Title	911 Emergency Calls Analysis
Student Names	Akhileshkumar Mutuguppe Subbarao Kadesh Basavaraj Huddar
Student IDs	19210140 19210323
Email Address	akhileshkumar.mutuguppesubbarao2@mail.dcu.ie kadesh.basavarajhuddar3@mail.dcu.ie
Programme	MCM – Full Time
Module Code	CA660
Date of Submission	12/14/2019

Disclaimer :

A report submitted to Dublin City University, School of Computing for module CA640 Professional and Research Practice , 2019/2020. We understand that the University regards breaches of academic integrity and plagiarism as grave and serious. We have read and understood the DCU Academic Integrity and Plagiarism Policy. We accept the penalties that may be imposed should we engage in practice or practices that breach this policy. We have identified and included the source of all facts, ideas, opinions, viewpoints of others in the assignment references. Direct quotations, paraphrasing, discussion of ideas from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the sources cited are identified in the assignment references. We declare that this material, which we now submit for assessment, is entirely our own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of our work. By signing this form or by submitting this material online we confirm that this assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study. By signing this form or by submitting material for assessment online we confirm that we have read and understood DCU Academic Integrity and Plagiarism Policy (available at: http://www.dcu.ie/registry/examinations/index.shtml)

Name(s) : Akhileshkumar Mutuguppe Subbarao

Kadesh Basavaraj Huddar

Date : 12/14/2019

911 Emergency Calls Analysis

Akhileshkumar Mutuguppe Subbarao
School of Engineering and Computing
Dublin City University
Dublin, Ireland
akhileshkumar.mutuguppesubbarao2@mail.dcu.ie

Kadesh Basavaraj Huddar
School of Engineering and Computing
Dublin City University
Dublin, Ireland
kadesh.basavarajhuddar3@mail.dcu.ie

***Abstract*— A 911 is the emergency number for citizens of United States of America which may be called in an event of medical emergency services, fire, traffic and other scenarios where immediate action is required. Strategists face difficult situations to provide decisions and resources for coping with these emergencies and sometimes face issues due to information scarcity. Aim of this paper is to perform exploratory data analysis of 911 emergency call received in Montgomery County, PA for the period 12/2015 – 11/2018. Exploring the main category, timestamp, geographic locations of the most frequent calls received is the cornerstone of this analysis. Python was used to explore, plot and visualize the data.**

*Keywords—**9-1-1 dataset , exploratory analysis, emergency services , visualization, hypothesis testing***

I. INTRODUCTION

Citizens in their daily lives face different types of emergencies. Situations that present an imminent risk to life, health, properties, and the environment require emergency intervention to prevent the situation from worsening. Every country has an emergency number which helps in the management of situation and tracking the incident at the earliest. In the United States, the emergency call system was invented as a' 9-1-1' network as a tool of providing easy and fast access to the answering point of public safety. The reason 911 was chosen because it was not widely used as an office / area / service code and also followed the telephone industry's switching configuration plans.[1]

A 9-1-1 system is based on the Public Switched Telephone Network (PSTN).The PSTN network directs 9-1-1 calls to the Public Safety Answering Point (PSAP) nearest to the source of the call. The PSAP operator collects emergency information and alerts the appropriate agency— the police, the fire department, or Emergency Medical Services (EMS), or sometimes all three. Delayed or negligent response to these emergency calls may lead to severe circumstances relating to crime or sometimes death. The popularity and increase in the IT services in present-day situations has led to better communications during emergency services. Bulk calls are received sometimes which cannot be analyzed easily. Modern analysis techniques need to be applied to get a clear understanding of the calls. With this ,it is possible to find places with a high rate of medical emergencies and fire incidents and to allocate resources accordingly. This paper mainly focuses on the exploratory data analysis of 9-1-1 emergency calls data compiled from Montgomery County in Pennsylvania[2] to look at the category of emergency, location with higher calls and other important insights.

This paper is organized as follows: Section 2 describes literature survey on analysis of 9-1-1 calls . Section 3 presents details of the dataset used and exploratory analysis of it and Section 4 details about the hypothesis and research questions. Section 5 describes the methods used for exploring data and Section 6 briefs about the results and findings. Section 7 draws conclusions about the research done.

