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

SDS 322E

20 April 2025

Sound Ordinance Permits in Austin

I. Introduction

Austin, Texas, is home to nearly 1 million people, and with increasing population density, it's important to minimize sound disturbances. Excessive noise can turn into noise pollution, which can cause sleep disturbances, stress, and cardiovascular problems in the population (Angelo, Dutchen). When organizations want to host events they believe will generate excessive sound, they apply for a sound ordinance permit, which permits them to do so by the city government. By analyzing sound ordinance permits, we can better understand how the city handles noise management. Sound is a part of everyday life, and if applicants can understand the factors that impact how sound ordinance permits are processed, they could better plan for events and create experiences meant to strengthen the community in Austin.

Our group wanted to investigate how factors like location, application dates, event type, and other considerations influence one's application and how quickly one can expect their permit to be issued. After exploring the data (as shown in our Data section), we developed three research questions that we believe will help us address these issues:

1. How do the organization applying and the event subtype influence how quickly a permit is issued?
2. What features influence issuance time?
3. Are there shared factors between subtypes that could help predict the likely subtype of an event?

II. Data Cleaning

In order to explore these research questions, we obtained Austin Sound Ordinance Permit Data from the City of Austin Open Data Portal. This dataset is reliable as it's directly sourced from the city government and updates every day. Our dataset originally had 67 columns with 6,730 rows, showing sound ordinance permit applications from 2009 to 2025. Features included dates on when the permit was applied and issued, when the event was meant to start, the subtype of the event, what time the event was supposed to start, the event's capacity, expected decibel level, who is applying for the event, and where the event takes place (location, zip code, and council district).

We spent a significant amount of time cleaning our data. We removed 24 columns we found irrelevant, redundant, or containing too few values. We also reclassified some columns to reveal different insights, such as making a column to indicate if a permit had multiple start dates.

We decided to convert all time-related data from strings into integers representing the hour of the day, creating a range of numbers from 0 to 23. We used this to create a new column, **time_diff**, which measured the difference between the issuance data and the application date. This represents the speed of permit issuance and was our target variable in multiple models.

There were several columns, such as capacity and decibel level, that contained mostly NA values but were still important for our project. Instead of dropping all NA values (which would have left us with 10% of the original dataset), we decided to run experiments in two ways: first, ignoring these columns, and second, by separately analyzing these columns. This approach allowed us to include them in our analysis without affecting how we analyze other features.

Additionally, we further preprocessed the dataset to ensure it was ready for modeling with certain models. Only focusing on two columns, applicant organization and subtype permits, we programmed a Python script to extract unique organization names and process them with Gemini 2.5, a large language model (LLM) with integrated web searching capabilities, to assign each generic label (e.g., 'Construction', 'Technology', 'Food Service'), creating a new COMPANY_TYPES column. We then merged these labels back into the main dataset, ensuring each organization was matched with its type and sorted alphabetically.

We cleaned string values in columns to make them more consistent – examples include cleaning street addresses to only include the street name and changing all the strings in the existing zoning column to only contain their abbreviation.

After data cleaning, we were left with 48 columns and 6,023 rows for us to explore and analyze.

III. Data Exploration

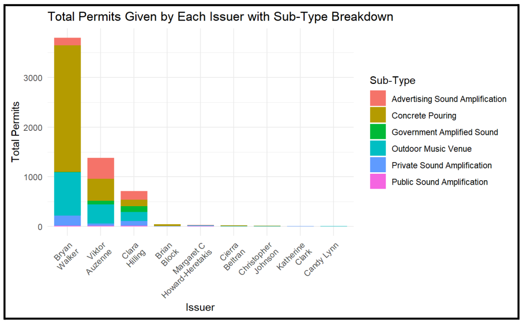
During data exploration, we identified several trends. First, we found that concrete pouring predominantly occurs during the Spring and Summer months (Figure 4) and typically takes place at night (Figure 5, 6) to leverage optimal weather conditions for concrete setting (Borkhataria). Our analysis also revealed that Austin appears to prioritize construction-related development, as evidenced by the significantly shorter approval times for construction permits compared to other subtypes, such as outdoor music events (Figure 9). Additionally, we observed that there was no meaningful correlation between population density and decibel levels across the city, suggesting that permit approvals are influenced more by event type than by local residential concentration (Figure 7). Lastly, we discovered that seasonal patterns impact the volume and subtype of permit submissions. For instance, advertising permits a spike in March near Q1 deadlines, and the government permits a rise in the Fall during election seasons (Figure 8). These findings highlight how environmental and economic factors all shape sound ordinances in Austin.

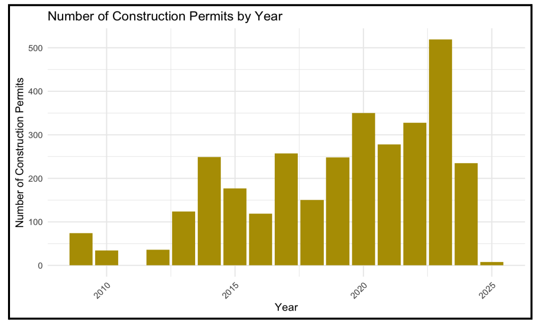
IV. Methods + Findings

Following our comprehensive data preprocessing, we proceeded to explore our hypotheses and validate them through both visualizations and predictive modeling. This section will outline the analytical methods we used and the key findings that emerged from our work.

A. Company Type and Event Subtype Effect on Issuance Times

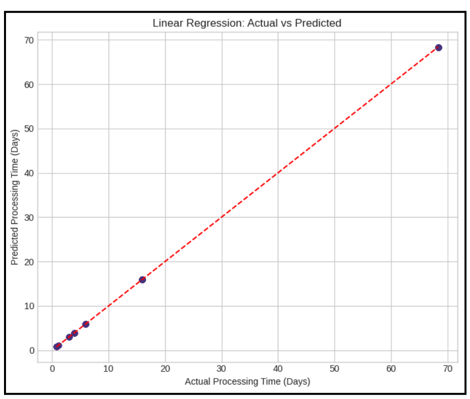
To investigate whether construction and sound permits are treated differently in Austin, we visualized permit data by subtype and issuer, finding that concrete pouring permits were the most frequently issued and nearly doubled from 2022 to 2023. Graphing average approval times revealed that concrete pouring permits were processed fastest, often on the same day. These results suggest that Austin increasingly prioritizes construction-related sound permits over music-related ones, reflecting a shift toward supporting rapid development.

