Exploratory Data Analysis

Jackson Gawarecki

Jgawarecki2@bellarmine.edu

2/6/2024

Introduction

Exploratory Data Analysis (EDA) is a process used by data scientists to analyze and investigate datasets and summarize their main characteristics. It involves the initial examination and visualization of the data to find patterns and relationships. Throughout this EDA, many data analysis tools will be employed, including Python and Tableau. The goal of this process is to better understand the relationship between the data and, in turn, better manipulate it.

This exploratory data analysis project utilizes data from the publicly accessible Louisville Metro Open Data website. It gives information about when, where, and why emergency calls were made and to which department. To begin working with the data, first all the non-numeric data had to be transformed into dummy variables. This mainly involved the different agencies and emergency event types. Along with these transformations, a response time column was added to the end of the dataset. After this process, the dataset was 167,586 rows long and 119 columns wide.

I chose this dataset because of the connection and interest I have in firefighting and the city of Louisville. I wanted to use real-life data and, with my dad being a former captain on the fire department, it seemed like the right fit. The size of the dataset also was a contributing factor as it holds over 150 thousand rows of data. Along with the years of data, there are many types of events that have happened over the years, some needing more of an explanation than others.

Data Set Description

Name	Data Type	Range of Values	NaN Percentage	Description
DATE	interval	3/1/2021-2/28/2023	0	day/month/year of each emergency
CREATE	interval	0:00:00-23:59:59	0	Hour:Minutes:Seconds of when each emergency was first called in
DISPATCH	interval	0:00:00-23:59:60	2.4608	Hour:Minutes:Seconds of when each emergency was disbatched to a fire department

ENROUTE	interval	0:00:00-23:59:61	17.6912	Hour:Minutes:Seconds of when the fire department was on their way to the scene of the emergency
ARRIVE	interval	0:00:00-23:59:62	11.8983	Hour:Minutes:Seconds of when the fire department got to the scene of the emergency
CLEAR	interval	0:00:00-23:59:63	0	Hour:Minutes:Seconds when the emergency was resolved/cleared
HOUR_OF	interval	0-2300	0	Specific hour during the day when the emergency event occurred in military time
PRIORITY	interval	0-9	0	Level of priority from 0-9. 9 being the least important, 1 being the most important/urgent.
Agency_Buechel FD	Nominal	0-1	0	Emergency will be responded by Buechel FD
Agency_Camp Taylor FD	Nominal	0-1	0	Emergency will be responded by Camp Taylor FD
Agency_Eastwood FD	Nominal	0-1	0	Emergency will be responded by Eastwood FD
Agency_Fairdale FD	Nominal	0-1	0	Emergency will be responded by Fairdale FD
Agency_Fern Creek FD	Nominal	0-1	0	Emergency will be responded by Fern Creek FD
Agency_Highview FD	Nominal	0-1	0	Emergency will be responded by Highview FD
Agency_Jeffersontown FD	Nominal	0-1	0	Emergency will be responded by Jeffersontown FD

Agency_Louisville Fire Department	Nominal	0-1	0	Emergency will be responded by Louisville Fire Department
Agency_Middletown FD	Nominal	0-1	0	Emergency will be responded by Middletown FD
Agency_Mutual Aid Fire District	Nominal	0-1	0	Emergency where there was an fire
Agency_Okolona FD	Nominal	0-1	0	Emergency will be responded by Okolona FD
Agency_Pleasure Ridge Park FD	Nominal	0-1	0	Emergency where there was an park
Agency_Shively FD	Nominal	0-1	0	Emergency will be responded by Shively FD
Agency_Special Rescue Team	Nominal	0-1	0	Emergency will be responded by Special Rescue Team
Agency_St Matthews FD	Nominal	0-1	0	Emergency will be responded by St Matthews FD
Event_Type_ACCIDENT--Auto_Accident_vs_Structure	Nominal	0-1	0	Emergency involves collision between a vehicle and a structure
Event_Type_ACCIDENT--Auto_Accident_w/Rescue	Nominal	0-1	0	Emergency involves auto accident requiring rescue operations to free individual(s) trapped inside
Event_Type_ACCIDENT--Auto_Rescue_/_Train	Nominal	0-1	0	Emergency involves vehicle collision with a train, requriing rescue of individuals involved
Event_Type_ACCIDENT--Injury_Accident	Nominal	0-1	0	Emergency where there was an injury
Event_Type_ACTIVE_AGGRESSOR	Nominal	0-1	0	Emergency where there is an active aggressor involved
Event_Type_AIRCRAFT--Aircraft_Crash_Alert	Nominal	0-1	0	Emergency where there is a potential aircraft crash prompting

