

Group 19

FINAL PROJECT REPORT

<https://buffalo.box.com/s/jgqv33uc5gnsxy7snvyzwprfzofyw6ir>

Akshayah Karuppayammal Chinnasamy	50419478
Sri Balaji Muruganandam	50414549
Abhishek Pughazhendhi	50414407

Title: 911 - What & How Well Do They Serve

Abstract:

This project analyses the increase/decrease in the efficiency of the 911 emergency helpline over the period of 2014 - 2020 in New York, USA using the datasets that are directly acquired from NYC's official open data, Exploratory Data Analysis, regression, and classification. Response time, time taken for dispatch, time taken to arrival, and the number of incidents are the major predictors of consideration.

Introduction:

- Every year, the 911 helpline handles tens of millions of calls and tends to the emergencies of people with various needs related to possible crimes, non-critical and critical medical urgencies, etc.,
- 911 collaborates with various agencies which include, EMS (Emergency Medical Services), FD (Fire Department), and NYPD (Police Department).
- The government records every single call that's been placed/handled by 911

Datasets:

- **Dataset 1:** [911 Open Data Local Law](#) (Rows - 6047 & Columns - 6)
- **Dataset 2:** [911 End-to-End Data](#) (Rows - 6348 & Columns - 30)
- **Dataset 3:** [Emergency - 911 Calls](#) (Rows - 663525 & Columns - 9)

Exploratory Data Analysis (Results):

- The maximum number of 911 calls were recorded for **“Non-Life Threatening Med Emergencies”** - 4.48 Million followed by **“Life-Threatening Med Emergencies”** - 3.25 Million and **“Possible Crimes”** - 2.24 Million. **(Figure 1)**
- **“NYPD (Non - CIP)”** division has handled the most number of 911 calls over 2014 - 2020 - 8.95 Million. **(Figure 2)**
- Most numbers of 911 calls were placed at the **17th hour** of the day (05:00 PM) - **44,119** & **Friday** of every week - **102942**. **(Figure 3 & 4)**
- The number of incidents per year has gradually increased over 2013 - 18 but declines abruptly from 2018 - 2020. **(Figure 5)**

- The total number of emergency calls made every month declines (**Figure 7**) but also remains between **~60,000 - ~50,000. (Figure 6)**
- The number of emergency calls per year gradually increases and peaks in 2018 (**Figure 8**) & but abruptly declines after 2019 (**Figure 9**).
- The **average response time** (the time it takes for a branch of 911 to respond to a recorded incident) has decreased over the years 2013 - 2020 from **~900 seconds (15 minutes) - ~600 seconds (10 minutes). (Figure 10)**

Regression Model:

- Using Scikit-Learn, linear & logistic regression models have been implemented to predict the response time and to classify the type of incidents respectively.
- X - Number of Incidents, Y - Response time were selected as the predictor and the target variable respectively to perform **single-variable linear regression** which yielded an **intercept - 270.0903** and **coefficient - 0.0101. (Figure 12)**
- The accuracy of the aforementioned model was just **18%**.
- **Multi-variable linear regression**, with a lot of added potential predictors was implemented to improve the model accuracy. (**Figure 13**)
- In order for us to do, the categorical variables had to be converted to numerical variables using **integer-encoding** and **one-hot encoding**.
- The improved model yielded an **intercept - 336.69** and an **accuracy - ~35%** which was significantly better than the previous model.

Classification Model:

- Before implementing logistic regression, the data had to be prepared through the removal of unnecessary columns, outliers, null values and formatting the required columns appropriately.
- **Integer and One-Hot Encoding** has been used to convert the categorical variables to numerical values followed by splitting the data into train (**80%**) and test (**20%**).
- The classification model has an **accuracy - 83%**.
- The model metrics of the aforementioned classification can be found in (**Figure 14**).

