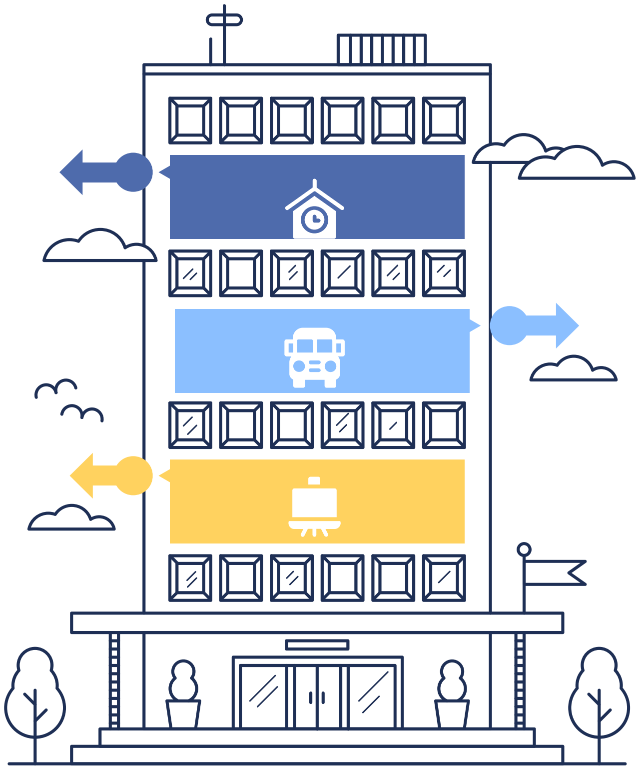
Sound Ordinance Permits in Austin Pt. 2

By: Eshi Kohli, Maadhav Kothuri, Daniel Lam, Nneoma Onochie, Greg Zachariah

Dataset Introduction

- Our data was sourced from the City of Austin open data portal and tracks **sound ordinance** permits in Austin, where entities request **permission to produce excessive noise**
- Our dataset had **67 columns** with **6,730 rows**, showing sound ordinance permit applications from **2009 to 2025**.
- Our data is directly **sourced from the city and updates everyday**, making it a credible and excellent source to use



Motivation + Research Question

What factors affect the speed of permit approval and permit subtype in Austin ?

- Sound ordinance permits affect local events, businesses, and city planning
- Understanding approval times could help applicants avoid delays and better allocate resources
- Exploring if approval time can be predicted using different factors
- Predicting subtypes could aid with resource prioritization

Our Data + Models

Data

- Used pre-processed data from exploratory analysis
- Computed the time difference for permit approval with the “IN_DATE” and “ISSUE_DATE”