				emergency preparedness
Event_Type_AIRCRAFT--Aircraft_has_Crashed	Nominal	0-1	0	Emergency where an aircraft has crashed
Event_Type_ALARM--Alarm_Sounding-BabyBox	Nominal	0-1	0	Emergency where a baby box alarm is sounding indicating a potential baby dropoff
Event_Type_ALARM--CO_Alarm_Sounding_with_Illness	Nominal	0-1	0	Emergency where Carbon monoxide alarm activation with reported illness, requiring immediate attention.
Event_Type_ALARM--CO_Alarm_Sounding_without_Illness	Nominal	0-1	0	Carbon monoxide alarm activation without reported illness, necessitating investigation.
Event_Type_ALARM--CO_Detector_Sounding	Nominal	0-1	0	Activation of a carbon monoxide detector, indicating a potential hazardous situation.
Event_Type_ALARM--Fire_Alarm_Sounding-Commercial	Nominal	0-1	0	Activation of a fire alarm system in a commercial property, requiring investigation and potential firefighting.
Event_Type_ALARM--Fire_Alarm_Sounding-Residential	Nominal	0-1	0	Activation of a fire alarm system in a residential property, necessitating investigation.
Event_Type_ASSIST--Assist_EMS_-_Code_1	Nominal	0-1	0	Assistance to Emergency Medical Services (EMS) with a lower-priority, non-life-threatening incident
Event_Type_ASSIST--Assist_EMS_-_Code_3	Nominal	0-1	0	Assistance to Emergency Medical Services (EMS) with a higher-priority,

				potentially life-threatening incident.
Event_Type_ASSIST--Assist_Police	Nominal	0-1	0	Assistance provided to law enforcement agencies for various situations.
Event_Type_ASSIST--Lap_In	Nominal	0-1	0	Assistance involving a person in need (Lap_In), requiring support or intervention.
Event_Type_ASSIST--Mutual_Aid_Response	Nominal	0-1	0	Collaborative assistance provided to another emergency response agency or jurisdiction.
Event_Type_ASSIST--Public_Assist	Nominal	0-1	0	Assistance provided to the public in non-emergency situations.
Event_Type_ASSIST--Safe_Place	Nominal	0-1	0	Assistance in providing a safe location for individuals in distress.
Event_Type_Arson_Investigation--Arson_Investigation	Nominal	0-1	0	Investigation of a fire incident suspected to be caused intentionally.
Event_Type_CHEMICAL--CHEMICAL_ODOR_IN_THE_AREA	Nominal	0-1	0	Report of a chemical odor in the area, requiring assessment and potential hazard mitigation.
Event_Type_CHEMICAL--Chemical_Spill	Nominal	0-1	0	Uncontrolled release of chemicals, necessitating containment and cleanup.
Event_Type_CHEMICAL--Chemical_Spill-Roadway	Nominal	0-1	0	Chemical spill on a roadway, requiring emergency response and road closure.
Event_Type_CHEMICAL--Chemical_Spill/Fire	Nominal	0-1	0	Chemical spill with a concurrent fire, requiring immediate

				response for containment and firefighting.
Event_Type_CHEMICAL--Mitigation	Nominal	0-1	0	Activities focused on reducing or preventing the impact of a chemical incident.
Event_Type_CHEMICAL--Wash-Off	Nominal	0-1	0	Cleaning or decontamination procedures following a chemical incident.
Event_Type_FIRE--Automobile_Fire	Nominal	0-1	0	Fire involving a motor vehicle, requiring extinguishment.
Event_Type_FIRE--Boat_or_Barge_Fire	Nominal	0-1	0	Fire incident on a boat or barge, requiring firefighting efforts.
Event_Type_FIRE--Controlled_Burn	Nominal	0-1	0	Planned and controlled burning of an area, monitored to prevent uncontrolled spread.
Event_Type_FIRE--Dumpster_Fire	Nominal	0-1	0	Fire involving a dumpster, necessitating firefighting to prevent spreading.
Event_Type_FIRE--Electrical_Fire-Structure	Nominal	0-1	0	Fire caused by an electrical issue in a structure, requiring immediate firefighting.
Event_Type_FIRE--Electrical_Odor-Structure	Nominal	0-1	0	Report of an electrical odor in a structure, indicating a potential fire hazard.
Event_Type_FIRE--Electrical_Vault_or_Station_on_Fire	Nominal	0-1	0	Fire involving an electrical vault or station, requiring immediate response.
Event_Type_FIRE--Explosion/Fire	Nominal	0-1	0	A combination of an explosion and subsequent fire, requiring emergency

