George Mason University

DAEN 690-001 Capstone

Fall 2020

Fairfax County Fire and Rescue Department/GMU Data Lab

Final Report

Team Patriots

Shiva Ram Kaushil Pabba (Scrum Master)

Andrew Bloom (Product Owner)

Doddanaik Basavaraj Vakkund (Developer)

Jahnavi Jonnalagadda (Developer)

9th December 2020

Table of Contents

Abstract

The Fairfax County Fire and Rescue department provides emergency and non-emergency services, such as fire suppression, technical and swift water rescue, hazardous materials, and Emergency Medical Services (EMS). As the department's data volume increases daily, there is a need to turn data into knowledge to manage data growth challenges and to plan the long-term by implementing strategic business decisions. In this project, we aim to explain a) is there any correlation between provider/ shift/ station / battalion and EMS outcome, b) Number of units dispatched within certain X (to be established) minutes of becoming available, c) how often are 2nd emergencies occurring within X (to be established) minutes by incident type (EMS/Fire/PSERVF) by first due, and d) the relationship of EMS procedures and medications to incident distances from the hospital via linear regression and Pearson's correlation. We found that the provider, shift, station, and battalion variables, though significantly dependent, are uncorrelated with EMS call outcome. The diazepam (valium) medication contributes the most to incident distance from the hospital, and the wound-burn care procedure contributes the most to incident distance from the hospital. The minimum frequency of units dispatched is 0.2 minutes and the maximum is 1,500,000 minutes. From the top 50 units, M414B, COM401, F426 are the busiest units with a frequency of 100,000 minutes and IV16, IV05 are the slowest units with a frequency greater than 1,300,000 minutes.

1 Introduction

1.1 Background

The Fairfax County Fire and Rescue Department (FCFRD) is a combination career and volunteer organization providing emergency and non-emergency services to protect the lives, property, and environment of the people of Fairfax County, such as fire suppression, emergency medical, technical rescue, hazardous materials, water rescue, life safety education, fire prevention and arson investigation services.

Like other first response service (fire, rescue, and emergency medical services (EMS)) organizations, the FCFRD is interested in turning to data analytics to knowledge to help its employees address challenges they face in their work every day and drive strategic decision-making processes and long-term planning, especially since they rely on good data and knowledge gleaned from the data to make critical life-and-death decisions. Emergency incidents such as wildfires, worsened by the expanding population and extreme weather conditions, have claimed lives, destroyed thousands of structures, prompted evacuations, and impacted localities, and the response to and recovery from these increasingly complex events impact municipal budgets, local economies, and the corresponding communities.

The current fire service concept does not focus so much on fighting fires as it does on doing EMS, HAZMAT, inspections, investigations, prevention, and other nontraditional but important tasks which are vital to the community. Balancing limited resources and justifying daily operations and finances, especially during tough economic times, is a scenario that every department is familiar with. For instance, for reports completed on incidents, a person would normally document the type of situation found, action taken, time of alarm, time of arrival, time completed, number of engines responding, number of personnel responding, and many other items. For fires, the list of documented items grows even longer to include area of fire origin, form of heat of ignition, type of material involved, and other related facts. Additionally, if civilian or fire-fighter injuries occur, other reports need to be completed as well.

Reports like the above mentioned are a legal requirement for documenting incidents, as victims, insurance companies, lawyers, and many others want copies of such reports. Indeed, fire departments maintain files for retrieval of individual reports, but the reports can provide a more beneficial service to fire departments by yielding insight into the nature of fires and injuries in their authority. Typically, basic information such as the number of fires handled, the number of fire-related injuries, and the number of fire deaths are tracked. However, if more probing questions are asked, such as the following, it could alter understanding of the data:

• What was the average response time to fires/other incidents?

• How much did response times vary by fire station areas?

• What was the average time spent at the fire/other incident scene?

• How much did the average time vary by type of fire and other incidents?

Also, there are three good reasons for thorough data analysis, which are the following: (1) to gain insights into fire problems, (2) to improve resource allocation for combating fires, and (3) to identify training needs. The most compelling reason is that analysis gives insight into fire problems, which in turn can affect departmental operations. For example, one may find that the average response time to fires in an area is 6 minutes, compared to less than 2 minutes overall, which may result in more equipment, more personnel, or justifying constructing another fire station. As an example of the second reason, which is improved resource allocation, statistically analyzing emergency medical calls can determine the impact of providing another paramedic unit in the field. Increasing the number of EMS units from four to five may, for example, decrease average response times from 5 minutes to 3 minutes. The last reason for analysis is to identify training needs. Most training in firefighting is based on a curriculum that has been in place for many years, and so observing how training matches the characteristics of fires in a particular authority is sensible. However, other training is also important, since an exception can always occur, and knowing more about the fires in an area can improve the training. Additionally, an analysis of firefighter injuries may indicate a need for certain types of training.

Because a community's health depends on whether its public safety agencies operate effectively, geographic information system (GIS) technology has often been suggested as a method of improving effectiveness. GIS can be used to analyze and measure data and share it with decision-makers, and the data can come from computer-aided dispatch (CAD) software, record management systems (RMS). community risk reduction (CRR) activities, or any data source that an agency deems appropriate. The paramount issue in public safety is determining how the data is used, visualized, and communicated, as problems arise due to a range of factors, which can disrupt proper data usage. For instance, the data can be incorrectly collected due to improper collection procedures. Also, software changes, proprietary vendor policies, and agency guidelines can obstruct data usage. Finally, software can be complicated, and visualization can be cumbersome. Using GIS technology can help users overcome those obstacles, and using GIS data to measure performance and focus resources can have wide-ranging benefits, such as the following:

1. Identifying populations at risk, as a community risk assessment (CRA), performed for the accreditation process and for risk reduction programs can identify high-risk occupancies (target hazards), at-risk populations, and community areas where there should be public education and prevention programs

2. Increased responder safety, as through a CRA, preplanning efforts can be focused on the highest-risk occupancies, and responders can identify these hazards before the response. Using GIS allows agencies to use a visual product to efficiently share these assessments with anyone deemed appropriate, and these plans can be communicated in near-real time as they are generated or updated

3. Support for the budgetary process, as elected officials and the community no longer want decisions based on gut feelings, as they did when anecdotal presentations were used to justify expenditures. GIS can allow data presentation such that gaps in service delivery, local at-risk areas, and the impact of growth on a community are easily identified.

Overall outcomes from out-of-hospital cardiac arrest (OHCA), both in terms of survival and neurologic and functional ability, are poor, as only 11 percent of patients treated by emergency medical services (EMS) personnel survive long enough to be discharged from the hospital, and although a few EMS systems have been able to significantly increase survival rates, a fivefold difference in survival-to-discharge rates still exists among communities in the United States. An EMS system is defined as a system that provides emergency medical care in response to serious illness or injury, such as a cardiac arrest, in the prehospital setting, that consists of a wide range of responders who provide critical services, such as response to 911 emergency calls, dispatch of medical personnel, triage, treatment, and rapid transport of patients by ground or air ambulances to appropriate care facilities. EMS staffing models for cardiac arrest responses often vary, partly because of differences in the availability and proximity of the nearest providers, resources, and call volume.