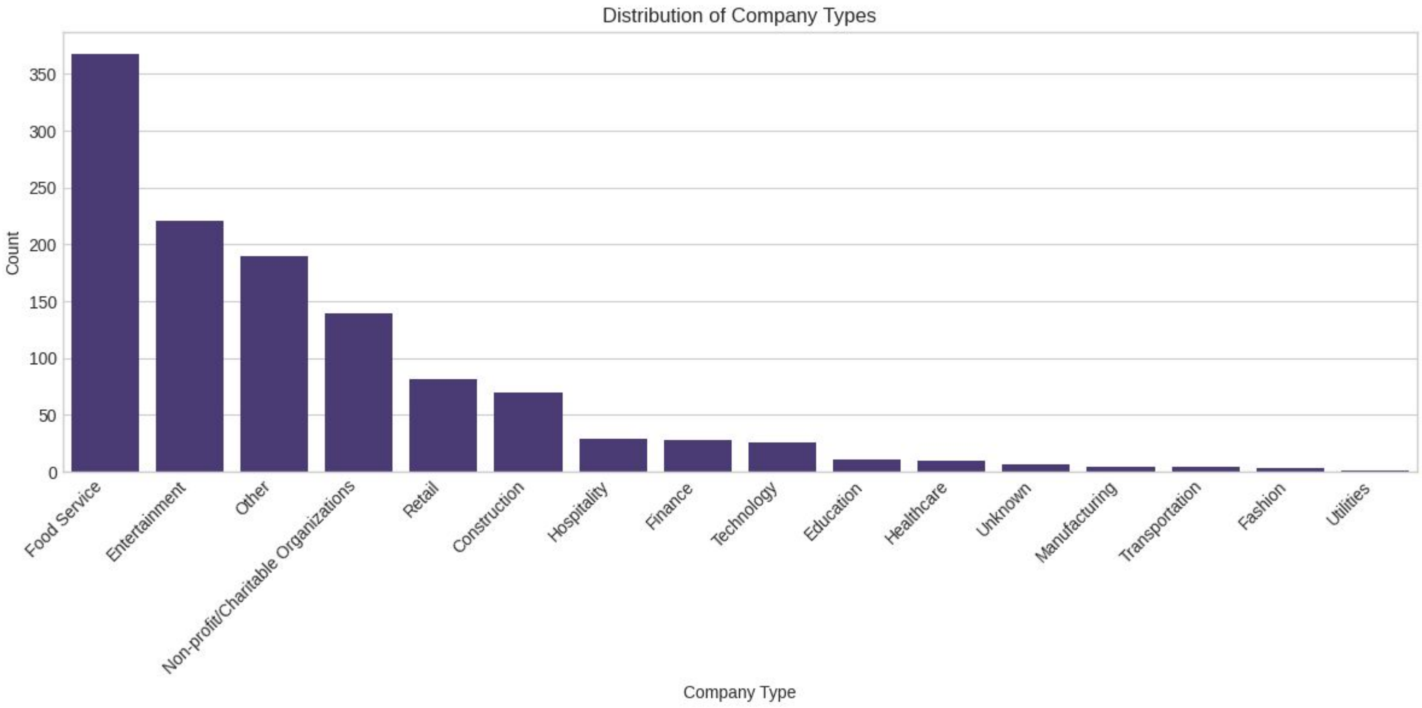
Models

- Linear/Polynomial Regression
- Random Forest
- K-Means Clustering

Factor 1

How Does Company Type and Subtype Permit Affect Application Times

Grouping Companies Into Generic Types



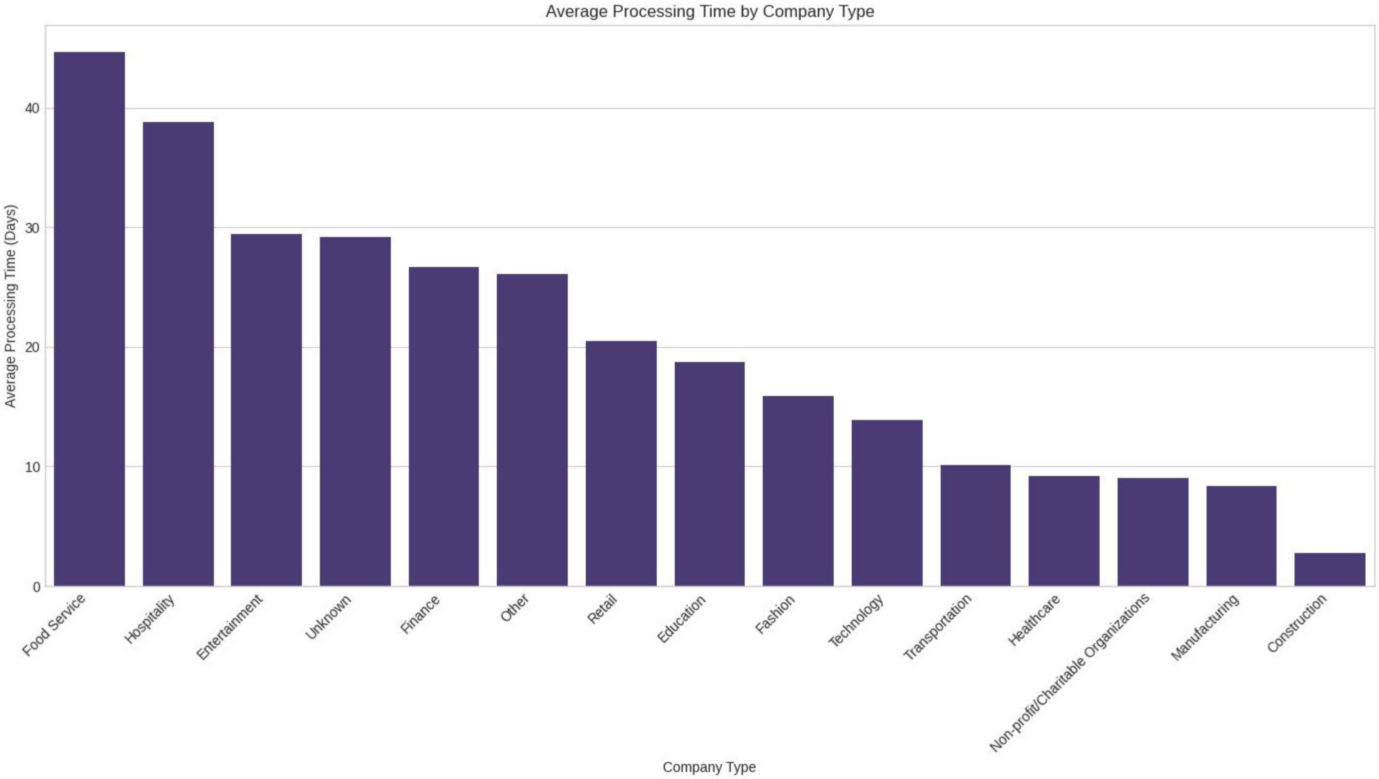
Adding Labels

Using **Gemini 2.5,** the LLM was fed the Company Names and was tasked to **label** the data itself into **Generic Types**

Combining Data

After labeling was done, I mixed the **Generic Labels** into a separate data set with **Processing Times** and **Subtype Permits**

Average Processing Times Per Company Type



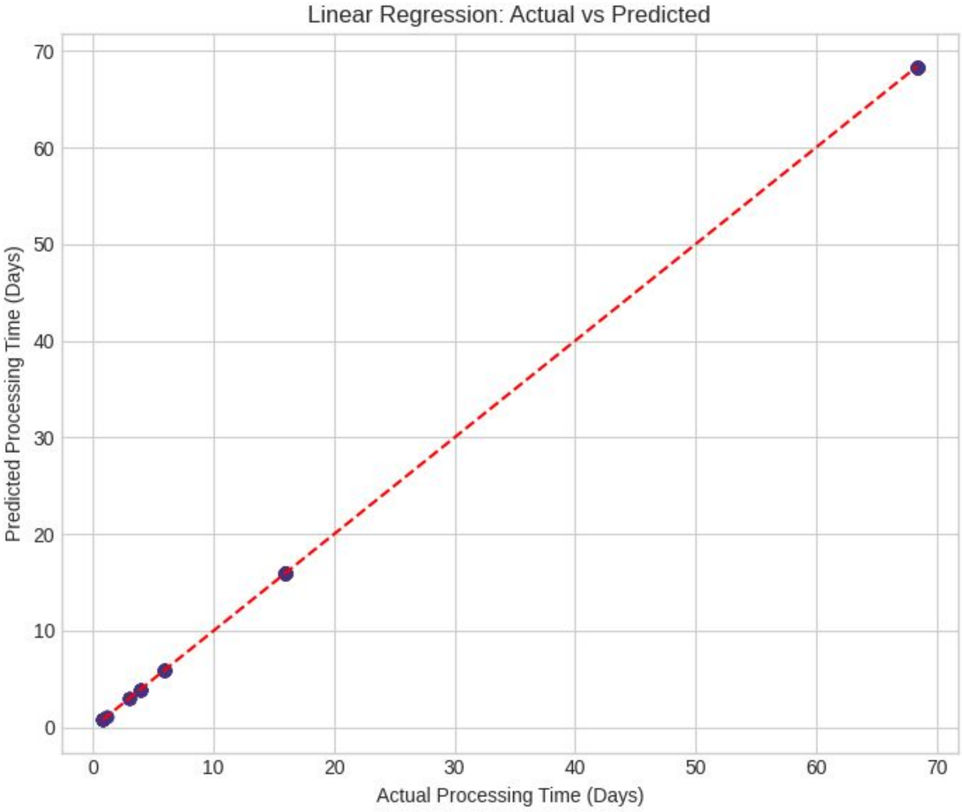
Generic Type x Time

This graph displays each generic company type with its **average approval time** for a sound permit:

Food Service and **Hospitality** being the **highest**

Construction and **Manufacturing** being the **lowest**

Using A Prediction Model



Training Model

I chose a **Linear Regression Model**, it was trained based on the **Company Generic Type**, **Sub Type Permits**, and **Average Time of Application**

Model Accuracy

The accuracy was test using **4 different methods**, Mean Square, Root Mean Square, Mean Absolute, and R^2. The model scored an average of **5% Error Rate**