				response and firefighting.
Event_Type_FIRE--FOOD_ON_THE_STOVE	Nominal	0-1	0	Fire incident involving food on the stove, requiring extinguishment.
Event_Type_FIRE--Field/Grass/Brush_Fire	Nominal	0-1	0	Fire incident involving fields, grass, or brush, requiring firefighting efforts.
Event_Type_FIRE--Fire-Type_Unknown	Nominal	0-1	0	Fire of unknown origin, necessitating investigation and firefighting.
Event_Type_FIRE--Fire/Close_to_Structure	Nominal	0-1	0	Fire incident dangerously close to a structure, requiring immediate intervention.
Event_Type_FIRE--Garage/Shed	Nominal	0-1	0	Fire incident involving a garage or shed, requiring firefighting efforts.
Event_Type_FIRE--SMOKE_IN_THE_AREA	Nominal	0-1	0	Report of smoke in the area, necessitating investigation for potential fire incidents.
Event_Type_FIRE--Structure_Fire-Commercial	Nominal	0-1	0	Fire incident in a commercial structure, requiring firefighting efforts.
Event_Type_FIRE--Structure_Fire-Rescue	Nominal	0-1	0	A fire incident involving a structure where there is also a need for rescue operations
Event_Type_FIRE--Structure_Fire-Residential	Nominal	0-1	0	A fire incident specifically occurring in a residential structure.
Event_Type_FIRE--Tanker_or_Train_Fire	Nominal	0-1	0	A fire involving a tanker or train, which may carry hazardous materials.
Event_Type_FIRE--Tractor_Trailer_Fire	Nominal	0-1	0	A fire incident involving a tractor-trailer, typically on a roadway.

Event_Type_FIRE--Transformer_on_Fire	Nominal	0-1	0	A fire incident specifically involving an electrical transformer.
Event_Type_FIRE--Trash_Fire	Nominal	0-1	0	A fire incident involving the burning of trash or refuse.
Event_Type_GAS_LEAK--CO_Leak_with_Illness	Nominal	0-1	0	A gas leak incident, specifically involving carbon monoxide, with reported illness.
Event_Type_GAS_LEAK--Gas_Leak_-_Inside	Nominal	0-1	0	A gas leak incident occurring inside a structure.
Event_Type_GAS_LEAK--Gas_Leak_-_Outside	Nominal	0-1	0	A gas leak incident occurring outside a structure.
Event_Type_GAS_LEAK--Gas_Leak_Outside_w/Fire	Nominal	0-1	0	A gas leak incident outside a structure with an associated fire.
Event_Type_GENERAL_HAZARD--Trees_Down	Nominal	0-1	0	A general hazard incident involving trees that have fallen.
Event_Type_GENERAL_HAZARD--Trees_on_Bldg	Nominal	0-1	0	A general hazard incident where trees have fallen onto a building.
Event_Type_LOCK_OUT/IN--Child_Locked/Car	Nominal	0-1	0	An incident involving a child locked inside a car.
Event_Type_LOCK_OUT/IN--Child_Locked/Car_in_Distress	Nominal	0-1	0	An incident involving a child locked inside a car in distress.
Event_Type_LOCK_OUT/IN--LOCK_OUT/IN_-_INJURY	Nominal	0-1	0	An incident involving a lockout or lock-in situation with reported injuries.
Event_Type_LOCK_OUT/IN--Lock-Out	Nominal	0-1	0	An incident involving a lockout where access is restricted.
Event_Type_LOCK_OUT/IN--Lock_In	Nominal	0-1	0	An incident involving a lock-in where individuals are unintentionally confined.
Event_Type_MEDICAL--MEDICAL_-_CPR	Nominal	0-1	0	A medical emergency requiring cardiopulmonary resuscitation (CPR)
Event_Type_MEDICAL--MEDICAL_-_MED_CALL	Nominal	0-1	0	A general medical emergency call.
Event_Type_MEDICAL--Medical_-_Alpha	Nominal	0-1	0	A medical emergency with an alpha-level response.