1.2 Problem Space

The main objective of the project is to revisit the practices of the Fire and Rescue department to help Fairfax county to make life-and-death decisions and improve decision making. Some of the questions we aim to answer from this project are

1) Is there any correlation between provider/shift/station/battalion to EMS call outcome (no EMS needed/refusal/transport/etc.)?
2) Examining the relationship of EMS procedures and medications to incident distance from the hospital for transported patients.
3) How often are second emergencies occurring within X minutes by incident type (EMS/Fire/PSERVF) by first due (X to be established)?
4) How frequently are the units dispatched with in X minutes of becoming available and the value of X need to be established in minutes. For example, X may be 20-30 minutes?

1.3 Research

Once the project was selected, there were many things that the team is not aware of like what operations and services fire and rescue team provides or how they function in a particular area. The initial research was done and found how fire departments in other counties provides their services to local communities. The data related to other counties for example Prince William County in Virginia has been observed. After going through a few online documents, we found that they have data related to types of incidents encountered, type of occupancies involved, timestamp, EMS: Abdominal pains, active shooter, allergic reactions, animal bite, vehicle accidents, vehicle leaking fuel and so forth.

1.4 Solution Space

We had answered our first two objectives using Cramer's-v test, chi-square test for independence, and linear regression to find out the linear relationship between the dependent and independent variables. For answering questions 3 and 4, no predictive algorithm is used as the frequency of occurrence of second emergency can be found by grouping and sorting the Incident by due dataset with respect to event type(E/F/P) and first due(station) and the frequency of units dispatched can be found by subtracting available time from next dispatch time.

1.5 Project Objectives

What does the team assume they will achieve as a solution when they finish this project?

The team members assume when the project is finished, solutions to the questions defined in the problem space will be provided, along with the real-time experience of collaborative working and achieving goals within a given period.

What does the team assume it will achieve in terms of understanding the problem after they finish this project?

The team members assume that after the project is finished, they will understand whether the questions can be answered with or without data analysis. For example, relationships between various attributes in the data can sometimes be found explicitly, meaning that a simple data query/search can give the relationship between attributes and hence does not require any data analysis or machine learning algorithms. However, some questions need certain statistical analysis to find meaningful insights from the data.

What does the team assume it will provide in value as a product of this project work to the world, targeted group, etc.?

The team members assume that they will provide better decision-making capabilities using the available data and save more lives. The ultimate objective is to convert data into useful information which can be used by the fire and rescue department to make better and informed decisions.

1.6 Primary User Story (-ies):

FCFRD provides services to the local communities that fall within the boundaries of Fairfax county. As the volume of the data grows steadily, converting the big data problem into useful information with data analytics can assist fire agencies to revisit their practices such as EMS outcome, provider fatigue etc. The meaningful insights drawn by data analysis not only assist in overcoming the challenges they face in everyday work but also guide them in making strategic business decisions and long-term planning.

1.7 Product Vision - Sample scenarios (why would someone want to use this)

After we complete this project, we will answer questions such as relationship between the EMS procedures and medications to incident distance from the transported patients, correlation between provider/shift/station/battalion to EMS call outcome, and how often are 2nd emergencies occurring within X minutes by incident type with certain statistical evidence. Below are the two scenarios on how the insights drawn from the data provide value to both firefighters and end users.

❖ Scenario #1

In general, when an incident occurs, the station closer to the incident is responsible for handling it. However, some stations always receive a high volume of calls whereas some of them receive calls less frequently. For example, Annandale Station receives a greater number of calls when compared to Clifton Station. In such cases, understanding the frequencies of occurrence of second emergencies by incident type helps the fire department to take necessary actions in handling the incidents in a more efficient manner. The insights obtained from analysis can also determine the impact of providing another unit in the field. For instance, increasing the number of units from four to five decreases average response time from 5 minutes to 3 minutes, a change that may save lives.

❖ Scenario #2

Examining the relationship between the EMS procedures and medications from the incident distance to the hospital helps the fire department agencies to understand how the paramedics are performing procedures. The outcome of this analysis can help identify whether the paramedics are performing the procedures regardless of the distance from the incident to the hospital. Critical incidents like cardiac arrest require immediate intervention even if the incident is closer to the hospital. Hence, the insights obtained from this analysis provide information to the fire department to caution the paramedics in performing procedures which may help in saving lives. Thus, the value achieved by this project will benefit both fire agencies and victims.

1.8 Definition of Terms:

Call Types

Fire calls: for car fires, building fires, or other subtypes of fire; something is on fire, or somebody suspects something is on fire, and there is a fire alarm

EMS (Emergency medical service) calls: for emergencies like not feeling well, car accidents, sudden sickness

Public service calls

P serve p: police department response

P serve f: fire department response

F bldg.: fire building

Stations: area surrounding a certain numbered station that the station is responsible for coverage if an emergency occurs in the area

Battalions: Group of stations. Generally, groupings of five or six stations are available.

2 Data Acquisition

2.1 Overview:

The dataset provided by the Fairfax County Fire and Rescue Department is a proprietary dataset stored in an SQL server database. It archives pure GIS data such as fire station location and vehicle location data and is updated each year. The four datasets provided are mainly EMS call outcomes, patient distances (time and distance transported, medications and procedures), incident by unit, and incident by due. The datasets contain information of incidents starting from January 1st, 2018 until August 31st, 2020. In the dataset, there are between 400,000 and 500,000 dispatches and 1,200,000-unit responses. The SQL server database is a reporting database, which will prove extremely useful for web-mapping applications. There are mainly three types of calls in the database: fire calls (for emergencies such as car fires and building fires), EMS (emergency medical service) calls, and public service calls (for problems such as a person having an accidental fall, a flooded basement or replacing a smoke detector or a car accident with no injuries).

2.2 Field Descriptions:

1) EMS Call Outcomes

#	Attribute	Description	Example
1	Patient_ID_Internal	uniquely identifies each individual patient in the database	482724
2	Performer_ID_Internal	uniquely identifies each individual healthcare provider in the database	21D3C99E-9E01-E211-B5F5-78E7D18CFD3C
3	Response_EMS_Shift	explains the shift during which a particular EMS call occurred	B - Shift
4	Response_EMS_Vehicle_Unit_Number	explains EMS vehicle unit number that responded to EMS call	M414
5	Station	represents the station that EMS vehicle unit which responded to EMS call belongs to	14
6	Battalion	represents the battalion that EMS vehicle unit which responded to EMS call belongs to	407
7	Disposition_Incident_Patient_Disposition	provides outcome of EMS call, either for patient or incident	Treated & Transported

2) Incidents by Due

#	Attribute	Description	Example
1	Incident #	uniquely identifies each individual incident in database	E180010006
2	Event Type Code	provides a shorthand abbreviation of the Event Type in the following column	ALS
3	Event Type	explains the type of event that occurred	ALS EMERGENCY
4	Event Category	provides a letter for each type of event category that occurred: E for EMS, F for fire, T for test events, P for public service, N for notifications and other events	E
5	Fire Box	contains a numbered ID for the fire box that contains the device for notifying the respective fire station of a fire outbreak	43503
6	First Due	refers to either the first apparatus arriving at the scene of a fire or the area in which a company is expected to be the first to arrive on the scene.	435
7	Battalion	represents the fire battalion that responded to the incident	407
8	Dispatch	displays the time that the unit was dispatched to the scene of the incident	1/1/2018