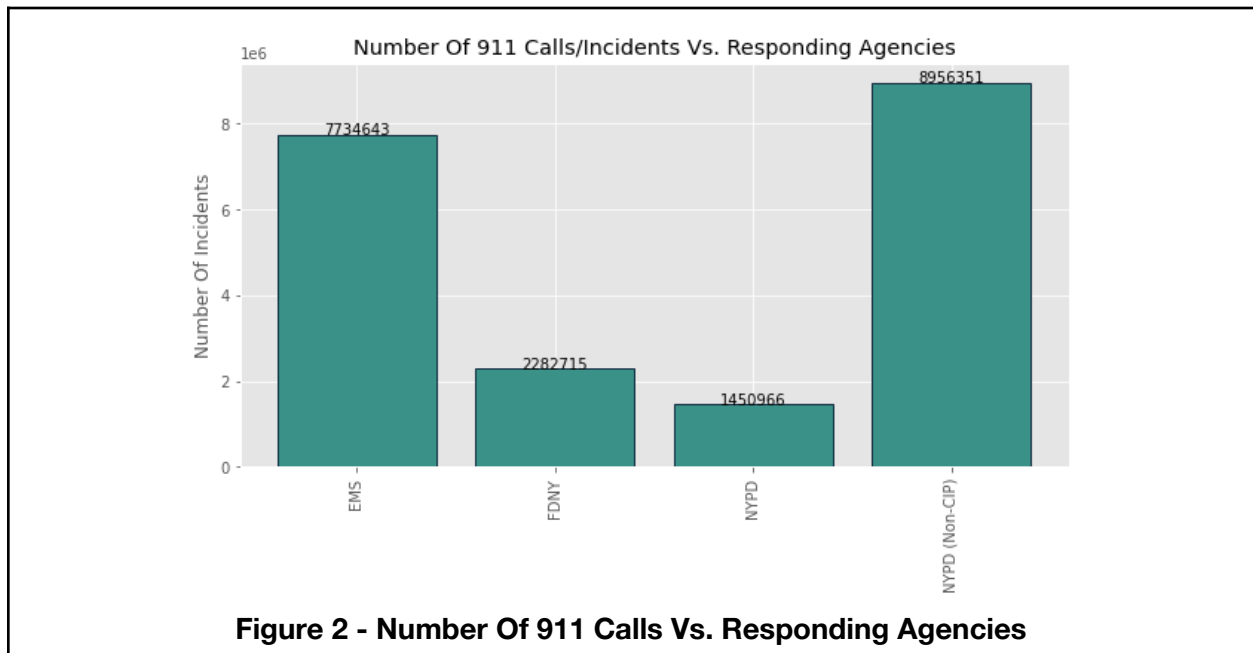
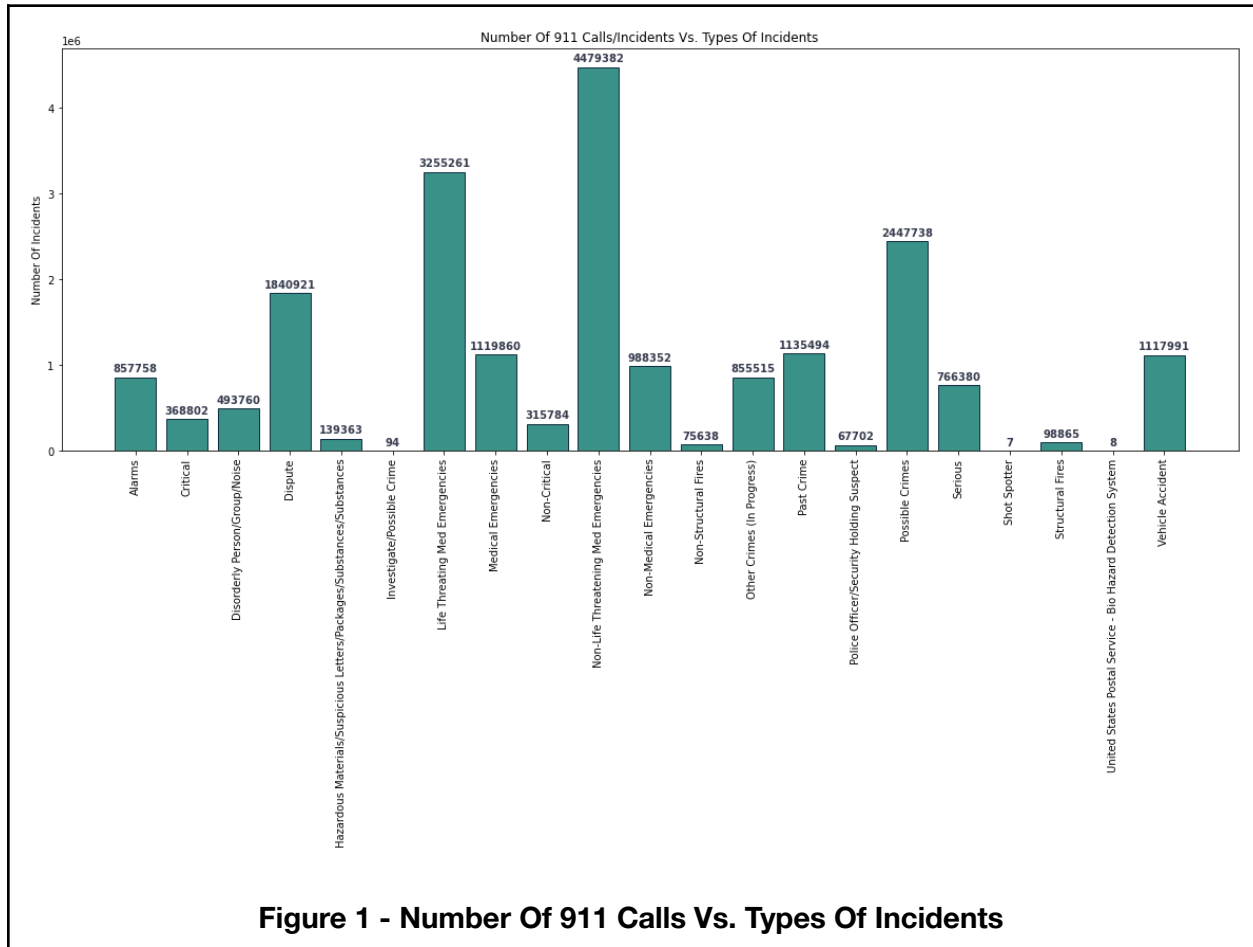
Future Research Directions:

- The models can be further trained to predict the response time (time it takes for a unit to respond to an incident) based on any given predictor (Type of incident, Geographic coordinates etc.).
- The trained model will be able to make multiple predictions (Eg: Time taken for dispatch, arrival, travel etc.) based on a single predictor (Eg: Type of Incident).

Conclusion:

- Considering the average response time, the efficiency of 911 has improved over the years of 2013 - 2020 from **~900 seconds (15 minutes) - ~600 seconds (10 minutes)**.
- The single variable linear regression model had an accuracy of **18%**, the multi-variable linear regression model had an accuracy of **35%** and the classification model had an accuracy of **83%**.

APPENDIX



Calls	
Hour	
17	44119
16	42797
15	42263
14	40346
12	40097

Figure 3 - Frequency Of Calls Vs. Hours Of A Day

Calls	
Day	
Friday	102942
Wednesday	99157
Thursday	98683
Tuesday	97684
Monday	96745

Figure 4 - Frequency Of Calls Vs. Days Of The Week

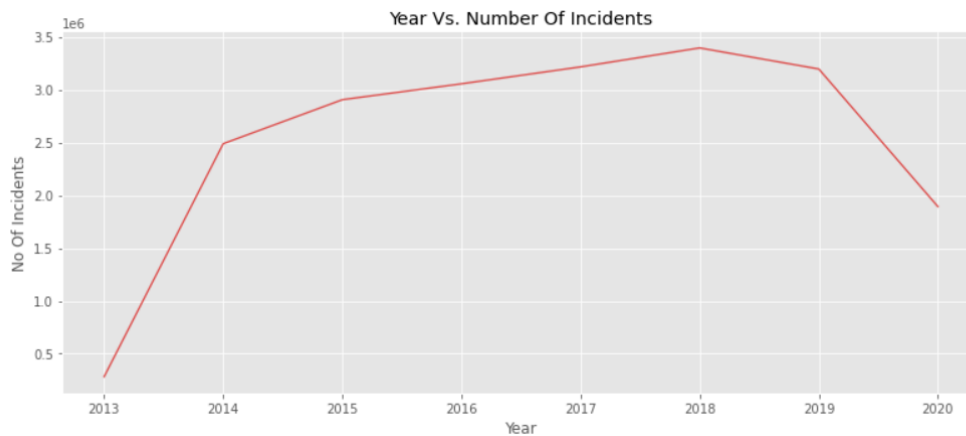


Figure 5 - Number Of Incidents Vs. Year

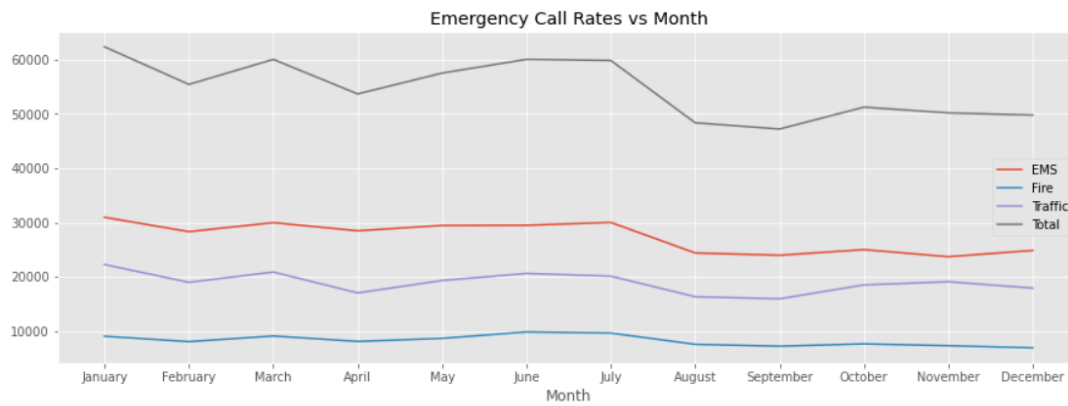


Figure 6 - Trends Of Number Of Emergency Calls Vs. Months

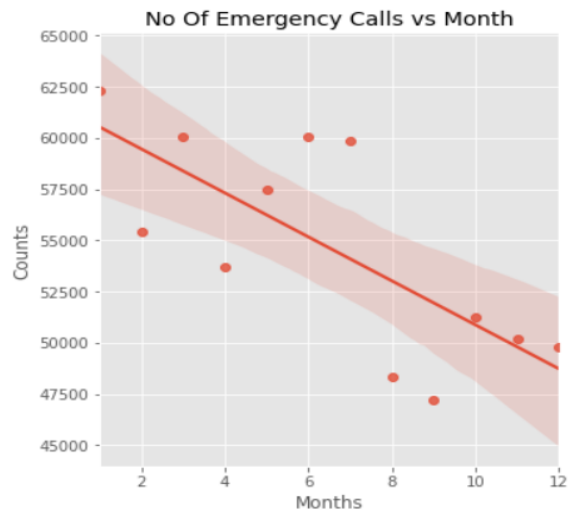


Figure 7 - Regression Plot (Number Of Emergency Calls Vs. Month)