Event_Type_MEDICAL--Medical_-_Bravo	Nominal	0-1	0	A medical emergency with a bravo-level response.
Event_Type_MEDICAL--Medical_-_Charlie	Nominal	0-1	0	A medical emergency with a charlie-level response.
Event_Type_MEDICAL--Medical_-_Echo	Nominal	0-1	0	A medical emergency with an echo-level response.
Event_Type_RESCUE--Elevator_Rescue	Nominal	0-1	0	A rescue operation involving individuals trapped in an elevator.
Event_Type_RESCUE--Extrication	Nominal	0-1	0	A rescue operation involving the extrication of individuals, often from vehicles.
Event_Type_SEARCH--Search-Missing_Person	Nominal	0-1	0	A search operation for a missing person.
Event_Type_SPECIAL--Collapse-Ceilling/Walls-Inside	Nominal	0-1	0	A special operation for a structural collapse, specifically involving ceilings or walls inside a building.
Event_Type_SPECIAL--Collapse-Structural	Nominal	0-1	0	A special operation for a general structural collapse.
Event_Type_SPECIAL--Confined_Space_Rescue	Nominal	0-1	0	A special rescue operation involving individuals trapped in a confined space.
Event_Type_SPECIAL--Hazardous_Materials_Incident	Nominal	0-1	0	A special operation involving a hazardous materials incident.
Event_Type_SPECIAL--High_Angle_Rescue	Nominal	0-1	0	A special rescue operation involving a high-angle scenario, such as cliffs or tall structures.
Event_Type_SPECIAL--Trench_Rescue	Nominal	0-1	0	A special rescue operation involving individuals trapped in a trench.
Event_Type_SPECIAL--Water_Rescue	Nominal	0-1	0	A special rescue operation involving individuals in water.
Event_Type_SPECIAL--Watercraft_Rescue	Nominal	0-1	0	A special rescue operation involving individuals on a watercraft.
Event_Type_STEAM_RUPTURE	Nominal	0-1	0	An incident involving a rupture or release of steam.
Event_Type_WATER_LEAK--Water_Leak	Nominal	0-1	0	There is a water leak incident, typically

				involving water discharge without a specific structural or electrical impact.
Event_Type_WATER_LEAK--Water_Leak-Structural-Inside	Nominal	0-1	0	A water leak occurring inside a structure, which may pose potential structural damage.
Event_Type_WATER_LEAK--Water_Leak/Electrical	Nominal	0-1	0	A water leak incident with a potential impact on electrical systems, highlighting the added risk of electrical hazards.
Event_Type_WEATHER--FLOOD_WARNING	Nominal	0-1	0	A warning indicating an imminent or ongoing flood situation, where rising water levels pose a threat to safety and property.
Event_Type_WEATHER--FLOOD_WATCH	Nominal	0-1	0	A watch indicating the possibility of flooding, suggesting conditions favorable for flooding and requiring monitoring.
Event_Type_WEATHER--SEVERE_THUNDERSTORM_WARNING	Nominal	0-1	0	A warning indicating the presence or imminent arrival of a severe thunderstorm, often involving intense rainfall, strong winds, and potential hazards.
Event_Type_WEATHER--THUNDERSTORM_WATCH	Nominal	0-1	0	A watch indicating the potential development of thunderstorms, urging vigilance and preparedness.
Event_Type_WEATHER--TORNADO_WARNING	Nominal	0-1	0	A warning signaling the presence or imminent formation of a tornado, indicating a high level of danger and the need for immediate protective actions.
Event_Type_WEATHER--TORNADO_WATCH	Nominal	0-1	0	A watch indicating conditions favorable for tornado formation, urging preparedness and monitoring.
Event_Type_WEATHER--WINTER_WEATHER_WARNING	Nominal	0-1	0	A warning for severe winter weather conditions, such as heavy snowfall, freezing rain, or

