Analysis of 911 Emergency Calls

Harshith Doppalapudi   Sarabesh NR   Sree Venkat Chintakula
Yamini Ane

Objective of the Analysis

The aim of the analysis is to uncover patterns and trends in emergency call data to inform decision-making for emergency services. Emergency call patterns can vary significantly throughout the day and across different periods of the year. By understanding these fluctuations, emergency services can optimize resource allocation, ensure adequate staffing levels, and ultimately improve response times and public safety.

Hourly Distribution of 911 Calls:

The analysis of average calls per hour serves a dual purpose. Firstly, it enables the identification of daily operational trends which are essential for short-term planning. Such analysis reveals the hours when demand on emergency services is at its peak and when it eases. This insight is vital for shift scheduling, ensuring that the workforce is aligned with the demand curve.

Secondly, it sheds light on human activity and societal behaviour as they correlate with emergency incidents. Understanding these patterns can help anticipate periods of high emergency call frequency, thus allowing for proactive management of emergency services.

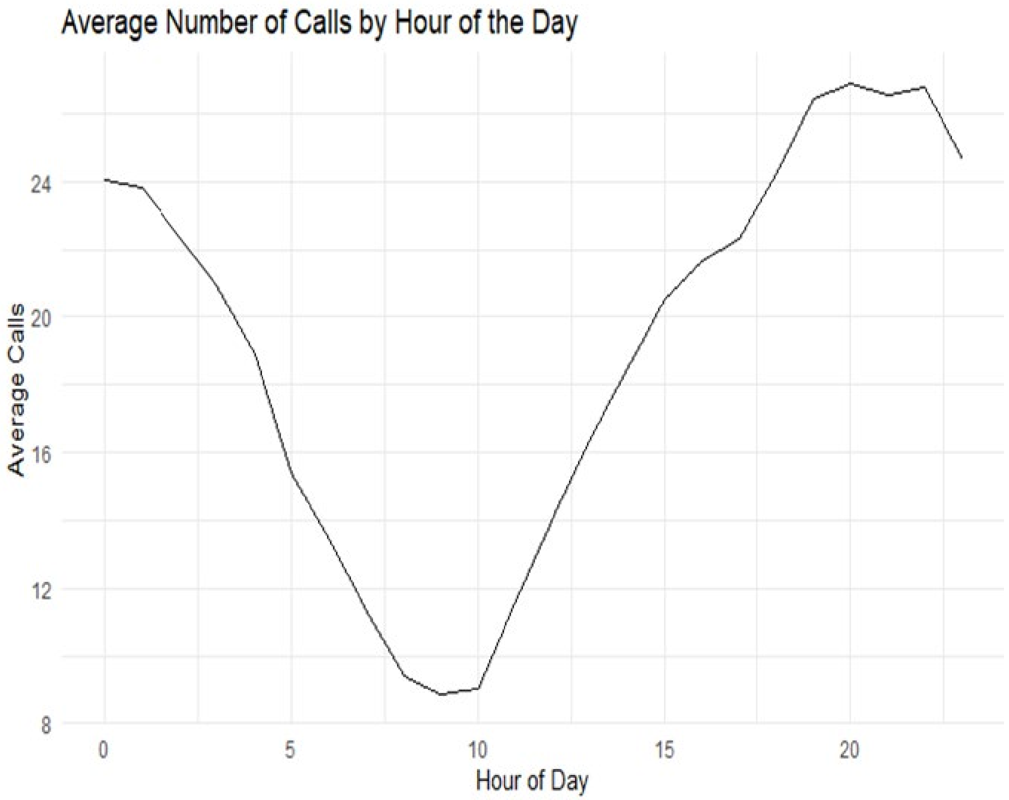

Figure 1: Hourly Distribution of 911 Calls

Figure 1 depicting the average number of calls received throughout the day. The lowest volume of calls consistently occurs in the morning hours around 9AM, indicating a general societal downtime. As the day progresses, call volumes increase from afternoon to evening peaking in the evening around 8PM. This peak is likely correlated with increased population movement and activity, including commuting, which can lead to more accidents or incidents requiring emergency assistance.

Annual Analysis: Average Calls per Hour

The aggregation of average calls per hour over the year presents a broader perspective, highlighting seasonal trends and anomalies over the 12-month period. These patterns can be influenced by a range of factors, including climate conditions, holidays, and large-scale events. For instance, in figure 2 the peak observed in June might be associated with summer activities that typically involve more outdoor events, leading to an increase in emergency incidents.