	Time		12:04:00 AM

3) Incidents by Unit

#	Attribute	Description	Example
1	Incident #	uniquely identifies each individual incident in database	E180010058
2	Event Type Code	provides a shorthand abbreviation of Event Type	GASIN
3	Event Type	explains the type of event that occurred	GAS LEAK: INSIDE
4	Event Category	provides a letter for each type of event category that occurred: E for EMS, F for fire, T for test events, P for public service, N for notifications and other events	F
5	Unit ID	explains the ID of the unit that was dispatched to the scene of the incident	E425
6	Unit Type	explains the type of unit that was dispatched to the scene of the incident	Engine
7	Dispatched	display the time that the unit was dispatched to the scene of the incident	1/1/2018 12:28:58 AM
8	Available	display the time that the unit arrived at the scene of the incident	1/1/2018 12:39:37 AM

4) Patient Distances - Procedures

#	Attribute	Description	Example
1	Patient_ID_Internal	uniquely identifies each individual patient in the database	482894
2	Procedure_Performed_Code	provides corresponding numerical code for value of Procedure_Performed_Description field	392230005
3	Procedure_Performed_ Description	describes the type of medical procedure that occurred	IV Start - Extremity Vein (arm or leg)

Patient Distances - Medications

#	Attribute	Description	Example
1	Patient_ID_Internal	uniquely identifies each individual patient in the database	494777
2	Medication_Given_RXCUI_Code	provides corresponding numerical code for value of Medication_Given_Description field	7806
3	Medication_Given_Description	Describes the type of medication given to the patient	oxygen

Patient Distances - Time and Distance Transported

#	Attribute	Description	Example

1	Patient_ID_Internal	uniquely identifies each individual patient in the database	479853
2	Incident_UnitLeftScene_To_PatientArrivedAtDestination_In_HHMMSS	provides length of time, in HHMMSS format, between when the incident unit left the scene of the incident and when patient arrived at destination	0:19:20
3	Incident_Unit_Left_Scene_Date_Time	displays date and time when the incident unit left the scene of the incident	1/1/2018 12:19:00 AM
4	Incident_Patient_Arrived_At_Destination_Date_Time	displays date and time when patient arrived at destination	1/1/2018 12:39:00 AM
5	Feet	gives distance between incident scene and hospital in terms of feet (ft)	37484.09521
6	Miles	gives distance between incident scene and hospital in terms of miles (mi)	7.099260456

2.3 Data Context:

After analyzing the EMS Call Outcomes, we found that the most common outcome that happened to incident patients was being treated and transported, which happened in 305,997 instances. The second, third, and fourth most common outcomes were patient refusal (AMA), which happened in 38,678 instances; no treatment/transport required, which happened in 33,754 instances, and canceled (prior to arrival), which happened in 31,351 instances. An exploratory data analysis has been done and the findings were shown visually using Figure 1 and 2. The left-hand side of Figure 1 illustrates the number of incidents handled by each battalion. It is clear that the battalions 404, 406 are the busiest one's while battalions 407, 401 are the slowest one's. The right-hand side of the Figure 1 illustrates the number of incidents handled by each shift. It is seen that A-Shift was found to be the slightly busiest shift which handled 68,574 unique incidents, followed by B- Shift with 69,501 unique incidents and C-Shift with 69,344 incidents. The Figure 2 gives the number of incidents belonging to each event category. It is observed that the event category emergency logs more incidents than other categories like fire and public service. It is also seen that event categories test and miscellaneous complaints have a smaller number of incidents as these categories are not the actual incidents rather, are dummy/test events.

From analyzing the Patient Distances-Medications dataset, we found that the most common medications given to patients in the area covered by the FCFRD were oxygen (given to 9,230 patients), normal saline (given to 6,755 patients), Ondansetron(Zofran) (given to 6,274 patients), and Fentanyl Citrate (Sublimize) (given to 5,909 patients), and the least common medications given were Tetracaine (given to 11 patients), Calcium Chloride (10%) (given to 12 patients), Cyanokit (Hydroxocobalamin) (given to 5 patients), and Diazepam (Valium) (given to 3 patients). From analyzing the Patient Distances-Procedures dataset, we found that the most common procedures performed on patients in the area covered by the FCFRD were of the category CV - ECG - 12 Lead Obtained (performed on 55,576 patients) and IV Start - Extremity Vein (arm or leg) (performed on 52,546 patients). The same findings have been visually shown using Figure 3. Upon closely analyzing the Patient Distances-Time and Distance Transported dataset, we found that for 419 incident units, it took 10 minutes to arrive at the hospital from the scene of the incident; also, 6 incident units left the scene of their respective incidents on October 5, 2018 at 10:43pm, 5 of which arrived at the hospital at 11:02pm that same day, and 6 incident units arrived at the hospital from their respective incident sites on January 13, 2018 at 12:08am.

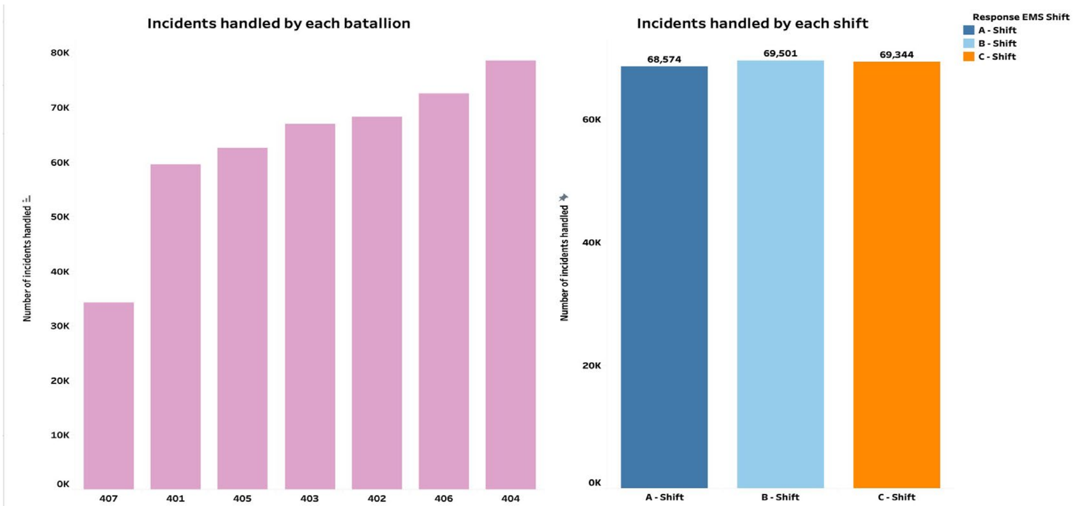


Figure 1: The number of incidents handled by each battalion (left) and the number of incidents handled by each shift (right).