				blizzards, requiring caution and preparation.
Event_Type_WIRES_DOWN--Wires_Down_-_Arcing	Nominal	0-1	0	Report of fallen electrical wires with visible arcing, indicating a hazardous situation that requires urgent attention.
Event_Type_WIRES_DOWN--Wires_Down_-_Imminient_Risk	Nominal	0-1	0	Fallen wires posing an imminent risk, suggesting a situation that demands immediate action to mitigate potential dangers.
Event_Type_WIRES_DOWN--Wires_Down_-_Low_Hanging	Nominal	0-1	0	Report of electrical wires hanging at a low height, indicating a potential hazard that needs to be addressed.
Event_Type_WIRES_DOWN--Wires_Down_-_Low_Risk	Nominal	0-1	0	Fallen wires with a lower perceived risk level, suggesting a situation that may require attention but is not immediately dangerous.
Event_Type_WIRES_DOWN--Wires_Down_-_Rescue	Nominal	0-1	0	A report of fallen wires requiring a rescue operation, indicating a situation where individuals may be trapped or in need of assistance.
Event_Type_WIRES_DOWN--Wires_Down_-_Wires_in_Trees	Nominal	0-1	0	Fallen wires entangled in trees, highlighting a potential hazard that requires careful handling and resolution.
Response_Time	Interval	0:02:10-1:45:00	11.8983	The total amount of time it took for emergency services to respond. (Arrival time – dispatch time)

Data Set Summary Statistics

For the dataset summary statistics, the programming language of python was used in the Jupyter notebook. The complete file is available via a GitHub repository; however, the most interesting models are included. The complete file is under 'Exploratory_Analysis.ipyn'. The use of python's libraries and stackoverflow aided in the data analysis.

To understand the data better, a numerical summarization was done to find the count, mean, std, min, and max. For most of the data, this is not very useful as it is comprised of dummy variables of 0 or 1, but for the few columns it may glean some information.

```
df.describe()
```

	HOUR_OF	PRIORITY	Agency_Buechel FD	Agency_Camp Taylor FD	Agency_Eastwood FD	Agency_Fairdale FD	Agency_Fern Creek FD	Agency_Highview FD	Agency_Jefferson
count	167586.000000	167586.000000	167586.000000	167586.00000	167586.000000	167586.000000	167586.000000	167586.000000	167586.00
mean	1308.261430	5.144851	0.021881	0.00136	0.000018	0.018235	0.037115	0.014637	0.03
std	621.016167	2.466530	0.146296	0.03686	0.004231	0.133802	0.189045	0.120096	0.18
min	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.00
25%	900.000000	2.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.00
50%	1400.000000	7.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.00
75%	1800.000000	7.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.00
max	2300.000000	9.000000	1.000000	1.00000	1.000000	1.000000	1.000000	1.000000	1.00

8 rows × 111 columns

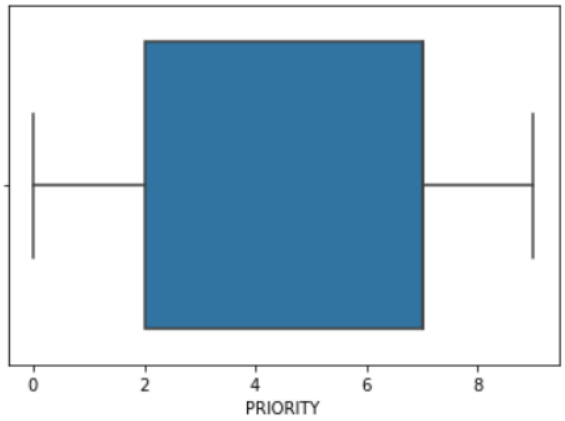
A correlation matrix between 'HOUR_OF', 'PRIORITY', and 'Response_Time' gives us a better understanding of how they are connected. This matrix shows that as the hours of the day go up, the response time should go down meaning that response is faster during the nighttime. It would also suggest that as the priority level increases (less important) the response time slightly increases. This is important because while the time can increase, we still want our emergency response time to be low no matter the event.