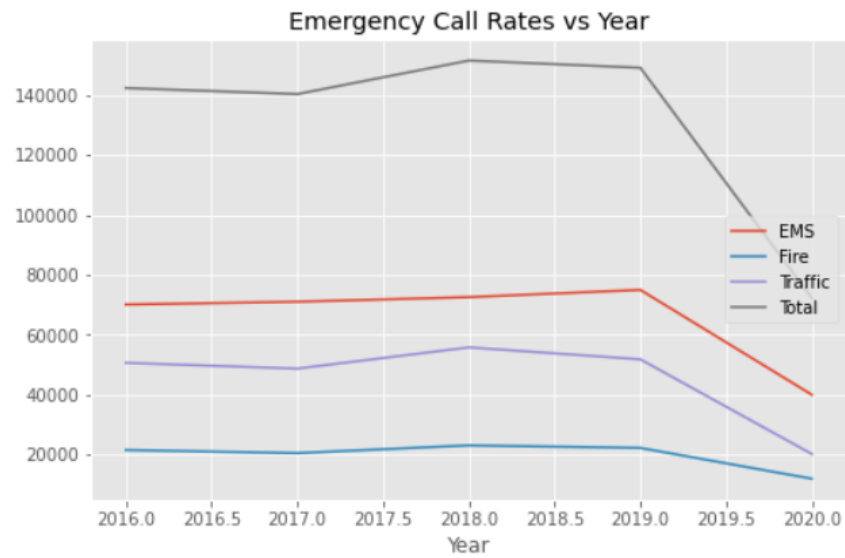


Figure 8 - Trends Of Number Of Emergency Calls Vs. Year

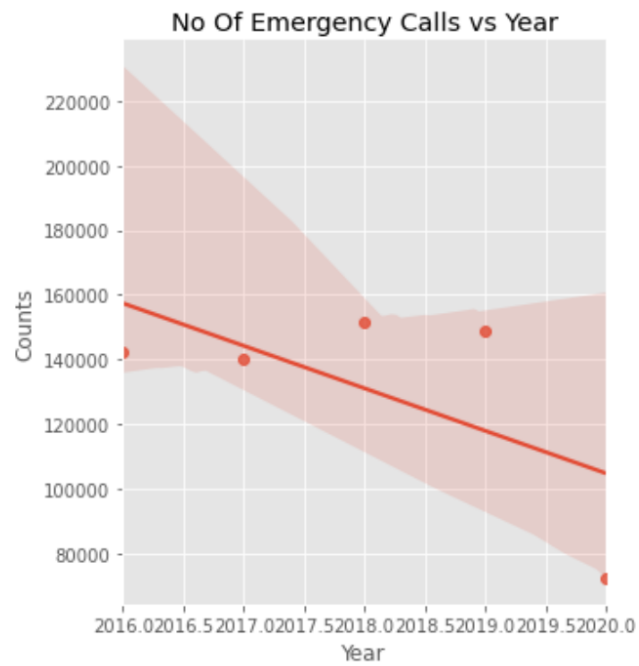


Figure 9 - Regression Plot (Number Of Emergency Calls Vs. Year)

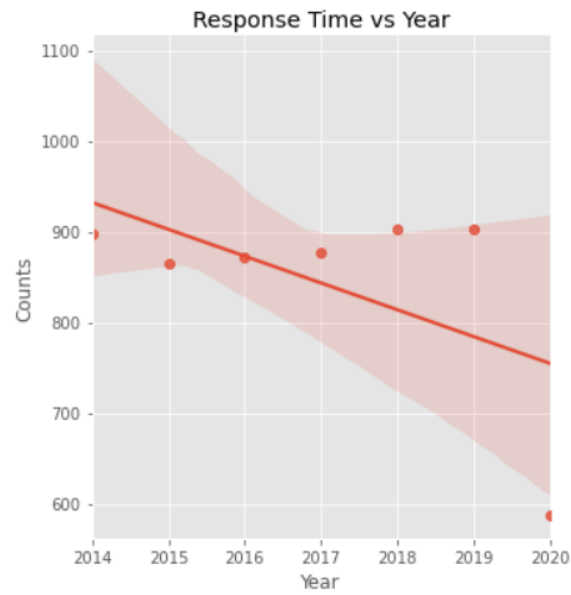


Figure 10 - Regression Plot (Response Time For A 911 Call Vs. Year)