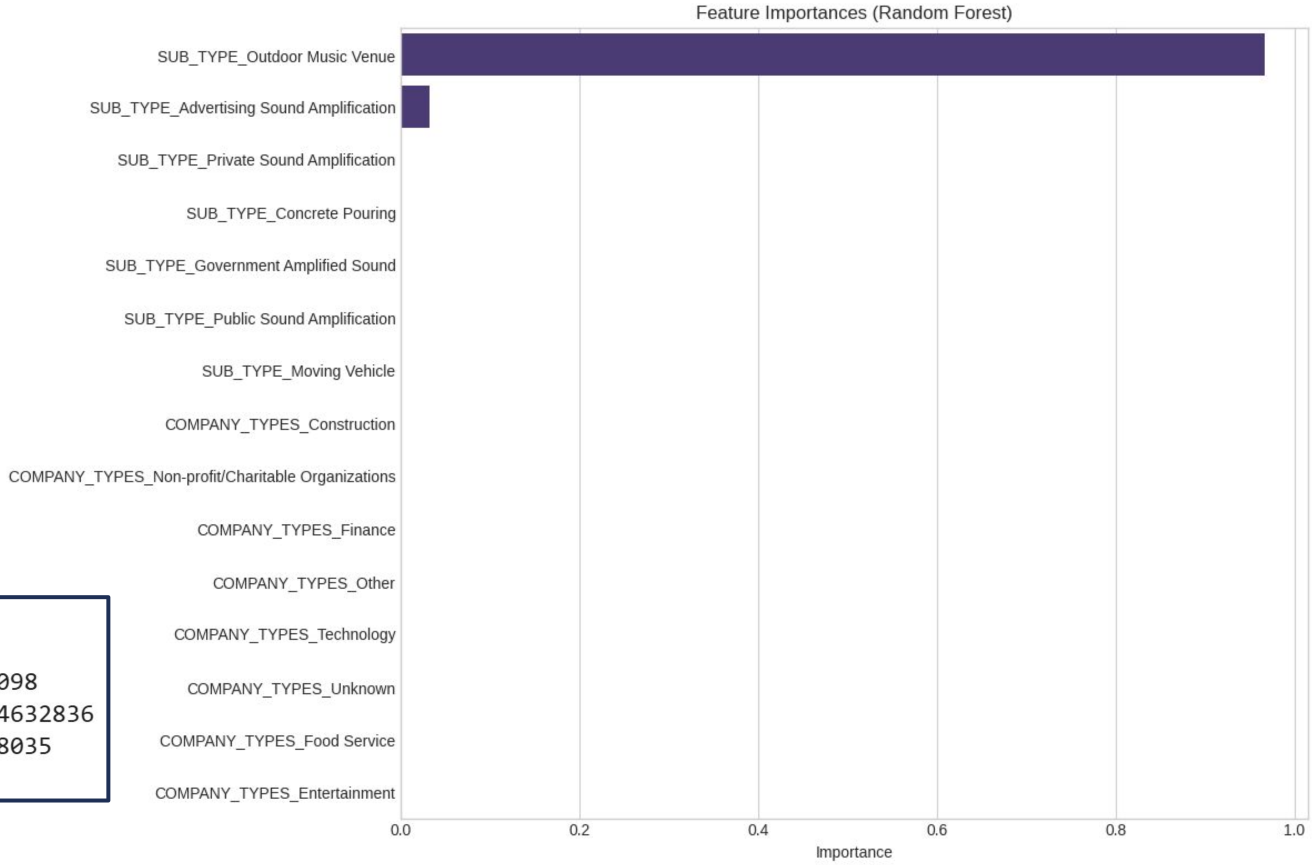
```
Training Linear Regression model...
Linear Regression Results:
Mean Squared Error: 8.172738523991911e-09
Root Mean Squared Error: 9.04031997442121e-05
Mean Absolute Error: 1.9597252664418304e-05
R² Score: 0.9999999999891184
```

Most Important Features

Calculating Importance

Using the **Random Forest Model**, it came out with the result of **Outdoor Music Venue** and **Advertising Sound Amplification** being the biggest factors for application speeds

```
Training Random Forest model...
Random Forest Results:
Mean Squared Error: 0.00019177930780396098
Root Mean Squared Error: 0.013848937794632836
Mean Absolute Error: 0.0017621251643018035
R² Score: 0.9999997446362289
```

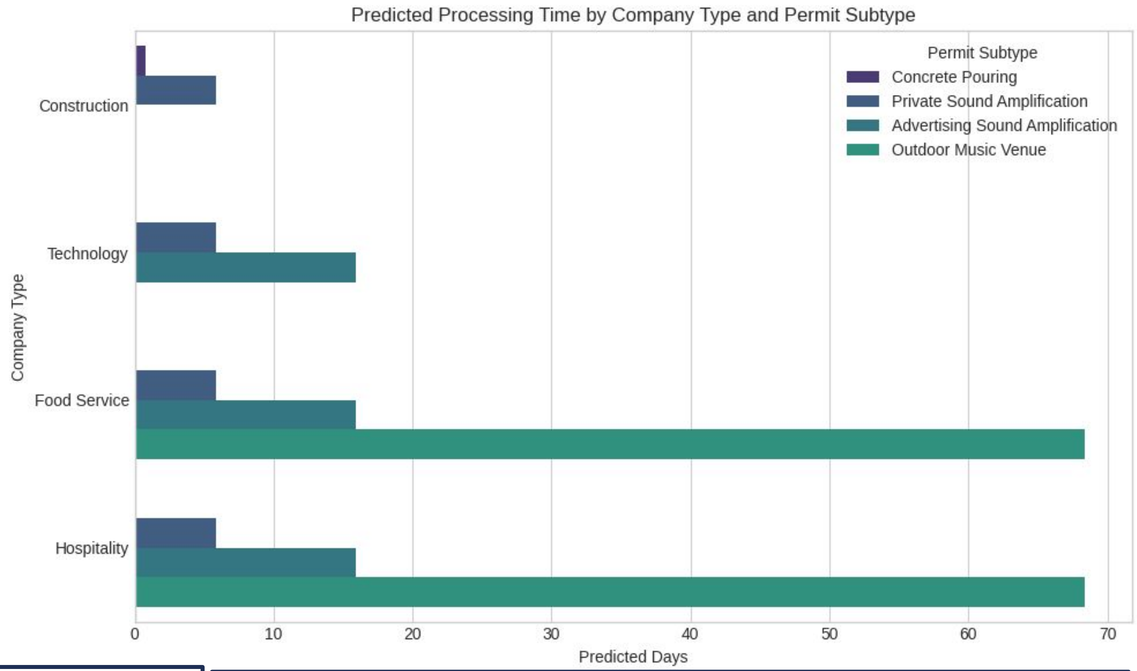


Predicting Processing Times

Testing Method

Using **randomized data**, I tested the prediction model to **calculate** the **average processing times** when inputting **generalized parameters** into the model