	HOUR_OF	PRIORITY	Response_Time
HOUR_OF	1.000000	-0.023982	-0.090863
PRIORITY	-0.023982	1.000000	0.003945
Response_Time	-0.090863	0.003945	1.000000

One interesting column that was chosen to be looked at further was the ‘PRIORITY’ column. After looking at priority number frequencies and proportions, we can see that it ranges from 0 to 9 however, 6 is never chosen. From the numerical summarization and this information, we can make a box and whisker plot to show the range and any outliers that there may be.

```
# Shows the possible outliers
numeric_column_to_analyze = 'PRIORITY'
sns.boxplot(x=df[numeric_column_to_analyze])
```

```
<AxesSubplot:xlabel='PRIORITY'>
```



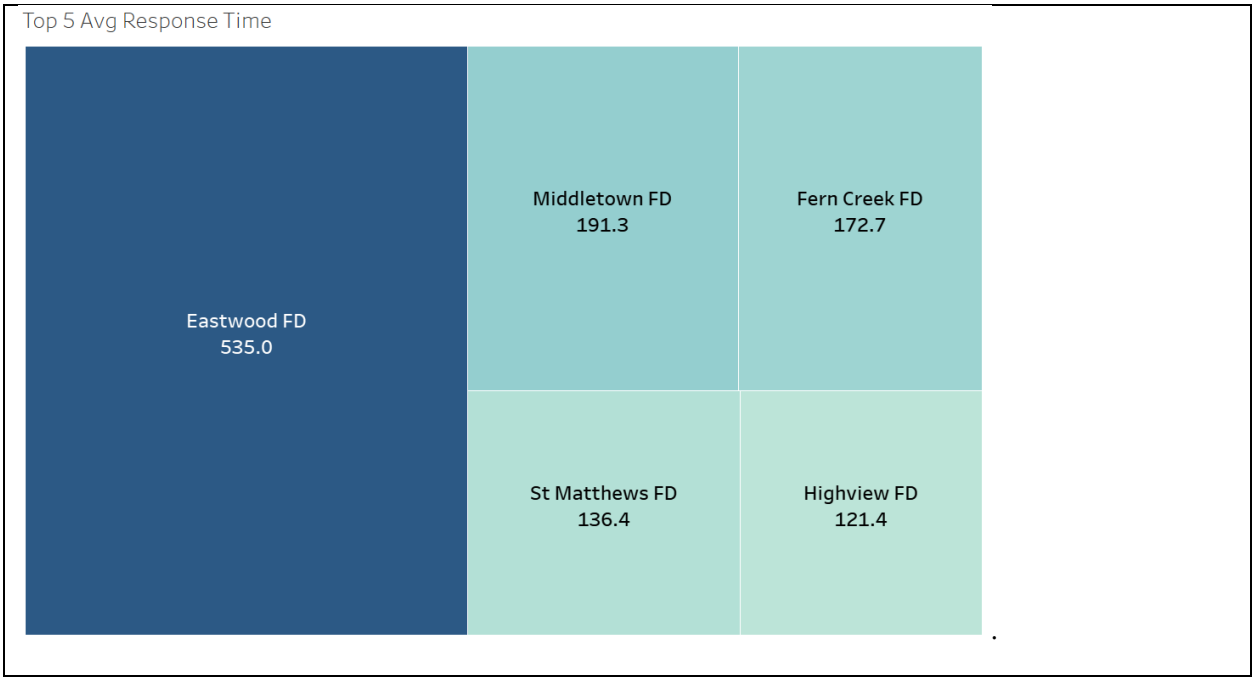

```
proportions
```

```
7    0.568556
2    0.248302
3    0.062463
1    0.058322
8    0.031393
4    0.012591
5    0.011027
9    0.007334
0    0.000012
```

In the GitHub repository, there is more to look through.