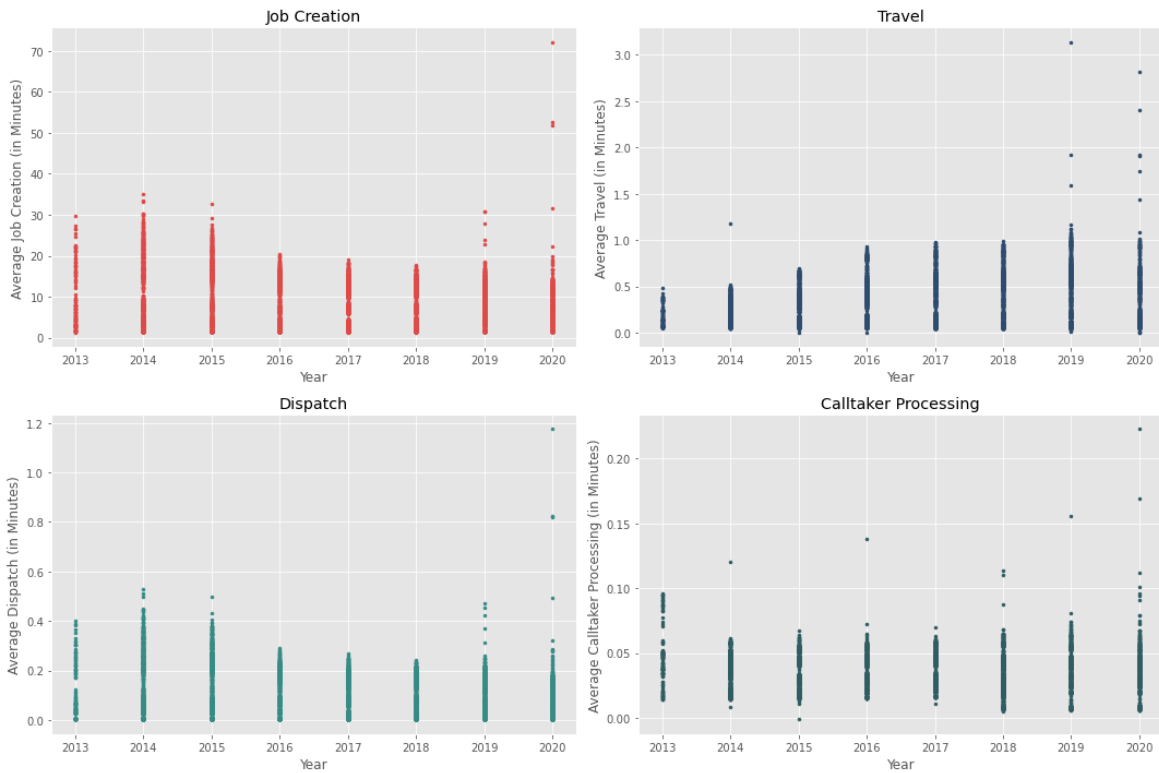
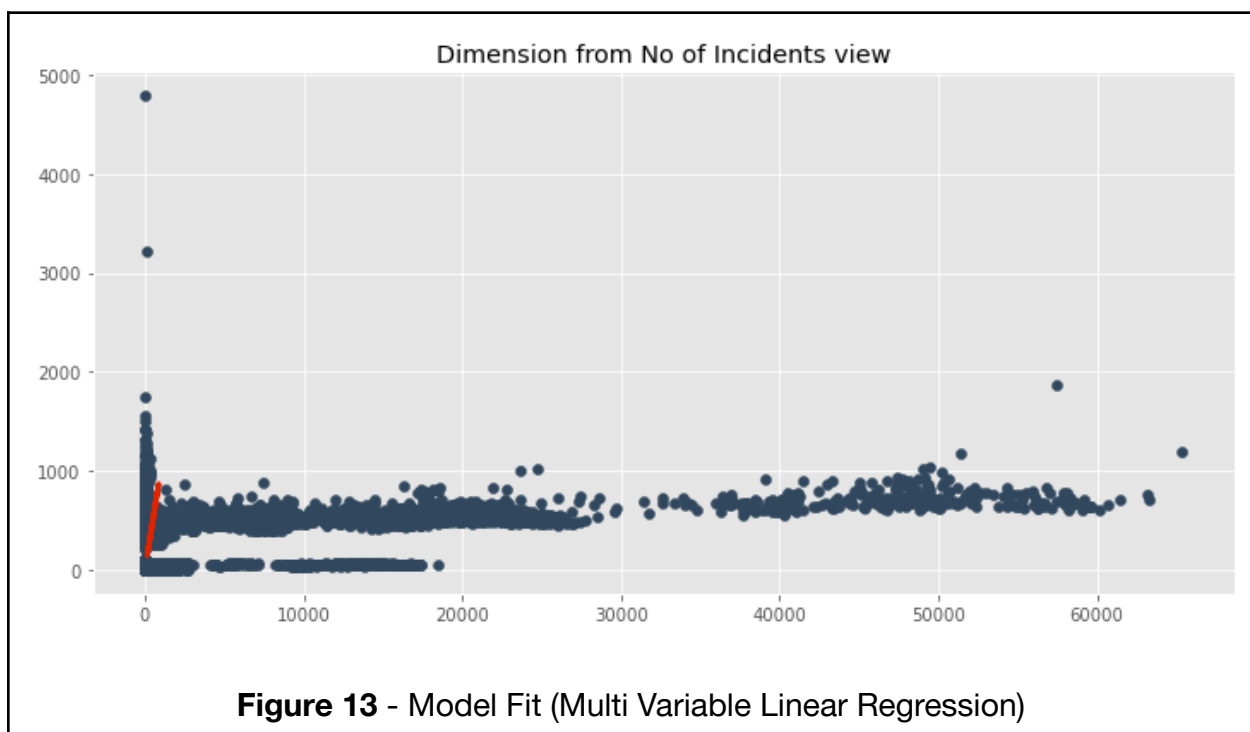