II. RELATED WORK

Gagandeep analyzes call patterns and checks the parts of the hotspot areas that uses most of 911 services. Natural language tools such as NLTK, Core NLP are used to mine and develop the template of call transcripts[3]. Authors here have developed a model using data mining methods to identify patterns based on the study of 9-1-1 call behavior characteristics. K-means clustering algorithm is used for clustering the dataset. Heat maps are used to illustrate different findings. Hot Spot Analysis is done which is a spatial cluster detection method that identifies statistically significant high and low spatial concentrations associated with a set of geographical characteristics[4].
Authors have taken 911 emergency call information from Portland Oregon metropolitan area was analyzed by the researchers. Hotspot analysis is carried out to determine high call volume areas and spatial analysis is carried out to understand factors that contribute to high call volumes. Authors believe that modelling approach helps shape policy making and offers a roadmap to equal distribution of resources to communities[5]. For the estimation of short-term 9-1-1 emergency call operation, researchers presented a multiple linear regression model. The model shows various capabilities of predictability for different times of the day. It has been shown that the model is successful in identifying anomalously large number of calls from two separate emergency incidents[6].

III. DATASET AND EXPLORATORY ANALYSIS

The dataset is chosen from Kaggle which is titled “Emergency – 911 Calls: Montgomery County, PA”. It contains the emergency calls made for the station from 12/2015 to 11/2018. Dataset is a CSV file with 371711 rows and 8 columns. Dataset is cleaned and columns are removed which are not required for the exploratory analysis. Description column had combination of different entities which are separated by the split function to create separate columns for each entity. The column timestamp is split into year, hour, month, day of week. Reason column is split into main and sub category.

Variable Name	Description
Latitude	Latitude of call region
Longitude	Longitude of call region
Description	Call description
Zip	Zip code of call location
Reason	Category of call received
Timestamp	Call time stamp
Township	Region from where call was placed
Address	Address of the call region

Table 1 : Dataset Attributes

Exploratory analysis for the dataset is as follows:
Loaded data is first checked for the details of the columns and its data type. Describe function is used to get overview of the data from statistical perspective. Mean, median, count, max, min, interquartile range of the data are explored.

Emergency calls are mainly categorized into 3 parts namely:

1. Emergency Medical Services (EMS) - calls for various medical emergencies and crimes(like Head injury, Burn victim, Cardiac arrest, Hit and run, Poisoning).
2. Fire – Calls related to fire accidents.
3. Traffic – calls for road emergencies.[4]

Exploratory analysis is carried out on different aspects of the data. Each categories of call are analyzed to check the maximum emergency recorded in the county station. Analysis on how the emergency calls are distributed in all the days of week and data is explored to see on which days the emergency call rates are maximum. Analysis is applied on categorical value like year to get insight on which month of the year had maximum number of emergency calls. Heatmap is used to check the co-relation between the attributes chosen accordingly. A cluster map is added to get the clear picture of the related attributes. A line graph is chosen to depict the relationship between the emergency call and month of each year. Finally, attributes like top 10 area with highest number of 911 calls and top 10 emergency description call (ex. Fire alarm, Head Injury) are analyzed. The call categories are divided into sub categories to get clear idea about emergency cause.

IV. RESEARCH QUESTIONS AND OR HYPOTHESES

1. What is the total count of different category of calls received at Montgomery County stations?
H0: Emergency calls category are equally distributed.
H1:Emergency calls category are not equally distributed.

2. How emergencies were distributed over the days in week?
H0: Emergency calls and days of week are independent.
H1: Emergency calls and days of week are dependent on each other.

3. How call frequency is distributed over the years?

4. What is the call frequency in different hour in days of week?
H0: Hours in a day and day of week are not related.
H1: Hours in a day and day of week are related.

5. How month and days of the week are co-related on the basis of calls?
H0: Month and day of week are not related.
H1: Month and day of week are related.

6. What is the average call count for different years?

7. Which are the top 10 area from which call are generated?

8. Which is the top emergency case in Montgomery County?

V. METHODS USED AND WHY

Sample data consists of 100 random values taken from total dataset.
For the first hypothesis we have taken null hypothesis as the emergency calls are equally distributed. Since it is categorical and has only one sample, Chi Square Goodness of Fit (One Sample Test) is used[8]. Formula used to calculate chi-square is

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Illustration from the taken sample is as follows. Out of 100 samples considered, 84 sample has emergency medical services, 6 are fire emergency and 10 being traffic emergencies. On calculating the chi square for the data, chi square, $\chi 2$ =43.18. On checking the chi square distribution table for the degree of freedom 2 and $\alpha = 0.05$ the probability level is 5.99. Since 43.18 > 5.99, the result does not fit in acceptance region, hence rejecting the null hypothesis. The same is proved in Figure 1 pie chart.

