Property Assessment in Madison, Wisconsin

STAT 451 GROUP PROJECT

By: Ahsan Fawwaz, Faris Hazim, Imran
Iskander, Nick Elias, and Tyler Kelly
Group 16

Overview

Understanding Madison's property dynamics

Key Areas of Exploration

Tax Variations

Properties & Market Valuation

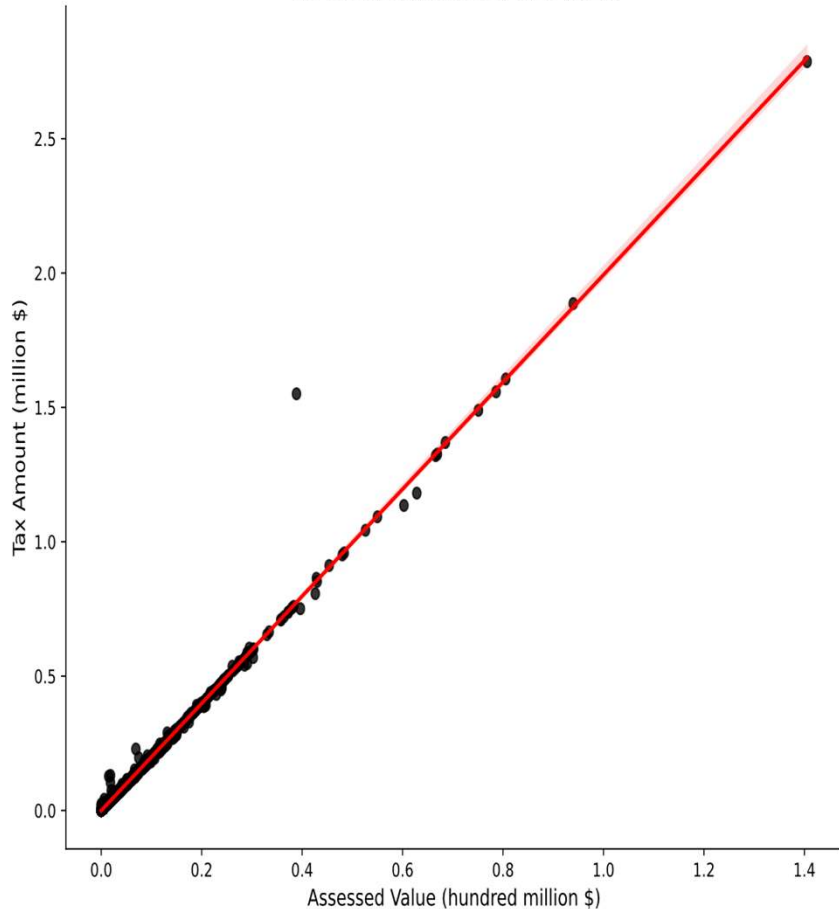Predictive Model Construction

Our Data

- Assessor Property Information (2022)

- Property Tax Rolls (2016 – 2022)

- Metro Transit Ridership by Stop (2022)

- Datasets obtained from City of Madison Open Data: https://data-cityofmadison.opendata.arcgis.com/

