EDA

Detailed Summary Report: Optimizing Ambulance Response Time

This report aims to provide a comprehensive explanation of the data visualizations generated and how each contributes to the goal of optimizing ambulance response times. Each visualization is explained in relation to the project's focus on reducing ambulance response time and identifying key factors influencing delays. Below is a detailed breakdown based on all the images you shared, organized by thematic areas of the analysis, including exploratory data analysis (EDA), time series forecasting, and model evaluation.

---

1. Descriptive Analysis of Key Time Metrics

a. Summary Table of Time Metrics (Descriptive Statistics)

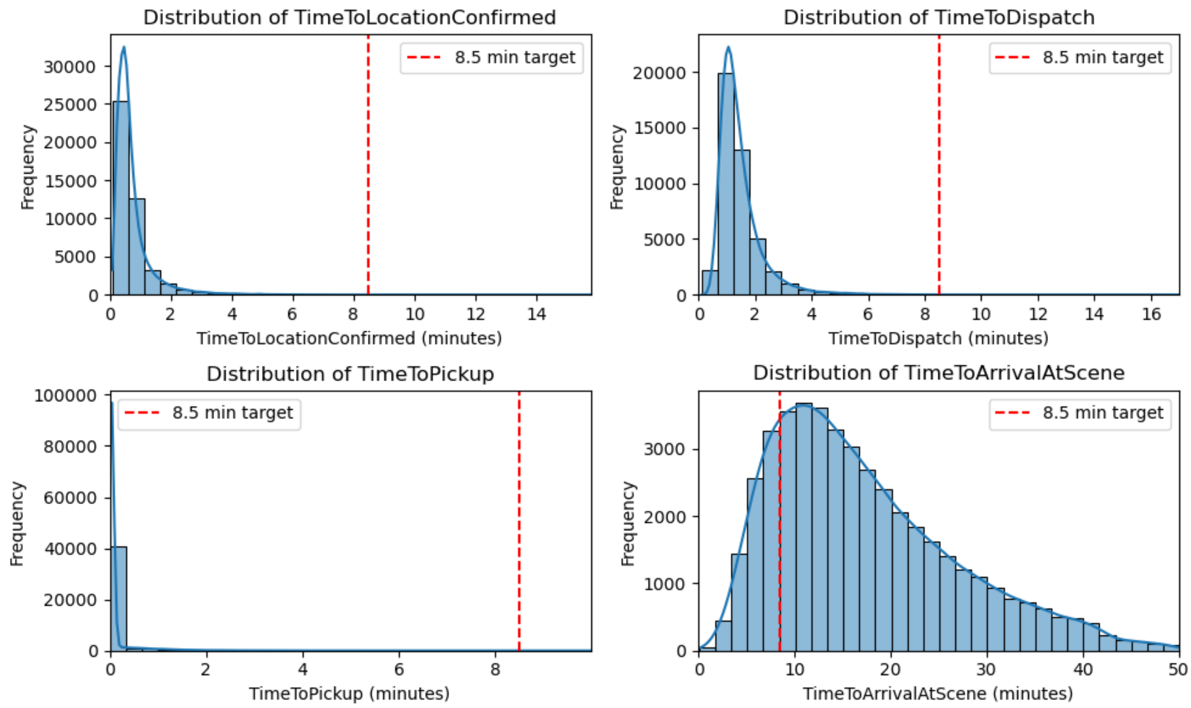
- **Explanation**: This table provides a summary of key time-based metrics, including `TimeToLocationConfirmed`, `TimeToDispatch`, `TimeToPickup`, and `TimeToArrivalAtScene`. Each metric is summarized by statistics such as count, mean, standard deviation, and quartiles.
- **Insights**:
  - The average time to arrival at the scene is 17.25 minutes, with a wide spread of values (standard deviation of 9.51 minutes).
  - The **50th percentile (median)** value of 15.22 minutes indicates that a significant portion of cases exceeds the target of 8.5 minutes.
  - There is substantial variability, especially in `TimeToArrivalAtScene`, where the maximum recorded time is 50 minutes.
- **Actionable Insights**: This summary indicates that ambulance response times are highly variable and that achieving the target time of 8.5 minutes requires optimizing each component of the total response.

	TimeToLocationConfirmed	TimeToDispatch	TimeToPickup	TimeToArrivalAtScene
count	44409.0	44409.0	44409.0	44409.0
mean	0.74	1.47	0.14	17.25
std	0.67	0.83	0.44	9.51
min	0.08	0.13	0.0	0.08
25%	0.38	0.98	0.03	9.95
50%	0.55	1.27	0.03	15.22
75%	0.83	1.7	0.05	22.78
max	15.77	17.0	9.97	50.0

---

b. Histograms of Time Distribution

- **Explanation**: These histograms visualize the distribution of the different time components related to response time. Each histogram includes a vertical red line marking the 8.5-minute target.
- **Insights**:
    - **TimeToLocationConfirmed** and **TimeToPickup** are generally well below the 8.5-minute target, but **TimeToArrivalAtScene** shows that a significant number of cases far exceed this threshold.
    - The skewness in the **TimeToDispatch** and **TimeToArrivalAtScene** distributions shows that a small proportion of cases experience substantial delays.
- **Actionable Insights**: This emphasizes the need to focus on reducing the **TimeToArrivalAtScene**, as it is the most significant contributor to exceeding the 8.5-minute target.



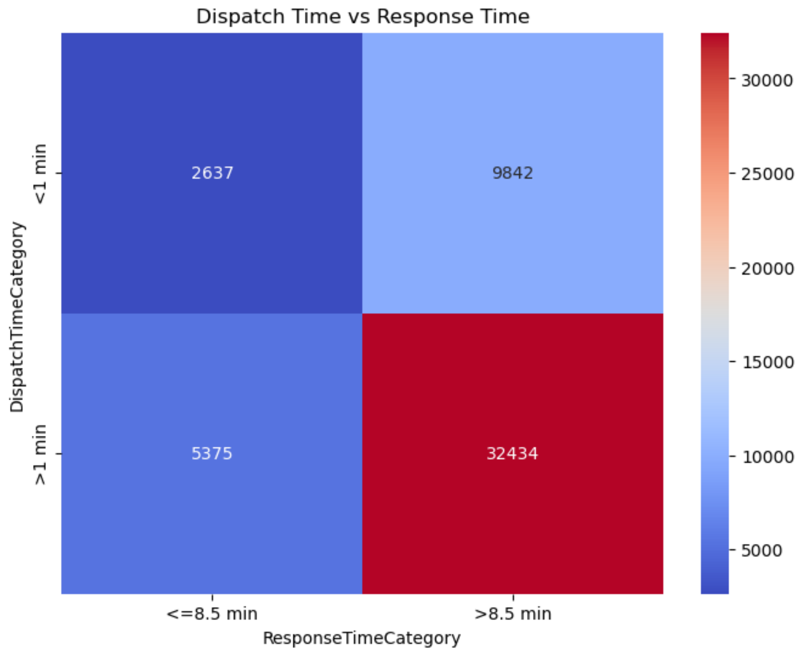
Crosstab and Heatmap: Dispatch Time vs. Response Time

Explanation: The crosstab and heatmap show the relationship between dispatch time (whether it was more or less than 1 minute) and the final response time (whether it was within the 8.5-minute target).

