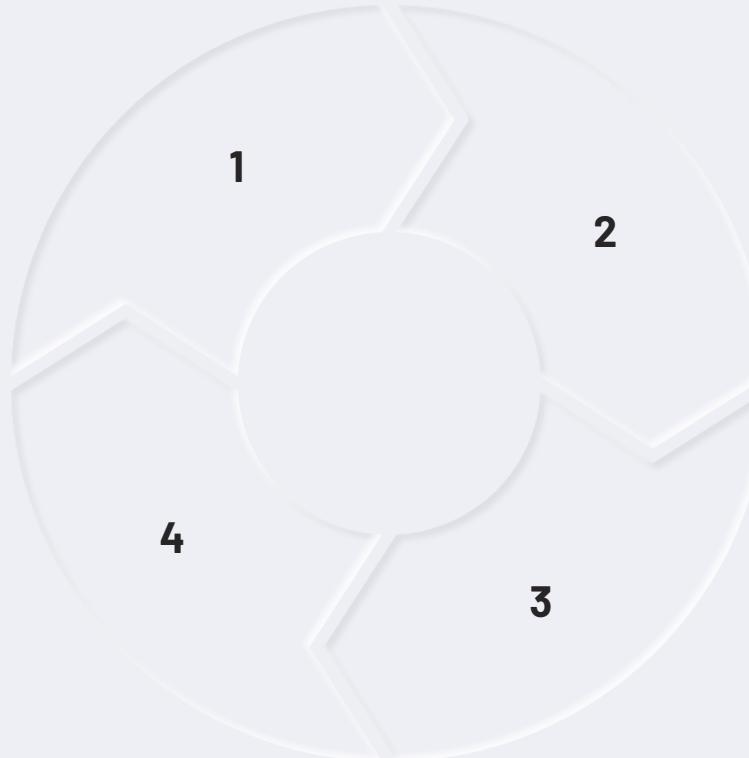
# Business Goals and Outcomes

## Historic Preservation

Identify and prioritize historically significant properties requiring immediate intervention to prevent deterioration.

## Value Creation

Generate economic and cultural value through the preservation and adaptive reuse of Buffalo's architectural heritage.



## Strategic Investment

Direct resources toward properties with the greatest potential for successful revitalization and community impact.

## Community Revitalization

Support neighborhood stability by addressing zombie homes and persistent code violations.

The expected outcome is a comprehensive property portfolio that guides investment in historic building stability, preserving Buffalo's unique cultural heritage while maximizing the impact of available resources.

# Conclusion



## Revitalizing Historical Properties

Goal: To revitalize historical properties in Buffalo.

## Informing Strategic Investments

Actionable insights for stakeholders.

## Preserving Cultural Heritage

Maximizing resource impact.

This project aims to revitalize historical properties, inform strategic investments, and preserve Buffalo's cultural heritage by providing actionable insights to stakeholders.