Heatmap of Property Values in Madison
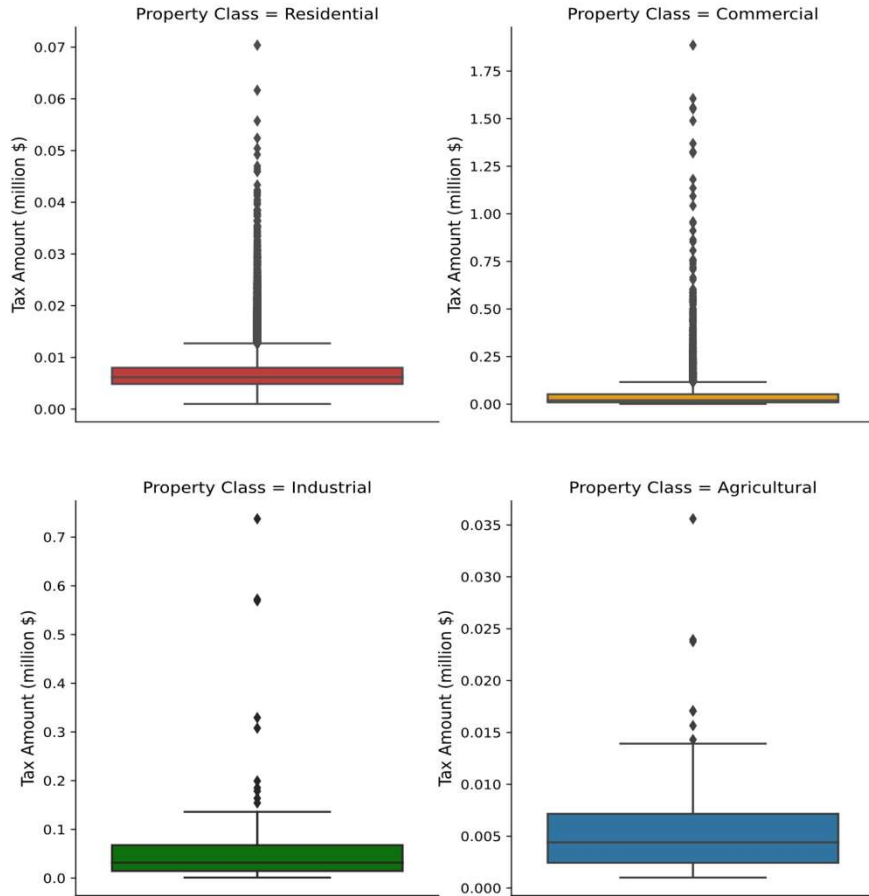
QUESTION 1 – TAXES

Tax vs. Property Assessed Value (2022)
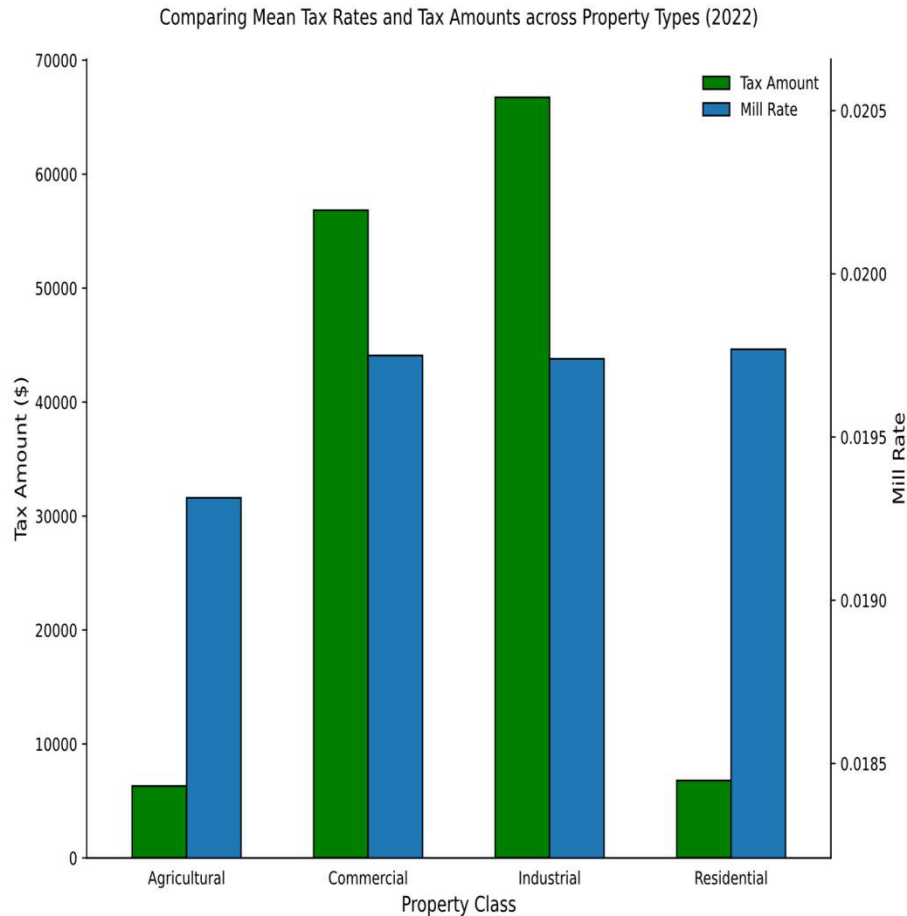
- **Does <u>property value</u> correlate with <u>tax amount?</u>**

- How do tax rates and tax amounts vary across different property types?