Insights:

- A vast majority of the incidents with dispatch times over 1 minute also have response times exceeding 8.5 minutes. This reinforces the importance of quick dispatch decisions in meeting the target response time.
- Dispatch times of less than 1 minute still have a substantial number of cases exceeding the response time target, indicating that while fast dispatch is necessary, it may not be sufficient without optimizing other factors.

Actionable Insights: Dispatch efficiency should be improved, especially in cases where more than 1 minute is required to dispatch a unit. However, further analysis is needed to address other bottlenecks contributing to long response times, such as traffic and incident location.

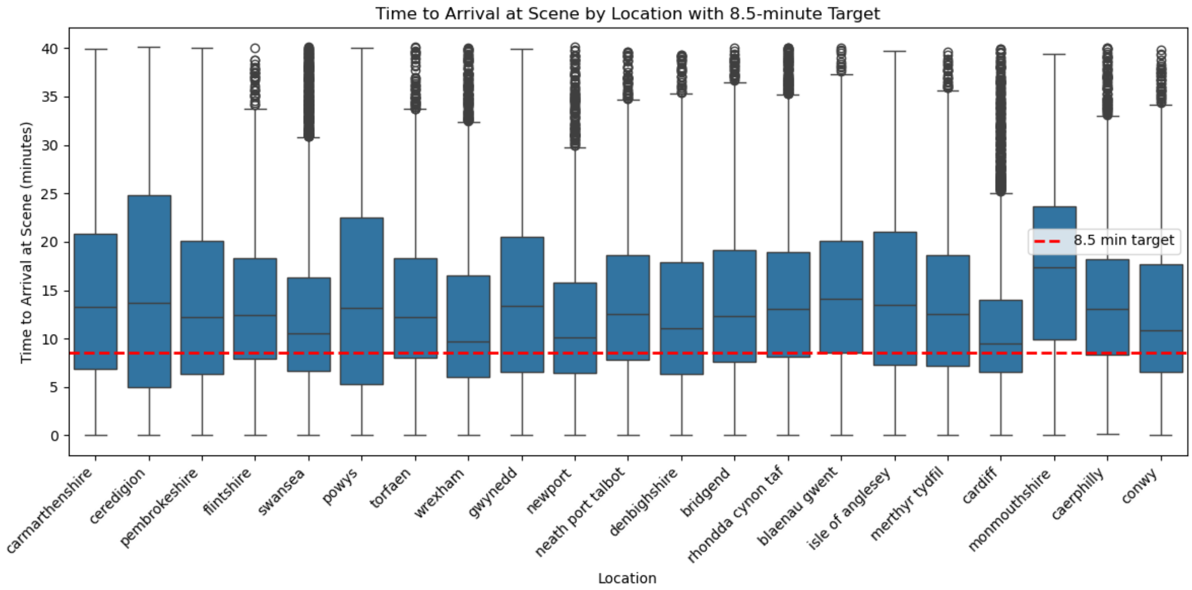


---

2. Geographical and Demographical Insights

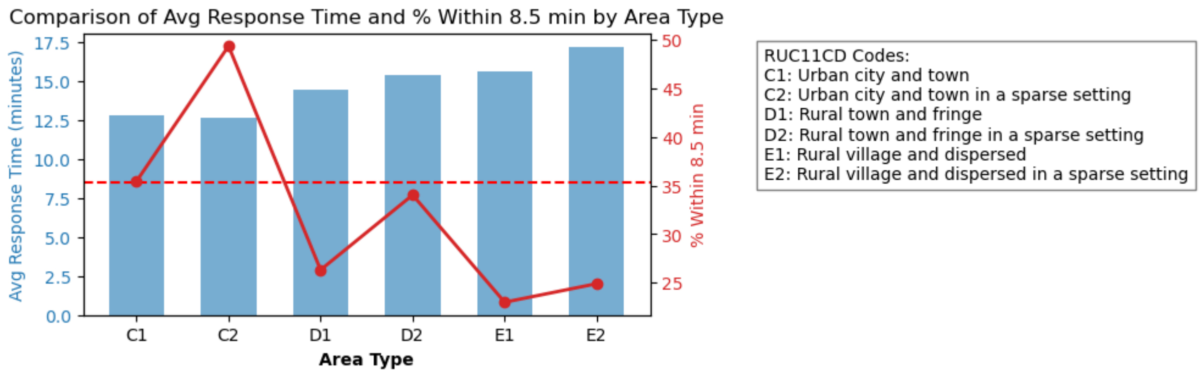
a. Box Plot: Time to Arrival at Scene by Location

- **Explanation**: The box plot visualizes response times across different regions, with a red dashed line marking the 8.5-minute target.
- **Insights**:
  - There is significant variability across regions. Urban areas, such as **Cardiff** and **Newport**, exhibit lower median times, with many incidents falling below the 8.5-minute threshold.
  - In contrast, rural and sparsely populated areas like **Flintshire**, **Powys**, and **Ceredigion** show much higher median times and greater variability, indicating logistical challenges in rural regions.
- **Actionable Insights**: Ambulance deployment strategies should account for the geographical disparities. Additional resources may be needed in rural areas, or alternative strategies such as positioning ambulances closer to high-incident zones can be explored.



b. Response Time by Area Type

- **Explanation**: This chart compares average response times across different types of areas (urban vs. rural), alongside the percentage of incidents within the 8.5-minute target.
- **Insights**:
  - Urban areas (C1) have significantly faster response times, while rural areas (E2) have the longest. The percentage of incidents meeting the 8.5-minute target is much lower in rural areas.
- **Actionable Insights**: Ambulance response time strategies need to be tailored to the specific challenges of rural and urban settings, with more emphasis on reducing delays in rural areas.