The graph displays the **biggest factors** to average application times, **Outdoor Music** being the **highest**




```
Predicted processing time for Food Service applying for Private Sound Amplification permit:
5.89 days (using Linear Regression)
Predicted processing time for Food Service applying for Advertising Sound Amplification permit:
15.89 days (using Linear Regression)
Predicted processing time for Hospitality applying for Outdoor Music Venue permit:
68.37 days (using Linear Regression)
Predicted processing time for Hospitality applying for Private Sound Amplification permit:
5.89 days (using Linear Regression)
Predicted processing time for Hospitality applying for Advertising Sound Amplification permit:
15.89 days (using Linear Regression)
```

```
Predicted processing time for Construction applying for Concrete Pouring permit:
0.77 days (using Linear Regression)
Predicted processing time for Construction applying for Private Sound Amplification permit:
5.89 days (using Linear Regression)
Predicted processing time for Technology applying for Advertising Sound Amplification permit:
15.89 days (using Linear Regression)
Predicted processing time for Technology applying for Private Sound Amplification permit:
5.89 days (using Linear Regression)
Predicted processing time for Food Service applying for Outdoor Music Venue permit:
68.37 days (using Linear Regression)
```

Conclusion

- Approval Times can be Predicted Based on **Subtype Permit** and what **Type of Company** is Applying
    - There is a bias towards **Construction** companies and **Construction permits** getting the **shortest approval times** (highest priority)
    - Any sort of **Bar** or **Food Service** wishing to do **Outdoor Music Venues** get the **longest approval times** (lowest priority)

Factor 2

What factors that affect permit approval time?

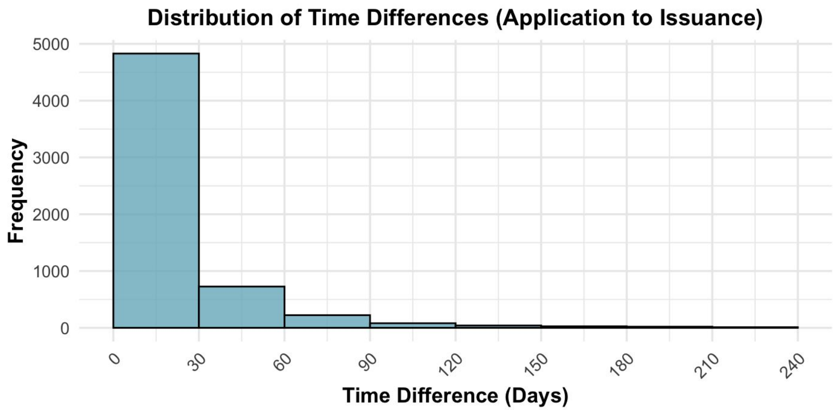
Model Selection - Linear Regression

- Approval times for sound ordinance permits typically take 0 days, but there are so many outliers, there may be different factors influencing time

- **Time_diff** (Issue date - Application Date)

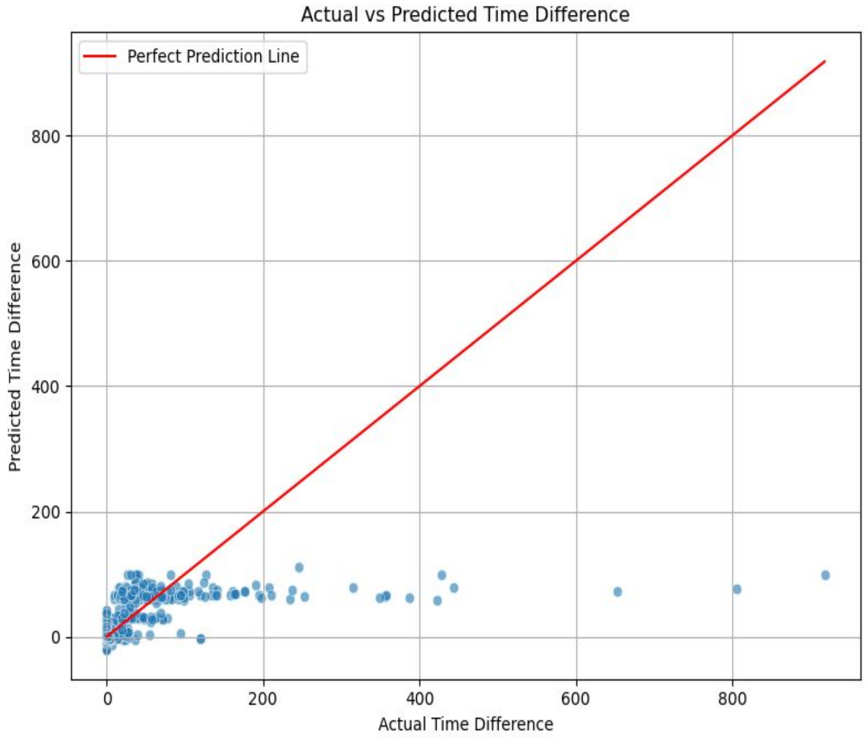
    - Target feature that Linear Regression Model would predict

- **Features Used:**
    - Event Year, Latitude, Longitude, Zipcode, Council District, Multiple Start Dates, Permit Subtype, If 51% of Permit's Establishment Depends on Food Sales, and its Amplified Sound District