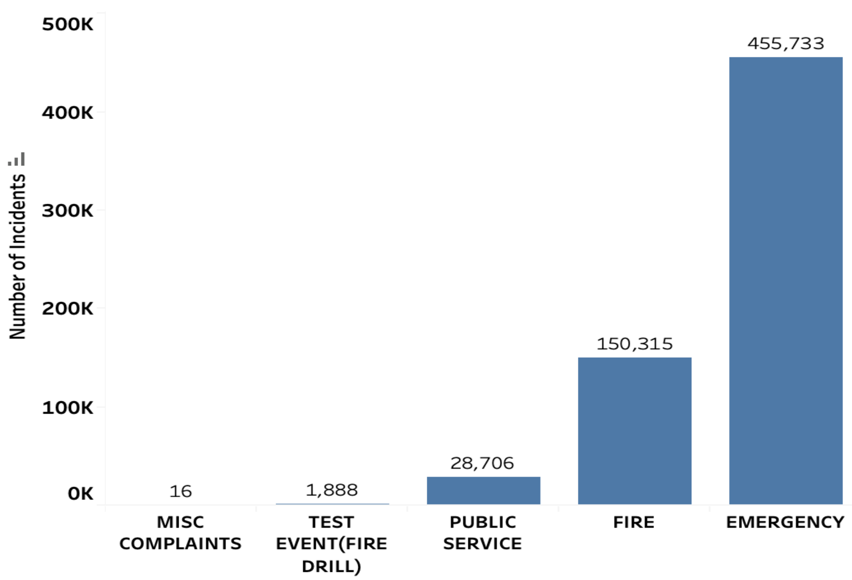


Figure 2: The number of incidents in each event category

From analyzing the Incidents by Due dataset, we found that the most common types of incidents that the FCFRD responded to were Advanced Life Support (ALS) Emergency events (100,209 events), Medical Emergency events (27,303 events), Basic Life Support (BLS) Emergency events (30,310 events), and Fire Alarm events (26,433 events). E (emergency medical service events) was found to be the code category with the highest number of events, with 198,946 events in the category, and N (notifications and other events) was found to be the code category with the least number of events, with 13 events in the category. Battalions 402, 406 and 403 were found to be the busiest battalions in the department, having responded to 47,286, 42,203, and 40,858 incidents, respectively. Battalions 407, 443 and 465 were found to be the least busy battalions, having responded to 18,296, 12,892 and 388 incidents, respectively.

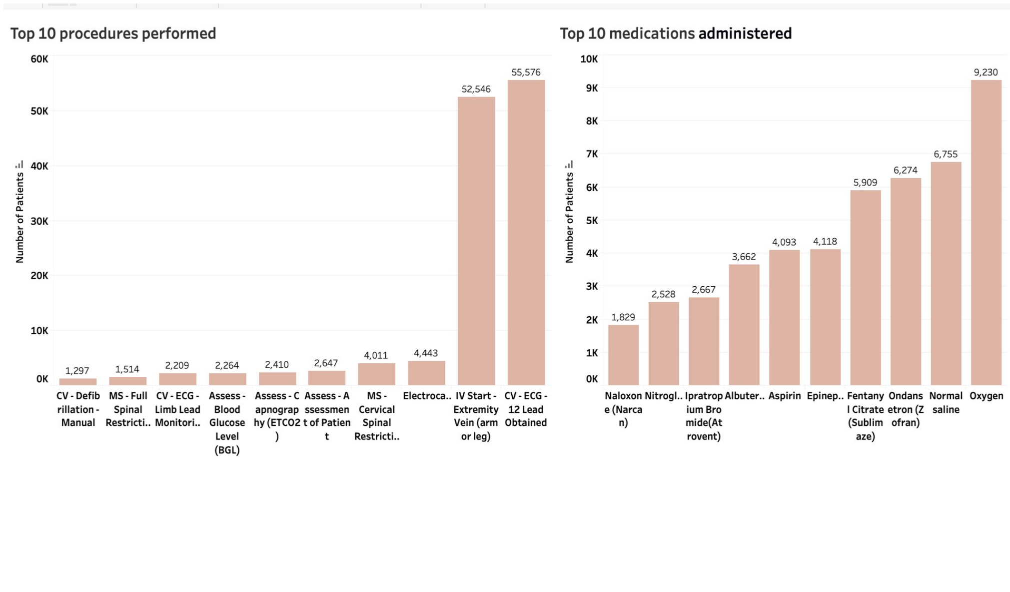


Figure 3: Top 10 procedures performed (left) and top 10 medications administered (right)

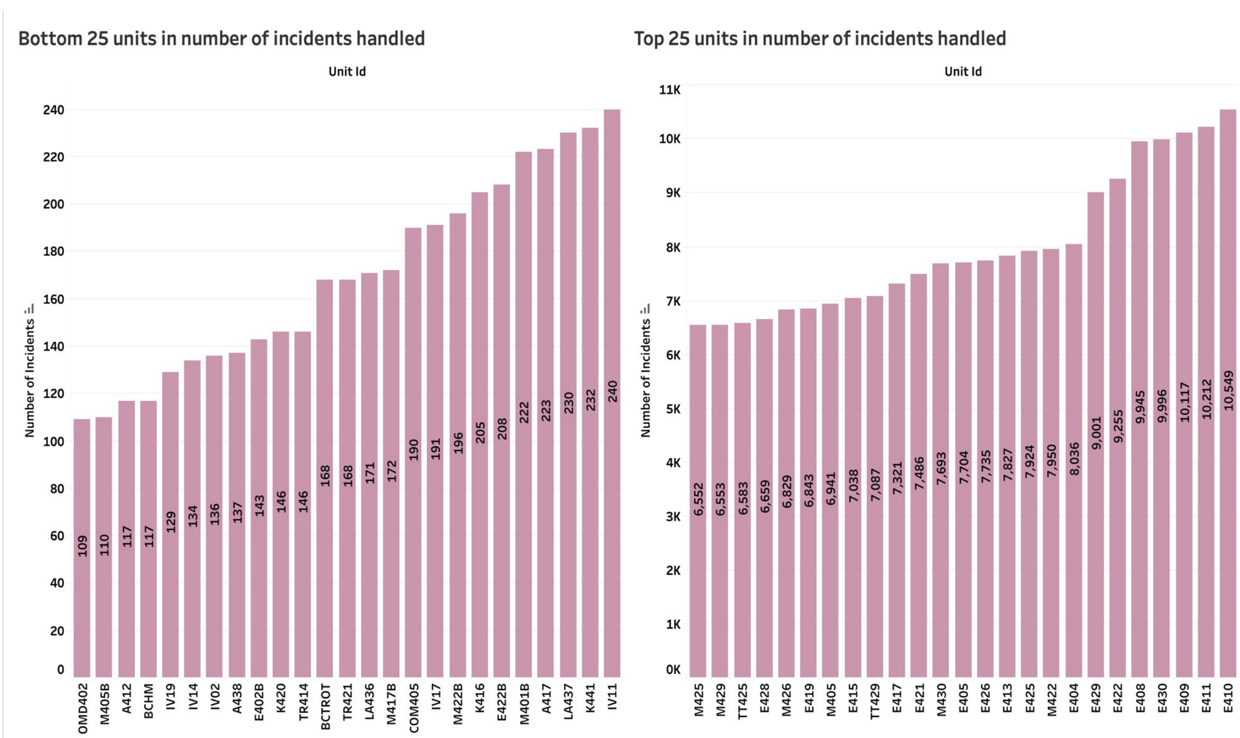


Figure 4: Bottom 25 units in number of incidents handled (left) and top 25 units in number of incidents handled (right)

From analyzing the Incidents by Unit dataset, we found that the most common types of incidents that the FCFRD responded to were Advanced Life Support (ALS) Emergency events (212,402 events), Medical Emergency events (57,133 events), Basic Life Support (BLS) Emergency events (50,932 events), and Fire Alarm events (50,835 events). E (emergency medical service events) was found to be the code category with the highest number of events, with 455,733 events in the category, and N (notifications and other events) was found to be the code category with the least number of events, with 16 events in the category. From right hand side of Figure 4, it is seen that units E410, E411 and E409 are found to be the busiest units in the department having responded to 10,549, 10,212, and 10,117 incidents, respectively and the units OMD402, M405B are the slowest units in the department having responded to less than 150 incidents (shown left). Engines and medic units were found to be the unit types that most responded to events, with 234,492 instances in the Engine category and 217,805 instances in the Medic category.