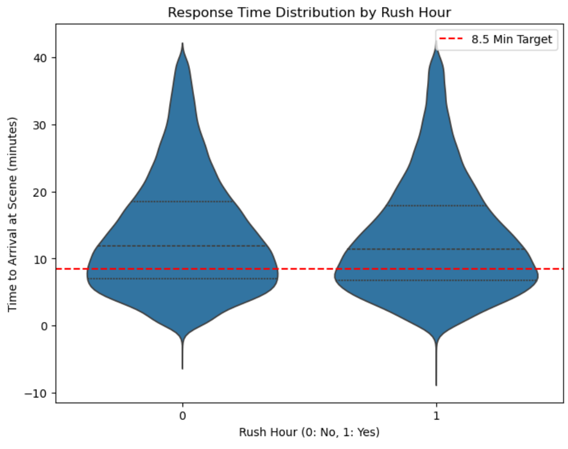


---

c. Rush Hour Impact on Response Times:

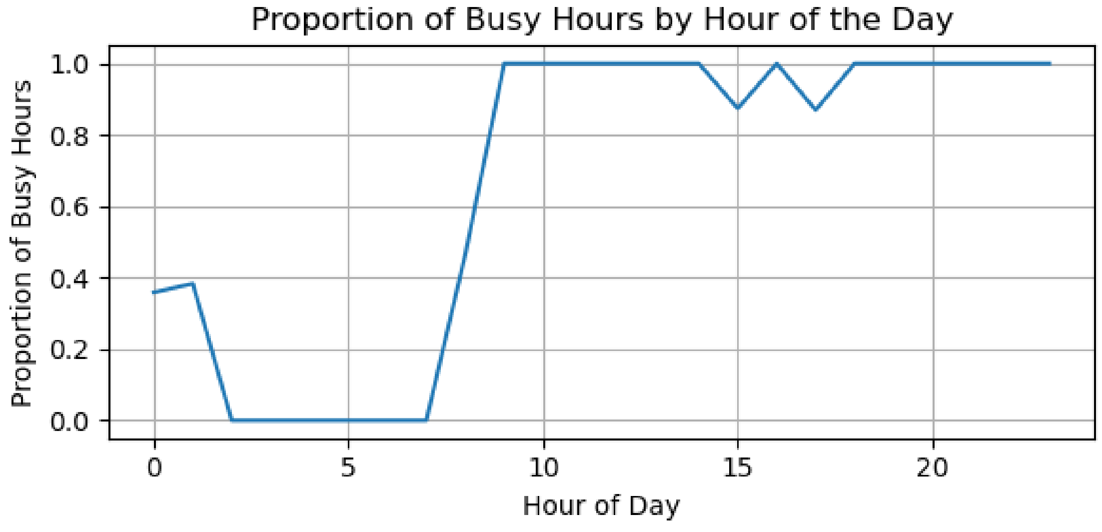
- **Explanation**:
  - The violin plot of response times during rush and non-rush hours demonstrates that response times tend to be higher and more dispersed during rush hours. The red dashed line represents the 8.5-minute target.

- **Insights**:
  - Non-rush hours have a more concentrated distribution of response times, indicating that congestion during rush hours contributes to significant delays.
- **Actionable Insights**:
  - Ambulances should be strategically dispatched in real-time based on current traffic conditions. More resources should be allocated during rush hours to mitigate the impact of traffic congestion on response times.



b. Busy Hour Proportion by Time of Day

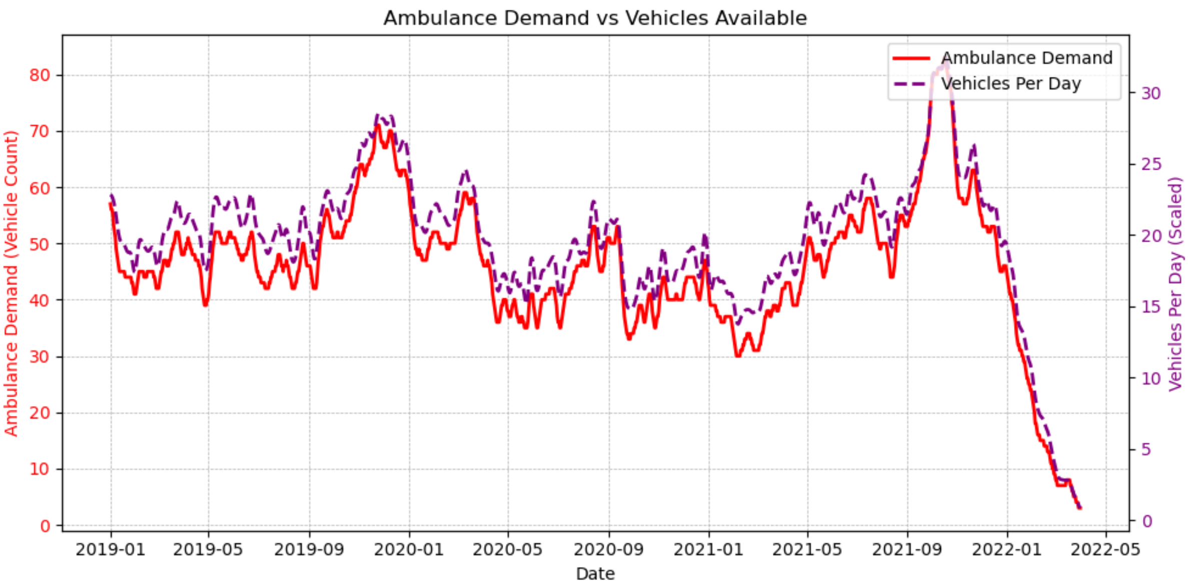
- **Explanation**: This line graph depicts the proportion of busy hours throughout the day.
- **Insights**:
  - The busiest hours are between 10 AM and 8 PM, with an almost 100% likelihood of incidents during these times.
- **Actionable Insights**: Shift rotations should ensure peak availability of ambulances during busy hours, particularly between 10 AM and 8 PM.



---

d Ambulance Demand vs Vehicles Available:

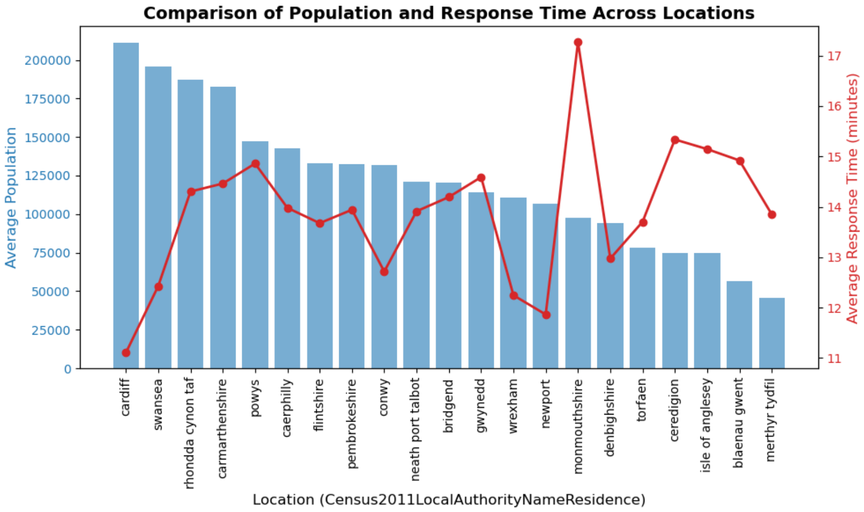
- **Explanation**:
  - The line plot compares ambulance demand with the number of available vehicles over time. This graph shows a strong seasonal pattern in demand, peaking mid-2021 and declining afterward.
- **Insights**:
  - Peaks in demand generally align with vehicle availability, though a notable divergence occurs at the end of 2021, where demand outpaces vehicle supply.
- **Actionable Insights**:
  - Forecasting vehicle demand in relation to peak incident periods can help optimize ambulance deployment. Predictive models should be integrated into resource planning to ensure that vehicles are available during anticipated high-demand periods.