- What is the distribution of the amount of tax payments per property class (Residential, Commercial, Industrial, Agricultural)?

Variability of Tax Payments for each Property Class (2022)

- Does property value correlate with tax amount?

- **How do <u>tax rates and tax amounts vary</u> across different property types?**

- What is the distribution of the amount of tax payments per property class (Residential, Commercial, Industrial, Agricultural)?

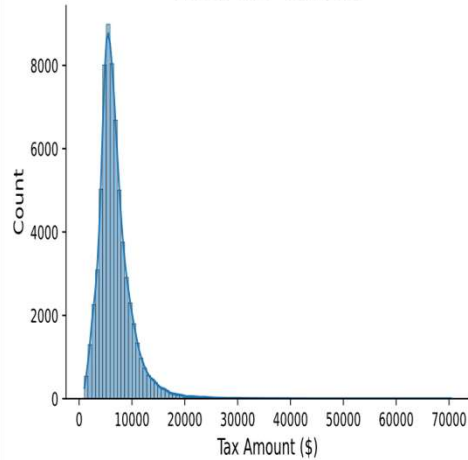Comparing Mean Tax Rates and Tax Amounts across Property Types (2022)

- Does property value correlate with tax amount?

- **How do tax rates and tax amounts vary across different property types?**

- What is the distribution of the amount of tax payments per property class (Residential, Commercial, Industrial, Agricultural)?
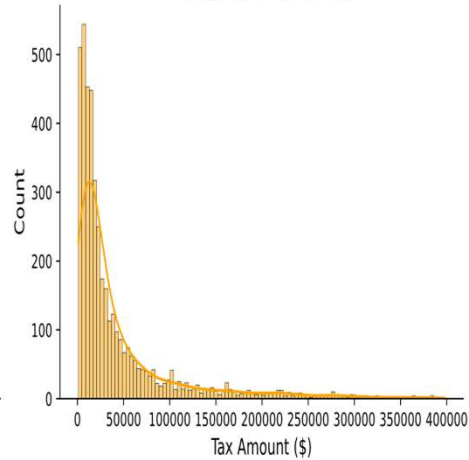
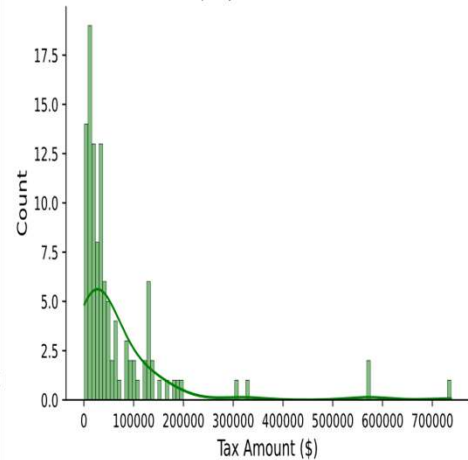Distribution of Tax across Different Property Types (2022)

- Does property value correlate with tax amount?
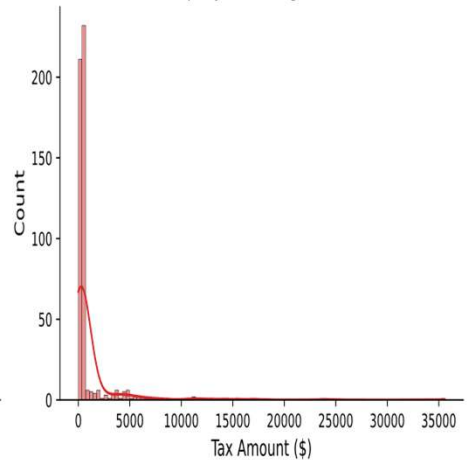
- How do tax rates and tax amounts vary across different property types?