For the second hypothesis we have taken null hypothesis as emergency calls and days of week are independent. Since both the attributes are categorical, Chi Square test of independence is used[7]. Formula used to calculate chi-square is

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Here is the illustration from the taken sample. On calculating the chi square for the data, $\chi 2 = 198.44$. On checking the chi square distribution table for the degree of freedom 12 (7 days in a week and 3 call category) and $\alpha = 0.05$ the probability level is 21.03. Since 198.44 > 21.03, the result does not fit in acceptance region, hence rejecting the null hypothesis. The same is proved in Figure 2 bar chart.

For the next Hypothesis:
H0: Hours in a day and day of week are not related;
One value is numerical and another one is categorical, we have used Z-test[9]. Formula used to calculate Z-test is

$$Z = \frac{X - M}{\frac{\sigma}{\sqrt{n}}}$$

Here is the illustration of the hypothesis. Mean value of the sample is μ =13.24 and standard deviation of the sample is 5.79, we have taken Friday as sample and the mean is $\mu 0$ = 15.18. On calculating the z value, we get values as -3.35. The critical region tells us to reject the null hypothesis at $\alpha = 0.05$ if $Z < -1.645$. Therefore, we reject the null hypothesis because $Z = -3.35 < -1.645$. Same is proved in cluster map in Figure 4.

For the hypothesis testing: H0: month and day of week are not related.
Here Kruskal-Wallis H Test is used as there is a need to check significant difference between the month and day of week based on emergency calls. Formula to calculate Kruskal-Wallis H Test is[10,11]:

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)$$

Here is the illustration of the hypothesis. Kruskal-Wallis H Test is used to check the hypothesis for the data. All calls received in each day of the week in two random months are ranked as per the model. With $\alpha = 0.05$ and degree of freedom 1 (two months are considered for testing), the critical value (α) from chi-square distribution is 3.841. But calculated results from the test is 5.45. Ustat = 5.45 > 3.841. Hence null hypothesis is rejected. Same is proved in cluster map in Figure 5.

VI. Results and Findings

First Exploratory analysis is conducted on three categories on which calls are segregated. Pie chart is used to analyze the calls registered on different categories and its percentage are calculated. From figure 1 calls received under Emergency Medical Services (EMS) is the highest amongst all with the percentage of 51.2% followed by traffic with 33.7% and Fire with 15.1% respectively.

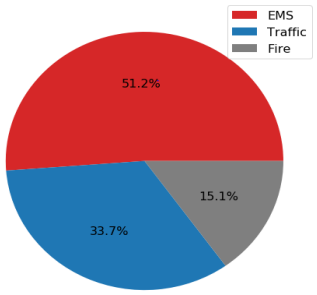

Figure 1

Figure 2 gives an analysis on how different category of calls are distributed over the days of week. On all days Emergency Medical Services (EMS) are the highest calls received with count of more than 25000 per day. Fire breakdown calls are less according to the study. Calls are minimum during the weekends since many of the people will be at home, whereas on Friday calls in all three categories are in huge amount.

Figure 2

Figure 3 illustrates the overall call frequency over the given years. By looking at the graph it is clear that calls have linear distribution, but there are few spikes in March 2018 and late 2018. Month of March in the Year 2018 has received highest number of emergency calls.

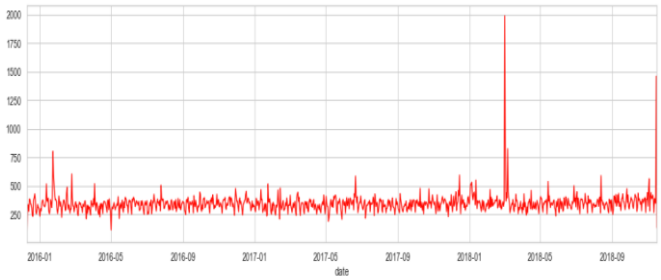

Figure 3

Figure 4 describes the heatmap of how call frequency is co-related among days of week and particular time. We can see that most of the calls are during day time with peak being around 15:00 to 18:00 hours. From night 20:00 hours to morning 06:00 there are least calls received. County stations have received most number of calls on Friday evening and Sunday being the lowest.

