Cameron Vinson

Joseph Sanchez

Jackson Griffin

An Nguyen

Advisors: Dr. Kyle Teller, Dr. Jiacheng Cai, Fire Chief Rob Frampton

Analyzing Costs Related to Structure Fires in Salisbury, MD, in 2023

May 18, 2025

Code: <https://github.com/camv2500/Analyzing-Costs-Related-to-Fires-in-Salisbury-MD>

Contents

[I. ABSTRACT 3](#_Toc198411395)

[II. INTRODUCTION 3](#_Toc198411396)

[III. RELATED WORK 4](#_Toc198411397)

[IV. METHODS 5](#_Toc198411398)

[V. RESULTS 7](#_Toc198411399)

[VI. FUTURE WORK 30](#_Toc198411400)

[VII. CONCLUSION 31](#_Toc198411401)

[VIII. AUTHOR CONTRIBUTIONS 31](#_Toc198411402)

[IX. REFERENCES 31](#_Toc198411403)

[X. APPENDIX 32](#_Toc198411404)

# ABSTRACT

**The economic impact of fire incidents extends beyond the losses of property and contents typically reported. Although the fire department keeps a list of property and content losses for each fire they respond to, the total losses from the given fires are not always understood. This project gathered data from available sources to create data sets to enhance understanding of the costs associated with structure fires in Salisbury, MD. Data included a collection of public data sets called NFIRS 2023 from FEMA, data sets from the City of Salisbury’s GIS Portal for geographic data, and weather data from Weather Underground. Using statistical analysis and models such as multiple linear regression and geographic tools such as QGIS, the project team analyzed and visualized the data sets we created to investigate and address issues related to economic costs associated with fires. One relationship we investigated was whether there was a correlation between the response time and the amount of the loss. We also checked for a correlation between the time to arrive and the number of firefighters to send in addition to how the length of time to extinguish the fire affects the overall loss amount. This study also tried to determine any correlation between times of day and amount of the loss. Future questions we proposed included: whether the presence of a smoke detector or proximity to a fire station had any effect on the overall expense incurred, whether any other geographic patterns are relevant, tax losses from building fires, and possible correlation between weather data and the amount of the loss.**

# INTRODUCTION

To understand the total losses from given building fires, fire cases from the FEMA dataset (Federal Emergency Management Agency, 2024) were used, as well as the City of Salisbury’s GIS portal (Cupp, 2025) for the year 2023 in Salisbury, Maryland. The primary variable that was investigated was the “Time to extinguish”. The time to extinguish house fires was analyzed to identify areas in the fire response process that can be changed to affect the overall cost. Relationships between time to extinguish and variables such as total loss, the number of personnel, and personnel costs, were included in this analysis. Geospatial visualizations to explain the cost were also included.

Our main analysis involved creating a model to show how the time it takes to extinguish a fire and number of personnel involved affects the overall cost of the loss. Then, we worked on geospatial visualizations to explain the findings. The linear equations used to drive our analysis were based upon:

,

, and

where , , and are the coefficients and , , and are the intercepts. The initial hypothesis was that an increase in the number of personnel would decrease the time to extinguish a fire and a lower time to extinguish a fire will decrease both the total loss and personnel cost.

# RELATED WORK

Previously published analyses of fire incidents include statistical data, national profiling, and the integration of environmental variables. Our study contributes to this body of work by evaluating the economic costs associated with structure fires, using municipal geospatial and weather data specific to Salisbury, MD.

Within the NFPA’s U.S. Fire Department Profile (NFPA, 2023) are incident frequencies, personnel statistics, as well as average response times across jurisdictions. This profile gives a national view of work done by various fire departments as well as listing various codes and standards. The work of the NFPA is thorough in its scope, but it aggregates data at a national level only. Regional or municipal specifics are not accounted for within this work. Our group used regional data from the City of Salisbury, Maryland (Cupp, 2025) as well as FEMA (Federal Emergency Management Agency, 2024). Extending this type of work, our project applies similar metrics like response time and personnel, but our analysis is specific to Salisbury, MD fire districts to assess how those variables impact local financial outcomes.

The U.S. Fire Administration’s state-level fire statistics portal (USFA, 2024) provides a more granular understanding into incident counts and losses broken down by state. These statistics can provide valuable baseline comparisons. Similar to NFPA, the data is descriptive and does not integrate multivariate statistics or predictive models. Our work differed with our use of regression models as well as machine learning models like Random Forest to quantify how features in the data like time to extinguish and number of personnel are significant and how they impact the economic loss associated with a building fire.