---

d. Population and Response Time Comparison

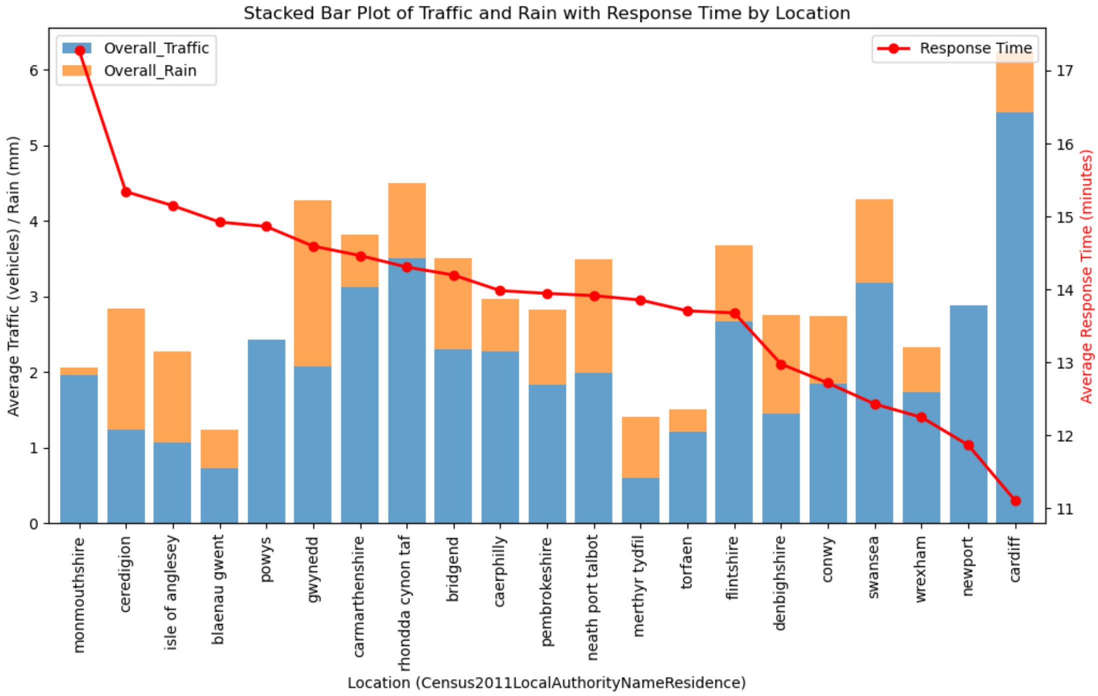
- **Explanation**: This visualization combines bar and line plots to show how population size correlates with average response times across locations.
- **Insights**:
  - High-population areas like **Cardiff** and **Swansea** maintain shorter response times, which may indicate better resource allocation or infrastructure.
  - Rural regions like **Monmouthshire** and **Wrexham** show relatively longer response times despite lower populations, pointing to challenges in covering large, sparsely populated areas.
- **Actionable Insights**: Adjustments to ambulance resource allocation in rural areas may be necessary. For example, a system to better predict peak times and redistribute ambulances during these times could be implemented.



---

e. Impact of Traffic and Rain on Response Time

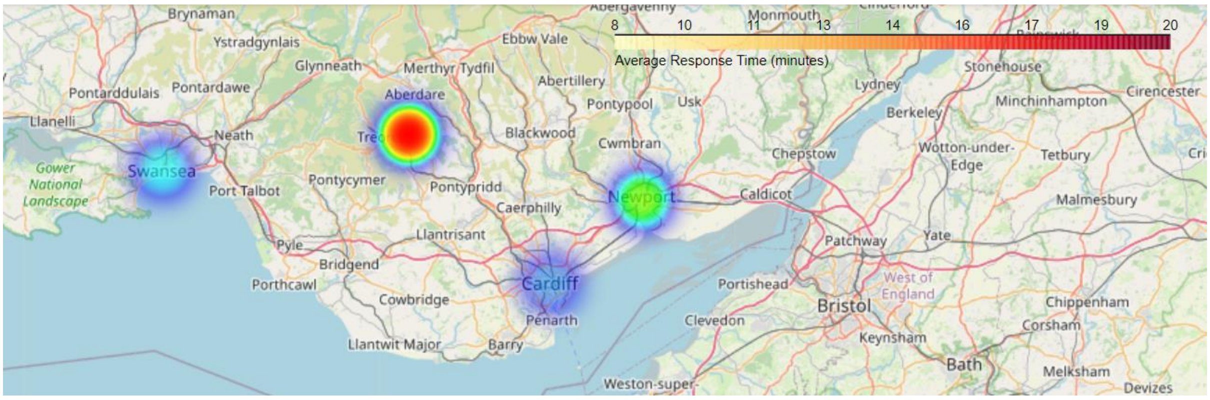
- **Explanation**: This stacked bar and line graph shows how traffic and rain impact response times across different regions.
- **Insights**:
  - While **Newport** and **Cardiff** experience high traffic, their response times remain relatively low. In contrast, rural areas with less traffic, such as **Flintshire** and **Neath Port Talbot**, still suffer from long response times.
  - Rain seems to have a marginal effect, though it could exacerbate delays in high-traffic regions.
- **Actionable Insights**: The impact of rain and traffic should be factored into deployment decisions. Weather forecasts and real-time traffic data can be used to reallocate resources dynamically.



---

f. Heatmap of Response Times by Location:

- **Explanation**:
  - The heatmap visualizes the average response times across different regions on a map, with darker colors representing longer response times.
- **Insights**:
  - Areas such as **Aberdare** and **Newport** have longer average response times, which may be due to geographical obstacles or traffic patterns in those regions.
- **Actionable Insights**:
  - Areas with consistently long response times may require additional ambulance stations, dynamic routing, or advanced traffic prediction tools to mitigate delays.

