



Deliverable 3

Boston Permitting

TEAM F

Aidan Ruvins
Timothy Evdokimov
Zachary Gou
Andre Lesnick
Akhil Kokkula



Background, Motivations, and Goals

Extension Project Background

We have analyzed *why* certain areas have a lower or higher permit approval rate, based on economic data (i.e. economic indicators over time, property value data, etc.)

Motivation

We observed in our analysis of the base questions that certain areas were heavily skewed in their approval/denial in certain areas, especially those with a reputation for being gentrified. We want to look deeper into this trend and understand the economic forces that cause it.

Goals

1. What economic indicators have a strong correlation with high approval rates, or approval rates changing drastically over time?
2. What features of the data about a given neighborhood, especially economic or demographic data, are best predictors of approval/denial rate?



Extension Project — Sources

Choice of Data: We added three new sources to analyze our extension projects, all relevant to the economic factors that might affect building permit approval

Property Assessment Data: Dataset of a representative sample of buildings, with data on number of bedrooms, square footage, roofing material (?), etc

Economic Indicators: Time series data of economic indicators Boston-wide for the past decade

Zoning Subdistricts: Data on what parts of Boston are zoned for residential, commercial, etc

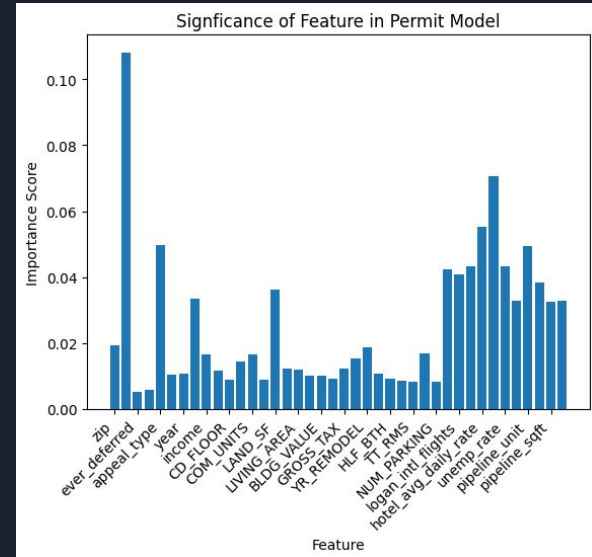
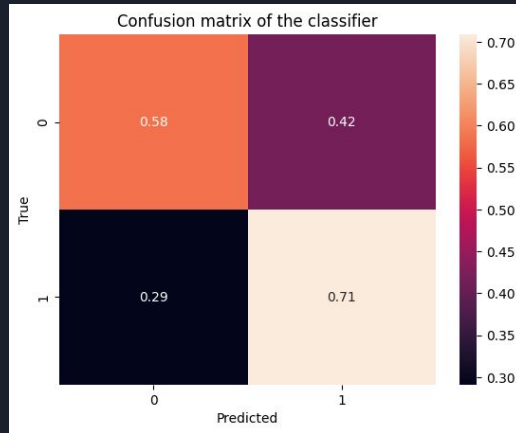
Cleaning the Data:

- Much data on economic indicators stopped being collected and was represented by “0” after the year collection ceased. This was filtered out in our data collection
- Property assessment data was mapped to permits based primarily on zip codes

Revised Model - Using Random Forest

Our “best” original model used a random forest classifier, inputting all the additional data and features yielded a balanced model with a marginal improvement in accuracy over the original one, with a 71% rate in correctly predicting denied permit but a lower rate of predicting approved ones.

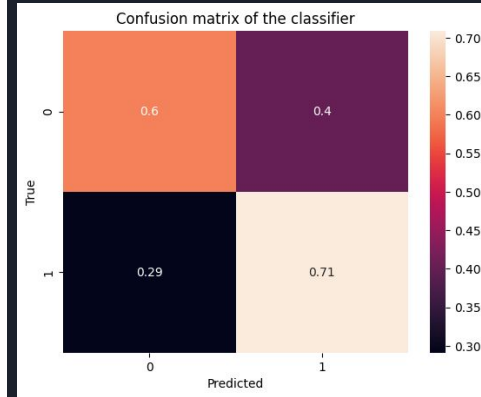
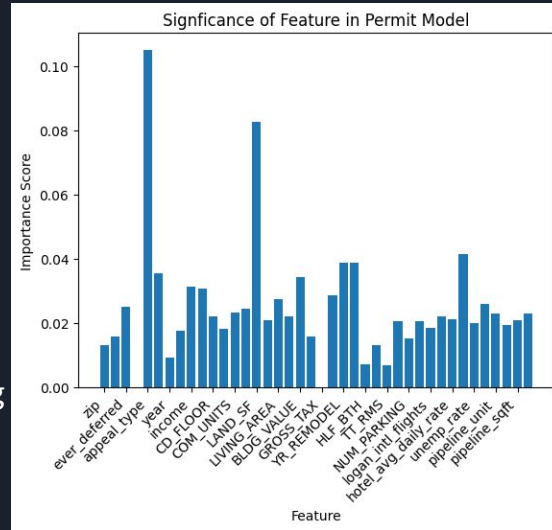
As before, the dataset we have is profoundly unbalanced, as rebalancing it produced a confusion matrix indicative of at least a somewhat functional model. The most significant features indicate the economic data is useful in predicting approvals/denials though