Fire risk modeling integrates exogenous factors. Weather represents one such factor, and its integration by recent studies has been explored. Warmer temperatures plus earlier spring onset relates to increased wildfire activity in the western U.S., as Westerling et al. (2006) showed. Fire impacts in urban areas may also worsen due to climate conditions because of this relationship. Our work included weather data coming from Weather Underground (2023) to explore and explain any significance, even though this study focused on building fires rather than wildfires.

A study was conducted by Gong et al. (2025) where machine learning was used for feature selection and regression techniques were used for forecasting fire emergency response times throughout Chinese megacities. They quantified how various factors predict time-sensitive outcomes with multiple modeling techniques (including Random Forest) similar to our approach. However, their study is focused upon optimizing logistics for emergency response, while our goal was for us to understand how time-related variables have an influence upon economic loss, not just efficiency.

For our approach, these works do together provide a rationale and a strong foundation. This research integrates municipal data with GIS, weather, plus district-level spatial analysis, concentrating on urban structure fires and local economic costs. For jurisdictions that are smaller like Salisbury, MD, this local and integrative approach makes our work valuable for municipal policy, fire resource allocation, and disaster planning.

# METHODS

*A. Data Collection and Simplification*

To select and transform data for the dataset, Microsoft SQL Server and Python were used. The FEMA USFA NFIRS data file (Federal Emergency Management Agency, 2024) was imported into a relational database table. Then, SQL was used to filter the FEMA data to include only rows for Salisbury, MD, only rows that had at least a property loss or content loss reported from a fire, and only building fires. The tables were exported to CSV using SQL Server tools. Python was used to join the tables into one dataset to be exported as a CSV. The dataset was copied and pasted into an Excel spreadsheet where we added features we needed and removed the features we did not need. Weather data, including wind direction, was merged into the data from Weather Underground from a combination of Python coding and manual edits (Weather Underground, 2023).

*B. Geospatial Data Analysis using Python and QGIS*

Google Maps was used to geocode the resulting data, and a new CSV data file with all the merged features was created from the Excel document. The Geopandas Python library was employed to convert the file from a CSV to a GeoJSON filetype that could be loaded into QGIS, an open-source GIS program. Using Python programming, we also provided statistics for each fire station and service district for the maps. Maps of all the different building fires in the data were created in QGIS. The fires were symbolized by the fire station that responded to the fire based on the FEMA data (2024), the geocoding, and the district polygons acquired from the City of Salisbury’s GIS portal (Cupp, 2025).

*C. Linear Regression*

We used a linear egression algorithm written in R to find relationships between the costs and time. The program was used to determine the significance of data features, to evaluate plots, to visualize the scatterplots, and to predict the costs of the building fires. This analysis was applied to the overall dataset, for the service response districts, and for the fire stations. As previously stated, we used this formula,

,

, and

where , , and are the coefficients and , , and are the intercepts. Transformations were also done using R, along with visualizing the models in red, the confidence intervals in blue, and the prediction intervals in green. We used 90% confidence and 90% prediction intervals for the transformations. The transformations were not done at the service response district, as service district 2 had only one data point.

*D. Random Forest Regression*

Random Forest Regression, a machine learning technique, was used to determine the importance of various features in the dataset and to predict the time it takes to extinguish the fire. Random Forest was also used to clean some of the data by predicting values for missing rows. The Scikit Learn library in Python was used to perform the Random Forest Regression. This was done for the dataset as a whole and repeated for each fire station.

*E. Grid Search Cross Validation Random Forest*

With the top 5 features determined from the first overall dataset random forest model plus the longitude and number of personnel features from the data, we used a cross-validation technique to predict the time to extinguish using 100 trees with the features determined from the overall feature importance along with longitude and number of personnel. One-hot encoding was used on the categorical features to convert them to numeric data that could be evaluated by the tree models. The tree with the lowest mean squared error was visualized for the overall dataset and each fire station. In addition, another random forest model was used on the same data but split with 80% training and 20% testing to evaluate the accuracy of the model using R2, MAE (Mean Absolute Error), MSE (Mean Squared Error), and RMSE (Root Mean Squared Error).

For the parameter grid, we used the following:

* max\_depth: [4, 5, 6]
* min\_samples\_split: [5, 10]
* min\_samples\_leaf: [5, 10]
* max\_features: ['sqrt', 'log2']

Fire Station 2 was a special case. Since the number of data points was much smaller, this parameter grid was used instead.