- **What is the distribution of the amount of tax payments per property class (Residential, Commercial, Industrial, Agricultural)?**

Tax Classification for Residential Properties

```
For DecisionTreeClassifier(), the best score is 0.936
Best parameters: {'criterion': 'entropy', 'max_depth': 3}

For KNeighborsClassifier(), the best score is 0.928
Best parameters: {'metric': 'euclidean', 'n_neighbors': 5}

For LogisticRegression(), the best score is 0.921
Best parameters: {'C': 1, 'max_iter': 5000}

For SVC(), the best score is 0.929
Best parameters: {'C': 1000, 'kernel': 'rbf'}
```

- Binned tax amounts into 5 classes: Very Low, Low, Intermediate, High, and Very High

- Trained 4 different models to predict a property's tax class based on its value.

- Models compared: Decision Tree, k-NN, Logistic Regression, and SVM.

- **Found that <u>Decision Tree</u> (criterion: entropy, max_depth: 3) has the best score of 93.6% on validation data.**
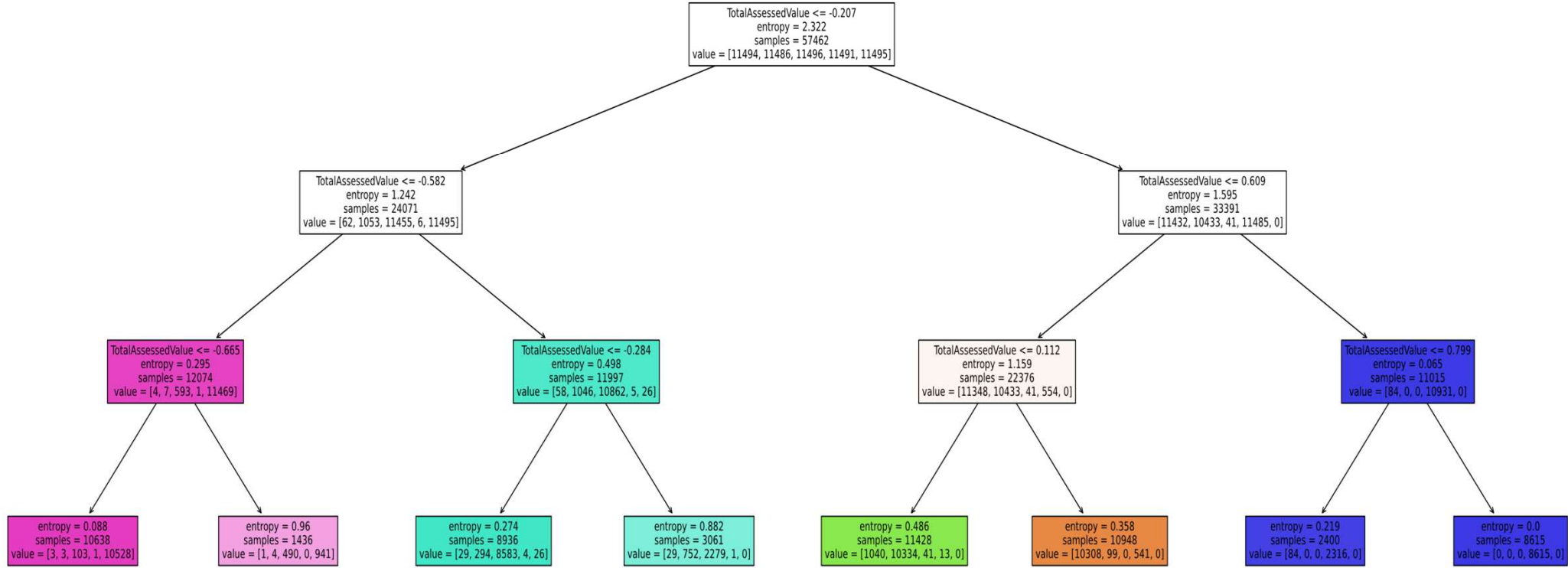
Decision Tree Diagram