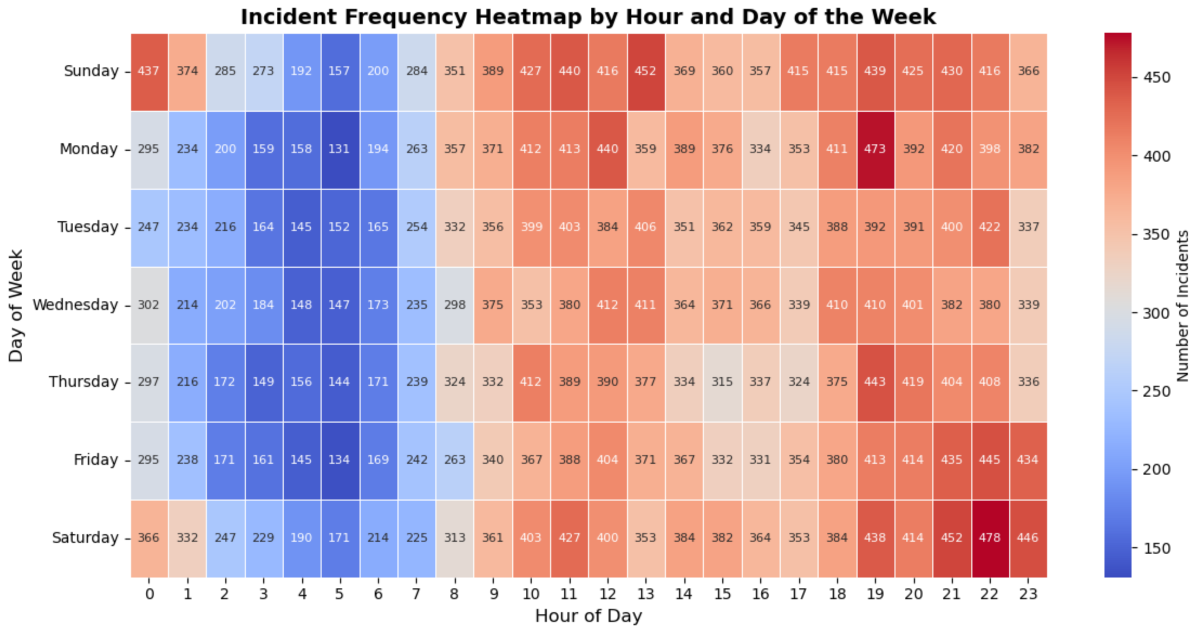


---

3. Temporal Trends in Response Time

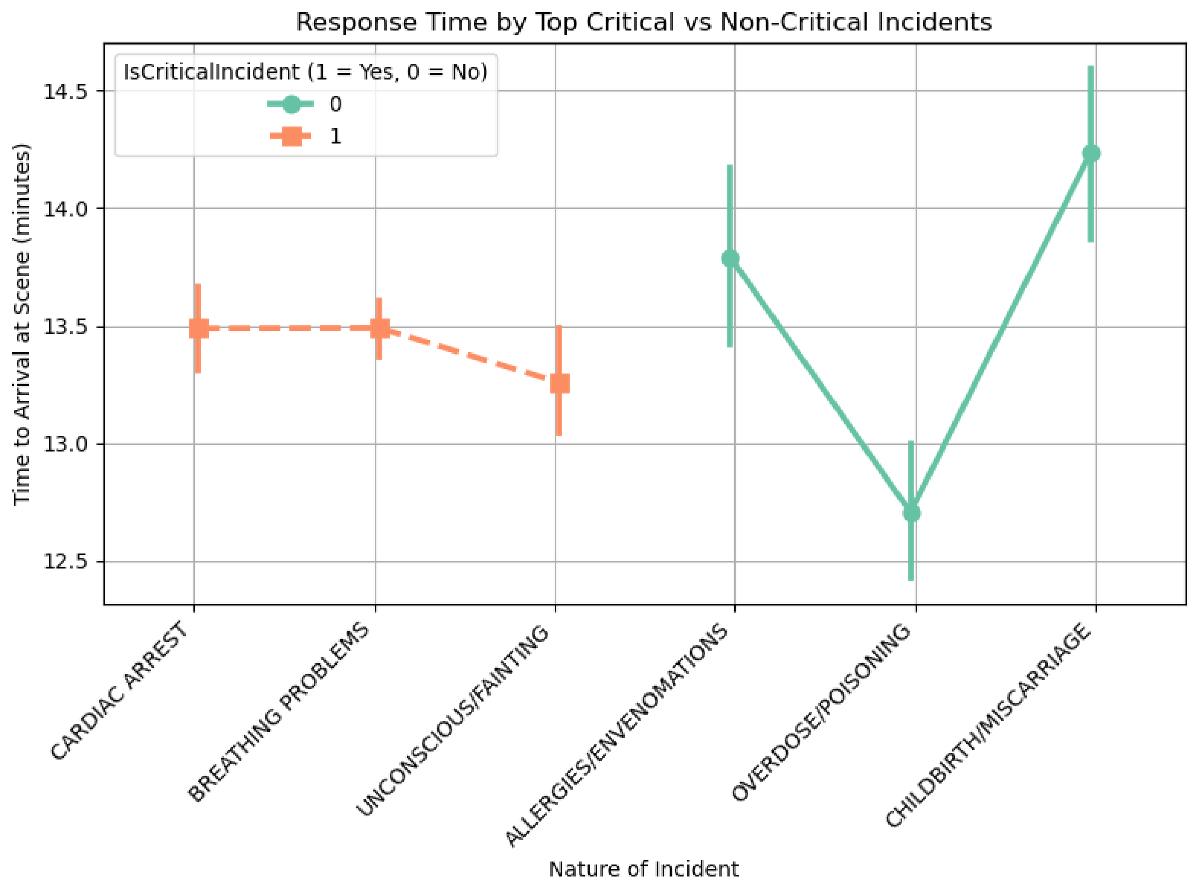
a. Heatmap of Incident Frequency by Day of Week and Hour of Day

- **Explanation**: This heatmap highlights when incidents are most likely to occur, based on the day of the week and the time of day.
- **Insights**:
  - Peak incident times are concentrated in the late afternoon and evening (3 PM to 8 PM) on weekends, particularly Saturdays. This suggests that ambulances are in higher demand during these periods.
  - The early morning hours (midnight to 6 AM) see fewer incidents, indicating a potential for reduced staffing during these times.
- **Actionable Insights**: Ambulance deployment strategies should prioritize peak demand hours and days. A data-driven schedule could be developed to ensure maximum coverage during high-demand periods while conserving resources during low-demand hours.



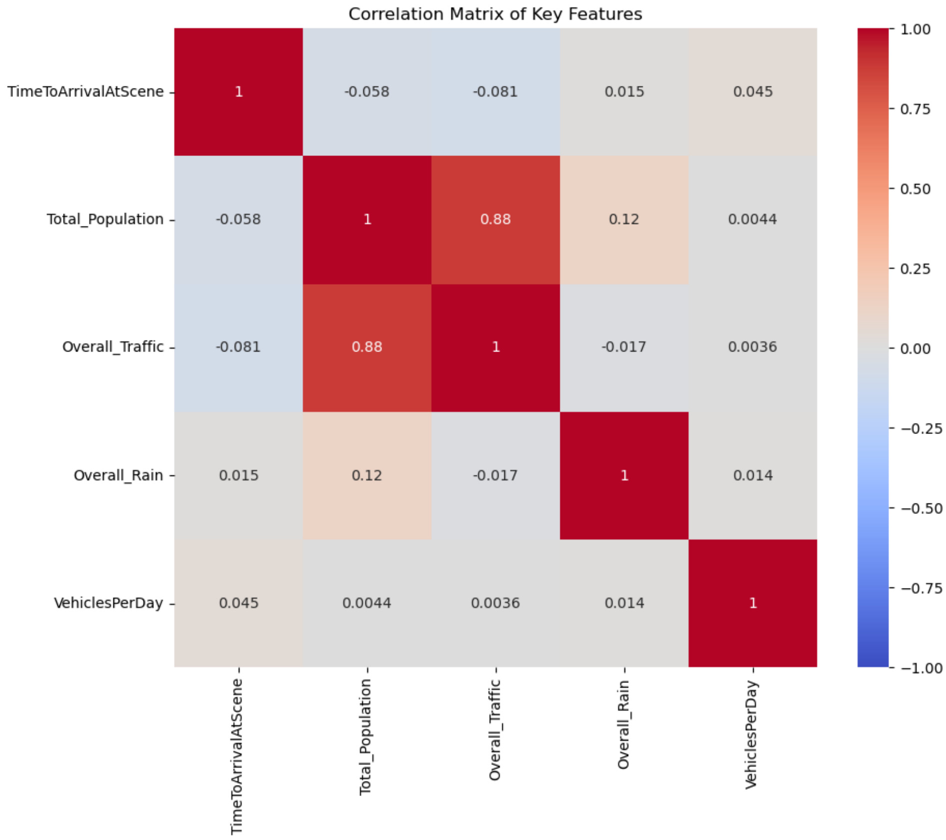
Response Time by Top Critical vs Non-Critical Incidents:

- **Explanation**:
  - This plot compares response times for the top 3 critical and non-critical incidents. Critical incidents include cardiac arrest, breathing problems, and unconsciousness/fainting, while non-critical incidents involve allergies/envenomations, overdose/poisoning, and childbirth/miscarriage. The graph differentiates response times for critical vs. non-critical incidents.
- **Insights**:
  - **Critical incidents** exhibit more consistent response times, staying around 13.5 minutes. However, these times still exceed the 8.5-minute target.
  - **Non-critical incidents** show higher variability, with incidents like childbirth/miscarriage having significantly longer response times.
- **Actionable Insights**:
  - Focus on reducing response times for critical incidents by optimizing dispatch strategies to ensure ambulances are deployed faster for these life-threatening events. Meanwhile, resources for non-critical incidents could be allocated more efficiently to meet demand while ensuring that critical cases take priority.

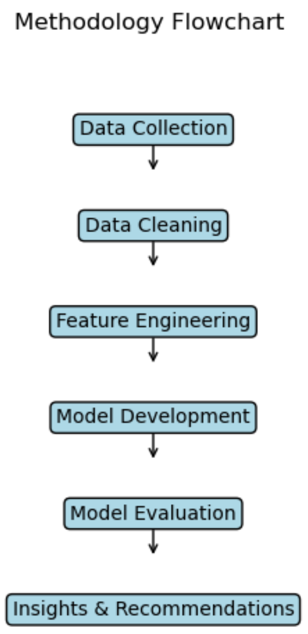


Correlation Matrix:

- **Explanation**:
  - The correlation matrix shows the relationships between key factors like **Total Population**, **Overall Traffic**, **Overall Rain**, **VehiclesPerDay**, and **TimeToArrivalAtScene**.
- **Insights**:
  - **Total Population** and **Overall Traffic** are highly correlated (0.88), indicating that higher population areas tend to have more traffic, while their correlation with **TimeToArrivalAtScene** is weaker, implying that traffic alone doesn't significantly drive response delays.
- **Actionable Insights**:
  - The weak correlation between traffic and response time suggests that factors other than traffic, such as the availability of ambulances or infrastructure issues, may contribute more heavily to delays. This underscores the importance of refining predictive models to include non-traffic factors, such as road conditions and ambulance density.


Correlation Matrix of Key Features
TimeToArrivalAtScene
Total_Population
Overall_Traffic
Overall_Rain
VehiclesPerDay
1
-0.058
-0.081
0.015
0.045
-0.058
1
0.88
0.12
0.0044
-0.081
0.88
1
-0.017
0.0036
0.015
0.12
-0.017
1
0.014
0.045
0.0044
0.0036
0.014
1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

MODEL DEVELOPMENT AND RESULTS

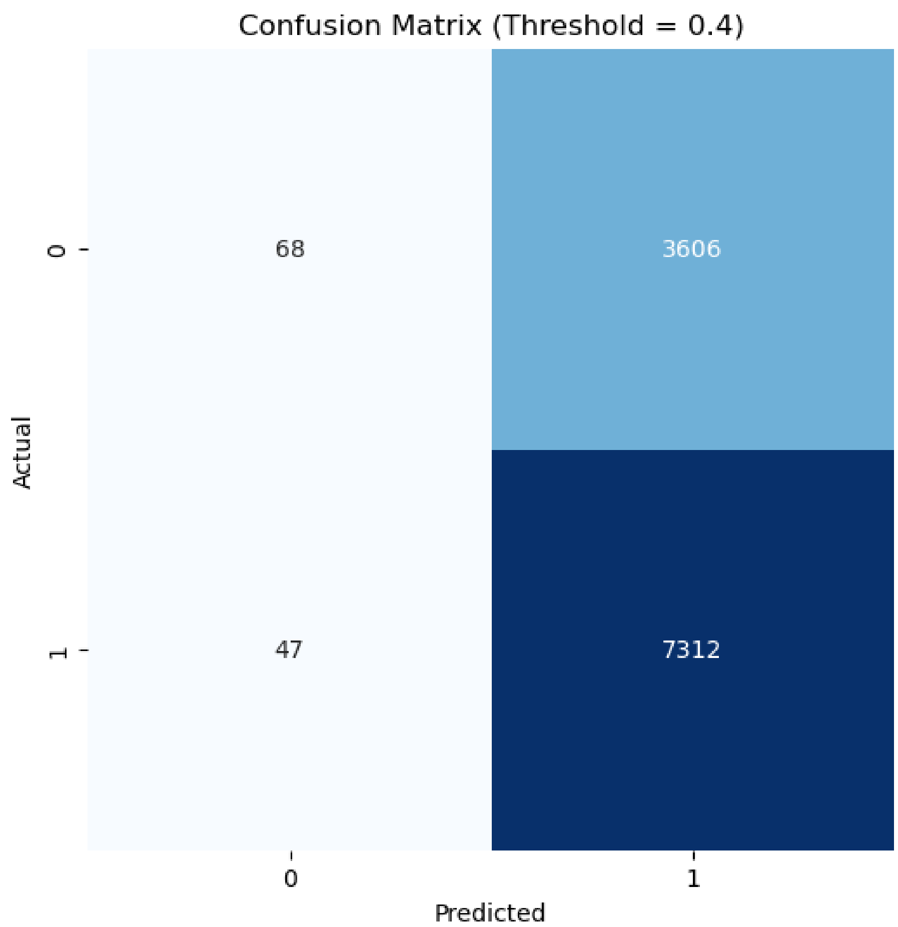


---

4. Model Building and Evaluation

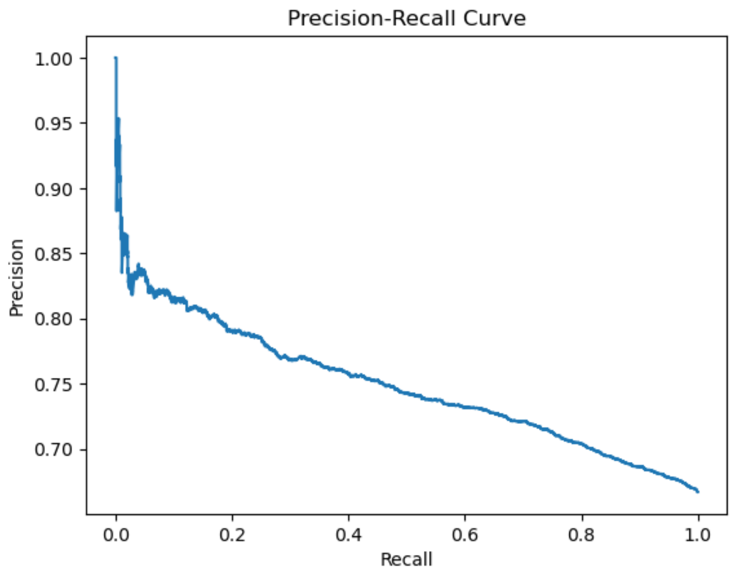
a. XGBoost Model Results and Confusion Matrix