* max\_depth: [4, 5, 6]
* min\_samples\_split: [2]
* min\_samples\_leaf: [1]
* max\_features: ['sqrt', 'log2']

*F. Correlation*

The Pandas and NumPy Python libraries were used to make correlation matrices to show the correlations for features that were significant to the overall cost associated with the fires.

*G. Visualization*

The Seaborn and Matplotlib Python libraries were used to create scatterplot visualizations. A Seaborn heatmap was used to visualize correlation matrices.

*H. Kernel Density Analysis and Zonal Statistics*

The team used a method called Kernel Density Estimation to estimate and visualize the density of the building fires. Zonal statistics were then used to get the mean density of each service district, and a map was created using QGIS.

# RESULTS

As seen in Figure 1, the relationship of the time to extinguish is significant with the total loss and personnel cost, and the number of personnel is not significant with the extinguish time. However, there are a few outliers. Fire Station 1 shows significance for the number of personnel with the time to extinguish. Service district 2 only has one data point and will return NaN, or missing values, as the result. For this reason, we decided to omit the service response districts from the random forest trees and the transformed linear regression analysis.

A map of a city

AI-generated content may be incorrect.

A close-up of a list

AI-generated content may be incorrect.

***Figure 1:*** *Map, statistics, and linear models of Fire Stations and Service Districts*

As shown in the correlation heatmaps in Figures 2-8, the total loss and the personnel cost are highly correlated with the time to extinguish in the positive direction. The number of personnel is weakly correlated with the time to extinguish in the positive direction. For Fire Station 1 (Figure 3) and Service District 1 (Figure 7), the number of personnel is strongly correlated with the time to extinguish in the positive direction. These findings agree with our map in Figure 1 and support the significance of the features that were examined.

A chart of heatmap for overlay

AI-generated content may be incorrect.

***Figure 2:*** *Correlation Heatmap for the overall dataset*

A screenshot of a graph

AI-generated content may be incorrect.

***Figure 3:*** *Correlation Heatmap for Fire Station 1*

A chart of different colors

AI-generated content may be incorrect.

***Figure 4:*** *Correlation Heatmap for Fire Station 2*

A screenshot of a graph

AI-generated content may be incorrect.

***Figure 5:*** *Correlation Heatmap for Fire Station 1*6

A screenshot of a graph

AI-generated content may be incorrect.

***Figure 6:*** *Correlation Heatmap for the Municipal District (Inside City Limits)*

A chart of the heatmap for service distancing

AI-generated content may be incorrect.

***Figure 7:*** *Correlation Heatmap for Service District 1*

A screenshot of a graph

AI-generated content may be incorrect.

***Figure 8:*** *Correlation Heatmap for Service District 16*

The model is linear for the personnel cost. Figures 9-15 show that the cost per extinguished minute rises for both the personnel cost and the total loss. The results of the total loss graphs were not linear. A logarithmic transformation was done as ln(Total Loss) = ArcTan(Extinguish Time) + b, where b is the intercept. The transformation was used to produce confidence intervals and prediction intervals (Figure 16).

A graph of a number of times

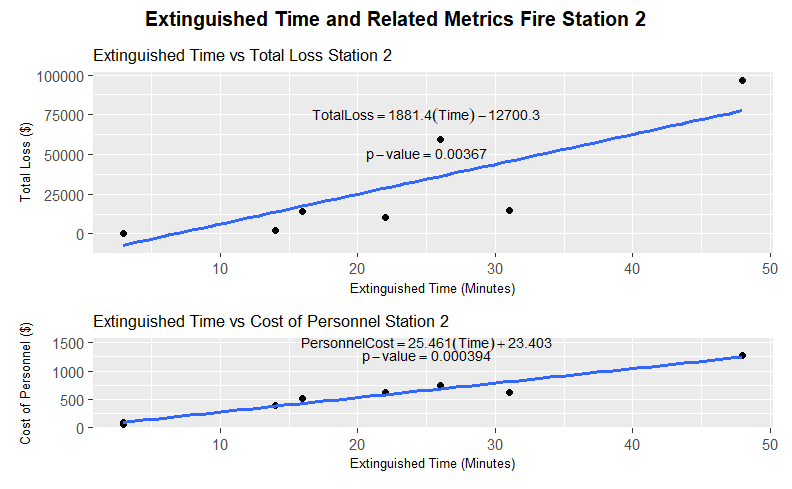
AI-generated content may be incorrect.