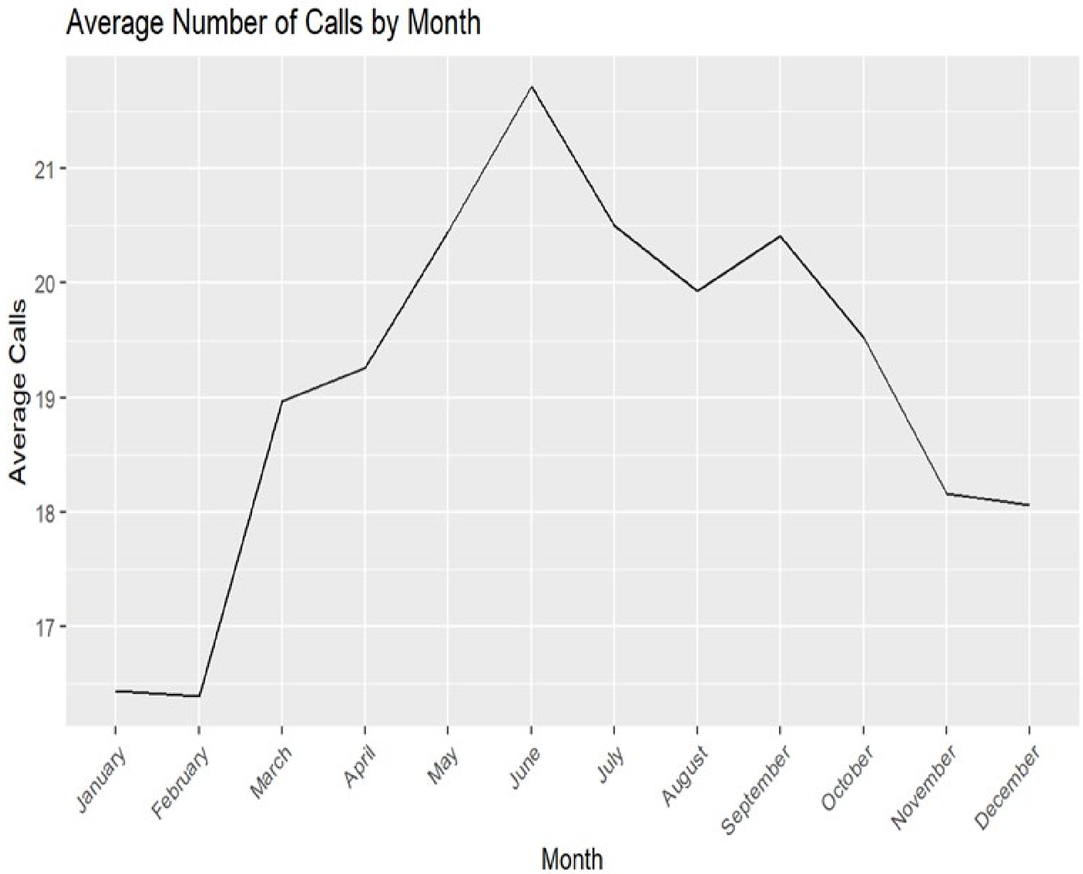

Figure 2: Seasonal Trends in 911 Call Volume

Conversely, the decrease in call volume in December could be attributed to a general societal slowdown during the holiday season, with many individuals taking time off work and spending more time at home, thus reducing exposure to situations that might necessitate an emergency call compared to June.

Daily and Hourly Call Patterns

Moving forward, an analysis of the hourly and daily patterns of average 911 calls was conducted. Figure 3 illustrates the number of 911 calls made at various hours throughout the week. It is evident that there is a consistent period each day, from late night until 10 AM in the morning, during which there is a decline in the average number of calls are received.

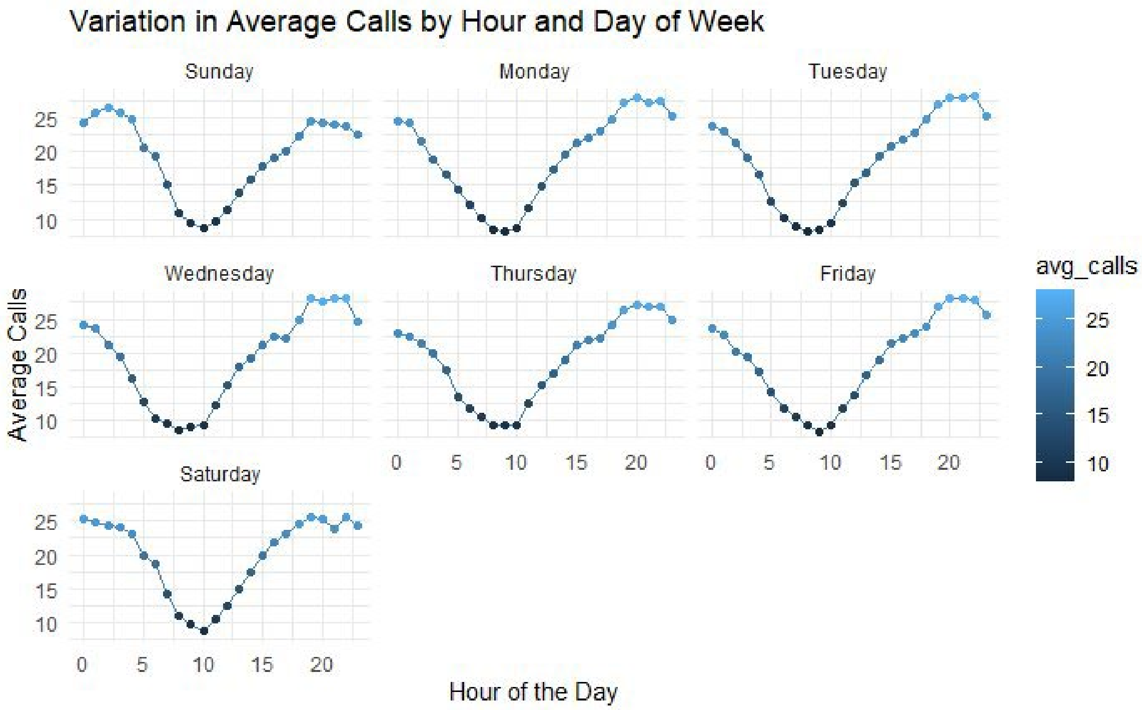

Figure 3. Variation in the average 911 calls by each hour of the day faceted by the day of the week

This trend may be attributed to individuals sleeping or being occupied with their daily routines. However, between 6 PM and 10 PM in the evening, the highest volume of calls is observed consistently. This recurring pattern indicates that most emergencies are reported during these post-work hours.

When examining the different days of the week, Monday to Friday appear quite similar, with a noticeable increase in activity in the evening. However, the pattern shifts during the weekend. On Saturday and Sunday evenings, there is a decrease in the number of calls, which may be attributed to individuals enjoying their weekends. Conversely, during late nights on weekends, there is a notable increase in call volume, possibly indicating heightened social activity. Furthermore, the early hours, specifically 9-10 AM, experience the lowest volume of calls, which can be attributed to individuals starting their day.

Monthly changes in hourly call patterns

Continuing the analysis to observe the hourly call patterns across different months of the year, a consistent daily pattern across all months, characterized by a peak in call volume during the evening hours was observed as shown in figure 4. This suggests that certain times of the day experience a higher frequency of calls, possibly aligning with rush hours or periods of heightened activity. Furthermore, a noticeable decline in the average number of calls is observed around 9AM to 10AM each day, indicating a lower incidence of events requiring emergency services during typical morning hours.

Additionally, while the shape of the graphs remains consistent, there are variations in call volume between months. Warmer months, from May to August, exhibit a higher overall average of calls compared to colder months.