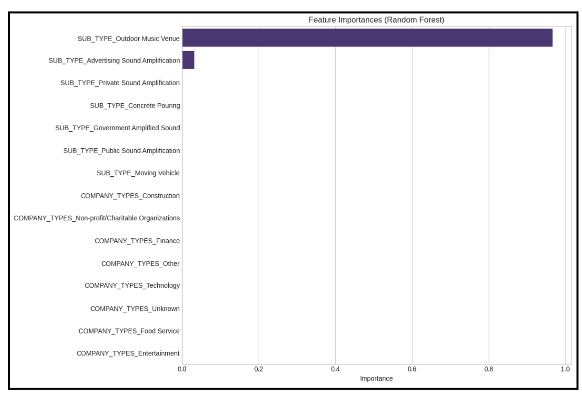




Building on these insights, we applied machine learning models to further test our ideas and quantify the relationships between applicant characteristics, permit types, and application processing times. We first performed additional preprocessing to create generic company labels (such as food service, retail, technology, healthcare, etc.) for each applicant in the dataset. We then calculated the average processing time for each generic company type, finding that food service companies experienced the longest average wait of about 45 days.

We then implemented two predictive models: linear regression and random forest. For the linear regression model, we trained on features including company type, permit subtype, and average approval times, creating a model that can predict average processing time for new applications. We then tested the model's error rate using metrics such as mean squared error, root mean squared error, mean absolute error, and R^2. Combining the averages, the model got an error rate of 5%. The random forest model was used to identify features that determine application times the most. This model highlighted that requests for outdoor music venues and sound amplification for advertising were the most significant indicators of long processing times. Measuring the error rate with the same methods as before, it scored an error rate of near zero.

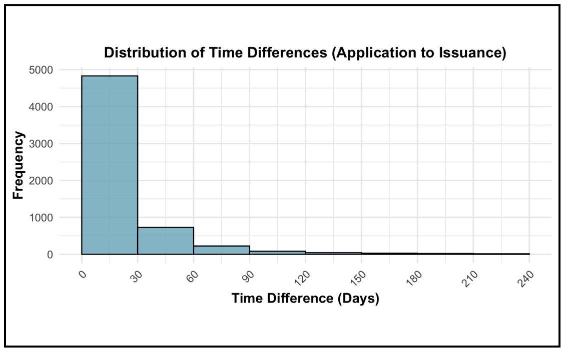




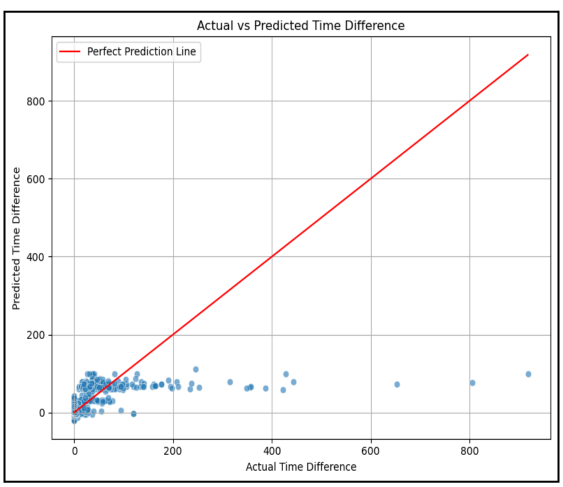
To further consolidate these findings, we developed a Python method to predict average processing times for hypothetical applications with randomized parameters. The model predicted that outdoor music sound permits, often requested by food service and bar establishments, had the highest average processing time of 68.37 days. On the other hand, construction-related companies applying for concrete pouring permits had a predicted processing time of just 0.77 days.

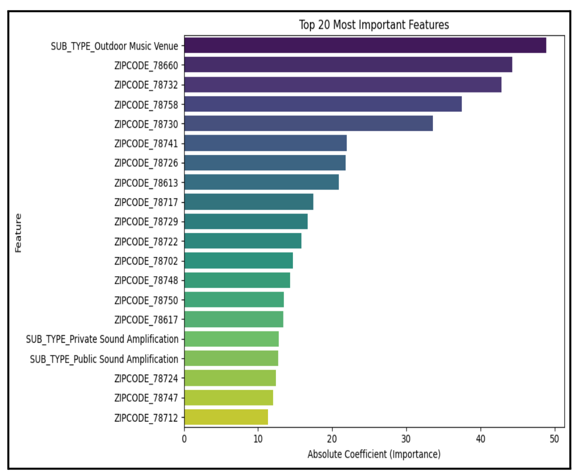
In conclusion, our modeling and visualization efforts confirmed that the type of permit and the applicant's generic company type influence application processing times. There is a clear bias favoring construction-related permits, which are processed more rapidly than those related to music and entertainment. This suggests that, despite Austin's identity as a live music hub, the city currently prioritizes construction activities over music-related events.

B. Determining Permit Factors' Effect on Issuance Times

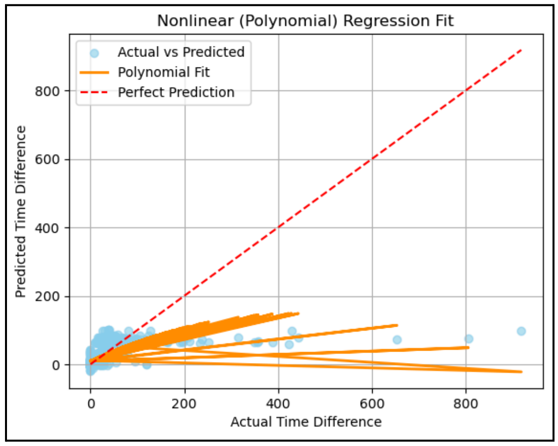


As seen in the graph above, the majority of the permits in our dataset were issued within a day of application (0 days), but several permits took longer to get approved. To determine what features impacted the speed of issuance, we created a linear regression model predicting permit issuance times. We used various features (i.e., event year, latitude/longitude, subtype, amplified sound district, if 51% of the permit's establishment depended on food sales). We discovered that linear regression was a poor model to evaluate. Our R^2 score indicated that our model only accounted for ~27% of the variability of the data, and our RMSE revealed our model had an average prediction error of ~50 days. The model reveals there is no linear relationship between the mentioned features and permit issuance time. While the model was imperfect, it highlighted zip code and outdoor music venue subtype as the most important features to predict the permit issuance times.