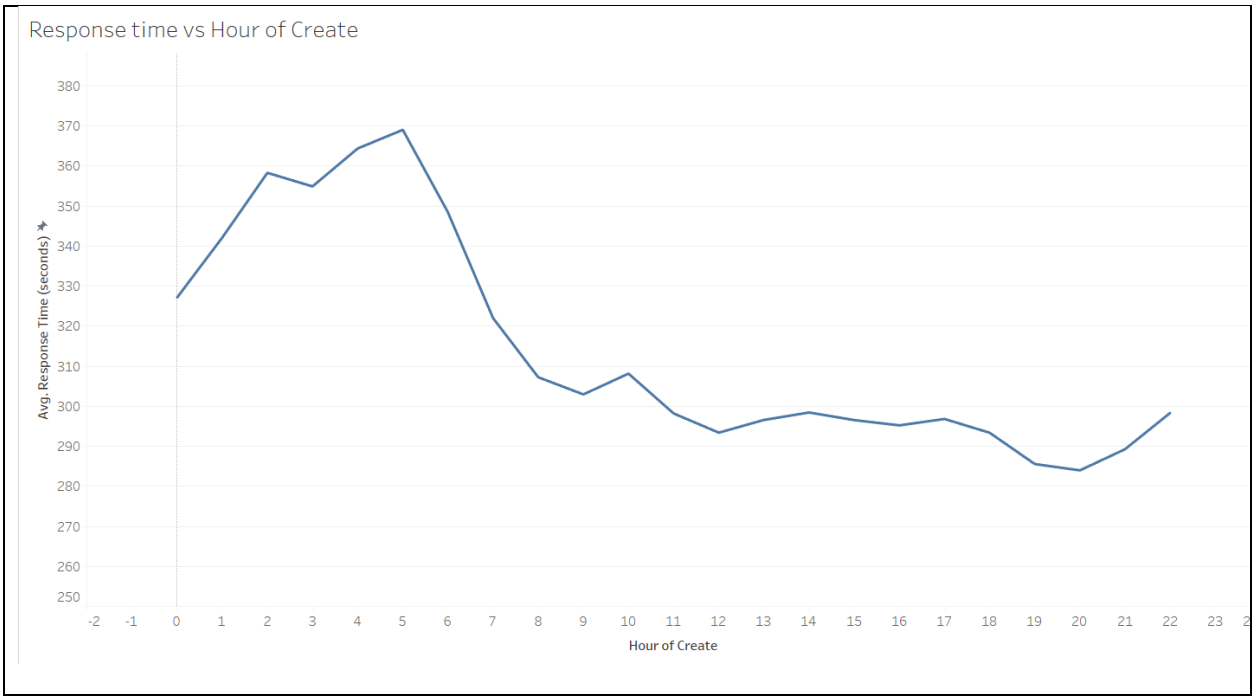
Data Set Graphical Exploration

All visualizations were done using Tableau software. The file for visualizations for the analysis can be accessed from the GitHub repository as well as all the charts and data visualizations. Using Tableau, we can better visualize key statistics in our data. For example,