2.4 Data Conditioning

To create the merged patient Distances dataset, merge () function in R was used to combine the patient distances-time and distance transported, procedures and medications datasets. The merging is done by utilizing common column which is patient id. Merging the Patient Distances-Medications and the Patient Distances-Time and Distance Transported datasets resulted in a dataset consisting of 53,027 entries (matched rows), while merging the Medications and Procedures datasets resulted in a dataset consisting of 124,694 entries (matched rows) and merging the datasets for Procedures and Time and Distance Transported produced a dataset consisting of 137,196 entries (matched rows). To create the merged Incidents dataset, we used the merge () function to combine the Incidents by Due and Incidents by Unit datasets, which both share the Incident Number, Event Type Code, Event Type, and Event Category variables, resulting in a dataset consisting of 608,094 entries (matched rows).

2.5 Data Quality Assessment:

	Dataset					
Quality	**EMS call outcome**	**Patient transport distance**	**Patient medication**	**Patient procedures**	**Incident by due**	**Incident by unit**
Completeness (Missing values)	0.0006%	0.003%	0.0164%	0%	0.003%	0.056%
Uniqueness	☑	☑	☑	☑	☑	☑
Integrity	☑	☑	☑	☑	☑	☑
Accuracy	☑	☑	☑	☑	☑	☑
Number of attributes with null values	4	1	2	0	1	2

Attributes to missing values	incident patient disposition; performer ID	incident left to scene	medication code; medication description	-	Battalion	event type; available date
Consistency	☑	☑	☑	☑	☑	☑
Conformity	☑	☑	☑	☑	☑	☑

2.6 Other Data Sources

No other datasets have been used apart from above mentioned four datasets.

3 Analytics and Algorithms

For the EMS Call Outcomes dataset, we used the Pearson's chi-square test on different pairs of variables in the dataset. First, we checked to see if all the variables in the dataset were of the data type "factor" and they were. A Pearson's chi-square test for independence is applied when you have two categorical variables from a single population, and is used to determine whether there is a significant association between the two variables, given by the Pearson correlation coefficient r, which measures the linear relationship between two variables X and Y, such that a correlation of 1 indicates the data points perfectly lie on a line for which Y increases as X increases and a value of -1 also implies the data points lie on a line for which Y decreases as X increases. This test procedure is appropriate when the following conditions are met: the sampling method is simple random sampling; the variables under study are each categorical; and if the sample data are displayed in a contingency table. The expected frequency count for each cell on the table is at least 5. This approach consists of the following four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. The null hypothesis states that knowing the level of Variable A does not help you predict the level of Variable B (the variables are independent), whereas the alternative hypothesis is that knowing the level of Variable A can help you predict the level of Variable B (the variables are dependent). The analysis plan describes how to use sample data to accept or reject the null hypothesis and should specify the following elements: a significance level (often, researchers choose significance levels equal to 0.01, 0.05, or 0.10, but any value between 0 and 1 can be used) and a test method (use the chi-square test for independence to determine whether there is a significant relationship between two categorical variables). Typically, the resulting P-value (the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct) from the chi-square test is compared to the significance level, and the null hypothesis is rejected if the P-value is less than the significance level. Because we knew that would not be enough to answer the question, we decided to use Cramer's V statistics to back up the chi-square test.

Cramer's V is used as a measure of association between two nominal variables, or as an effect size for a chi-square test of association. Cramer's V varies from 0 to 1, with a 1 indicating a perfect association and 0 indicating no association. Following Cramer's V test, we put the results in a correlation plot to simultaneously investigate the dependence between multiple variables and highlight the most correlated variables in the dataset. In the below visual, correlation coefficients are colored according to the value.

For the output2 dataset, which combined the three separate Patient Distances datasets, we ran a linear regression model to examine the relationship of EMS procedures and medications to incident distance from the hospital for transported patients (both in terms of feet and miles). We used the medications given description and procedure performed description factor variables from the output2 dataset. The values of the medications given description coefficients are compared level-by-level to the omitted reference level medications given description Acetylsalicylic Acid (Aspirin, ASA), and the values of the procedure performed description coefficients are compared level-by-level to the omitted reference level procedure performed description Airway - Clear/Suction.

For the Incident by Unit dataset, the data is sorted by event category, unit id and dispatch time. Using this sorted data, we subtract the available time column from the next dispatch time column which gives the frequency of occurrence of the next emergency. Now, the data is grouped using the event category, unit id and dispatch time to obtain the most frequent dispatch time. The average of the most frequent dispatch time for the total number of days in the dataset gives how frequently units are dispatched after becoming available.

Using the Incident by due dataset, our objective was to find the frequency of occurrence of second emergency by event type and first due. To achieve this, we grouped the dataset into the event category and sorted this grouped data by first due and dispatch time. Now, we calculated the difference between dispatch time of any two consecutive rows of the sorted data where the value of first due is equal for both the rows. Since we are interested only in finding the occurrence of a second emergency, we keep only the second values for each day, event type, and first due and discard the other values. In the last step, we find the average of second emergencies daily by event type and first due.

4 Visualization

The results from the Crammer's v test are visually displayed using Figure 5. We can see that all features with respect to each other have almost no correlation. Only battalion and station is seen to have positive and high correlation (equals to 1) because a station is inside battalion and its value doesn't change.

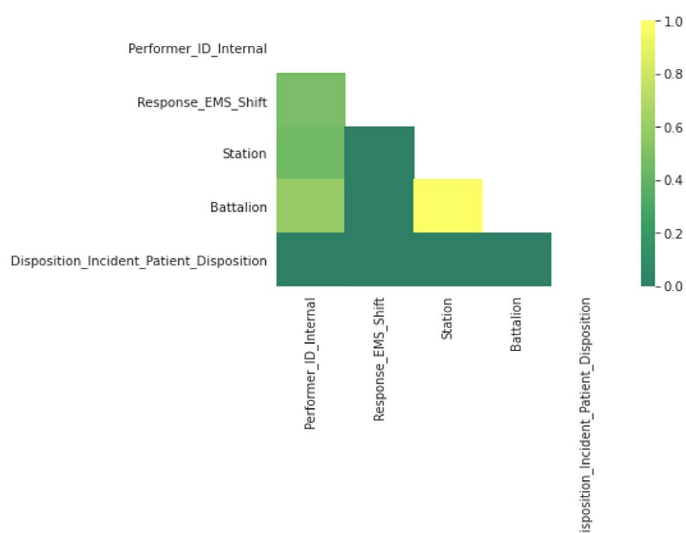


Figure 5: Correlation plot of EMS call outcome with respect to station/shift/battalion/provider