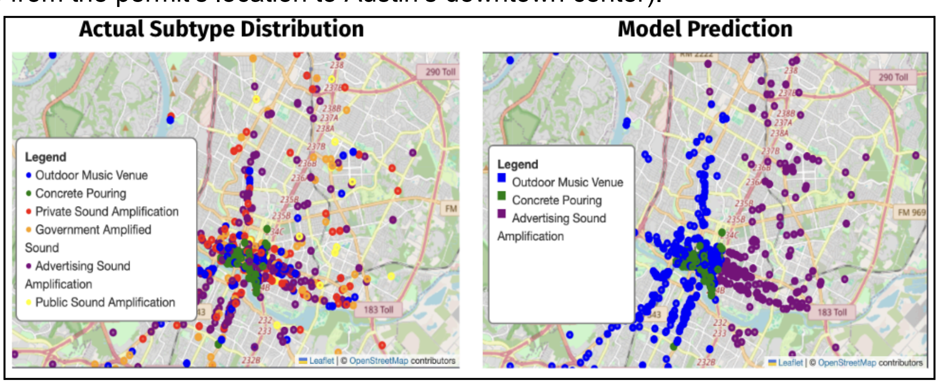
After noting the nonlinear shape of the actual permit issuance time, we decided to use a nonlinear regression model instead. We used polynomial regression, we felt its shape better represented our permit issuance data. We got much better results, as the polynomial model had an R^2 score of 0.4926 and an RMSE of 20.5864 days. Our findings revealed that while polynomial regression models were better for predicting issuance times, the factors we included did not help predict issuance times, especially when the majority of permits were issued in 0 days.



C. Geographical Factors' Effect on Issuance Times

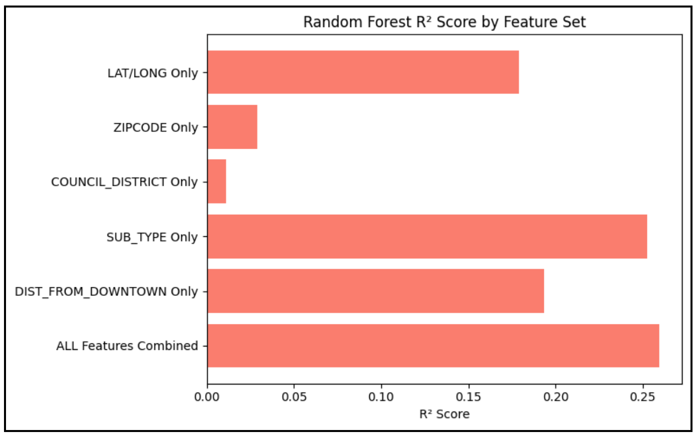
Realizing that there wasn't a strong relationship using all of our features to predict permit approval time, we wanted to begin focusing on a specific subset of features. We wanted to explore whether geographic features could affect permit approval times. Our motivation behind this primarily lies in our intent to explore if certain regions of Austin have a cluster of one specific permit subtype. We wanted to then use these geographic clusters to compare permit approval times to inform our findings for our hypothesis.

Starting out, we opted to build a K-Means model, which would ideally detect 6 different clusters — one cluster for each sound ordinance permit subtype (i.e., Advertising Sound Amplification, Concrete Pouring, Government Amplified Sound, Private Sound Amplification, Public Sound Amplification, Outdoor Music Venue). Our model was constructed using primarily geographic features, as alluded to earlier; these included: Latitude, Longitude, Zipcode, Council District, and Distance from Downtown (a binary feature that we engineered by calculating the distance from the permit's location to Austin's downtown center).

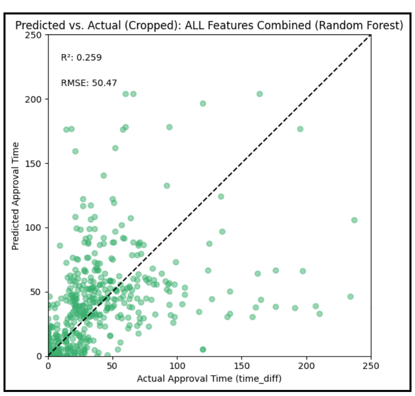


Extrapolating from the K-Means model's clustering of our data points (shown above), we were very quickly able to tell that it performed quite poorly. The model only identified 3 clusters: Outdoor Music Venue, Concrete Pouring, and Advertising Sound Amplification. We attributed this behavior to the composition of our dataset, which overwhelmingly contained permit submissions for those 3 mentioned subtypes. Analyzing the accuracy in classifying each subtype, we can see that there was a high accuracy for Concrete Pouring and Outdoor Music Venues, leading us to believe that they make up the majority of permit submissions in Downtown and West Downtown. Evidently, no cluster detection of government, private, and public sound permits.

Due to the poor performance of our K-Means model, we weren't able to proceed with comparing the average permit submission times between the clusters. To rectify this, we decided to make a second attempt and build a Random Forest Model. Unlike K-Means, Random Forest can model complex, nonlinear relationships between geographic features and approval time. The greater flexibility offered by this model would enable us to explore the relative importance of each geographic feature. We decided to use the same geographic features for this model as we did with our K-Means model.



From the model, we can see that the Random Forest model performed better than K-Means, but still had relatively poor performance with all the features combined, explaining only 26% of the variation in approval time. To measure the performance of each of the features, we trained the Random Forest model on differing combinations of features to see which yielded the highest performance. Comparing these geographic features against the permit subtype, we noticed that a permit's subtype is a significantly better predictor of approval time than any of the geographic features.