***Figure 9:*** *Scatterplots and model visualizations for the overall dataset*

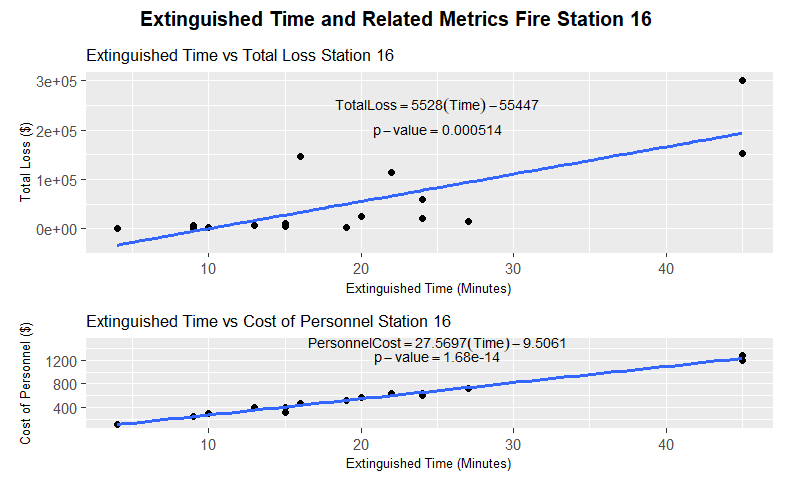
A graph of extinguishing time and a fire station

AI-generated content may be incorrect.

***Figure 10:*** *Scatterplots and model visualizations for Fire Station 1*



***Figure 11:*** *Scatterplots and model visualizations for Fire Station 2*



***Figure 12:*** *Scatterplots and model visualizations for Fire Station 1*6

A graph of a number of numbers and a line

AI-generated content may be incorrect.

***Figure 13:*** *Scatterplots and model visualizations for the Municipal District*

A graph of a service

AI-generated content may be incorrect.

***Figure 14:*** *Scatterplots and model visualizations for Service District 1*

A graph of a number of times

AI-generated content may be incorrect.

***Figure 15:*** *Scatterplots and model visualizations for Service District 16*

For the overall dataset, Figure 16 shows the transformed model for predicting the total loss with the extinguish time along with the confidence and prediction intervals.

A graph with a green and blue line

AI-generated content may be incorrect.

***Figure 16:*** *Model (red), Confidence (blue), and Prediction (green) regions for the total loss vs extinguish time for the overall dataset*

The first random forest model for the overall dataset (Figure 17) determined feature importances for predicting the time to extinguish. The top 5 most important features based on the model were FIRE\_SPRD (Fire Spread), OTH\_APP (Other Apparatus), HEAT\_SOURC (Heat Source), Latitude, and Wind Direction. To keep everything consistent, we used these 5 features plus the number of personnel (SUP\_PER) on both the overall analysis and the analysis for each fire station and added Longitude to allow the model to use the entire geographical coordinate if necessary. Since response time was not found to be as significant, we decided not to analyze the response time for this project.

A graph with a blue line

AI-generated content may be incorrect.

***Figure 17:*** *Exploratory Random Forest Model showing the features from most important to least important for the overall dataset*

For the categorical features used:

FIRE\_SPRD – “The extent of fire spread in terms of how far the flame damage extended. The extent of flame damage is the area burned or charred and does not include the area receiving only heat, smoke, or water damage” (U.S. Fire Administration, 2015). The values for the fire spread are explained from the U.S. Fire Administration. (2015) as follows:

* 1 means that the fire spread is confined to the object of origin,
* 2 means that the fire spread is confined to the room of origin,
* 3 means that the fire spread is confined to the floor of origin,
* 4 means that the fire spread is confined to the building of origin, and
* 5 means that the fire spread is beyond the building of origin.

OTH\_APP – Other Apparatus “includes specialized or non-standard vehicles and equipment that don't fall under primary categories like engines, ladders, or rescue units.” (OpenAI, 2024; Federal Emergency Management Agency, 2024).

Other apparatuses are used for specific purposes. Examples of these types of vehicles include drones, boats, or rescue support vehicles (OpenAI, 2024; Federal Emergency Management Agency, 2024). In our dataset, the values represent the number of other apparatuses that were onsite.