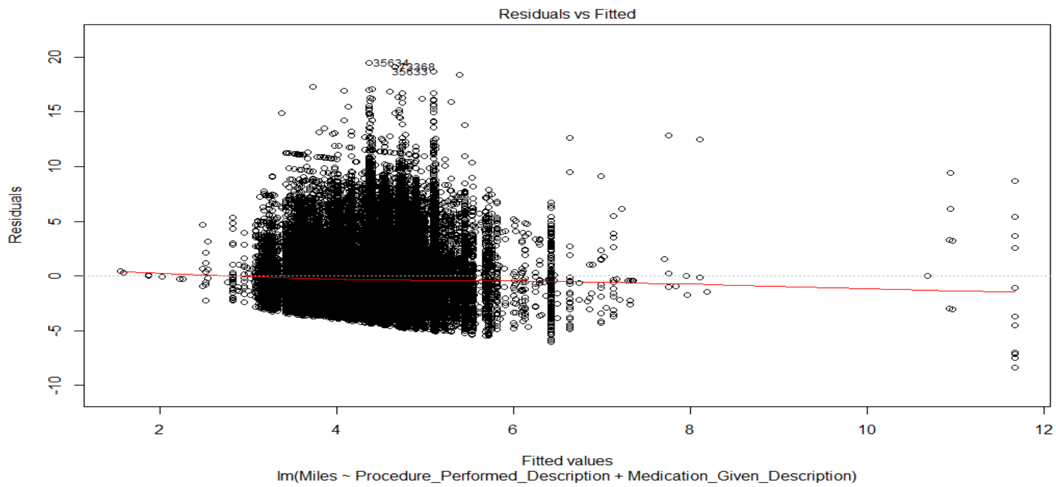


Figure 6: Residuals vs Fitted graph for medications and procedures performed with respect to incident distance from hospital in miles

In Figure 6, the red line is expected to be close to the horizontal line centered at 0 rather than below the line, to show that there is no indication of non-constant variance. For the above plot, the red line must be a horizontal line close to the center which shows that the assumption of linearity is violated. Based on the above responses the model assumptions appeared to be violated. The Linearity of Data assumption can be assessed using Residual vs fitted plot seems to be violated in the given linear model and data.

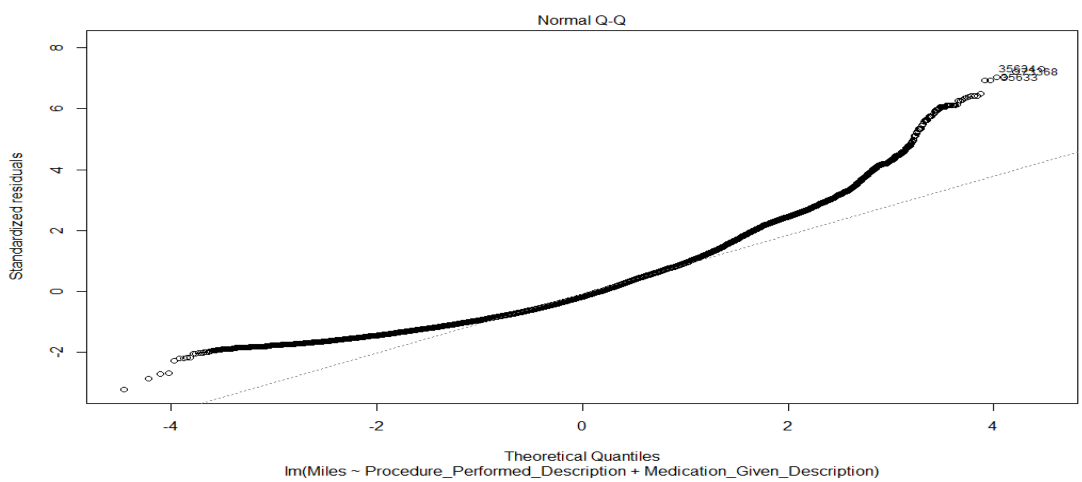


Figure 7: Q-Q plot for medications and procedures performed with respect to incident distance from hospital in miles

In Figure 7 showing Q-Q plot, the residuals must be perfectly aligned with the diagonal line to satisfy the assumption that the residuals are normally distributed. In the above plot the residuals are not perfectly aligned with the diagonal line. Only a part of the residuals can be observed on the diagonal line while a sizable number of residuals can be seen above the diagonal line showing the upper and lower tails. We can assess the Normality of residuals assumption using the Q-Q plot which seems to be violated.

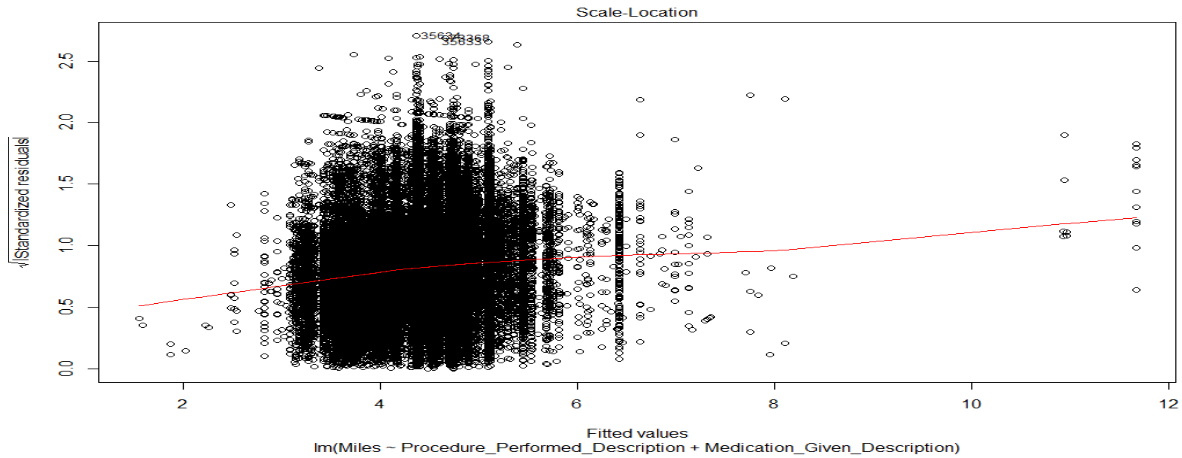


Figure 8: Standardized Scale-Location plot for medications and procedures performed with respect to incident distance from hospital in miles

In Figure 8, a horizontal line with equally spread points is expected which indicates good homoscedasticity. In the above plot a diagonal straight line can be observed which indicates the residuals are spread uniformly. The assumption of homoscedasticity of residuals that is homoscedasticity can be assessed using the scale-location plot.

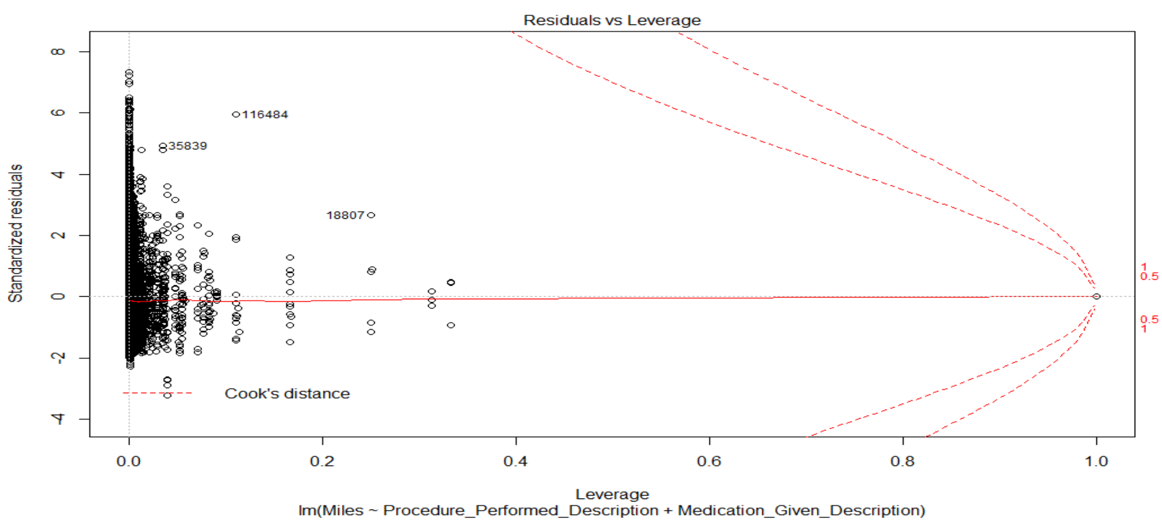


Figure 9: Standardized Residuals vs Leverage graph for medications and procedures performed with respect to incident distance from hospital in miles to show outliers.