When comparing our Random Forest approval time predictions against the actual times, we can see that most of the points are clustered near the origin. Furthermore, the model seems to be satisfactory at predicting short approval times given geographic features, but this likely could be skewed by the class imbalance of Concrete Pouring permits. We noticed that the model was likely to underestimate approval time for permits that took a long time to get approved. Overall, there was noticeable scatter in the graph displayed above, further reinforcing the low accuracy and high RMSE (50.47 days) of our model.

When considering our findings from both of our models, we concluded that geographic features alone are not a strong predictor of sound permit approval times. We realized that approval times can vary regardless of what location the sound ordinance permit is for. Thus, there is no clear “bias” towards approving sound permits coming from one specific area in Austin. Rather, our findings have shown that the subtype could be far more critical in predicting the approval time of a permit.

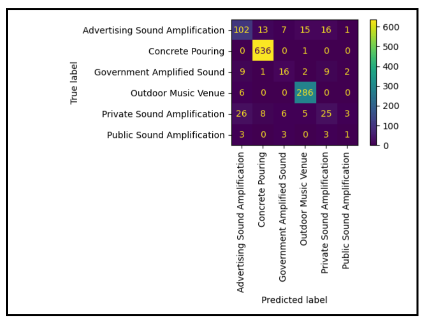
D. <u>Determining Factors for Subtype Prediction</u>

Given the findings that subtype was quite important in predicting approval times, we further investigated this attribute in the data. Specifically, we were interested in understanding predictive factors for subtype. Given that the distribution of permits to subtypes was *not* uniform, determining such factors could provide insights into how to best allocate resources for each subtype. In this way, policymakers could improve the permit approval process for a larger portion of subtypes.

To understand predictive factors for subtype, we chose to develop a random forest model. This model was ideal for our purposes because our data contained a mixture of categorical and numeric values, which this type of model can handle easily. Furthermore, random forests are good at capturing non-linearities, expanding the types of relationships that we could capture. To prepare our input data, we imputed missing numerical values with the median value and missing categorical values with the most frequent category. Furthermore, categorical variables were one-hot encoded.

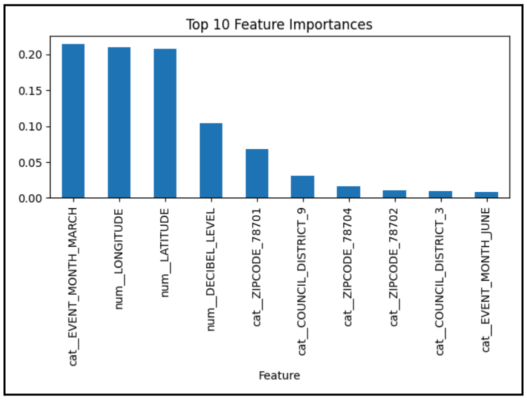
To build our model, we utilized the following features: latitude, longitude, zip code, council district, decibel level, and event month. Then, we developed a random forest made up of 100 estimators with a maximum depth of 6 decision nodes. This ensured that we could capture complex relationships while limiting overfitting to the training data.

```
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
```



	precision	recall	f1-score	support
Advertising Sound Amplification	0.70	0.66	0.68	154
Concrete Pouring	0.97	1.00	0.98	637
Government Amplified Sound	0.50	0.41	0.45	39
Outdoor Music Venue	0.93	0.98	0.95	292
Private Sound Amplification	0.47	0.34	0.40	73
Public Sound Amplification	0.14	0.10	0.12	10
accuracy			0.88	1205
macro avg	0.62	0.58	0.60	1205
weighted avg	0.87	0.88	0.88	1205

After training our model, we reached an overall accuracy of 88.4%. As seen above, the categories with the highest F1 scores were Concrete Pouring and Outdoor Music Venue, while the category with the lowest F1 score was Public Sound Amplification. Given the high F1 scores for Concrete Pouring and Outdoor Music Venue, we wanted to ensure that our high performance wasn't solely from an imbalance in data. To verify this, we analyzed the distribution of data points per subtype, as shown above. Since our accuracy was higher than the proportion of the best predicted categories in the data, it's clear that the model was learning valuable relationships. As a result, our model's performance cannot be attributed solely to imbalances in class labels.



To understand predictive factors, we calculated the importance of each feature in the model. In doing so, we found that the event month being March, longitude, and latitude were the biggest factors that influenced subtype. Upon investigation of the data, we found that Concrete Pouring was the most densely clustered subtype, while Outdoor Music Venue was the least dense subtype. Since these were the most common subtypes, location made a large impact on subtype classification. Similar trends were found when March was the event month.

V. Overall Model Discussion

Our models provide important insights into how sound ordinance permits are processed in Austin and shed light on larger patterns that impact the city's cultural and economic landscape. We found that approval times are biased by event subtype, with construction-related permits being consistently prioritized for faster processing, while outdoor music venue permits experience the longest delays. This trend suggests that as Austin grows, city operations are skewing toward facilitating urban development over cultural events. Since geographic location had minimal influence on issuance times, our findings reinforce the idea that prioritization strategies should be built around permit subtypes rather than location-based factors. Practically, this means policy changes could focus on dynamically prioritizing different types of permits during peak seasons. This could lead to favoring construction in spring when projects typically surge, and supporting OMV permits during the summer and fall when outdoor events are most common. These strategies would help balance the city's development goals with the need to maintain Austin's culture.

However, our analysis was limited by the quality and completeness of the available data. Missing values in key fields like decibel levels and capacity information reduced the precision of our models and likely introduced some bias. Additionally, while we focused on predicting approval times, we were unable to build a model predicting whether a permit would ultimately be approved or denied. This was due to the dataset lacking a clear status for each application. If future datasets included more consistent tracking of permit statuses, such as whether an application was pending, approved, rejected, or withdrawn, we could develop more nuanced models that not only forecast processing times but also estimate approval likelihood. Improving record-keeping practices would significantly improve the practical applications of data-driven policy planning for Austin's future.