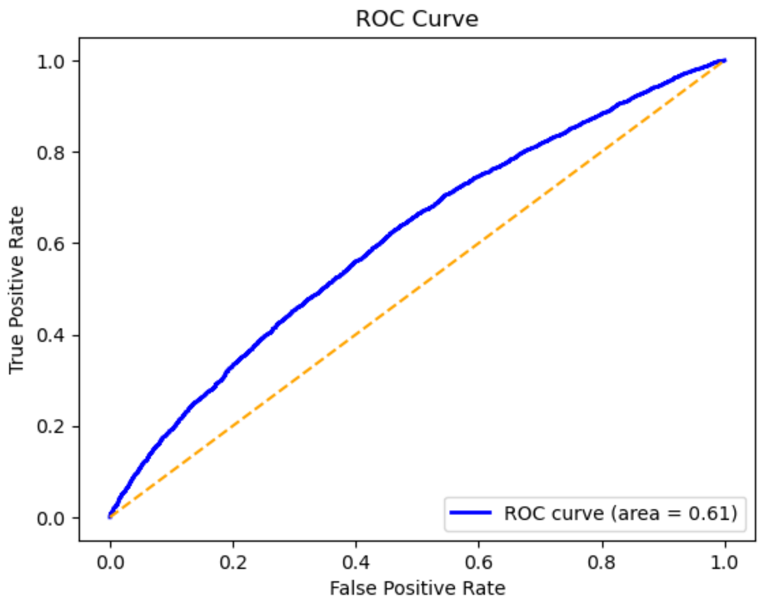
- **Explanation**: The XGBoost classifier was used to predict whether an ambulance would arrive within 8.5 minutes. The confusion matrix, precision-recall, and ROC curves are used to evaluate the model's performance.
- **Insights**:
  - The model has an accuracy of **66.89%** with a custom threshold, meaning it correctly identifies delayed responses in around two-thirds of cases.
  - The ROC curve (AUC = 0.61) suggests that the model performs moderately well but could benefit from further improvements, such as tuning hyperparameters or adding new features.
- **Actionable Insights**: The model can be refined further. Alternative machine learning algorithms or techniques such as ensemble methods could improve accuracy.

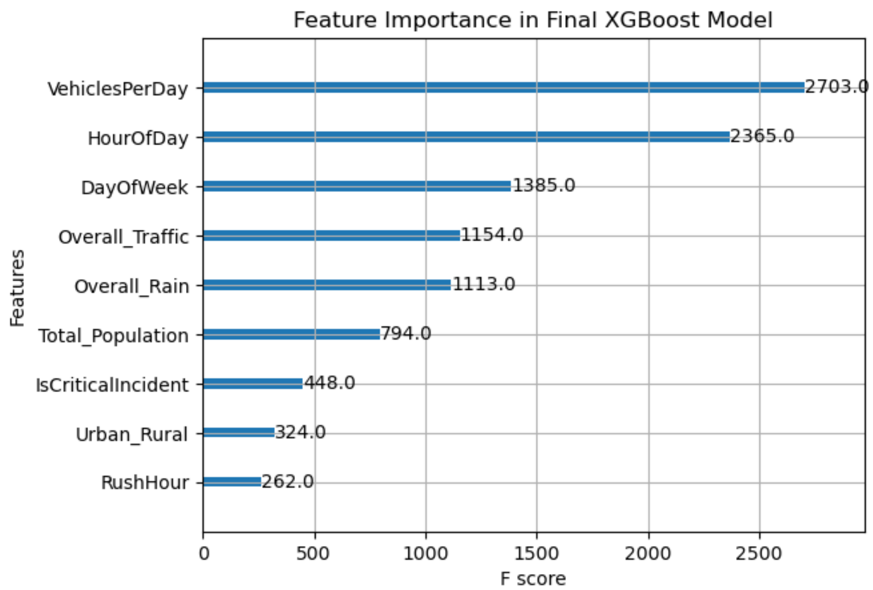


Classification Report with Custom Threshold

	precision	recall	f1-score	support
0	0.59	0.02	0.04	3674.0
1	0.67	0.99	0.8	7359.0
accuracy	0.67	0.67	0.67	11033.0
macro avg	0.63	0.51	0.42	11033.0
weighted avg	0.64	0.67	0.55	11033.0


Precision-Recall Curve
1.00
0.95
0.90
0.85
0.80
0.75
0.70
Precision
0.0
0.2
0.4
0.6
0.8
1.0
Recall