HEAT\_SOURC – “The heat source that ignited the Item First Ignited (Block D3) to cause the fire” (U.S. Fire Administration, 2015). Codes for heat source, which is explained from the U.S. Fire Administration. (2015) are:

* 10 – 14: caused from operating equipment,
* 40 – 43: caused from a hot or smoldering object,
* 50 – 56: caused from explosives or fireworks,
* 60 – 69: caused from other open flame or smoking materials,
* 70 – 74: caused from chemical or natural heat sources,
* 80 – 84: caused from a heat spread from another fire; not caused from operating equipment,
* 97: caused by multiple heat sources,
* 00: caused by other heat sources,
* UU: undetermined.

Wind Direction (Weather Underground, 2023) – The direction the wind was blowing when the fire happened. The directions are abbreviated. For example, SSW. CALM means there was no wind blowing when the fire happened.

Latitude and Longitude use the decimal degrees system.

Random forest tree models, such as the one in Figure 18, which had the highest accuracy of all the trees produced, were used to predict the time to extinguish future fires. The representative tree shown in Figure 18 for the overall data set can be interpreted as follows. If the fire spread\_value\_4.0 value (confined to building of origin – value 4 in the chart above) is less than or equal to 0.5, the fire is not confined to the building (the left branch of the tree labeled “True”). If the spread is confined to the building, the right branch of the tree is taken, and the predicted time to extinguish is around 33.833 minutes. Using the data in the chart generated from the linear regression analysis in Figure 16, we can be 90% confident that the average total loss will be between $16,290.74 and $43,318.05, and we are 90% confident that the predicted total loss will be between $2,172.67 and $324,799.40. The personnel cost may be calculated as 29(33.833) – 24.63, which is $956.53. If the fire is not confined to the building, the longitude is checked. If the location is east of -75.575 degrees, which is east of the Wicomico Civic Center, it predicts 18.5 minutes. For that time, we are 90% confident that the average total loss will be between $10,283.37 and $24,329.37, and we are 90% confident that the predicted total loss will be between $1,307.65 and $191,326.50. The personnel cost may be calculated as 29(18.5) – 24.63, which is $511.87. If the location is east of -75.6 degrees, which is east of Sea Gull Stadium, 11.5 minutes is the predicted time to extinguish, and we are 90% confident that the average total loss will be between $5,305.07 and $11,767.26 and 90% confident that the predicted total loss will be between $656.71 and $95,059.14. The personnel cost may be calculated as 29(11.5) – 24.63, which is $308.87. Otherwise, 15.2 minutes is predicted (the final leaf node of the left branch of the tree). For this time, we are 90% confident that the average total loss will be between $8,174.10 and $18,641.29, and we are 90% confident that the predicted total loss will be between $1,023.68 and $148,850.40. The personnel cost may be calculated as 29(15.2) – 24.63, which is $416.17.

For the tree model used in figure 18, the max depth (maximum level of the tree) is 4, the max features uses sqrt (square root of the total number of features), min samples per leaf (minimum number of samples per terminal node) was 5, and the minimum number of samples per split was 5. The accuracy of the model when splitting the data is:

* R-Squared: -113.60378599687122
* Mean Absolute Error: 7.191525840504645
* Mean Squared Error: 113.60378599687122
* Root Mean Squared Error: 10.658507681512981

The average of all the trees will give an approximate time to extinguish based on the significant features identified and fed into the algorithm.

A diagram of a algorithm

AI-generated content may be incorrect.

***Figure 18:*** *Random Forest Tree with the Lowest Mean Squared Error for the Overall Dataset*

For the features used, fire spread with codes 4, 1, and 2, longitude, latitude, other apparatus, number of personnel, no wind, south winds, and heat source with codes 00, 10, and 40 were important. This is shown in Figure 19.

A graph with numbers and text

AI-generated content may be incorrect.

***Figure 19:*** *Feature importances for the overall dataset for the model used in Figure 18*

We then applied the same algorithm and significant features from the overall analysis to each fire station. For Fire Station 1, Figure 20 shows the transformed model for predicting the total loss with the extinguish time along with the confidence and prediction intervals.

A graph with a green line

AI-generated content may be incorrect.

***Figure 20:*** *Model (red), Confidence (blue), and Prediction (green) regions for the total loss vs extinguish time for Fire Station 1*

Figure 21 from the tuned random forest regression shows that if the number of other apparatuses is fewer than 4, the time to extinguish the fire for Fire Station 1 will be 12 minutes. Using Figure 20’s data from the linear regression, we are 90% confident that the average total loss will be between $2,979.68 and $16,051.29 and are 90% confident that the predicted total loss will be between $330.40 and $144757.10. The personnel cost may be calculated as 36.397(12) – 130.137, which is $306.63. Otherwise, the time to extinguish will be 23.714 minutes, with a 90% confidence that the average total loss will be between $5,988.30 and $43,876.60, and a 90% confidence that the predicted total loss will be between $739.50 and $355,302.20. The personnel cost may be calculated as 36.397(23.714) – 130.137, which is $732.98.