VI. RRI Discussion

A. Stakeholder Involvement

In the context of stakeholder involvement, key stakeholders for our project include city officials, policymakers, venue owners, and community members. These members could play a more substantial role in shaping the direction and application of our findings. For instance, Austin city planners and policymakers could use our findings from metrics like decibel level and permit subtype to make more informed decisions that balance the city's rapid urban growth with public well-being. Within the context of Austin's rapid growth, stakeholders need to be able to apply trends found in our data, such as the overwhelming majority of sound ordinance permits originating from construction, to make Austin's expansion sustainable. One notable insight we obtained in our research was the stagnation in outdoor music venue sound ordinance applications, which could signal broader effects that trickle into other parts of Austin's economy, such as weakened tourism and small business revenue. Such information could be used to equip city officials and congressional representatives with tangible data that could be used to assess the health of the local economy and the wider implications of sound ordinance permit applications. Stakeholder involvement in Austin could be aided through data-sharing platforms, where sound ordinance data can be cross-referenced against other city metrics, ensuring that Austin policies and directives remain grounded in real-time, relevant data.

B. <u>People Affected</u>

The people most affected are musicians, event organizers, and fans of outdoor music events, as outdoor music permits take an average of 70 days to process. This can make it difficult to plan live events and potentially diminish Austin's music scene. Downtown residents are also impacted, since construction-related sound permits are typically approved the same day, leading to frequent construction noise and roadblocks. This prioritization of construction over music not only affects businesses but also the daily life of residents and visitors, highlighting the need for policies that balance urban development with support for Austin's cultural identity.

C. <u>Training and Equipping</u>

To equip city officials for better consistency in record-keeping around sound ordinance permits, tailored training and ongoing support are essential. Team members need training on accurate and complete data entry practices, particularly for fields where we observed significant missing values, such as expected decibel levels, venue capacity, and permit status information. Providing structured workshops on standardized reporting protocols and best practices for digital record management would help officials and partners improve the overall reliability of public datasets. Formal education initiatives, such as onboarding modules for new employees, could be paired with informal options like periodic refresher sessions and data quality audits. By strengthening education and support systems, Austin can ensure that future permit datasets are more complete, making it easier for policymakers and researchers to analyze trends and make informed, equitable decisions for the city's development.

D. <u>Potential Conflict</u>

As seen from our analysis, there is an overwhelming bias towards construction permits when compared to other subtypes of permits. Construction Pouring makes up many of the permits and has a much lower approval time than any other permit. As a result, city officials may try to emphasize other permit types to ensure that all subtypes are properly addressed. While this is important for the approval of all subtypes, it may create a conflict with construction-related entities. Approval of sound permits is necessary for them to start working, and with the accelerated rate of construction in Austin, speed is very important for these companies. Any delays in permit approval because of a shift in approval priorities could lead to resistance, so navigating these issues will be vital for any policy decision to be effective.

VII. Conclusion

Through our comprehensive analysis of Austin's sound ordinance permit data, we found that permit issuance times are primarily influenced by event subtype and company type rather than geographic location. We also realized that construction-related permits seem to receive prioritization, revealing a systemic bias toward urban development over cultural activities. Our machine learning models, including random forest and polynomial regression, supported these findings and demonstrated that while some seasonal and organizational trends can predict issuance times, most permits are issued rapidly with little variation based on location.

While our findings highlight important trends that can guide future policymaking, like adjusting permit prioritization to balance out the subtype of permits being issued, they also reveal the importance of having public data that is complete, interpretable, and correct. Everyone should understand how local government affects their events, and we believe there would be a positive impact if there were better data management. Austin is a city of music and is the live music venue capital of the world – we want business owners to be able to utilize this moniker, utilize the sound ordinance permit process, and connect the city through sound. Improving the quality of public permit data would enable the city to make more informed, equitable decisions that balance the needs of rapid development with preserving Austin's cultural identity as a live music hub.

VIII. Acknowledgment

Eshi Kohli	100%
Maadhav Kothuri	100%
Daniel Lam	100%
Nneoma Onochie	100%
Greg Zachariah	100%
LLM Usage	Used LLM to label "generic company type"

IX. References

1. Angelo, Angelina Sophia. “Could Everyday Noise Be Affecting Your Health?” *Environmental Health Sciences Center*, 4 Feb. 2025, environmentalhealth.ucdavis.edu/blog/could-everyday-noise-be-affecting-your-health.
2. Borkhataria, Guillaume. “What Is the Best Weather to Pour Concrete - Explained By Converge.” *Converge*, 29 July 2024, https://www.converge.io/blog/blog-best-weather-to-pour-concrete.
3. City of Austin, Texas. “Sound Ordinance Permits: Open Data: City of Austin Texas.” *Data.AustinTexas.Gov - The Official City of Austin Open Data Portal*, 15 Apr. 2025, data.austintexas.gov/Recreation-and-Culture/Sound-Ordinance-Permits/ryu3-tuin/about_data.
4. Dutchen, Stephanie. “Noise and Health.” *Harvard Medicine Magazine*, 14 Mar. 2025, magazine.hms.harvard.edu/articles/noise-and-health.