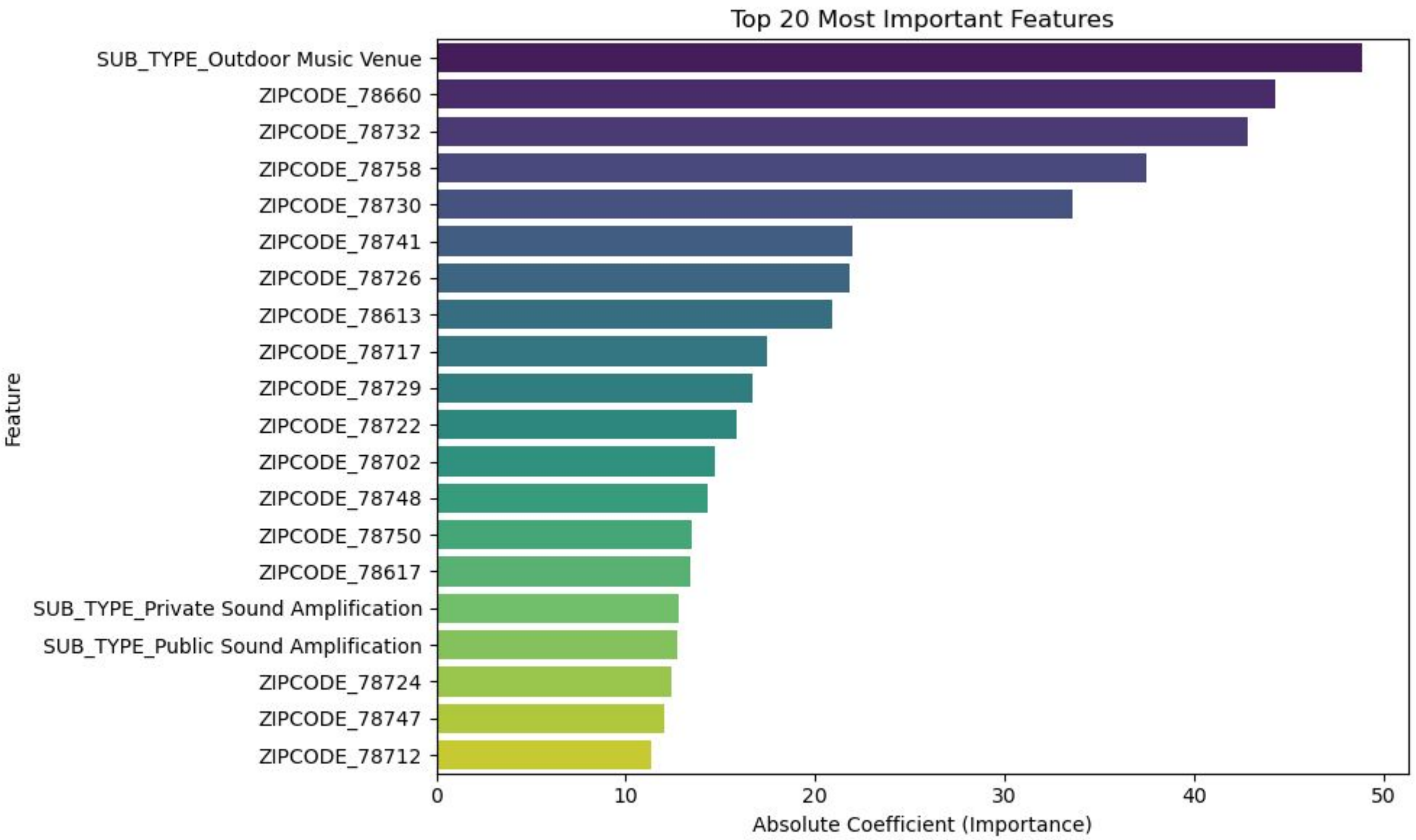


Linear Regression Evaluation and Insights

- Using these features, preprocessing them, and running it through a Linear Regression model, we used Root Mean Squared Error and R^2 Score
    - **R^2**: Linear regression model only explains a small portion of the variation of the model
    - **RMSE**: The model has an average prediction error of ~50 days
- **There is no strong linear relationship with permit features and issuance time**

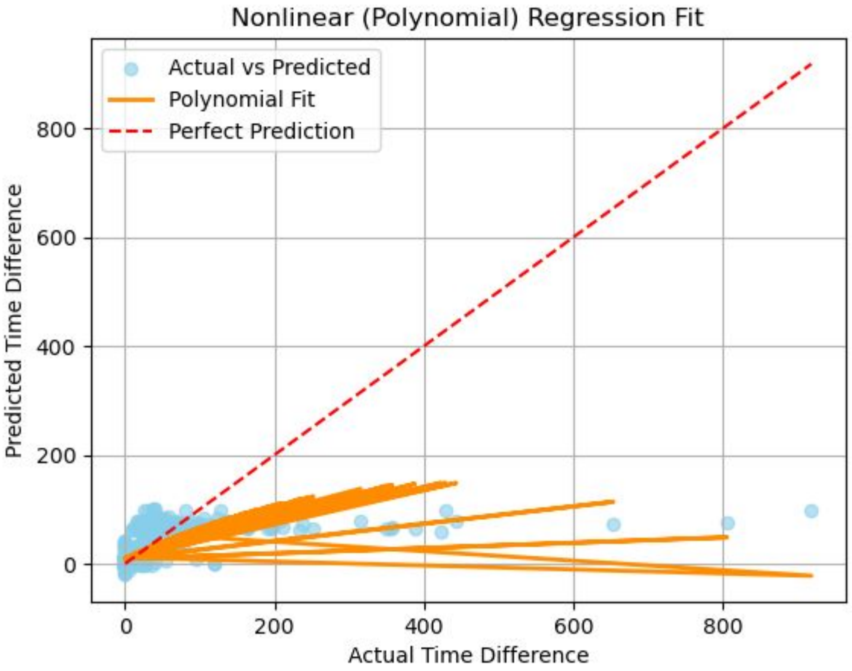


R^2 Score: 0.2761
Root Mean Squared Error: 49.8974


Top 20 Most Important Features
SUB_TYPE_Outdoor Music Venue
ZIPCODE_78660
ZIPCODE_78732
ZIPCODE_78758
ZIPCODE_78730
ZIPCODE_78741
ZIPCODE_78726
ZIPCODE_78613
ZIPCODE_78717
ZIPCODE_78729
ZIPCODE_78722
ZIPCODE_78702
ZIPCODE_78748
ZIPCODE_78750
ZIPCODE_78617
SUB_TYPE_Private Sound Amplification
SUB_TYPE_Public Sound Amplification
ZIPCODE_78724
ZIPCODE_78747
ZIPCODE_78712
Feature
0
10
20
30
40
50
Absolute Coefficient (Importance)

Model Selection - Polynomial Regression

- After looking at the distribution of the predictions, decided to look into **nonlinear models**
- **Polynomial regression** is a useful model for data that has a big rate of change change
- Because most of the permits were issued within a day, the nonlinear nature of data works best with polynomial
    - **R^2:** Polynomial model explains 50% of the variance of the data
    - **RMSE**: The model has an average prediction error of ~20.5 days (improvement from 50 days using Linear Regression)



R^2 Score: 0.4926
Root Mean Squared Error: 20.5864

Findings

- Many of the factors that go into how long sound ordinance permits take to be approved do NOT have a linear relationship, shown by the poor Linear Regression model

- Polynomial is not the best but much better than linear, accommodating for the nonlinear nature of the data

- In the future, we believe that log-based modelling might better fit the data

Factor 3

How do geographic features affect permit approval time?