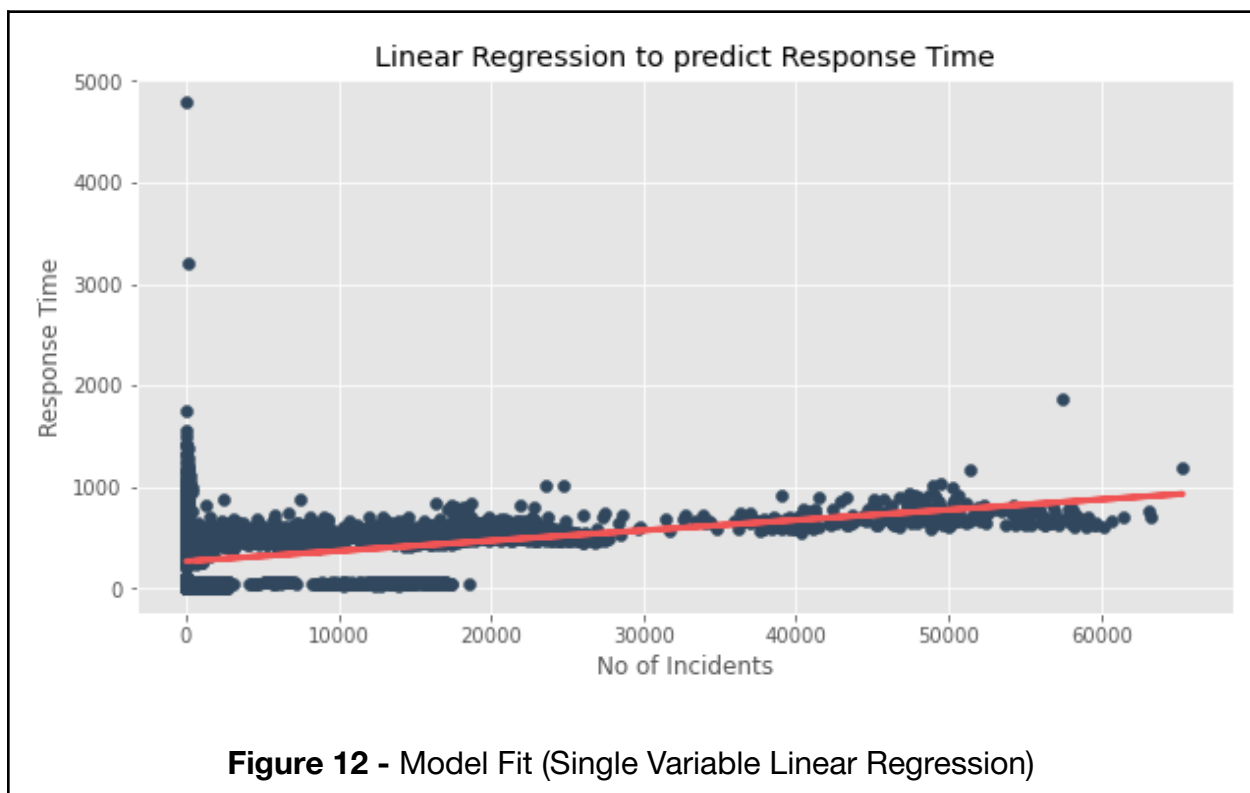