In the Residuals Vs Leverage graph there should be no evidence of outliers. The dashed curves, i.e., Cook's distance lines, should not appear in the plot. None of the data points should come close to having both high residuals and leverage. In the above plot we can observe the cook's distance lines with an influential observation that has not been identified. Also, three observations (#18807, #35839, and #116484) have been identified as being close to the Cook's distance lines. The assumption of independence of residual error terms can be assessed using the Residuals Vs Leverage plot.

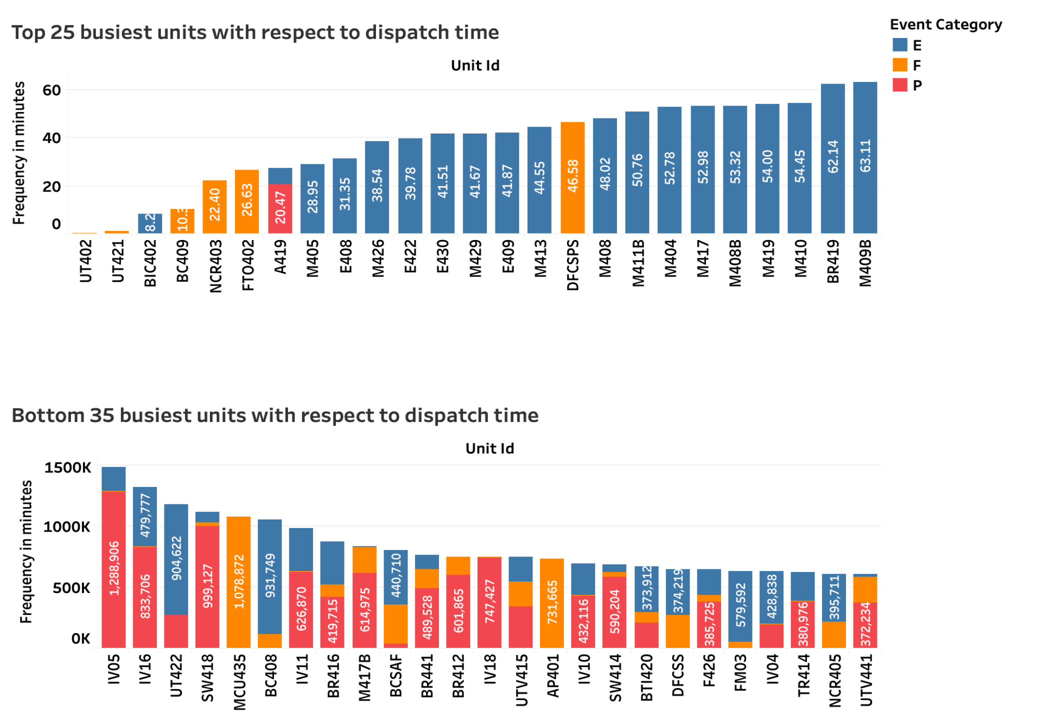


Figure 10: Top 25 busiest units (top) and bottom 25 (bottom) busiest units with respect to dispatch time

Figure 10 shows the top 25 busiest units and bottom 25 busiest units (slowest units) with respect to dispatch time. It is seen that UT402, UT 421, BIC 402 are the busiest units while IV16, IV05 are the slowest units with the frequency greater than 900 days (~ 1,300,000 minutes).

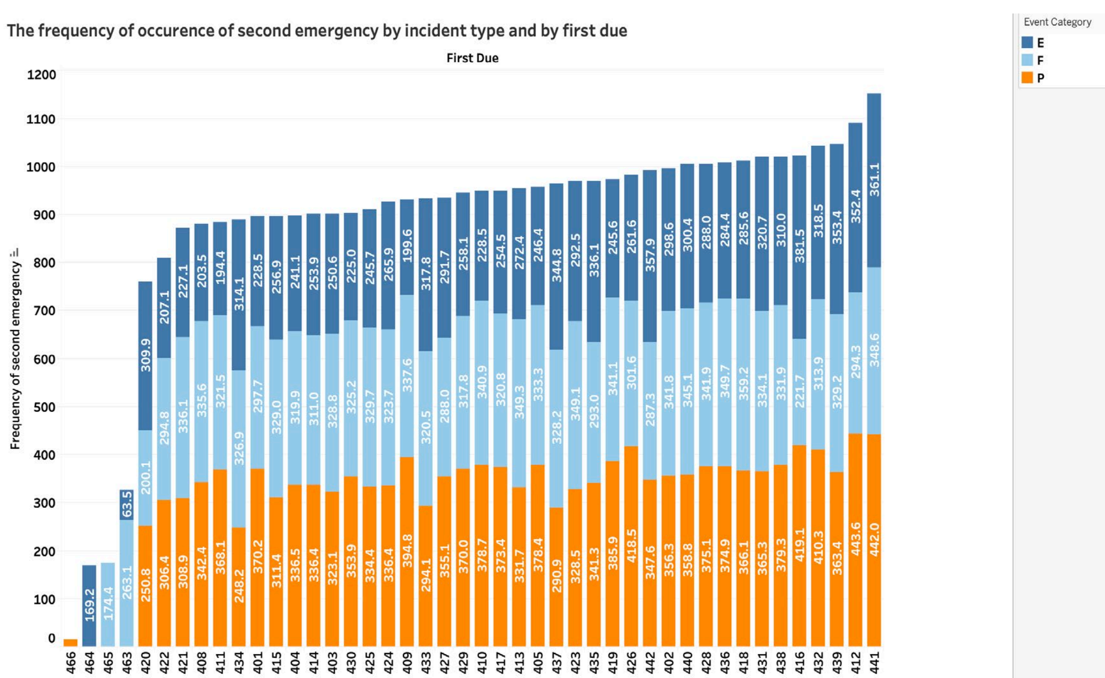


Figure 11: The frequency of occurrence of second emergency by incident type and first due

From Figure 11, it is evident that the stations 463,464,465,466 are the busiest stations that receive second emergencies frequently when compared with other stations. However, these stations handled only specific event categories but not all. The busiest stations that handle all three categories are 420,421, 422 while the slowest stations being 412, 439, 441.

5 Findings

Upon performing chi-square tests on the variables in the EMS Call Outcomes dataset, we found that the chi-square tests resulted in extremely low p-values, indicating that the variables tested have a significant relationship and are therefore dependent. We started with the null hypothesis that the variables in the dataset are independent, but because low p-values result in the null hypothesis being rejected, we found that all the variables with incredibly low p values are significant and thus are dependent. After calculating Cramer's V statistics for the variables in question, however, we found that the correlation between the variables in question and the outcome variable was exceptionally low, and thus the variables we tested (provider, shift, Station, and Battalion) have little to no correlation with the dependent variable (Ems call outcome).