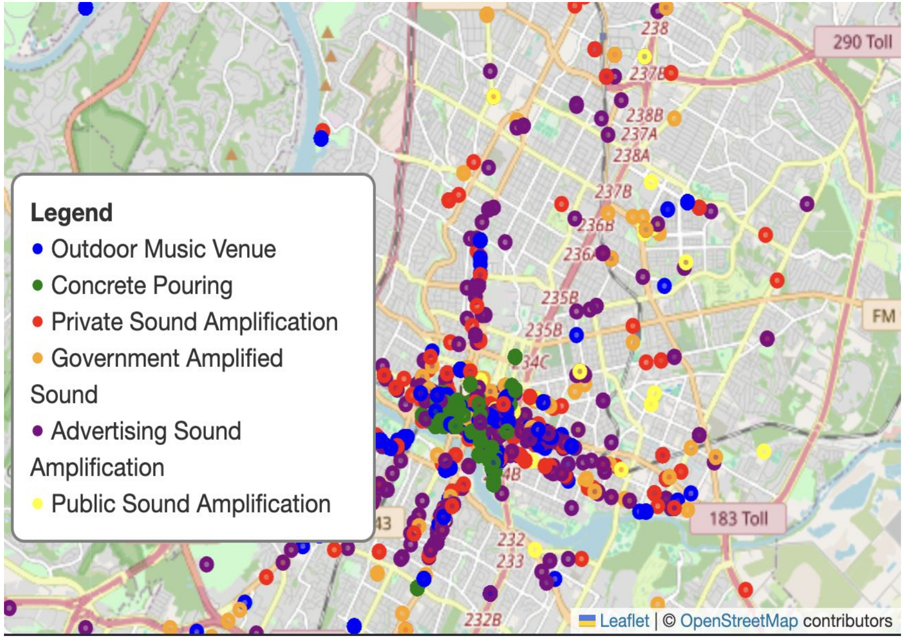
Model Selection (Attempt 1)

- **K-Means (k = 6 → one cluster per subtype)**
  - **Advertising Sound Amplification**
  - **Concrete Pouring**
  - **Government Amplified Sound**
  - **Private Sound Amplification**
  - **Public Sound Amplification**
  - **Outdoor Music Venue**
- **Motivation:**
  - Explore if certain regions of Austin have clusters of one specific subtype
  - Use geographic clusters to compare permit approval times
  - Utilizing the following geographic features:
    - Latitude + Longitude
    - Zip Code
    - Council District
    - Distance from Downtown

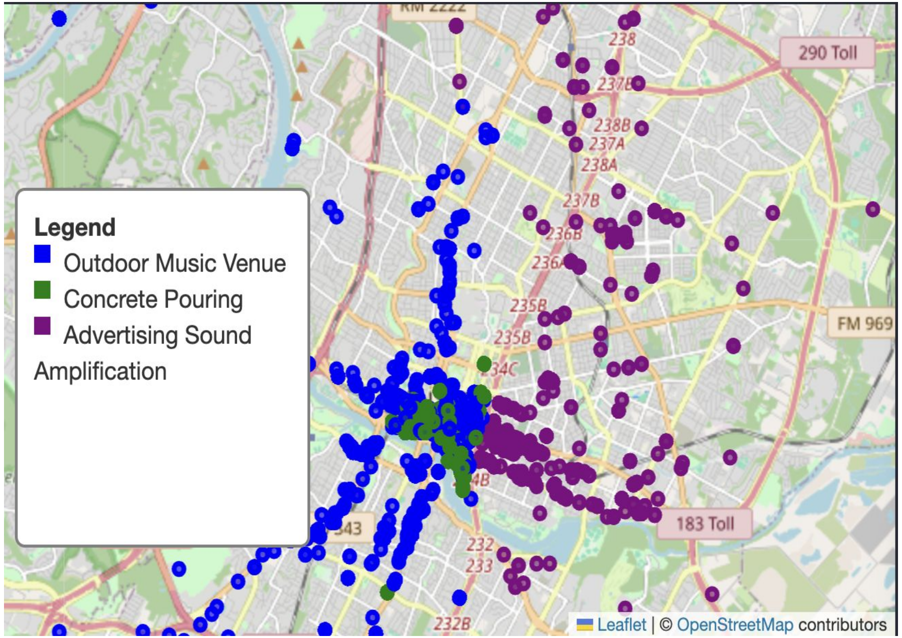
Model Selection (Attempt 1)

K-Means (k = 6 → one cluster per subtype)

Actual Subtype Distribution



Model Prediction



Evaluation (Attempt 1)

K-Means was not meaningfully predictive

- **Only identified 3 clusters (concrete, outdoor, advertising)**
  - These subtypes comprise the majority of our data, leading to greater majorities in cluster centroids
- High accuracy for **concrete pouring** and **outdoor music venue** sound permits
  - Explained by majority of concrete pouring permits being clustered in Downtown
- No cluster detection of government, **private**, and **public** sound permits

```
Accuracy by SUB_TYPE:
SUB_TYPE
Concrete Pouring                    100.00%
Outdoor Music Venue                  85.87%
Advertising Sound Amplification      42.97%
```

```
Accuracy by SUB_TYPE:
SUB_TYPE
Government Amplified Sound            0.00%
Private Sound Amplification           0.00%
Public Sound Amplification            0.00%
```

Model Selection (Attempt 2)

- **Random Forest**
- **Motivation:**
    - Unlike K-Means, Random Forest can model complex, nonlinear relationships between geographic features and approval time
    - More flexible than K-Means and allows us to explore relative importance of each feature
    - Utilizing the following geographic features:
        - Latitude + Longitude
        - Zip Code
        - Council District
        - Distance from Downtown

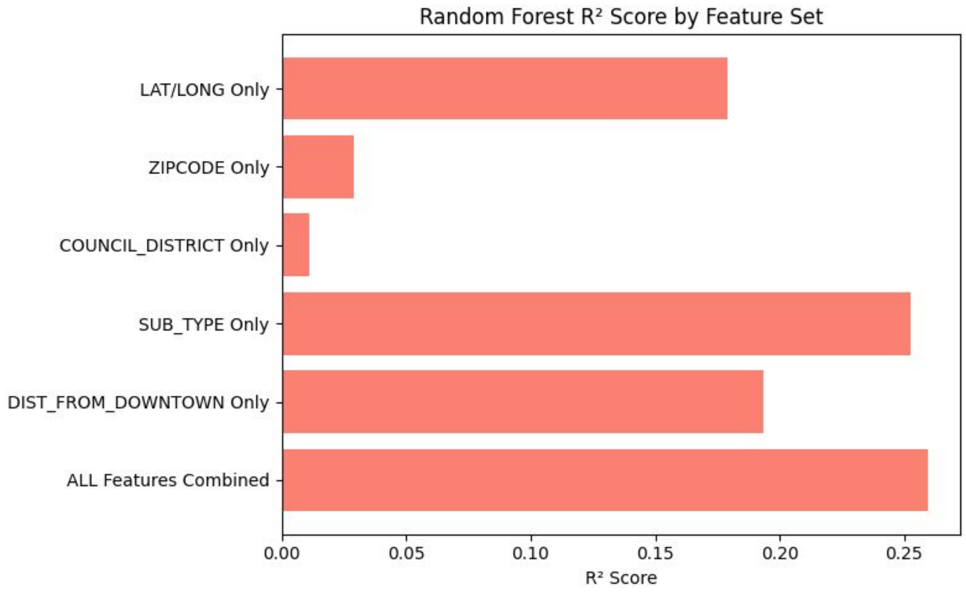
Evaluation (Attempt 2)