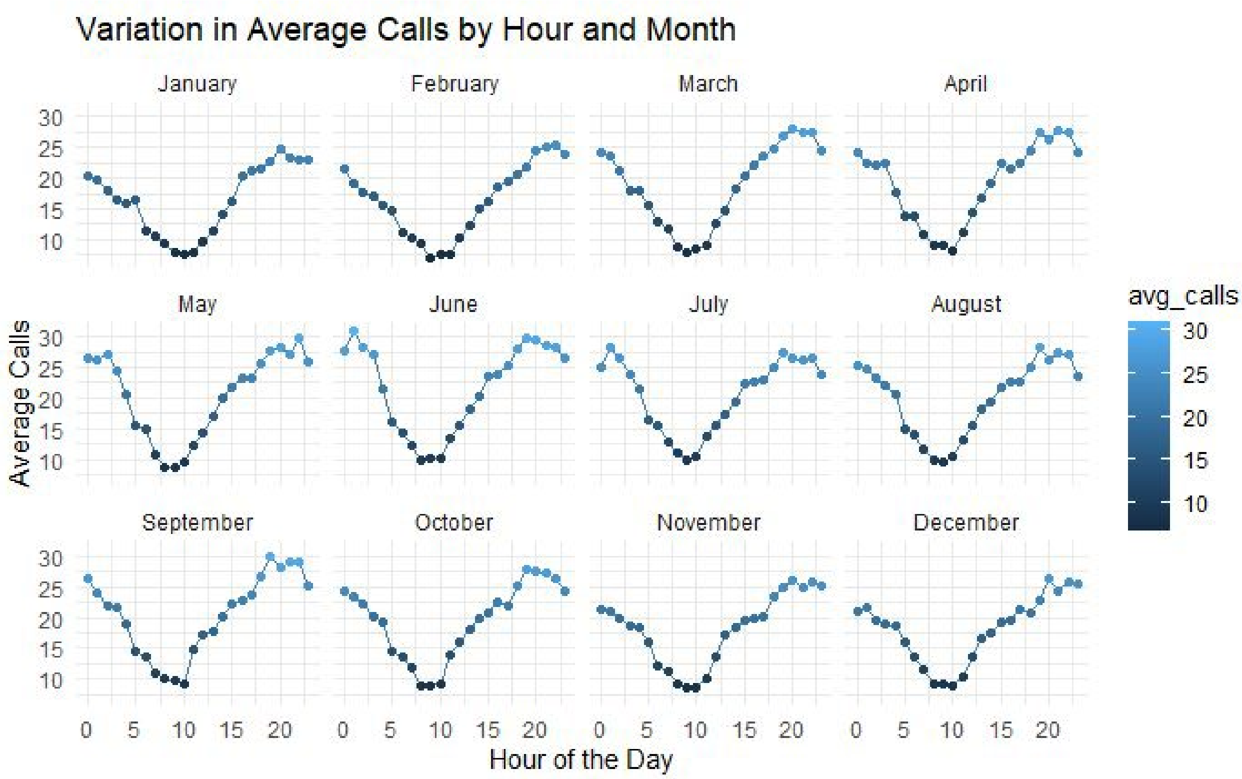

Figure 4. Monthly variation in the hourly average of 911 calls

The lowest average calls per hour consistently occur in the morning, around 9-10 AM, across all months. This trend suggests a seasonal variation, with summer months showing a higher frequency of calls compared to winter months. This correlation may be attributed to increased outdoor activities and associated incidents during warmer weather.

Models

Linear Modelling

In our exploratory data analysis (EDA) phase, we initiated our modelling process by considering a linear framework. We aimed to forecast the hourly TotalCalls, a measure indicative of activity, leveraging predictors such as Month, Day of the Week, and Hour of the day.

Initially, we explored the performance of linear models using both raw TotalCalls and log transformed TotalCalls. Surprisingly, our analysis revealed that the linear model yielded superior performance when TotalCalls were subjected to a logarithmic transformation. This transformation enhanced the model's predictive capability, highlighting the non-linear nature of the relationship between the predictors and the TotalCalls. Considering the discrete nature of our predictors, we meticulously evaluated models with and without interaction terms. Intriguingly, the information criteria metrics, namely the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), favoured the model devoid of interaction terms. This underscored the notion that the linear model, unadorned by interactions, provided the most parsimonious representation of the data.

Hence, our selected model, depicted as follows:

```
lm(log_TotalCalls ~ factor(Month) + factor(Day) + factor(Hour), data=hourly_calls)
```

To visually assess the model's performance, we present graphs juxtaposing the predicted TotalCalls against the actual observations across different Hour, Day, and Month segments.

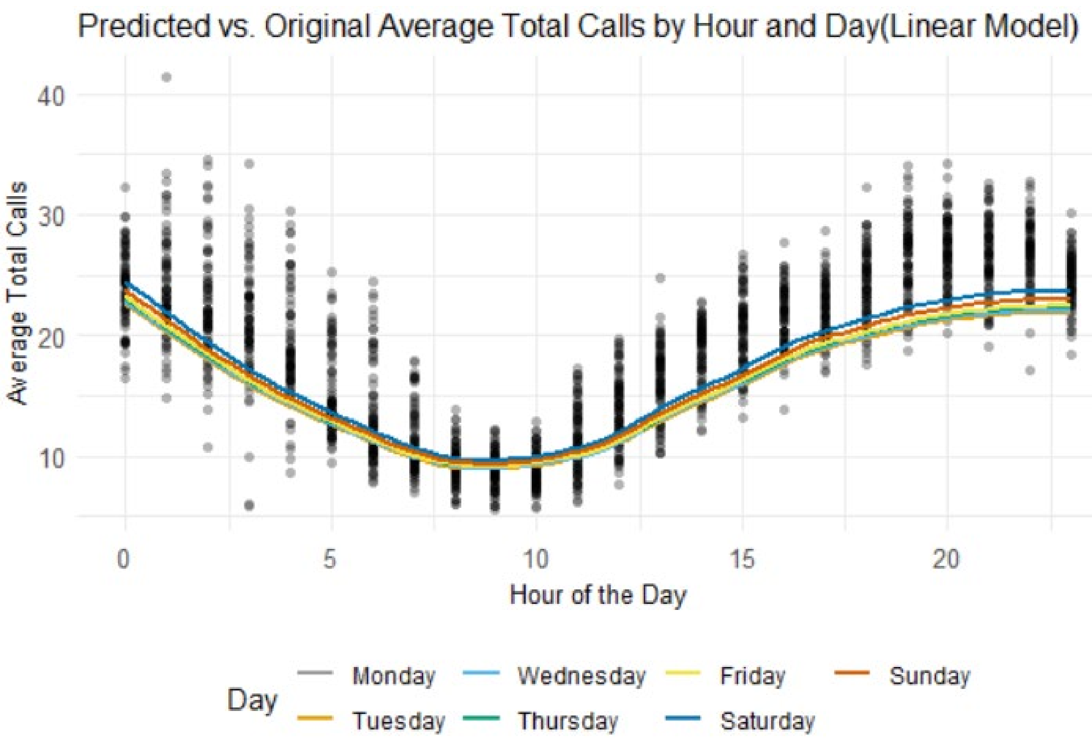

Figure 5. Comparision between actual and predicted average total calls per hour by the linear model.