For the tree model used in figure 21, the max depth (maximum level of the tree) is 4, the max features uses sqrt (square root of the total number of features), min samples per leaf (minimum number of samples per terminal node) was 5, and the minimum number of samples per split was 5. The accuracy of the model when splitting the data is:

* R-Squared: -64.29294933333337
* Mean Absolute Error: 7.96266666666667
* Mean Squared Error: 64.29294933333337
* Root Mean Squared Error: 8.018288429168246

A diagram of a number of samples

AI-generated content may be incorrect.

***Figure 21:*** *Random Forest Tree with the Lowest Mean Squared Error for Fire Station 1*

For the features used for Fire Station 1 as shown in Figure 22, the number of personnel, Longitude, Latitude, other apparatus, and fire spread with code 2 are important. The fact that the number of personnel is the most important for only this fire station can explain the outlier on the map in Figure 1 since the model did not find it as important in other stations or in the overall analysis.

A graph with blue bars

AI-generated content may be incorrect.

***Figure 22:*** *Feature importances for Fire Station 1 for the model used in Figure 21*

For Fire Station 2 Figure 23 shows the transformed model for predicting the total loss with the extinguish time along with the confidence and prediction intervals.

A graph with a green and blue line

AI-generated content may be incorrect.

***Figure 23:*** *Model (red), Confidence (blue), and Prediction (green) regions for the total loss vs extinguish time for Fire Station 2*

To explain Figure 24, if the heat source is caused by electrical arcing, the time to extinguish is 3 minutes, and we are 90% confident that the average total loss will be between $30.54 and $505.10, and 90% confident that the predicted total loss will be between $7.45 and $2,071.00. The personnel will cost 25.461(3) + 23.403, which is $99.79. If the fire spread is not confined to the object of origin, the wind direction is not WNW, and two or less other apparatus is sent, then the time is 3 minutes. If the other apparatus count is more than two, then the time is 15 minutes. We are 90% confident that the average total loss will be between $4,883.48 and $30,986.61, and we are 90% confident that the predicted total loss will be between $906.06 and $167,012.30. The personnel cost may be calculated as 25.461(15) + 23.403, which is $405.32. If the wind direction is WNW, then the time is 31 minutes. We are 90% confident that the average total loss will be between $8,052.47 and $64,691.01, and we are 90% confident that the predicted total loss will be between $1,608.59 and $323,837.30. The personnel cost may be calculated as 25.461(31) + 23.403, which is $812.69. If the fire spread is confined to the object of origin, and the heat source value is other, then the time is 45 minutes. We are 90% confident that the average total loss will be between $9,268.53 and $80,654.73, and we are 90% confident that the predicted total loss will be between $1896.51 and $394,172.30. The personnel cost may be calculated as 25.461(45) + 23.403, which is $1,169.15. If the heat source value is not equal to other but undetermined, then it will take 16 minutes. We are 90% confident that the average total loss will be between $5,197.09 and $33,810.33, and we are 90% confident that the predicted total loss will be between $972.02 and $180,772.90. The personnel will cost 25.461(16) + 23.403, which is $430.78. Otherwise, it will take 26 minutes to put out the fire. We are 90% confident that the average total loss will be between $7,374.08 and $56,513.09, and we are 90% confident that the predicted total loss will be between $1,452.01 and $287,003.10. The personnel will cost 25.461(26) + 23.403, which is $685.39.

For the tree model used in figure 24, the max depth (maximum level of the tree) is 4, the max features uses sqrt (square root of the total number of features), min samples per leaf (minimum number of samples per terminal node) was 1, and the minimum number of samples per split was 2. The accuracy of the model when splitting the data is:

* R-Squared: -1100.28125
* Mean Absolute Error: 29.125
* Mean Squared Error: 1100.28125
* Root Mean Squared Error: 33.17048763584883

A diagram of a fire safety system

AI-generated content may be incorrect.

***Figure 24:*** *Random Forest Tree with the Lowest Mean Squared Error for Fire Station 2*

Out of all features used for Fire Station 2 as shown in Figure 25, Latitude, other apparatus, Longitude, the number of personnel, wind direction going west, west of northwest, northwest, south of southwest, west of southwest, north of northeast, no wind, fire spread with codes 4,1, and 2, and heat source with codes 0, 10, 66, 40, 13, and undetermined are important.