Random Forest better than K-Means, still poor performance

- **Best performing model:** all geographic features combined + subtype

- **Most predictive individual feature:** subtype
  - When compared against geographic features, subtype still takes the lead in determining approval times

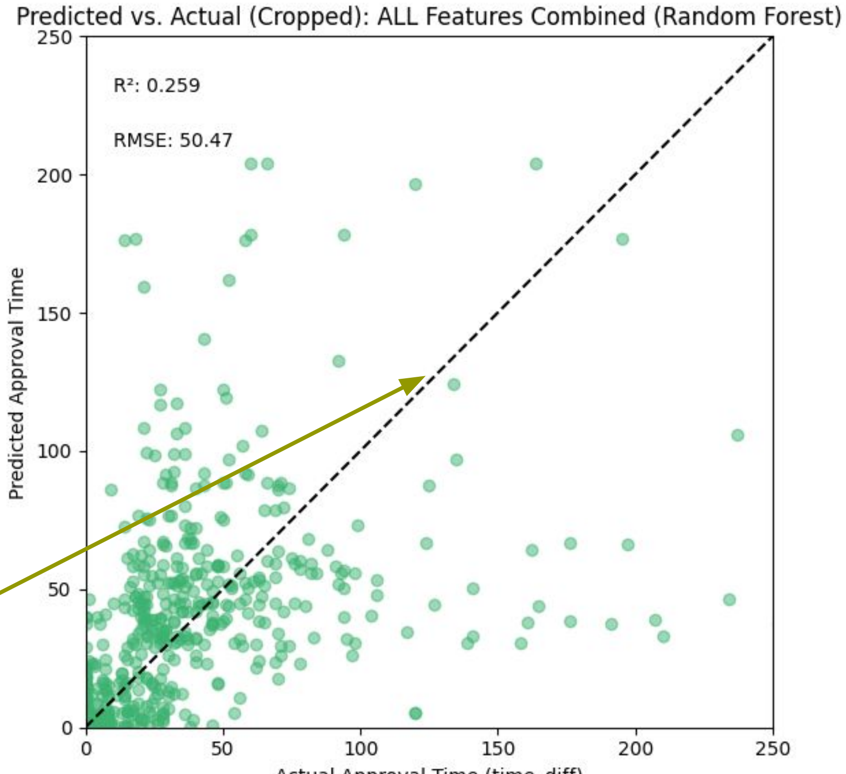
- All features combine explain only 26% of the variation in approval time



Evaluation (Attempt 2)

Random Forest better than K-Means, still poor performance

- Most points clustered near origin
  - Model is generally ok at predicting short approval time given geographic features (likely due to class imbalance of concrete permits)
  - More likely to underestimate approval time for permits that actually took a long time
- Noticeable scatter
  - Reinforces low accuracy and high RMSE



Dotted line represents perfect prediction

Findings

- Geographic features alone are not strong predictors of approval times
  - Approval time can vary regardless of where the sound ordinance permit originates from
  - There is no clear “bias” towards approving sound permits coming from one specific area in Austin

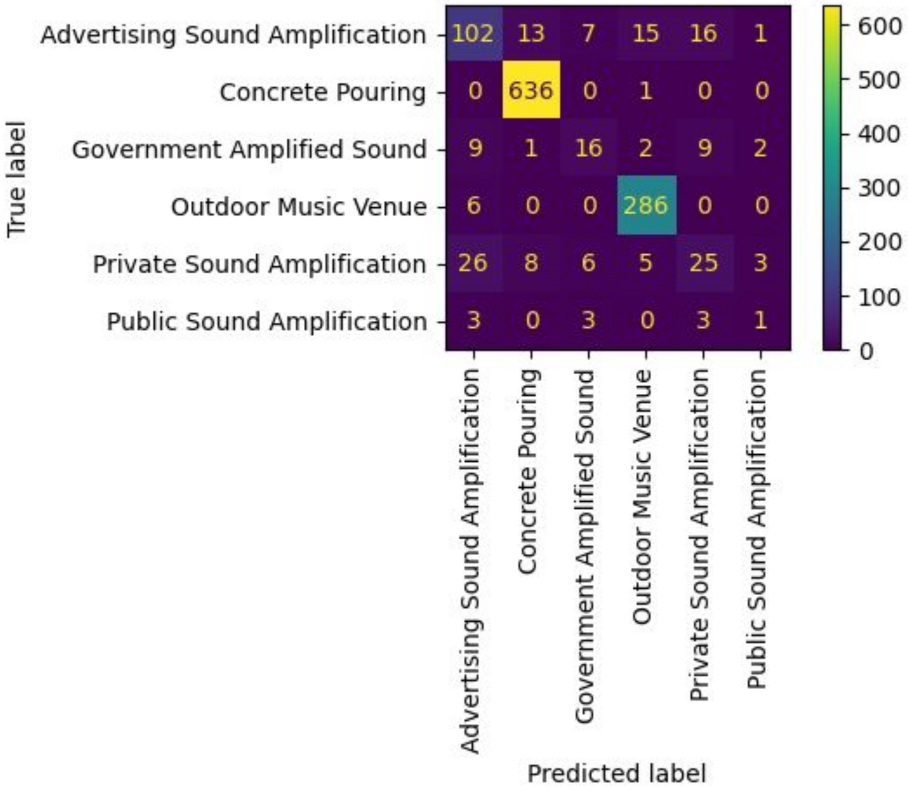
Factor 4

What Factors Can Be Used to Predict Subtype?

Model Selection - Random Forest

- As mentioned earlier, random forest is very flexible and helps us model nonlinear relationships
- Allows for exploring feature important while handling categorical and numeric data
- Imputed missing numerics with median and missing categorical with most frequent
- Features to be Included:
    - Latitude + Longitude
    - Zip Code
    - Council District
    - Decibel Level
    - Event Month (one-hot encoded)

