Title: Improved Property Sales in Milwaukee, WI
Name: Ethan Hood

Abandoned homes, vacant lots, and condemned buildings will often fall into city or municipality ownership. If these properties are not needed for public use, cities like Milwaukee may auction them off to private investors or individuals. However, many of these properties require expensive upkeep and remodeling before they are in any state to be sold. If these management costs exceed the final sale price of the property, then the city and by extension the taxpayers must pay the difference. The goal of this project is to understand what factors may control the final sale price of improved properties in order to make this a more profitable system.

The city of Milwaukee has released 4 years of improved property sales data through their OpenData portal which can be found here. Each dataset has 7 columns including 2 ids, a property type, dates acquired and sold, and expenses and profits associated with each property. I combined the datasets from each year and determined a continuous 'days owned' column from date arithmetic on 'dates acquired/closed'. Additionally, I created discrete 'year closed' and 'season closed' columns where March, April, and May correspond to Spring; June, July, and August correspond to Summer; September, October, and November correspond to Fall; and December, January, and February correspond to Winter.

I begin my analysis by mapping the Sale Price of each property to its Total Management Expenses as seen in **Figure 1**. For extra dimensionality and to observe possible trends, the Season Sold of each property is represented by a unique color. Each point is slightly transparent to better demonstrate the clustering of values at the bottom of the graph. In order to observe as much data as possible in the most useful frame, the x-axis is scaled by a logarithmic factor and two outlying points are dropped (one property that supposedly costed $100,000,000,000,000 to maintain and sold for $2,625 in the Fall and another that cost $43,580 to maintain but sold for $1,300,000). This graph also includes a dashed y=x line to provide reference; any point found above the line was sold for a profit and any point beneath was sold at a loss. In total, 1415/1897 or about 75% of the properties recorded were sold for a profit.

I next perform unsupervised learning on the data by imposing a principal component analysis on 4 numerical features (management expenses, sale price, year closed, and days owned) as seen in **Figure 2.** When first passing the data through a standard scaler, 3 of the 4 features contribute just under 25% of the explained variance with the final adding 26%. Without using this pipeline, however, a single feature is found to contribute almost entirely to the variance of the dataset.

But which of these features most influences sale price? **Figure 3** shows the results of performing a Linear Regression on several features determining closing price. To quantify season and property type, separate columns were first created for truth values for each season and property type was split into residential or non residential categories. The results demonstrate that 'days owned' and 'total management expenses' have a negligible bearing on closing price while residential properties greatly decrease sales. Properties auctioned off in Spring appear to sell for higher prices.

To Test the legitimacy of this model, I partitioned the data evenly into a train/test split. The model initially achieved an accuracy of around 5%. After shifting to a 75%/25% train/test split and first piping the data through a standard scaler, the accuracy increased to a max of 40%.

This project sought to identify the factors that could best determine the closing sale price of Milwaukee owned improved properties. However, selling improved properties is not as much of a profit seeking operation as it is a responsibility of the city. It's not unusual for properties to be sold at a loss out of necessity and consequently difficult to identify patterns in the available data that could improve profit as evidenced by the low performing models. Nevertheless, it was determined that this is still a largely profitable system as seen in **Figure 1** and management expenses do not have a significant bearing on the closing price of properties as seen in **Figure 3.** In the future, Milwaukee should seek to sell properties quickly and without expensive improvements.