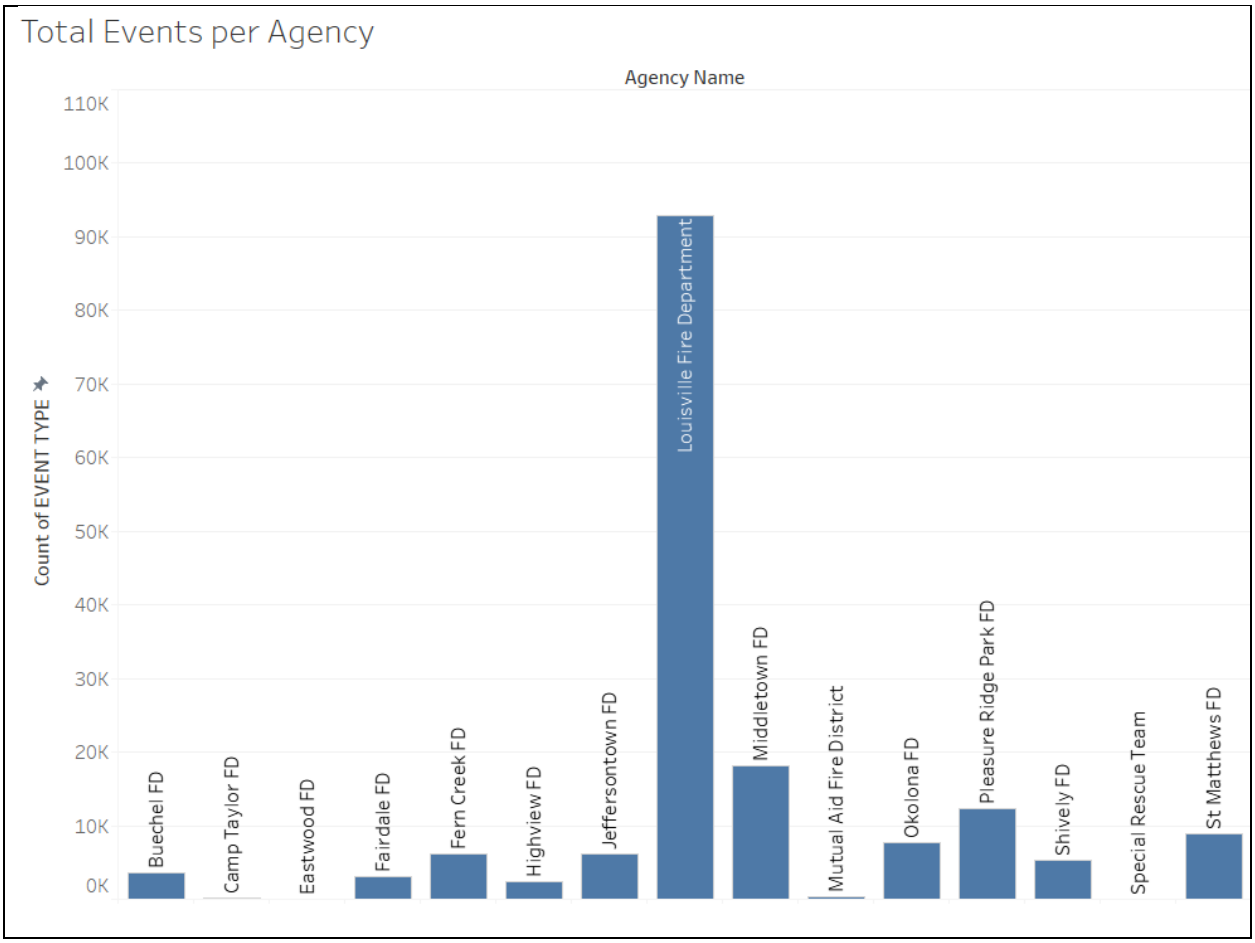
Above are the top 5 longest average response times labeled with the name of the agency, or fire department, and the average response time in seconds. This is important when judging how 'good' a fire department is at their job because we want the response time to be low in most cases. It may also lead us to investigate the types of calls these departments are receiving because they could be low priority events that don't require quick action.

Another visualization that may be helpful is to find out what hour of create (time when call was made) has the lowest response time, and how it fluctuates during the day.



Here you can see that the time of day with the lowest response time is 8 pm with times typically lower during the nighttime compared to the daytime. This confirms the information from the correlation matrix.

Another potentially meaningful fact to know is how the different agencies compare on number of calls. Below there is a bar graph with the count of event types as the y axis and the name of the agency as the x axis. LMFD (Louisville Metro Fire Department) has the most events. This is because it covers the most populated and largest area.



More visualizations and potential questions about how these data points interact are in the GitHub repository under a .twbx file.

Summary of Findings

Throughout the analysis, I have realized that there are too many dummy variables to get any idea of how many of these points go together or are related. Multicollinearity needs to be addressed because the presence of high correlations between these variables can lead to instability in regression coefficients and potentially obscure the true relationships between predictors and the response variable. I also ran into visualization issues when using Tableau because of the high number of variables. To solve this, I kept the original columns in so that they can easily be used. Managing over 100 dummy variables is not easy or efficient, I plan to remedy this problem by grouping event type variables into more generalized fields. Some more important fields would be weather, fire, wires down, and medical call. Because most of them are already ranked in some way, a potential solution could be to re-group them into three

categories increasing in severity each time. This may cause some issues with the ranking system already in place though.

In terms of missing data, there is not much. The most missing in a single row is about 17% for the ‘ENROUTE’ column which is not extremely important as it just means they are traveling. The more important columns, ‘DISPATCH’, ‘ARRIVE’, and ‘Response_Time’, are missing about 2.5%, 12%, and 12% respectively. This 12% missing can be fixed with a simple mean of ‘Response_Time’ because it does not make up a significant part of the data.

Overall, through this EDA I was able to better understand the nature and structure of my data. I think that condensing the amount of columns is an important next step, but with the way it is already labeled there should be no problem in finding a way to organize it.