The visual depiction portrays the relationship between the actual average total calls per hour and those predicted by the linear model. Specifically, it focuses on the average total calls versus the hour of the day. The scatterplot features black data points representing the original averaged data. Superimposed are smoothed LOESS lines derived from the model's predicted values. Notably, each day of the week is distinguished by a unique colour line, totalling seven in representation.

Observing the graph, we discern a close alignment between the lines and the average of the original scatter points. Furthermore, minimal disparity is evident among the lines, signifying reduced variability between different days. As a result, this model indicates a consistent pattern: the least average total calls are predicted during the 10th hour of the day, while the highest occur during the 0th and 24th hours. Notably, Saturdays exhibit the highest predicted total calls, whereas Mondays demonstrate the lowest, suggesting a distinct weekday influence on call volumes.

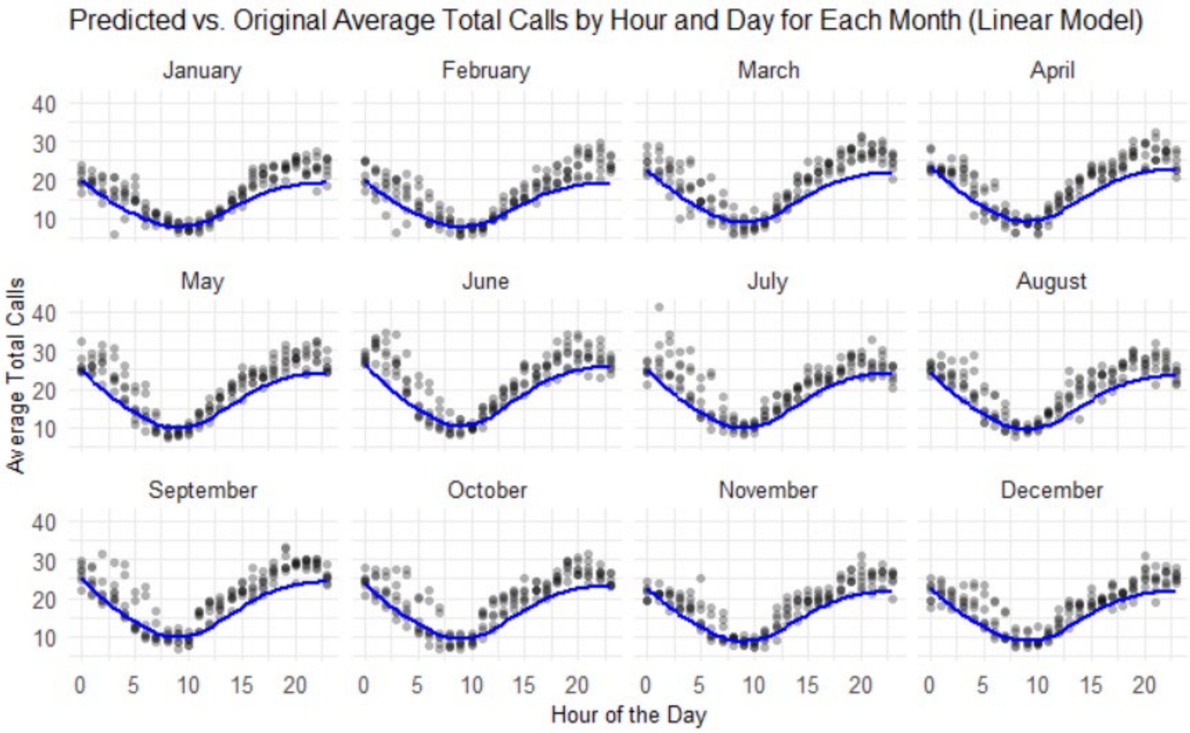

Figure 6. Graphs representing predicted and actual average total calls faceted by month

The graph illustrates the relationship between the predicted line, depicted in blue via LOESS smoothing over model-predicted points, and actual data points represented by black scatter points. It offers insights into how average total calls fluctuate across different months and evaluates the model's ability to capture these variations.

Notably, the original data exhibits slight variations, as evidenced by the scatter points, which the linear fit endeavours to replicate. While the model closely aligns with the majority of these minor variations, it occasionally deviates from the actual data. For instance, during the 20th hour of March and April, the predicted line falls notably below the average. Despite these discrepancies, the linear model generally succeeds in accommodating the variance observed across different months, albeit with some missed calls for select months.

GAM Modelling

In the exploratory data analysis (EDA) stage, our attention turned to Generalized Additive Models (GAM) to account for the non-linear complexities observed in the 911 call volume data. We sought to enhance our forecasting ability for the hourly total number of calls by incorporating flexible smooth functions of time alongside categorical variables for day and month.

Our iterative modelling process began with a base model and progressively integrated interaction terms, allowing us to scrutinize the interplay between different temporal factors. We were particularly interested in how the calls varied with the hour of the day, days of the week, and across months. The models' performances were quantitatively evaluated using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), with the second model, gam_model_2, showing a remarkable balance between complexity and fit, as evidenced by its AIC and BIC scores.