QUESTION 2 –

PROPERTIES & MARKET VALUES

Average Assessed Property Value in Madison (2016-2022)
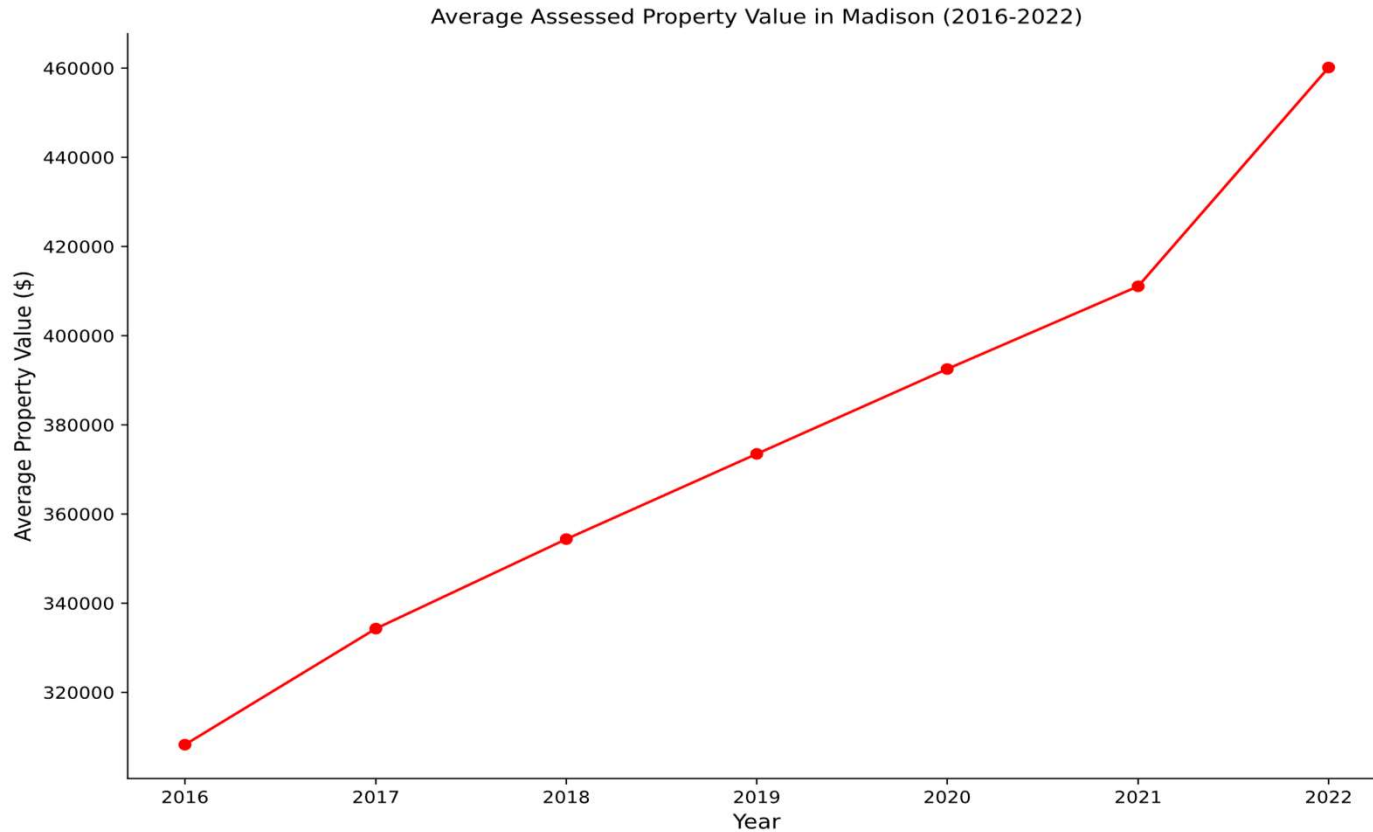
Average Property Values ('16 – '22)
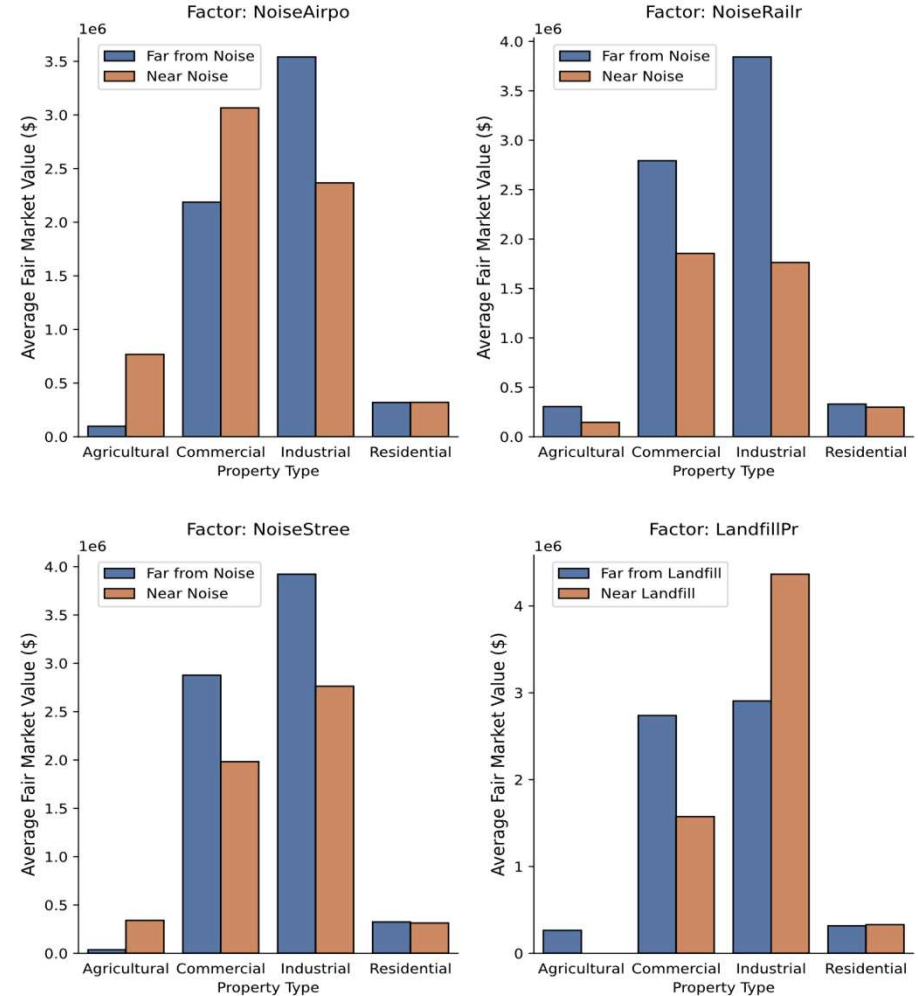
Environmental Impacts on Property Market Value