Figure 11 - Time Taken For Dispatch, Travel, Job Creation & Call Processing Vs. Year



	precision	recall	f1-score	support
Alarms	0.82	0.88	0.85	67
Critical	0.94	0.90	0.92	72
Disorderly Person/Group/Noise	0.77	0.81	0.79	81
Dispute	0.93	0.82	0.87	80
Hazardous Materials/Suspicious Letters/Packages/Substances/Substances	0.98	0.97	0.98	65
Investigate/Possible Crime	0.67	0.57	0.62	14
Life Threatening Med Emergencies	0.73	0.48	0.58	79
Medical Emergencies	0.88	0.82	0.85	71
Non-Critical	0.79	0.76	0.78	80
Non-Life Threatening Med Emergencies	0.58	0.78	0.67	69
Non-Medical Emergencies	0.85	0.85	0.85	60
Non-Structural Fires	0.91	0.83	0.87	84
Other Crimes (In Progress)	0.69	0.82	0.75	66
Past Crime	0.78	0.78	0.78	59
Police Officer/Security Holding Suspect	0.95	0.96	0.95	77
Possible Crimes	0.84	0.93	0.88	69
Serious	0.86	0.88	0.87	64
Structural Fires	0.84	0.94	0.89	54
United States Postal Service - Bio Hazard Detection System	0.00	0.00	0.00	2
Vehicle Accident	0.91	0.86	0.88	57
accuracy			0.83	1270
macro avg	0.79	0.78	0.78	1270
weighted avg	0.83	0.83	0.83	1270

Figure 14 - Model Metrics Of Logistic Regression For Classification