We describe gam_model_2 as follows:

gam(log_TotalCalls ~ s(Hour) + factor(Day) + factor(Month) + s(Hour, by=Day), data = hourly_calls)

To critically evaluate the effectiveness of gam_model_2, we crafted a series of visual comparisons, juxtaposing the predicted values against the actual call data across various timescales:

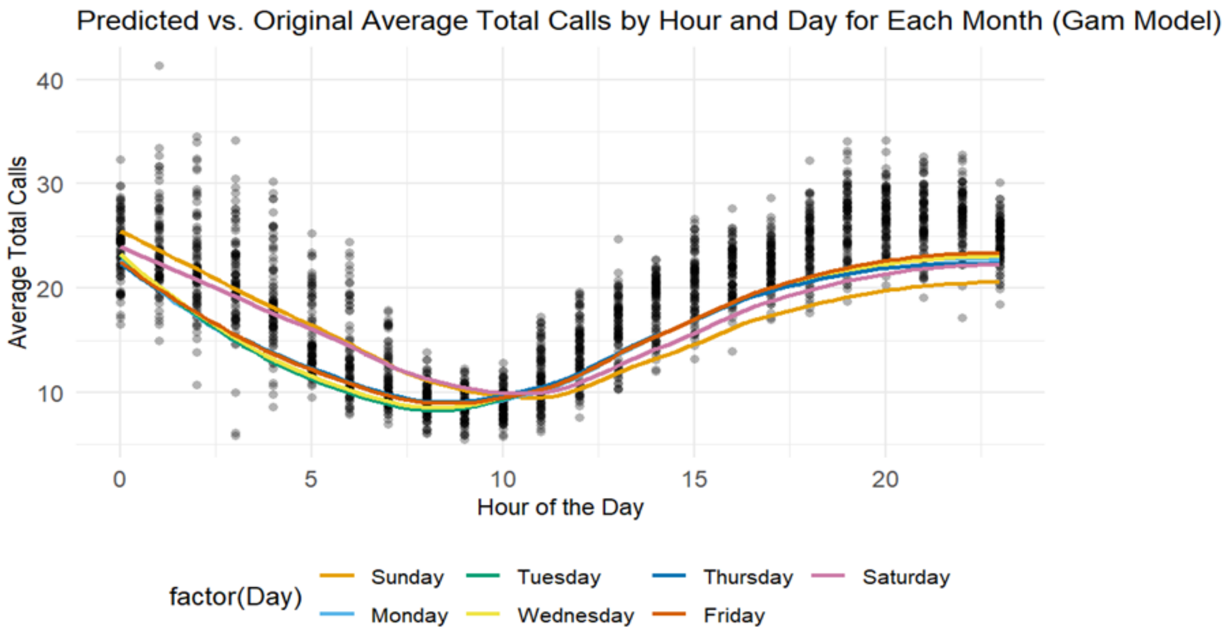

Figure 7. Comparison between actual and predicted average total calls per hour by the GAM model.

This figure displays the alignment of predicted and actual average call volumes across hours and days. With different colors indicating each day of the week, the plot reveals a diurnal pattern, closely matching the peaks and troughs of the actual call data with the model's predictions.

In the Figure 8 next page, we visualize how the daily call patterns fluctuate each month. The GAM model's predictions are overlaid with actual data points, illustrating the model's capacity to replicate the seasonal variance in call volume. The model's proficiency is apparent as it tracks the observed data with considerable accuracy, despite minor deviations at certain hours, like the underprediction noted during early evenings in the cooler months.

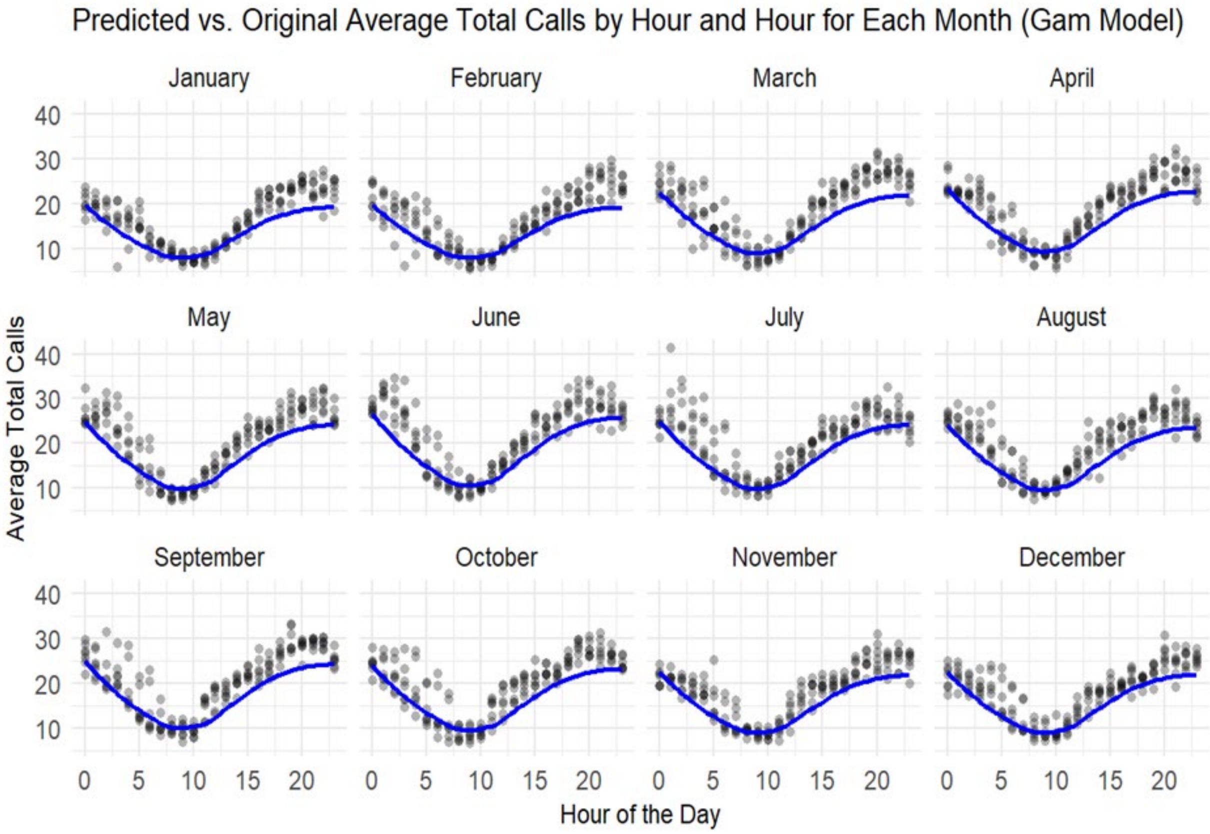

Figure 8. Graphs representing predicted and actual average total calls faceted by month