X. Supporting Materials

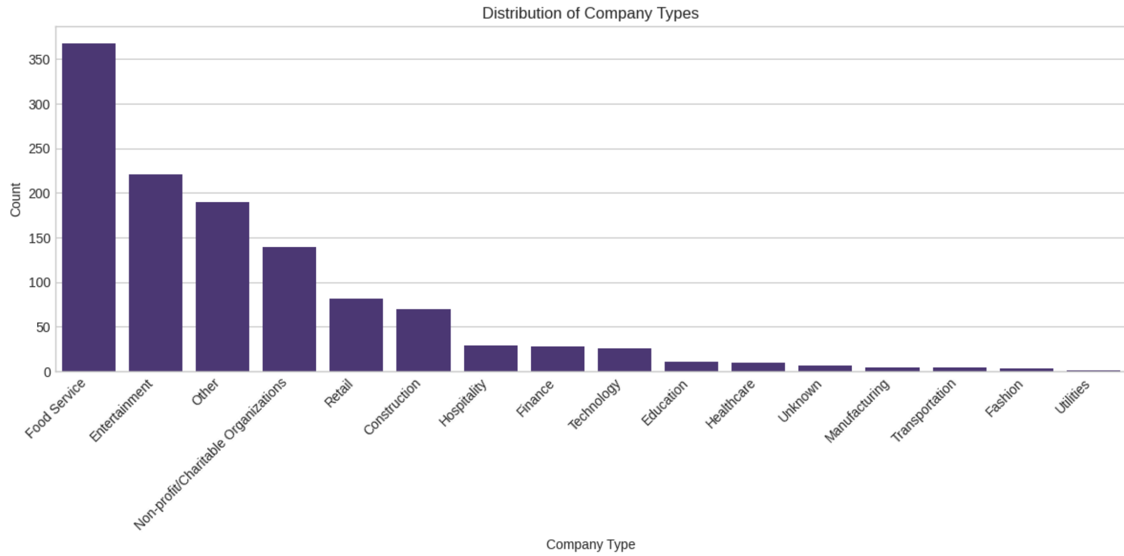


Figure 1: A bar graph that shows the distribution of generic types within the data

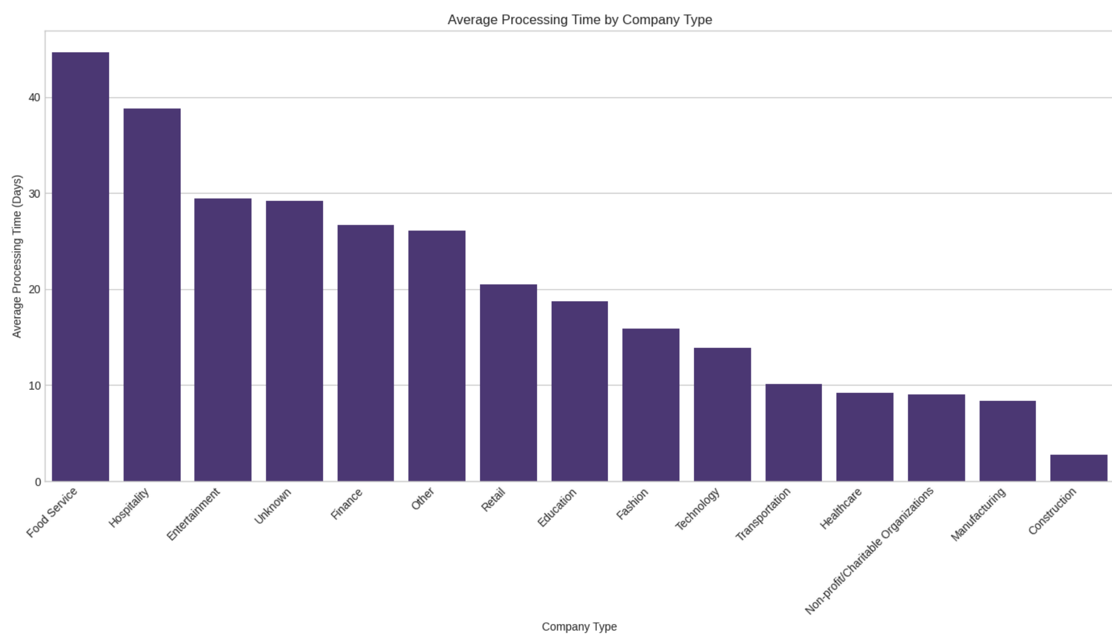


Figure 2: A bar graph that shows the average processing times based on generic company type.

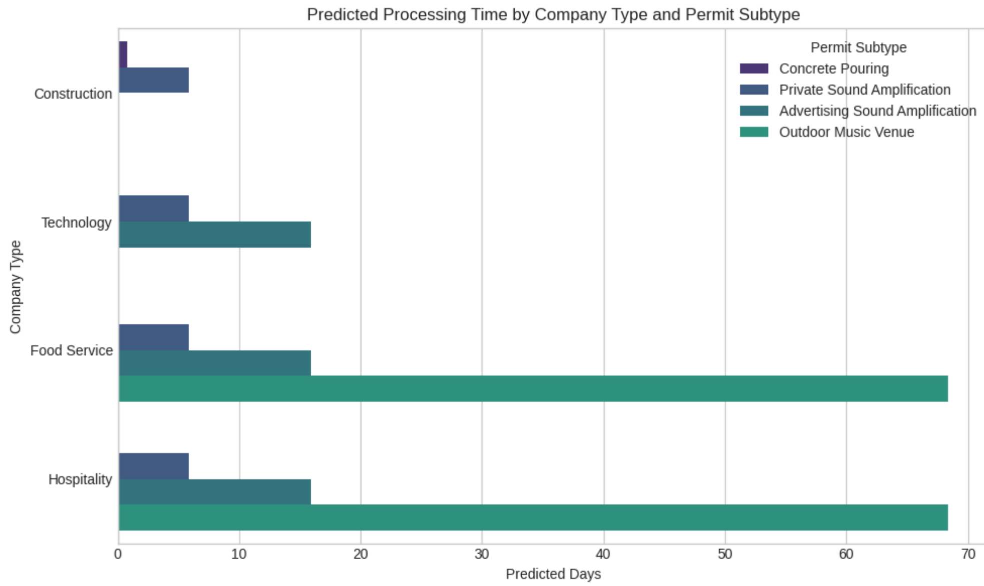


Figure 3: A bar graph that shows the average processing times on randomized parameters using the linear prediction model we taught.