Figures

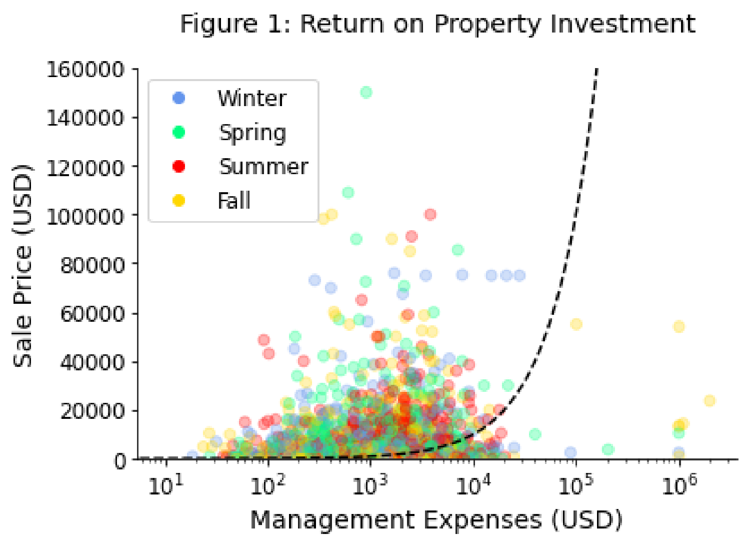

Figure 1: Return on Property Investment
Winter
Spring
Summer
Fall
Sale Price (USD)
Management Expenses (USD)
160000
140000
120000
100000
80000
60000
40000
20000
0
10^1
10^2
10^3
10^4
10^5
10^6

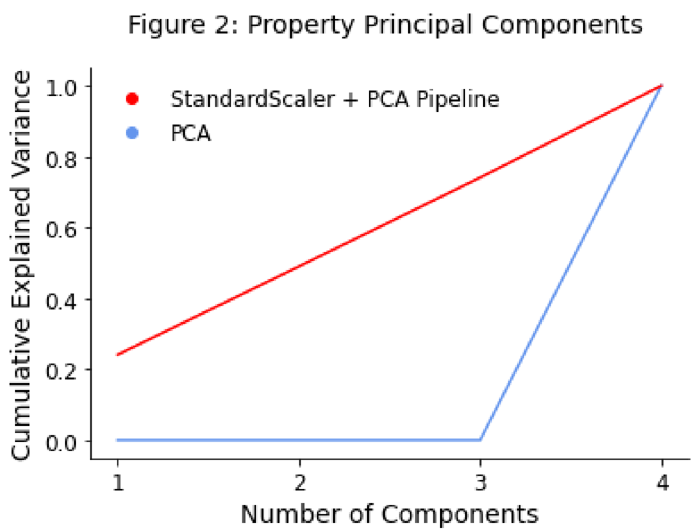

Figure 2: Property Principal Components
StandardScaler + PCA Pipeline
PCA
Cumulative Explained Variance
Number of Components
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4

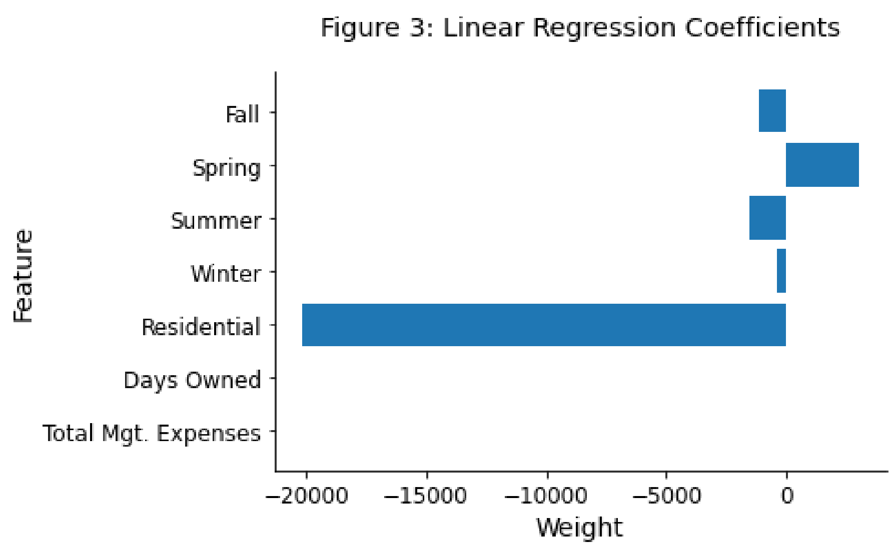

Figure 3: Linear Regression Coefficients
Fall
Spring
Summer
Winter
Residential
Days Owned
Total Mgt. Expenses
Feature
Weight
−20000
−15000
−10000
−5000
0