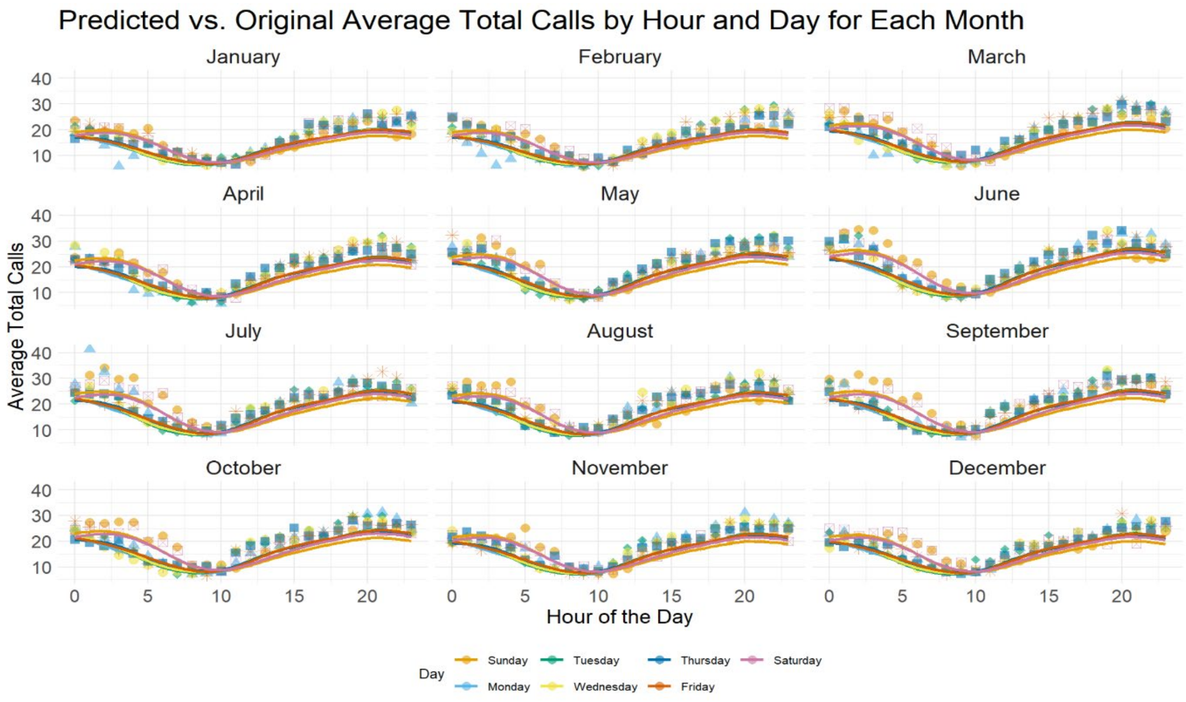

Figure 9. Comprehensive Yearly Overview

In the last figure, Figure 9 we present a faceted view that combines the insights from the previous figures, offering a month-by-month and day-by-day visualization of call volumes. This detailed depiction underscores the adaptability of gam_model_2 to the intricate patterns in the data, showcasing its predictive prowess throughout the calendar year.

Conclusion and Future Work:

The utilization of both linear models and Generalized Additive Models (GAM) has deepened our comprehension of the temporal dynamics within emergency call data. The linear model laid the groundwork, indicating the potential of transformation and paving the way for more sophisticated analyses. It was the progression to gam_model_2, however, that unveiled a richer tapestry of time-related interactions and their impact on call volumes. Our visual evaluations lend credence to the model's robustness in forecasting call demands, suggesting a path to revolutionize emergency service operations.

As we contemplate the evolution of our analysis, several paths lie open for exploration and enhancement: Integration of External Factors: Augmenting our models with variables like weather conditions or local events could yield a finer calibration and enhance predictive precision.

Spatial Analysis: By incorporating geographic information, we could gain a more granular understanding of call distributions, which is essential for targeted resource deployment.

Model Exploration: The exploration of alternative statistical methods and machine learning techniques may unearth additional layers of insight or superior predictive capabilities. Additionally, fitting separate models based on call priority—high, medium, and low—could refine our predictions, enabling a more nuanced anticipation of needs and a more strategic allocation of resources.

Operational Implementation: The true test of our models lies in their application within the operational frameworks of emergency services, offering a practical measure of their impact and utility. Exploring alternative methodologies, such as basic loess modeling, introduced new challenges. Loess modeling's requirement for non-categorical values posed difficulties, as direct conversion to numerical values yielded suboptimal results.

Our continued efforts in these areas aim to not only build on the solid analytical foundation established by this study but also to contribute substantively to the optimization of emergency response mechanisms and the safeguarding of community well-being.

Appendix

Appendix I: Detailed Visualization of Call Volume by Hour and Priority
This appendix includes the R code and the resulting visualization that details the distribution of 911 emergency calls throughout the day, segmented by the priority level of the calls.