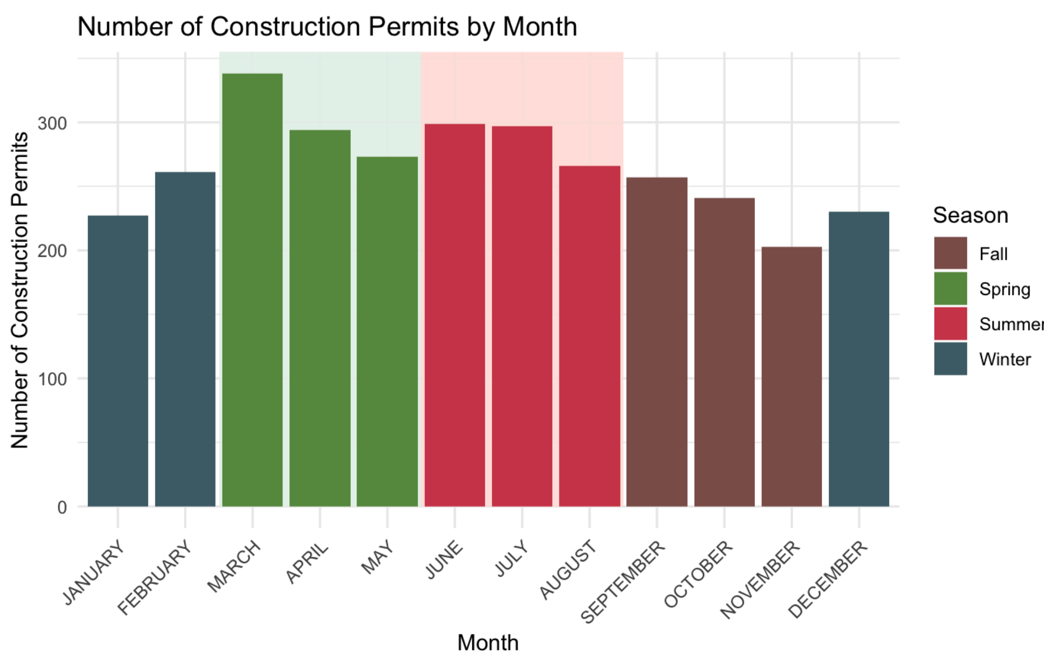


Figure 4: A bar graph showing the number of construction permits issued by month, categorized by season of the year