After doing linear regression models for the output2 dataset, we found that regarding procedures performed, procedure wound -burn Care contributes the most to incident distance from the hospital relative to procedure Airway - Clear/Suction, with a coefficient value of 36544.857 (meaning that an EMS call involving the wound – burn care procedure increases incident distance from the hospital relative to an EMS call involving an airway – clear/suction procedure by 36544.857 feet) followed by procedure Enviro - Hypothermia Induction Therapy, with a coefficient value of 15029.565 (meaning that an EMS call involving enviro – hypothermia induction therapy increases incident distance from the hospital relative to an EMS call involving an airway – clear/suction procedure by 15029.565 feet). Regarding medications given, medication (Valium) contributes the most to incident distance from the hospital relative to medication Acid (Aspirin, ASA), with a coefficient value of 32419.578 (meaning that an EMS call involving diazepam (valium) increases incident distance from the hospital relative to an EMS call involving aspirin by 32419.578 feet)), followed by medication Cyanokit (Hydroxocobalamin), with a coefficient value of 16932.678 (meaning that an EMS call involving Cyanokit (hydroxocobalamin) increases incident distances form the hospital relative to an EMS call involving aspirin by 16932.678 feet). In summary, the coefficients indicate either an increase or a decrease in incident distance from the hospital relative to the baseline variable.

Upon calculating the Time Length variable in the Incidents by Unit dataset, which was in terms of days, we then calculated Time Length in terms of seconds and minutes. After running descriptive statistics on the Time Length column, we found that the average response time to an incident, regardless of unit, is 31.54355171 minutes, or 1892.613102 seconds (about 31 and a half minutes). Further analysis by event category in the Incidents by Unit dataset indicated that the average response time for the category E (EMS) is 34.2715525008889 minutes; for the category F (fire), it is 24.6484871656653 minutes; for the category T (test events), 22.403125 minutes; for the category P (public service), 26.1925733319382 minutes, and for the category N (notifications and other events), 3.64970338983051 minutes.

The stations 463, 464, 465, and 466 are the busiest stations that receive second emergencies frequently when compared with other stations. However, these stations handled only specific event categories but not all. The busiest stations that handle all three event categories (E/F/P) are 420,421, and 422 while the slowest stations being 412, 439, and 441.

The minimum frequency of units dispatched is 0.2 minutes and the maximum value is 1,500, 000 minutes. UT402, UT 421, BIC 402 are the busiest units while IV16, IV05 are the slowest units with the frequency greater than 900 days (~ 1,300,000 minutes).

6 Summary

In summary, we came to the following conclusions. First, we found that the provider, shift, station, and battalion variables, though significantly dependent, are uncorrelated with EMS call outcome. Second, we found that the diazepam (valium) medication contributes the most to incident distance from the hospital relative to aspirin, and the wound-burn care procedure contributes the most to incident distance from the hospital relative to an airway – clear/suction procedure. Third, EMS and public service events have very long response times compared to notifications and other events and test events. Finally, stations 463, 464, 465, and 466 are the busiest stations that receive second emergencies frequently when compared with other stations, with regards to handling only specific event categories but not all. In contrast, the busiest stations that handle all three event categories (E/F/P) are 420,421, and 422 while the slowest stations being 412, 439, and 441. The units UT402, UT 421, BIC 402 are the busiest units while IV16, IV05 are the slowest units with the frequency greater than 900 days (~ 1,300,000 minutes).

7 Future Work

In this project, we solved four out of five problem statements that Fairfax fire and rescue department are willing to know about, and those who want to continue the work can solve the last problem statement which is "based on Automatic Vehicle Location (AVL) vs Mobile Computer Terminal data, i.e., how accurate are our personnel provided timestamps? Should AVL ping rates be shortened? Are our members clicking enroute significantly before the vehicle moves?". Another potential future work is to see if including an additional unit in the battalion can reduce the response time of handling an incident.

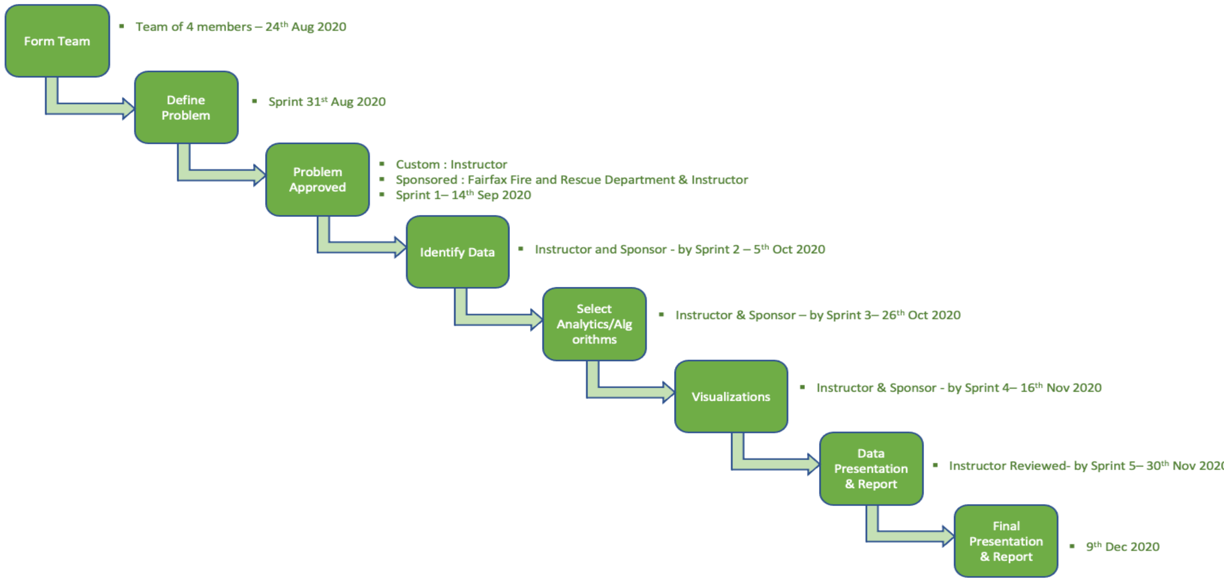
8 Appendix

Link for Project code: https://github.com/asbloom630/Project-DAEN-690

9 Appendix

No risks or mitigations found

10 Appendix



11 Appendix

1. Department, F. c. (n.d.). Retrieved from Fairfaxcounty.Gov: https://www.fairfaxcounty.gov/fire-ems/

2. NFPA. (2015). Data *in the Fire Service.* Retriever from https://nfpa.org/-/media/Files/News-and-Research/Resources/Fire-service/Responder-Forum/2015-NFPA-Responders-Forum-Fire-Service-Data.ashx

3. FEMA. (2004). *Fire Data Analysis Handbook Second Edition*. Retrieved from https://www.usfa.fema.gov/downloads/pdf/publications/fa-266.pdf

4. ESRI. (n.d.). Retrieved from https://www.esri.com/en-us/industries/fire-rescue-ems/overview

5. ArcGIS. (n.d.). Retrieved from https://solutions.arcgis.com/local-government/fire-service/