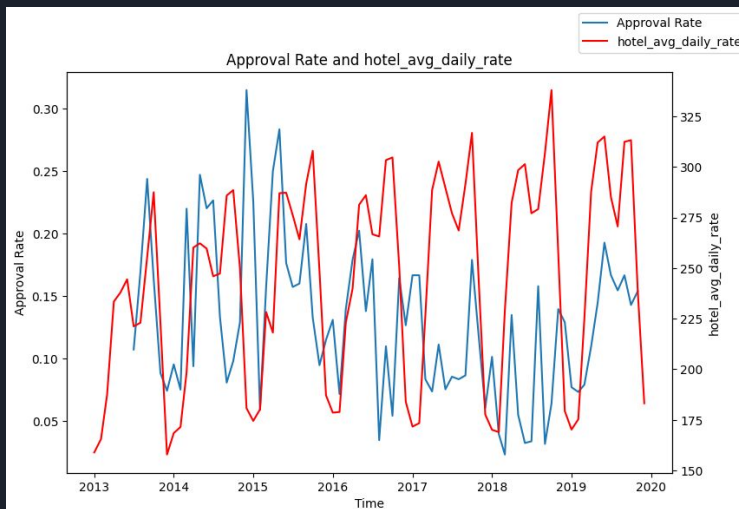
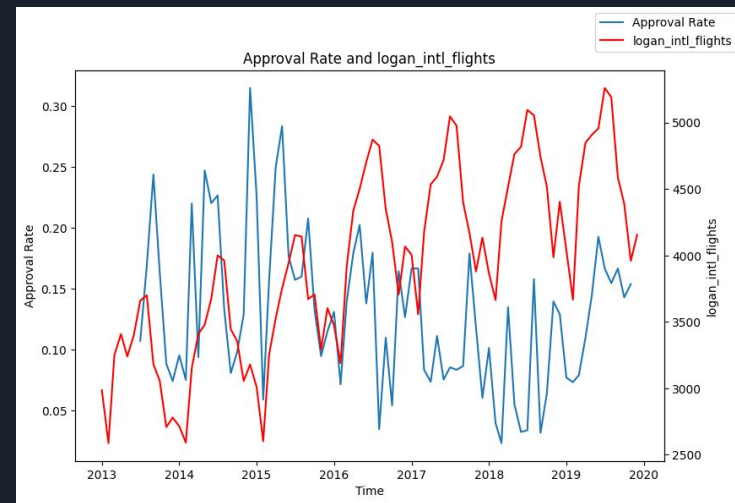
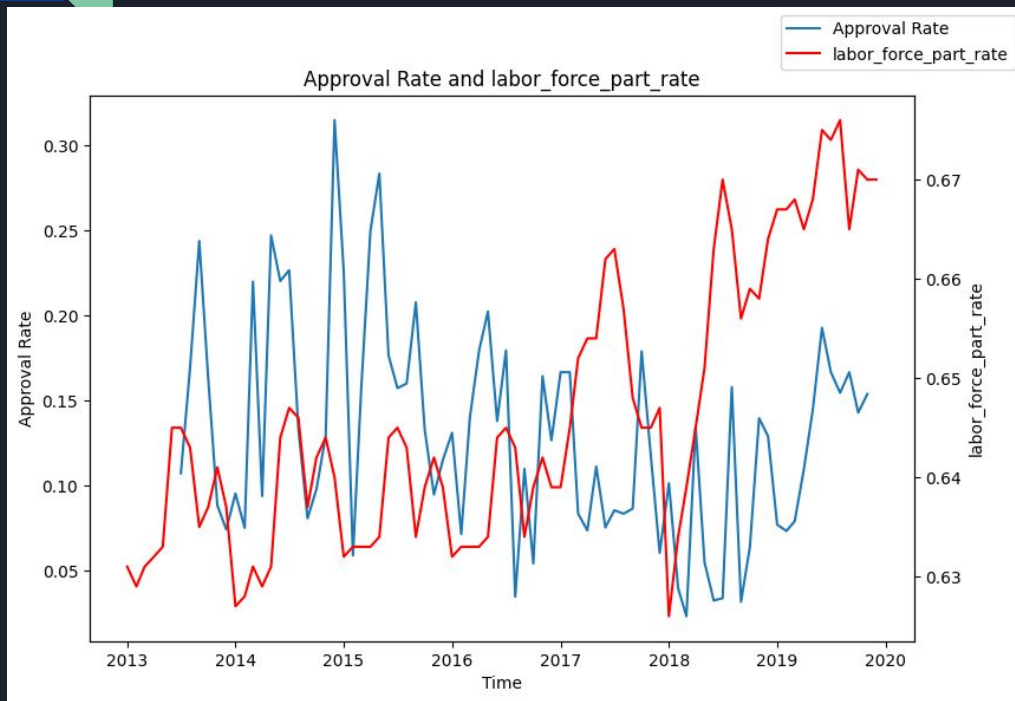
Revised Model — Using XGBoost