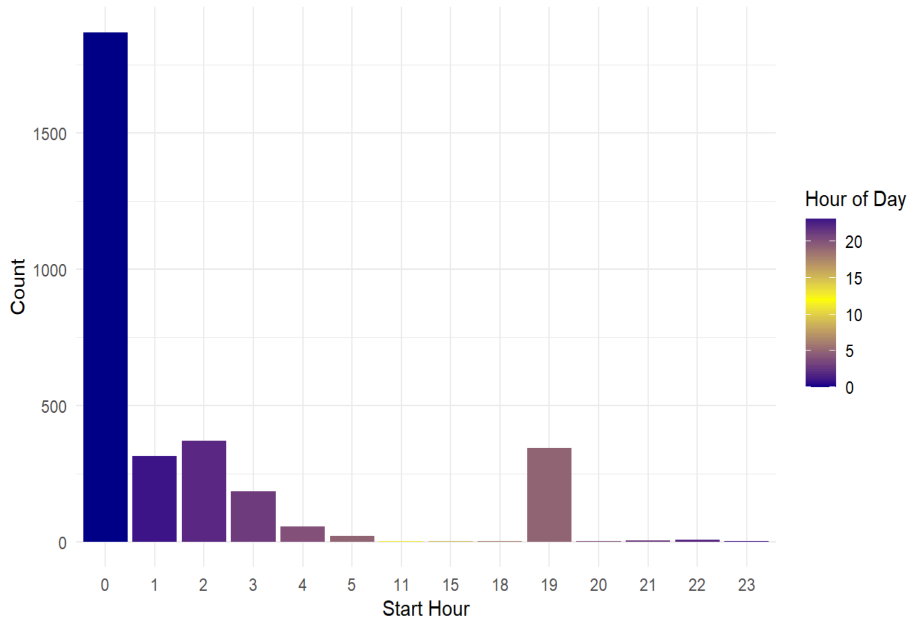


Figure 5: A bar graph showing the start hour for a concrete pouring permit

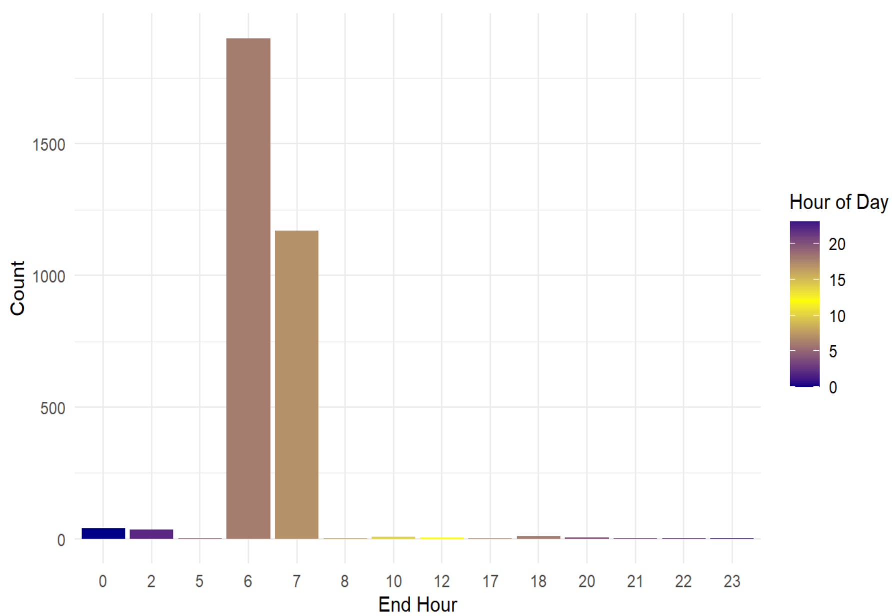


Figure 6: A bar graph showing the end hour for a concrete pouring permit

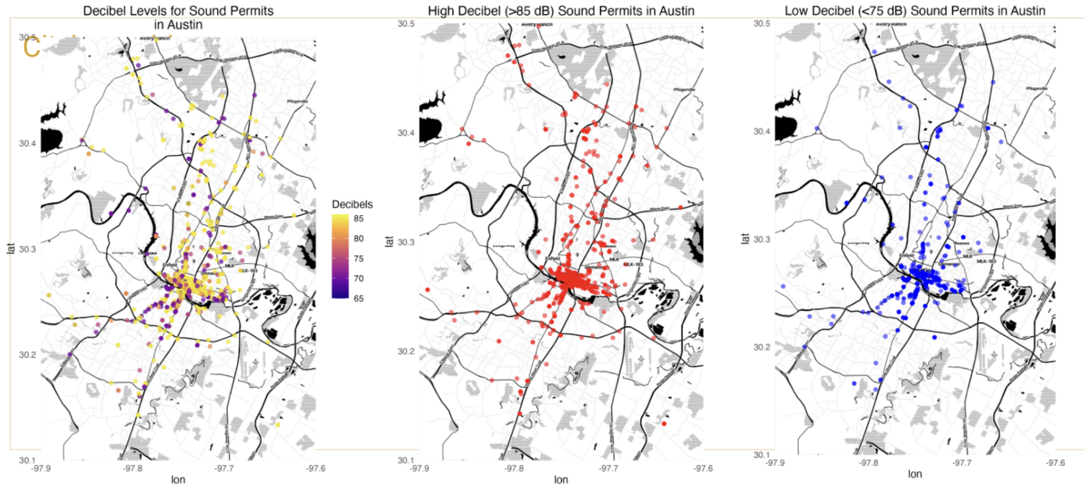


Figure 7: Point distribution map overlaid on Austin, which shows the location of permits categorized by their decibel level

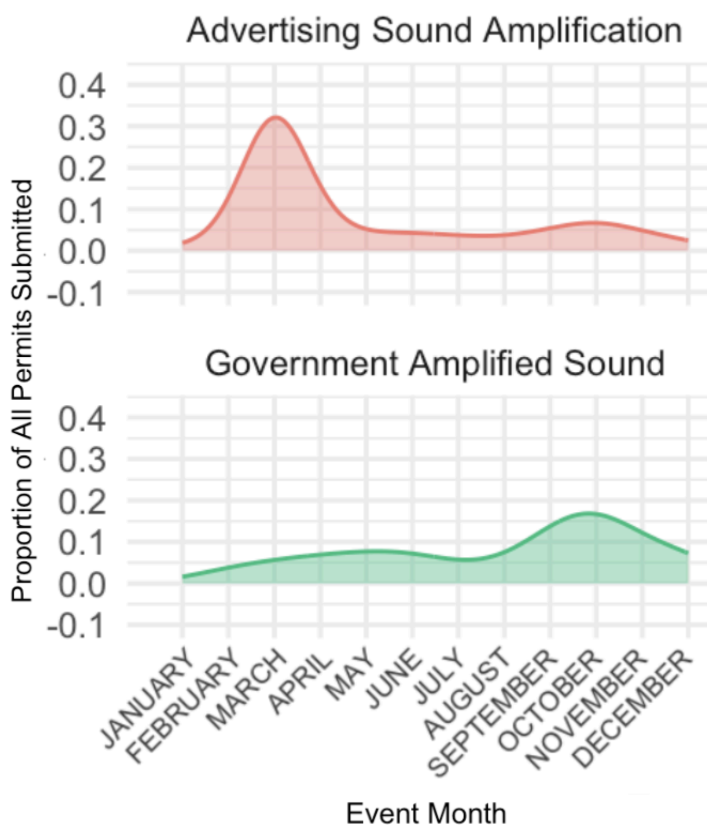


Figure 8: Density plot of the proportion of all permits submitted during the year for the Government and Advertising subtypes

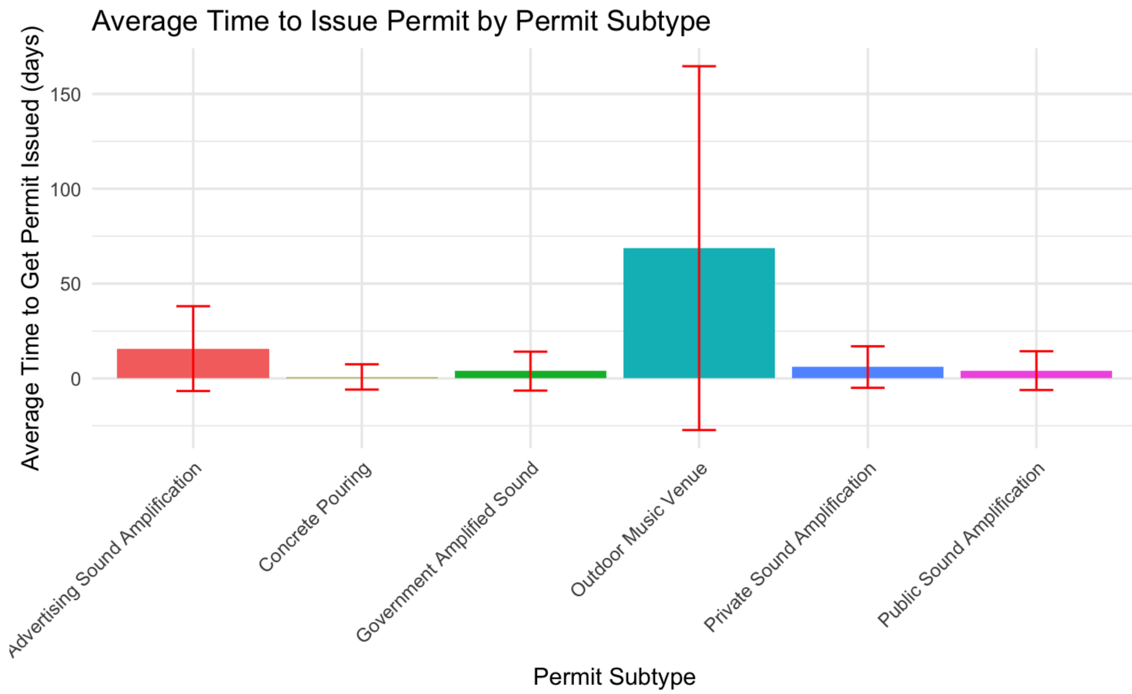


Figure 9: Bar plot of the average time it took each permit subtype to get the permit issued with standard error bars