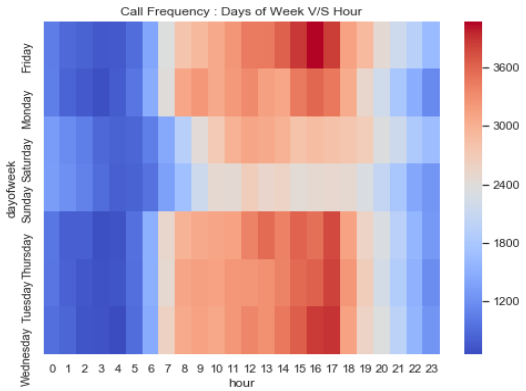

Figure 4

Figure 5 is a dendrogram which depicts the relationship between month and days of the week. From the analysis it is clear that Friday's in month of march have received highest emergencies. November month being the lowest reported month.

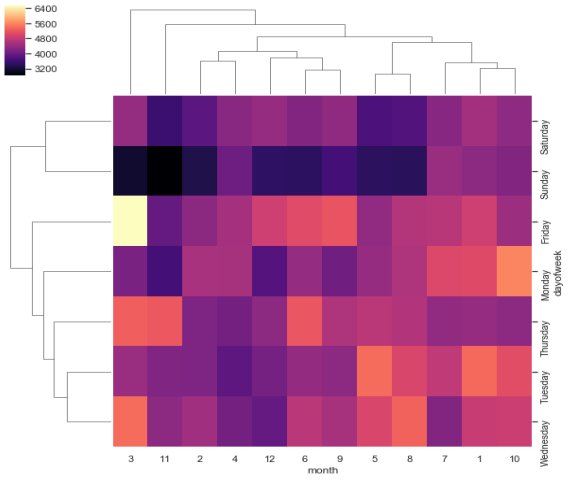

Figure 5

According to the overall analysis of the figure 6 , we can say that average calls in each month of all years is around 10000. Emergency calls are increasing every year and highest spike was received on March 2018 which is confirmed by above cluster map. Since the dataset has data till November 2018 there is a sudden drop in the graph. February 2017 has the lowest calls amongst all years.

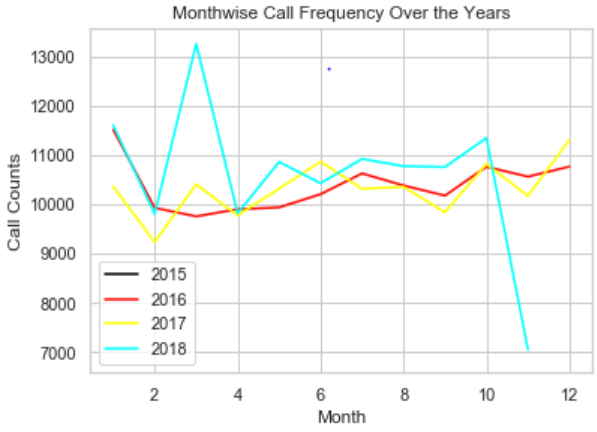

Figure 6

After exploring the data, we have analyzed the top 10 areas from which people have faced the highest number of emergencies. Figure 7 illustrate those top 10 townships. Lower Merion is top one area which has more than 30000 emergency situations. Average amount of calls received in Montgomery County, PA are around 15000.

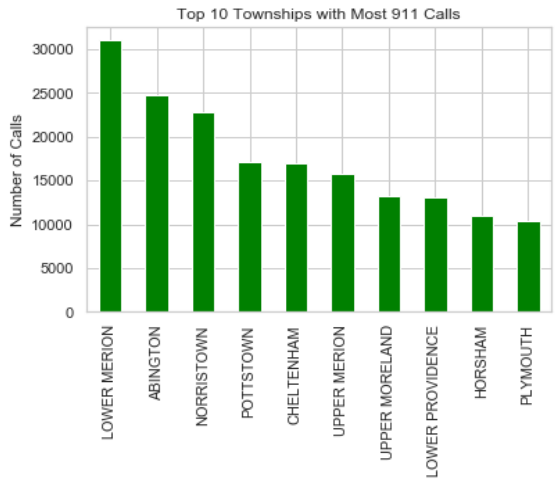

Figure 7

From figure 8, we are analyzing top 10 incidents for 911 calls. Main category reason is sub divided into many different categories. On analyzing these sub categories, 911 emergency calls are mostly due to Vehicle accident with a count of around 90,000 cases. Lowest being recorded for head injuries in the top ten list. After looking at the graph there is a huge gap between vehicle accident and other sub categories, which is a serious concern.