- Insignificant impact on residential properties

- In general, properties built far from noise factors are valued higher

- Industrial properties are valued higher near landfill

Environmental Impacts on Property Market Value (2022)

Classifying Property Classes
based on the Environmental Aspects

Flow:

- Random sampling of 30000 residential properties and merging back with other types

- 80% train/ 10% validation/ 10% test split

- Oversampling training data

- Hyperparameter tuning

- Model comparison: Logistic regression, SGD, decision tree, kNN, and stacking [SGD + kNN]

Result:

- **kNN** classifier with **metric = 'euclidean', n_neighbors = 2** obtained the highest accuracy on validation data

Classifying Property Classes
based on the Environmental Aspects

Evaluating kNN with best hyperparameters on test data:

Accuracy	Precision	Recall
0.834	0.766	0.834

Classifying Property Values (high/low) based on Physical Features

Data Setup:
Select physical features,
Clean/ merge columns

Feature Selection:
Find best alpha,
Filter out irrelevant features

Data Split:
80% training/
10% validation/
10% test

Model Selection:
Fit to decision tree/ kNN/ logistic regression

Ensemble Learning and Final Comparison:
Fit to gradient boosting,
Compare final accuracy

Classifying Property Values (high/low) based on Physical Features

Accuracy score on validation data:

Decision Tree	kNN	Logistics	Gradient Boosting
0.843	0.845	0.853	0.868

∴ Final model → Gradient boosting classifier

∴ Parameters → 'learning_rate': 0.1, 'max_depth': 9, 'n_estimators': 100

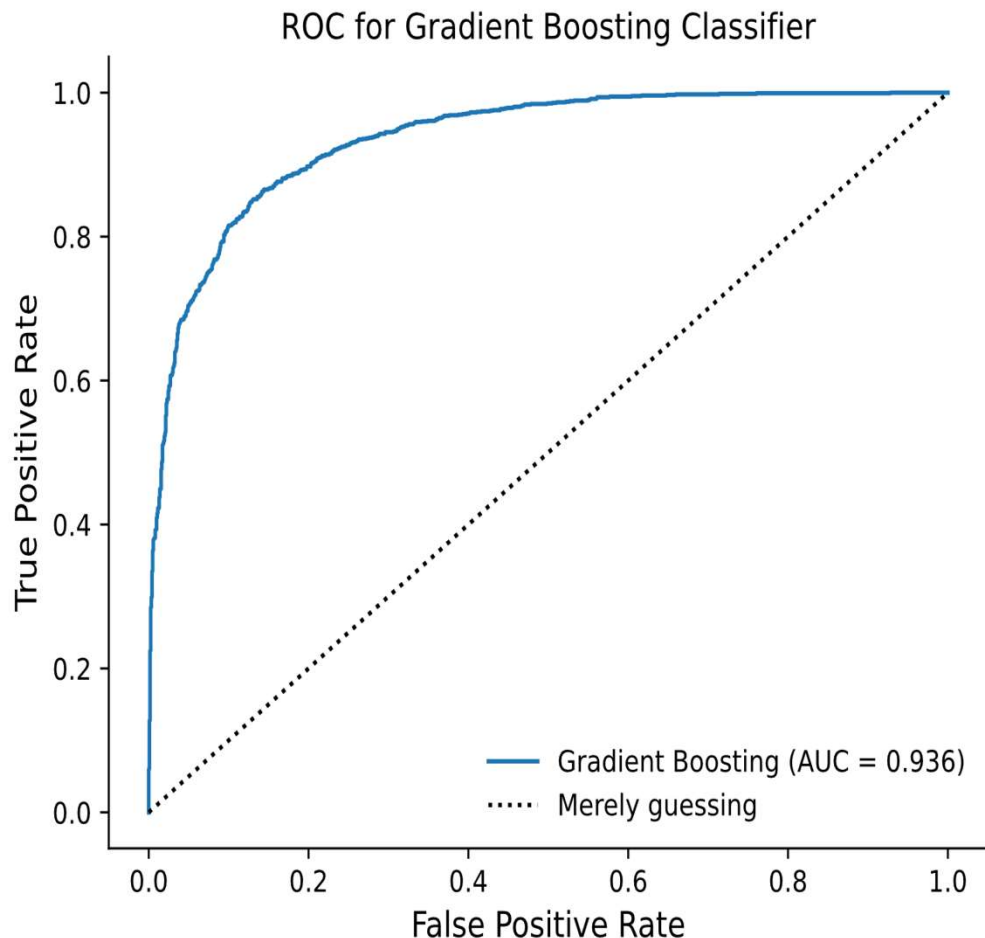# Classifying Property Values (high/low) based on Physical Features

How <u>good</u> is Gradient Boosting on test data?