Factor 4: Model Results



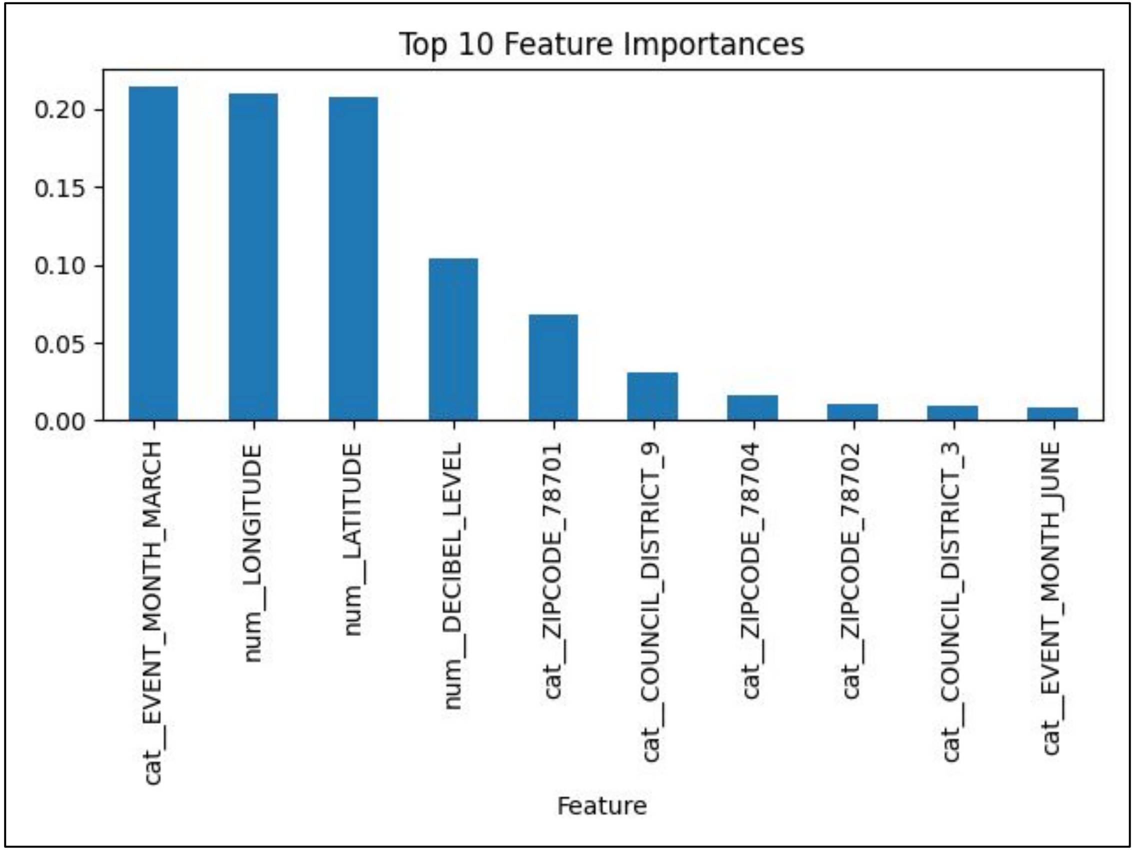
Calculating Importance

Using the **Random Forest Model**, we had the best F1 score for Concrete pouring (0.98) and Outdoor Music Venue (0.95). We reached an overall accuracy of 88.4%.

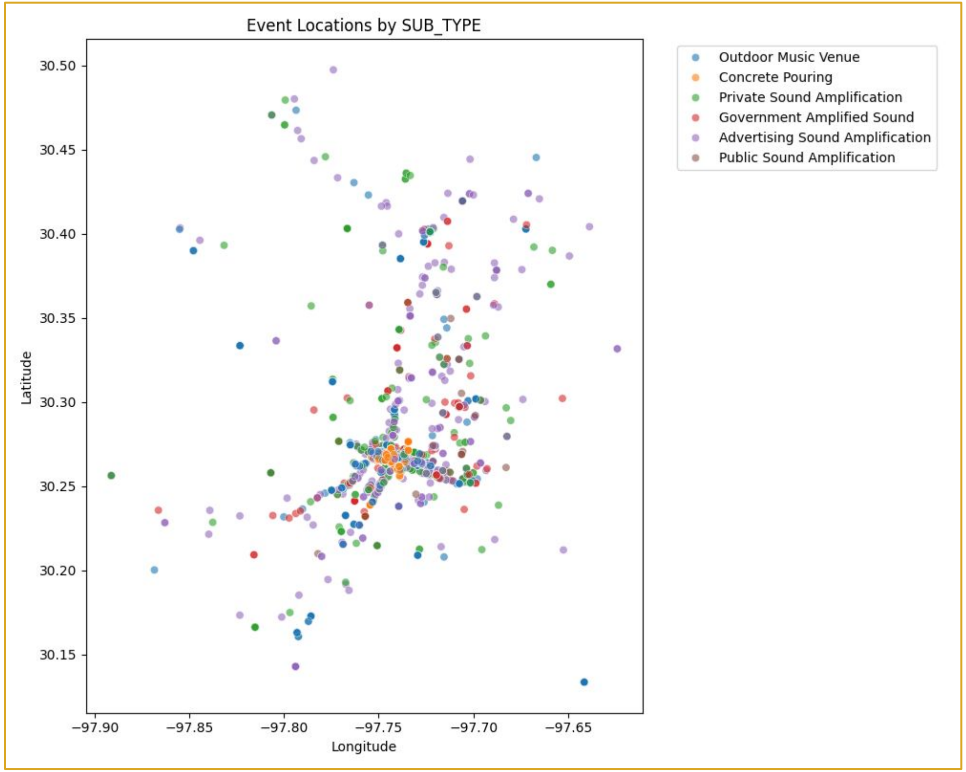
Distribution of Subtypes

```
SUB_TYPE
Concrete Pouring                   0.528972
Outdoor Music Venue                0.242072
Advertising Sound Amplification    0.127511
Private Sound Amplification        0.060435
Government Amplified Sound         0.032210
Public Sound Amplification         0.008800
Name: proportion, dtype: float64
```

Factor 4: Feature Importances



Factor 4: Feature Explanations



Concrete Pouring is clustered in one area, and while Outdoor music venue is too, it's more spread out than any other subtype. Since these make up ~70% of the data, this explains why this feature is important.

Similarly, Concrete Pouring and Advertising were the most common subtypes in March, which explain why this feature was also significant.

How We Can Use This IRL

- Bias in approval times exist (as shown by subtype prediction accuracy)
  - Construction related permits are prioritized with the shortest processing time
  - Outdoor Music Venues (OMV) have the slowest approval time
  - **Policy Suggestion: Prioritize subtypes during peak times**
    - **(ie. Construction in March, OMV during in summer and fall months)**
- Geographic Location has minimal influence
  - Reinforces notion that we should prioritize by permit subtype

Peer Contribution Table

100%	100%	100%	100%	100%
Eshi Kohli	**Maadhav Kothuri**	**Daniel Lam**	**Nneoma Onochie**	**Greg Zachariah**
Explored Linear Regression model with capacity and decibel features	Determined factors for predicting subtype using Random Forest	Determined factors for application times using Linear Regression	Determined factors for application time using Linear and Polynomial Regression	Determined factors for application time using K-Mean and Random Forest

Questions?