A graph of a number of different types of data

AI-generated content may be incorrect.

***Figure 25:*** *Feature importances for Fire Station 2 for the model used in Figure 24*

For Fire Station 16, Figure 26 shows the transformed model for predicting the total loss with the extinguish time along with the confidence and prediction intervals.

A graph with a green line

AI-generated content may be incorrect.

***Figure 26:*** *Model (red), Confidence (blue), and Prediction (green) regions for the total loss vs extinguish time for Fire Station 16*

Fire Station 16’s tree model showed something a bit different as location was much more significant. The tree with the lowest Mean Squared Error as shown in Figure 27. The tree can be explained as follows. If the Latitude is north of 38.367, which is north of US 50 Business, the time to extinguish is 13.143 minutes, and there is 90% confidence that the average total loss will be between $5,519.22 and $17,144.88, as well as 90% confident that the predicted total loss will be between $946.23 and $100,003.40. The personnel cost may be calculated as 27.5697(13.143) - 9.5061, which is $352.84. Otherwise, the time is 26.444 minutes, with a 90% confidence that the average total loss will be between $19,610.94 and $75,730.85, and a 90% confidence that the predicted total loss will be between $3,642.14 and $407,769.10. The personnel cost may be calculated as 27.5697(26.444) - 9.5061, which is $719.55.

For the tree model used in figure 27, the max depth (maximum level of the tree) is 4, the max features used sqrt (square root of the total number of features), min samples per leaf (minimum number of samples per terminal node) was 5, and the minimum number of samples per split was 5. The accuracy of the model when splitting the data is:

* R-Squared: -103.47770727040817
* Mean Absolute Error: 8.5
* Mean Squared Error: 103.47770727040817
* Root Mean Squared Error: 10.172399287798733

A diagram of a number of error

AI-generated content may be incorrect.

***Figure 27:*** *Random Forest Tree with the Lowest Mean Squared Error for Fire Station 16*

The important features as seen in Figure 28 for Fire Station 16 are Longitude, Latitude, fire spread with codes 1 and 2, and other apparatus.

A graph with blue bars

AI-generated content may be incorrect.

***Figure 28:*** *Feature importances for Fire Station 16 for the model used in Figure 27*

To give an idea about the overall density of fires in Salisbury, the following two maps were created. The heatmap shown in Figure 29 shows that the hotspots of building fires are in the center of Salisbury and eastern Salisbury on US 50 Business.

A map with many colored circles

AI-generated content may be incorrect.

***Figure 29:*** *The heatmap of Building Fire density in Salisbury, MD in 2023*

When visualizing by service response districts as seen in Figure 30, the municipal district has the highest fire density, followed by district 16 and district 1. As expected, district 2 has the lowest fire density since there was only one fire in our dataset for district 2.

A map of a city

AI-generated content may be incorrect.

***Figure 30:*** *Density of 2023 Building Fires visualized by Service Response Districts in Salisbury.*

# FUTURE WORK

For future work, additional years such as 2024 data (not fully available at the time of the study) and prior years can be used to see if the models are correct for those years. This can lead to a time series analysis. The analysis can be repeated to reduce and predict the response time. The analysis by fire station could be redone with the top 5 important features for each fire station rather than the analysis that was used in this study, which was the significant features of all fire stations overall. The feature importances for each fire station were shown in the appendix (Figures A2, A3, and A4). An analysis using tax data might show how the properties have changed over time and show whether a loss in tax revenue happened. Aerial imagery of Salisbury over time could show how these fires impacted the structures. Other areas can be added like Fruitland, Delmar, Snow Hill, Laurel, and Seaford, to see how the models differ or how more rows of data improve or change predictions.

Future questions that could be researched include how response time might affect costs. This was only shown to be a significant feature for one fire station, but additional data from other years or including additional areas might yield other results. Another question might be whether weather patterns correlate with the cost. Our study found that wind direction was a significant factor, but there may be other weather data that could be examined, such as wind speed, humidity levels, and rainfall amounts that could yield significance. Since the ranges for the overall cost predictions were large, we also believe that the models could be run again with addition of other features with lower significance to see if those features produce different predictions.