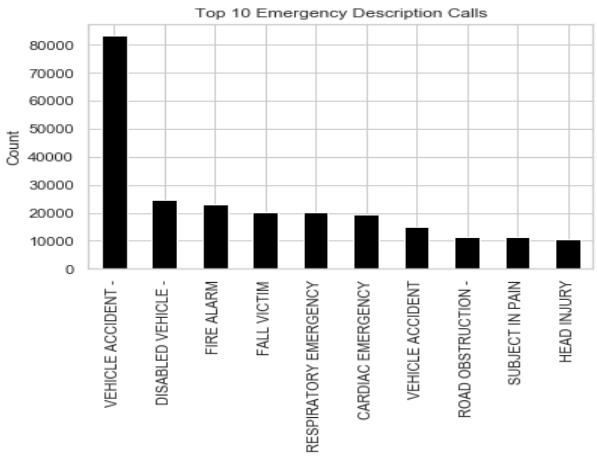

Figure 8

VII. Conclusions

The Exploratory Data Analysis provides insight into the data through statistical and visualization techniques in order to focus on important aspects of the data for further analysis. This paper provides analysis of 911 emergency calls for Montgomery County, PA and following conclusions are obtained. Emergency Medical Services category has the greatest number of calls and the majority of incidents have occurred during evening time from 15:00 – 18:00 with 00:00 – 07:00 being calm hours. Weekends are most placid days of the week since maximum number of people stay at home. Majority of calls are made on Friday of the week, March and October months considering overall years. 2018 has received a greater number of calls in a given year. Maximum calls have originated from Lower Merion township and 19401 zip code making government to concentrate more on these regions. After analyzing the overall records, vehicle accidents are the main reason for most of the emergency calls. Hypothesis testing is done along with exploratory data analysis to get more crisp results which is of utmost importance in emergency situations.

References

[1] 9-1-1 origin amp history - national emergency number association. 2017.

[2] Montgomery County, PA: 911 Calls [Online]
Available: https://www.kaggle.com/mchirico/montcoalert

[3] Malhotra, G. Analyzing 911 data using Natural Language Processing.

[4] Selvam, A., Thiagarajan, B. and T. K., T. (2016). Mining Patterns from 9-1-1 Calls Dataset. International Journal of Applied Information Systems, 11(7).

[5] Cramer, Dan & Brown, Albert & Hu, Gongzhu. (2011). Predicting 911 Calls Using Spatial Analysis. Studies in Computational Intelligence. 377. 15-26. 10.1007/978-3-642-23202-2-2.

[6] Hector Jasso. Tony Fountain. Chaitan Baru. William Hodgkiss. Don Reich. Kurt Warner. Prediction of 9-1-1 Call Volumes for Emergency Event Detection. The Proceedings of the 8th Annual International Digital Government Research Conference.

[7] McHugh, Mary. (2013). The Chi-square test of independence. Biochemia medica. 23. 143-9. 10.11613/BM.2013.018.

[8] Singhal, Richa & Rana, Rakesh. (2015). Chi-square test and its application in hypothesis testing. Journal of the Practice of Cardiovascular Sciences. 1. 10.4103/2395-5414.157577

[9] https://www.pitt.edu/~super1/ResearchMethods/Arabic/HypothesisTestingpart1.pdf

[10] Vargha, Andras & Delaney, Harold. (1998). The Kruskal-Wallis Test and Stochastic Homogeneity. Journal of Educational and Behavioral Statistics - J EDUC BEHAV STAT. 23. 10.2307/1165320.

[11] Ostertagova, Eva & Ostertag, Oskar & Kováč, Jozef. (2014). Methodology and Application of the Kruskal-Wallis Test. Applied Mechanics and Materials. 611. 115-120.