MGS 618: Experiential Learning

Third Estate Ventures: Zombie Houses

Team:

- **Rahul Gaur**
- **Mugdha Deshmukh**
- **Pratik Shetty**
- **Ojas Shinde**
- **Afua Kwakye-Berko**
- **Suakshay Bahal**

Problem Definition

Third Estate Ventures is an organization focused on community improvement. Through utilizing data to comprehend and resolve issues, the company improves living and health outcomes for those it serves.

Our team was brought on to assist the company predict zombie houses in Buffalo, NY. A zombie house is created when a mortgage lender starts the foreclosure process, prompting the homeowner to move out of the house. However, due to red tape, backlogs of other problematic properties, etc. – the foreclosure process was never completed. Thus, leaving an unoccupied and decaying property in the neighborhood.

Governor Cuomo said, "For many New Yorkers, homes are our single most important investment, but that investment can be undermined by the blight of neglected and abandoned properties (Lane)." Zombie houses can lower surrounding property values, possess mold and wild animals, and/or become home to squatters; potentially causing significant problems for a neighborhood. Additionally, zombie houses prevent anyone else from moving into and enriching a community. New York State has laws that hold mortgage lenders and servicers accountable for maintaining zombie houses. This legislation streamlines the rehabilitation, repair, and improvement of the vacant and abandoned zombie properties. If these entities are found to not be properly upkeeping these vacant and abandoned properties, they will face a civil penalty of $500 per day per property (Lane).

We have joined up with Third Estate Ventures to generate an algorithm that accurately predicts zombie houses. Although beginning in Buffalo, the hope is to be able to use the algorithm to provide relief to all communities and neighborhoods across the United States. Through being able to forecast these zombie houses, Third Estate Ventures can better hold mortgage lenders and servicers responsible for the law. Further, revitalizing, uplifting and empowering the communities the company serves.

How did we go about it?

Since we needed to predict that whether the house is Zombie or not, we have to train our Machine Learning model first & then predict which comes in the realm of supervised learning. We have approached the government organizations to give us the Les Pendens list so that we can use that to form our dataset for the machine learning algorithm. We have also used the code violations list

from the open buffalo dataset for the dataset purpose. We also did web scrapping to fetch the Taxpayer information which can be used as independent variable for the model as well.

Alpha -1 Design:

We began with analyzing the problem and decided to build the Logistic Regression Model since this involves the Supervised learning & the expected output is categorical which is binary. We have used Python in Jupyter Notebook for building the logistic regression model & since the output is categorical & binary, we are considering the output as binary: 1 for "Zombie" & 0 for "Not a Zombie". We improvised on the model with the inputs from the client through a scheduled weekly meeting & came up with the best independent variables as factors affecting the prediction. We have also generated the correlation matrix to investigate the dependence between multiple variables at the same time. We have also calculated the Variance Inflation Factor (VIF) for all the independent variables & the VIF values are good in order to proceed. Once our model was complete, we ran some set of 1159 data points from the year 2020 & predicted that only 27 data points are most likely to be the Zombie houses.

Also, we have built a Decision tree & random forest models to cross-check the most important attribute affecting the prediction for being the Zombie house.

Alpha -2 Design:

After predicting the data points for Zombie houses in 2020, we have showcased the visualizations of Zombie houses using tableau to get faster insights. We have shown the zombie houses in buffalo using Satellite streets so that it will be easier to locate the areas of zombie houses on a street map of buffalo. The houses can be filtered based on code violations and area wise zip codes. After visualizations, we found the following interesting insights about Zombie houses:
• The greatest no of Zombie houses in Buffalo are in the area - 'University, 14215'.
• Among all the code violations, 'IMPC-301.3' is most likely to be violated by almost all the Zombie houses.
• More than 70% of the Zombie houses are still in 'Active' status.
• Each house has multiple records associated with multiple code violations.

The street view of every house can be seen with the 'Google map street view' on the tableau. This will be helpful to get an idea of the locality before visiting the site.

Impediments faced:

Following are few of the impediments faced during the project:

1. **Problem**: The violation codes used as input variables were biased to one single code (IPMC 301.3) which meant that our model would be biased. Numerically, of our dataset of 84 Zombie House records, 80 of them were violating IPMC 301.3 (vacant structures and land).

   Solution: In our training dataset, we checked if our list of Zombie Houses were violating any other codes except the codes for abandonment. We found that there were 3 codes which were generally violated by these houses viz. IPMC-302.4 (Weeds on the property), IPMC-304.7 (Roofs and drainage), IPMC-308.1 (Accumulation of rubbish or garbage) which was an obvious maintenance failure for an empty house.

2. **Problem**: To predict if a house is Zombie, we need several input variables. The most challenging variable were the mortgage details which needed to be scraped from the website in time as an independent variable for the analysis.

   Solution: We built a scraping tool in python to automatically scrape the data based on the list of addresses that were provided in an excel format. This was an automated process as the user had to only upload the excel file and run the code.

3. **Problem**: The data was not in a readily consumable format as an input for the model and web scraping. For example, the addresses in the Les Pendens list didn't have a consistent format across the years. House No and streets were collated in some documents and in some they were not.

   Solution: We made the documents consistent in a single format and performed quality checks wherever it was necessary.

Summary:

This project was eye-opening as it exposed us to some of the challenges communities face. The use case for the algorithm we have created will prove to be extremely valuable to individuals, neighborhoods, communities, financial entities, and governmental bodies – among others. Third Estate Ventures will utilize our algorithm to ensure that mortgage lenders and servicers are doing the duties required of them by law. Through predicting which houses are most likely to be zombie houses, we can ensure that unoccupied houses are maintained, and the neighborhoods in which they reside in are sustainable for those who live in it. This project enabled us to gain experience in generating data models, matrices, algorithms, web scrapping codes, and Tableau visualizations.

Overall, it was a good learning experience for the entire team.