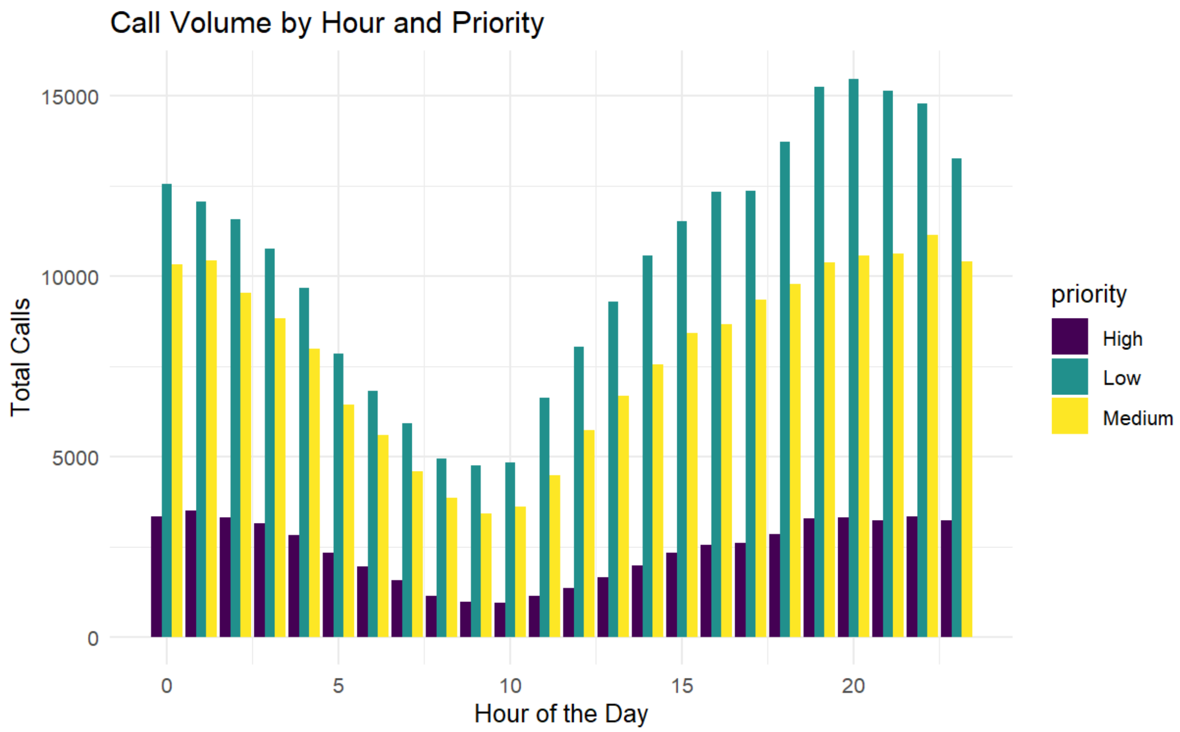

Figure Description: The graph illustrates the total call volume by hour, segmented into three priority categories: High, Medium, and Low. This visualization helps in identifying peak hours for different types of emergency calls, which is crucial for resource allocation and operational planning in emergency response services.

Appendix J: Detailed Visualization of Call Volume by Hour, Day, and Priority
This appendix provides the R code for generating a visualization that displays the log-transformed average number of emergency calls by hour, further categorized by day of the week and priority level.

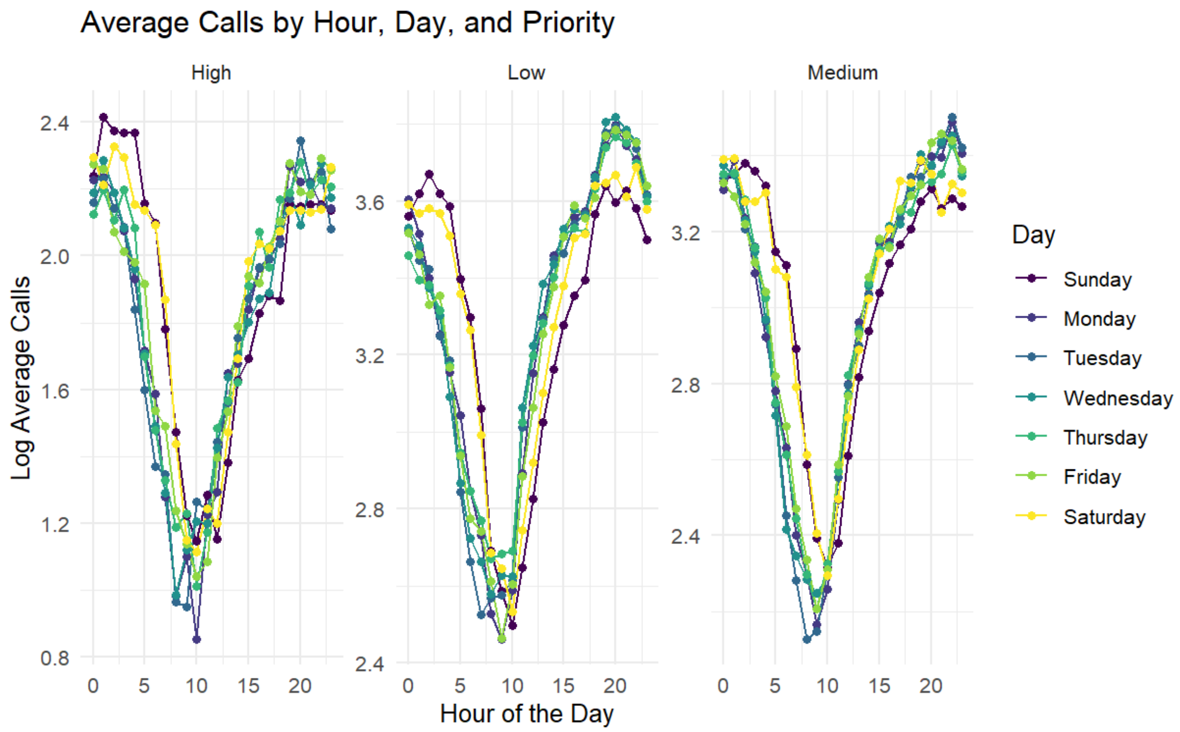

Figure Description: The graph illustrates the interaction effects between the time of day, day of the week, and call priority on the log-average number of calls received. Each facet represents a priority category, and within each facet, different days are color-coded. This level of granularity offers detailed insights into when certain types of calls are most likely to occur, which can inform staffing and resource allocation decisions.

Appendix K: Residual Plots for Model Diagnostics:

This appendix provides residual plots for the linear and Loess models used in our analysis. These plots are essential for evaluating the fit of the models and identifying any patterns in the residuals that might indicate model inadequacies.