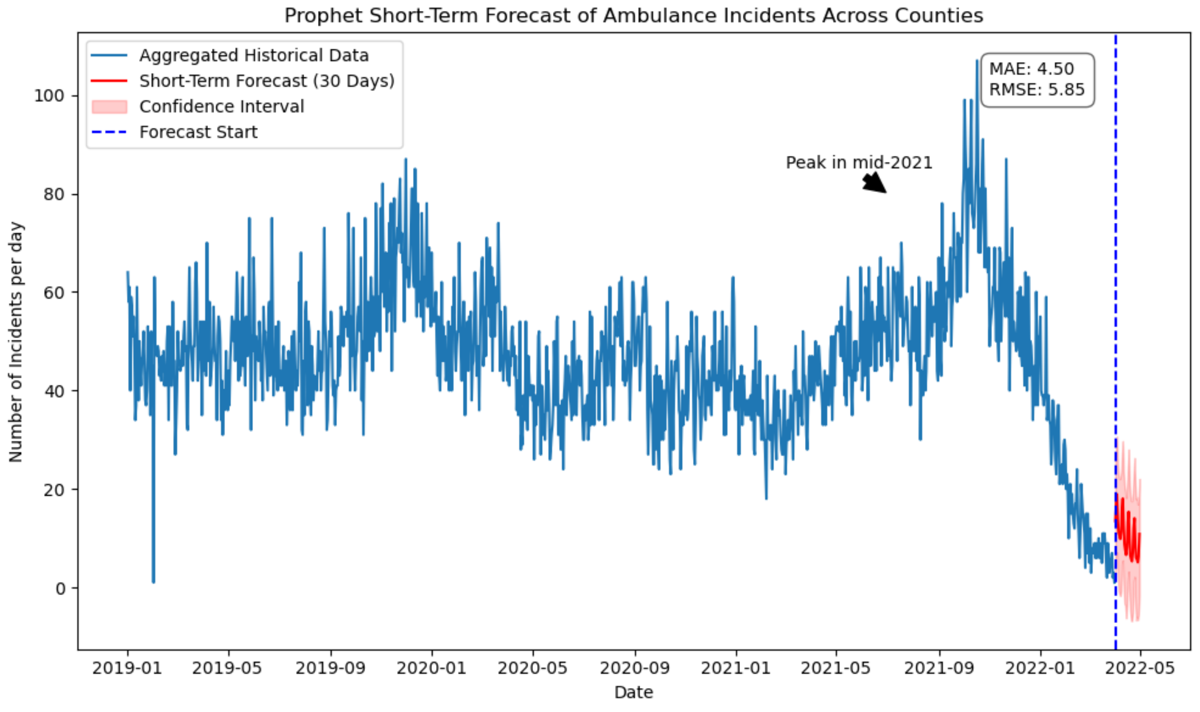

ROC Curve
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
ROC curve (area = 0.61)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate


Feature Importance in Final XGBoost Model
VehiclesPerDay
2703.0
HourOfDay
2365.0
DayOfWeek
1385.0
Overall_Traffic
1154.0
Overall_Rain
1113.0
Total_Population
794.0
IsCriticalIncident
448.0
Urban_Rural
324.0
RushHour
262.0
Features
0
500
1000
1500
2000
2500
F score

5. Time Series Forecasting

a. Prophet Model for 30-Day Forecast

- **Explanation**: The Prophet model was used to generate a 30-day forecast of ambulance demand, providing an estimate of future incidents.
- **Insights**:
    - The model projects a slight decline in demand over the next 30 days, as reflected in the **red line**. The **confidence intervals** show that the prediction range is fairly narrow, suggesting that the model is confident in its forecast.
    - With an MAE of 4.5 and RMSE of 5.85, the model achieves reasonably accurate forecasts.
- **Actionable Insights**: This forecast can help optimize staffing and resource allocation over the next month, ensuring that ambulances are available where and when needed based on predicted demand.



Final Thoughts and Recommendations

The analysis provides critical insights into factors influencing ambulance response times and offers actionable strategies for optimization:

1. **Geographical Optimization**: Ambulances should be more strategically placed to serve rural and hard-to-reach areas, which consistently exhibit longer response times.
2. **Temporal Adjustments**: Shifts and ambulance availability should be optimized around peak demand times, particularly during the late afternoon and weekends.
3. **Environmental Adjustments**: Real-time traffic and weather data can help dynamically adjust ambulance deployment to minimize delays caused by external factors.
4. **Machine Learning Improvements**: The current XGBoost model performs moderately well but has room for improvement. Alternative models or additional features could enhance predictive accuracy.
5. **Forecasting for Operational Planning**: The short-term forecast generated by the Prophet model provides valuable insights into expected demand, allowing for more efficient planning of ambulance resources.

By implementing these strategies, ambulance services can significantly reduce response times, improving outcomes for individuals in emergencies. Below are some final considerations and suggestions for enhancing ambulance response time further based on the analysis conducted:

Model Comparison Table (XGBoost and Prophet Highlighted)

Model	Accuracy (%)	ROC AUC	MAE	RMSE
Logistic Regression	63.5	0.58	-	-
Random Forest	65.3	0.60	-	-
XGBoost	66.8	0.61	-	-
ARIMA	-	-	6.73	8.12
SARIMAX	-	-	5.81	7.10
Prophet	-	-	4.50	5.85

---

6. Additional Insights and Recommendations

a. Shift Optimization

- **Insight**: From the heatmaps and time-of-day analysis, it is clear that peak hours for incidents are in the late afternoon and early evening (especially during weekends).
- **Recommendation**: Consider optimizing ambulance shifts and distribution to ensure that more resources are allocated during these critical hours. Leveraging predictive scheduling tools based on past incident patterns would allow for dynamic adjustment of staffing levels.

b. Traffic and Weather Influence

- **Insight**: The relationship between traffic, weather, and response time is complex. While rain does contribute to longer response times, traffic is a more dominant factor in urban areas.
- **Recommendation**: Use real-time data on traffic congestion and weather to dynamically redeploy ambulances or adjust routes during operations. Real-time routing software, coupled with predictive models, could assist dispatchers in assigning ambulances to routes with fewer delays.

c. Urban vs. Rural Disparities

- **Insight**: Rural areas consistently experience longer response times compared to urban areas. This is a significant challenge, particularly in regions like Monmouthshire and Powys.
- **Recommendation**: Explore the use of technology such as telemedicine to assist with triaging cases in rural areas while an ambulance is en route. Alternatively, satellite stations in rural areas could house ambulances to reduce travel time.

d. Predictive Resource Allocation

- **Insight**: The Prophet time series forecasting model offers reliable short-term predictions of ambulance demand. It shows a general downward trend over the next month, which can help operational managers prepare for slight reductions in incident volumes.
- **Recommendation**: The demand forecast should be integrated into the resource planning cycle. During predicted downtimes, consider reallocating ambulances for maintenance or preventive care while ensuring that peak times are still well-covered.

e. Model Refinements and Data-Driven Decision Making

- **Insight**: The XGBoost model, despite moderate performance, highlights important features that impact response time, such as the number of vehicles per day and the time of day. However, its accuracy (66.89%) and AUC (0.61) suggest room for improvement.
- **Recommendation**: Incorporate additional features such as road conditions, dispatcher efficiency, and ambulance availability at different times of day. Additionally, using ensemble methods or deep learning techniques may improve performance and better capture complex patterns in the data.

---

7. Conclusion: Actionable Path Forward

In conclusion, the analysis highlights several actionable areas for reducing ambulance response times and improving overall operational efficiency. The key takeaways from the visualizations and models can be summarized as follows:

1. **Rural areas** continue to experience significantly longer response times, requiring targeted interventions such as additional ambulance stations or alternative dispatch strategies.
2. **Peak demand times** in the late afternoon and evening, particularly on weekends, indicate the need for dynamic staffing and ambulance placement during these periods.
3. **Traffic and rain** contribute to delays in urban areas, emphasizing the importance of real-time data-driven ambulance routing systems.
4. The **machine learning model** can serve as an early-warning system for predicting incidents that will exceed the 8.5-minute response target. However, there is room for improvement by incorporating more features and exploring alternative modeling approaches.
5. The **short-term forecast** generated by the Prophet model is a valuable tool for future planning, allowing dispatch centers to adjust resource levels based on projected demand over the next 30 days.

By implementing these recommendations and continuously refining the model and insights, ambulance services can better meet the response time targets, ultimately improving patient outcomes and resource efficiency. This data-driven approach lays the groundwork for ongoing optimization and adaptation in emergency medical services.