We also attempted a newer form of modelling using the XGBoost classifier, which proved effective, improving on the decision tree model very slightly, being about 2 percentage points more accurate and generally better at predicting approved permits than the tree model.

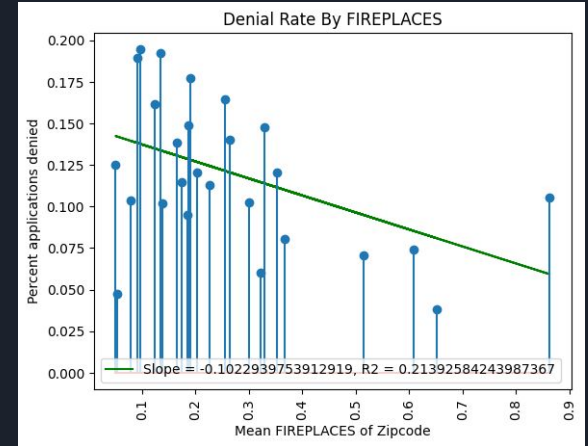
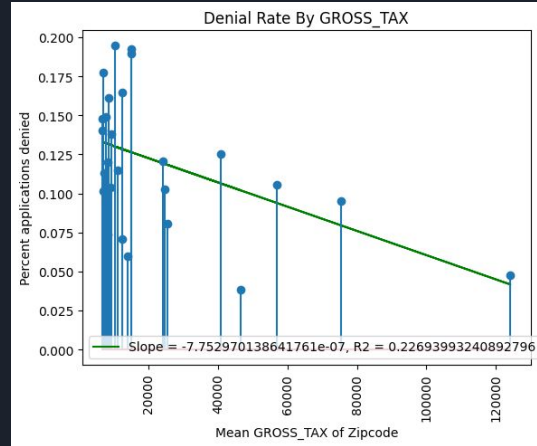
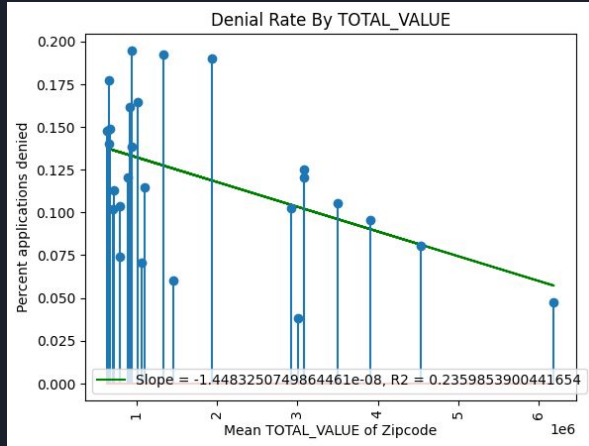
It also did a better job of highlighting important features, further suggesting that of the new data we used for our extension project, square footage, building value, and for some unknown reason, *hotel occupancy rate*, are predictive of permit approval



Notable Economic Indicators



Approval/Denial by Housing Feature



Left to Right: Total Land Value, Tax Paid on Property, Mean number of fireplaces per property



Individual Contributions

Aidan Ruvins: Helped work on the final report and the README file for deliverable 4. Helped work on the deliverable 3 slides

Timothy Evdokimov: Found extension project data, cleaned extension project data, and developed decision tree model and feature highlights.

Zachary Gou: Developed XGBoost model of permit approval process.

Andre Lesnick: Devised a system to find latitude and longitude information for permit appeals and other economic data by mapping to addresses in the MA voterfile

Akhil Kokkula: Helped create the deliverable 3 slides, started working on the final report, and the README file for the project