The residual plot for the following linear model which we found to give best metric scores.

lm(formula = log_TotalCalls ~ factor(Month) + factor(Day) + factor(Hour), data = hourly_calls)

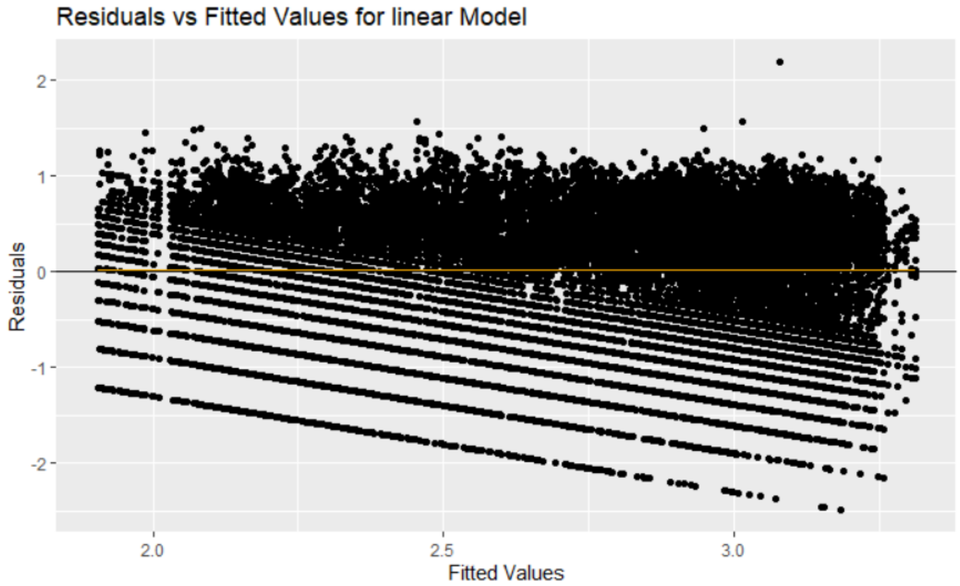

Figure 10. Residual Plot for linear_model1

The residual plot for the following Loess model which we found to give best metric scores.

loess(formula = log_TotalCalls ~ Month_Number + Day_Number + Hour + Month_Number * Day_Number * Hour, data = hourly_calls)

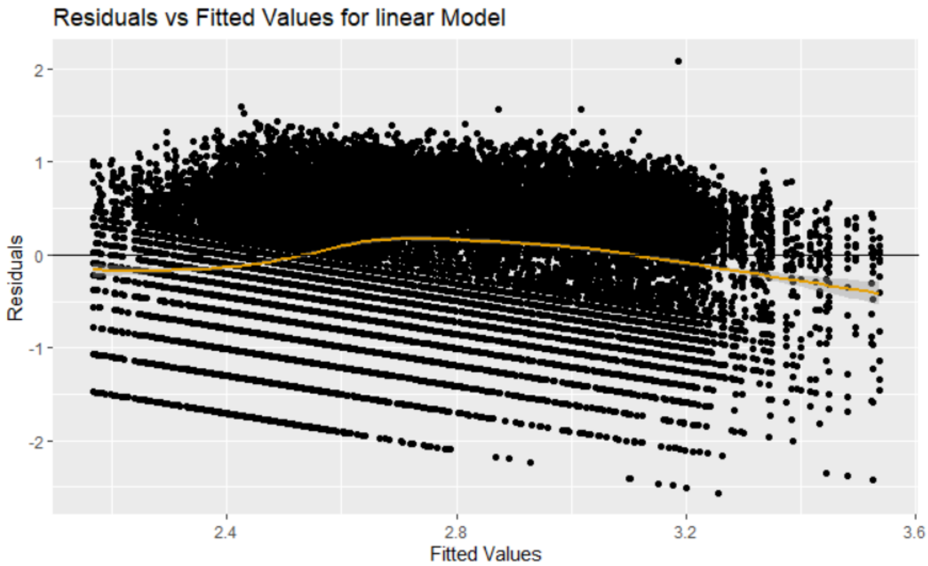

Figure 11. Residual Plot for loess_model2

The residual plot for the following GAM model which we found to give best metric scores.

gam(log_TotalCalls ~ s(Hour) + factor(Day) + factor(Month) + s(Hour, by=Day), data = hourly_calls)

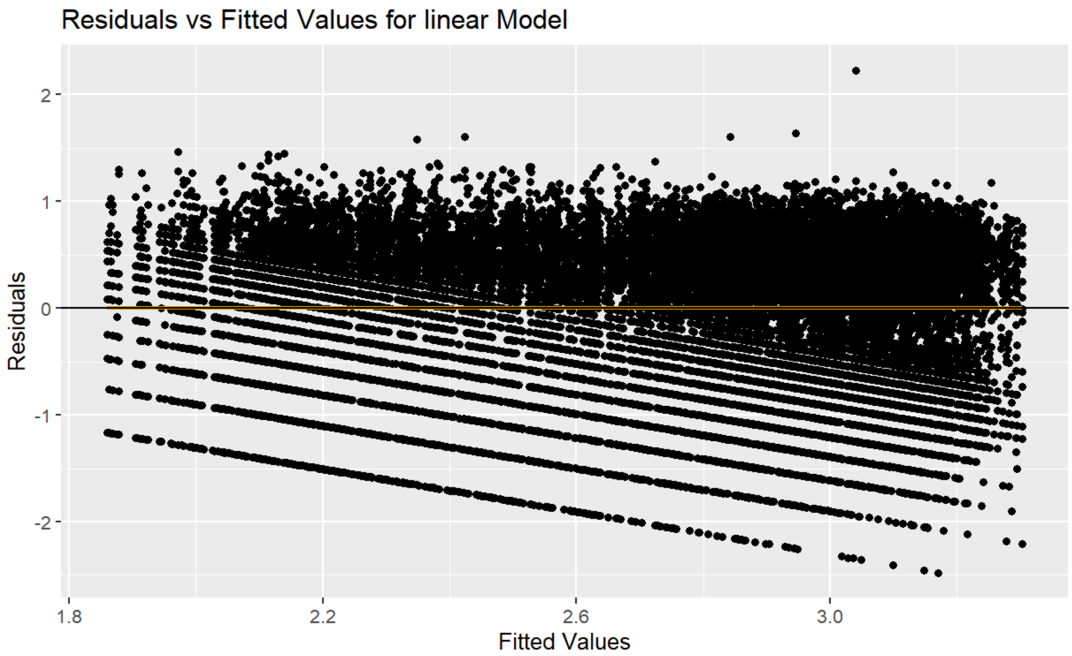

Figure 11. Residual Plot for gam_model2