Precision: 0.868
Recall: 0.852
Accuracy: 0.858
AUC: 0.936

ROC for Gradient Boosting Classifier

- Gradient Boosting (AUC = 0.936)
- Merely guessing

QUESTION 3 –

BRINGING IT ALL TOGETHER

Predicting Market Values for Residential Properties

Data Setup:
Use all features,
Clean/ merge columns

Feature Selection:
Find best alpha,
Filter out irrelevant
features

Data Split:
80% training/
10% validation/
10% test

Model Selection:
Fit to decision tree/ kNN/
linear regression/ SGD

**Ensemble Learning and
Final Comparison**:
Fit to random forest and
gradient boosting,
Compare final R^2

Predicting Market Values for Residential Properties
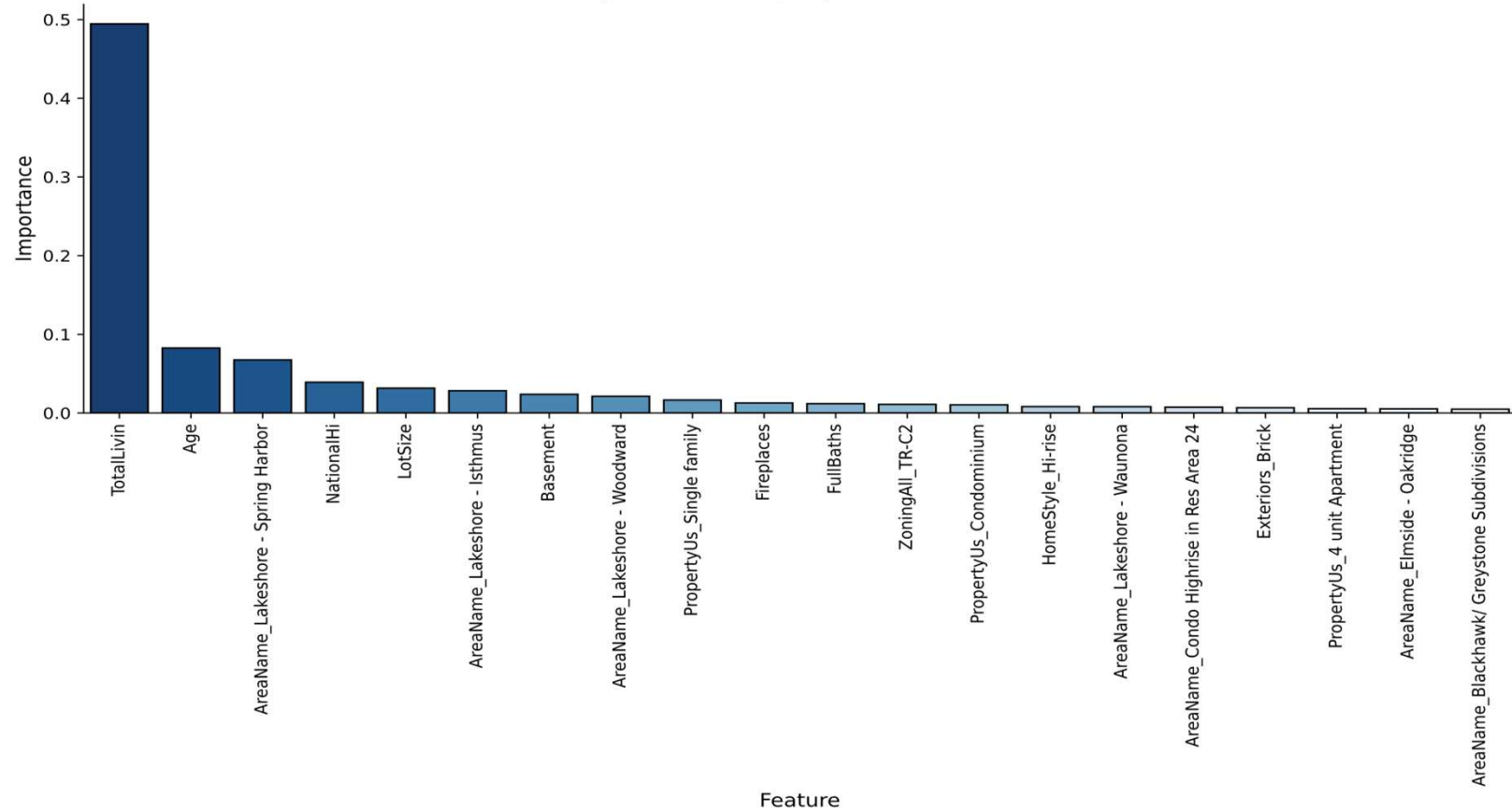
R^2 score on validation data:

Decision Tree	kNN	Linear Regression	SGD	Random Forest	Gradient Boosting
0.750	0.767	0.842	0.826	0.795	0.870

∴ Final model → Gradient boosting regressor

∴ Parameters → 'learning_rate': 0.1, 'loss': 'squared_error', 'max_depth': 9, 'n_estimators': 100

∴ R^2 score on test data → 0.881

Top 20 Most Important Features for
Estimating Residential Property Market Value in Madison

Feature Importance