# CONCLUSION

The number of personnel alone cannot determine the time to extinguish. The fire spread, other apparatus, Latitude, Longitude, heat source, and wind direction can determine the time to extinguish along with the number of personnel. Of those variables, only other apparatus and number of personnel can be controlled. Therefore, it will be important to weigh the number of other apparatus and the number of personnel that are sent to each fire vs. the number of each necessary to fight the fire. Fire Station 2 had the worst random forest accuracy, while Fire Station 1 had the best accuracy based on the errors. The overall dataset had a better mean absolute error than Fire Station 1. The municipal service district had the most fires and highest fire density based on the data examined. Overall cost predictions were in large ranges for most of the time to extinguish predictions.

# AUTHOR CONTRIBUTIONS

Jackson did the research for the existing literature and wrote the related work and accompanying bibliography. Jackson worked on the initial models for districts 2 and 16. Joseph worked on the models for the fire stations. Joseph also researched and acquired some tax data that may be included in a future work effort. An worked on the statistical analysis using R and created the final linear regression models. She worked on the overall analysis as well as working on the fire station models. She did the transformations for the total loss data and associated visualizations related to those models. Cameron cleaned the initial datasets, did the SQL database imports and exports, geocoded the data, created the QGIS maps, performed the Random Forest Analysis and accompanying visualizations, and worked the initial models for the municipal district and service district one. Cameron, An, and Joseph worked on the correlation matrices.

# REFERENCES

Cupp, K. (2025). *Service Response Districts* [Shapefile]. City of Salisbury GIS Open Data Hub. <https://salisbury-gis-open-data-hub-salisbury.hub.arcgis.com/datasets/83ae60edda95483a8d65802eb222a206_1>

Federal Emergency Management Agency. (2024). *FEMA USFA NFIRS 2023 All Incidents (Annual)* [Data set]. U.S. Department of Homeland Security. <https://www.fema.gov/about/openfema/data-sets/fema-usfa-nfirs-annual-data>

Gong, J., Liu, J., & Zheng, Y. (2025). *Improving urban fire emergency response times using machine learning techniques*. Environmental Modelling & Software, 169, 105798. <https://doi.org/10.1016/j.envsoft.2024.105798>

National Fire Protection Association. (2023). *U.S. Fire Department Profile*. <https://www.nfpa.org/education-and-research/research/nfpa-research/fire-statistical-reports/us-fire-department-profile>

OpenAI. (2024). *ChatGPT* (May 2024 version) [Large language model]. <https://chat.openai.com/chat>

U.S. Fire Administration. (2015). *National Fire Incident Reporting System complete reference guide* (Version 5.0). U.S. Department of Homeland Security, Federal Emergency Management Agency. <https://www.usfa.fema.gov/downloads/pdf/nfirs/nfirs_complete_reference_guide_2015.pdf>

U.S. Fire Administration. (2024). *State Fire Statistics*. <https://www.usfa.fema.gov/statistics/states/>

Weather Underground. (2023). *Salisbury, MD weather history (KSBY)*. The Weather Company. <https://www.wunderground.com/history/daily/us/md/salisbury/KSBY>

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). *Warming and earlier spring increase western U.S. forest wildfire activity*. Science, 313(5789), 940–943. <https://doi.org/10.1126/science.1128834>

# APPENDIX

In Figure A1, we decided to visualize the confidence and prediction intervals with a maximum of 30 minutes instead of 80 minutes.

A graph with green and blue lines

AI-generated content may be incorrect.

***Figure A1:*** *Model (red), Confidence (blue), and Prediction (green) regions for the total loss vs extinguish time for the overall dataset from 1 minute to 30 minutes*

To see how the feature importances change using the exploratory random forest model to pick the top 5 important features, here is how the features change across each Fire Station.

For Fire Station 1, the top 5 important features are Latitude, heat source, other apparatus, supplementary apparatus, and humidity. The top 3 features line up with 3 of the top 5 most important features in the overall dataset.

A graph with a number of different colored lines

AI-generated content may be incorrect.

***Figure A2:*** *Exploratory Random Forest Model showing the features from most important to least important for Fire Station 1*

For Fire Station 2, the top 5 important features are total square feet, property value, factor of ignition, building width, and content value. None of the features here line up with Fire Station 1 or the overall dataset.

A graph with a blue line

AI-generated content may be incorrect.

***Figure A3:*** *Exploratory Random Forest Model showing the features from most important to least important for Fire Station 2*

For Fire Station 16, the top 5 important features are fire spread, cause of ignition, structure status, detector, and dew point. Only fire spread is also one of the most important features in the overall dataset.

A graph with a blue line

AI-generated content may be incorrect.

***Figure A4:*** *Exploratory Random Forest Model showing the features from most important to least important for Fire Station 16*