Analyzing Weather Impact on Ambulance Response Time

[1]Rensselaer Polytechnic Institute, Tetherless World Constellation, Troy, NY, United States

Shrey Jain – jains6@rpi.edu

Introduction

- In this world of rapid development, having a perfect health directly means leading a productive life. While prevention is better than cure, there are some cases where some serious, unexpected things happen which require immediate action. In those scenarios it is crucial to provide immediate responses to such emergencies as they could turn out be a lifesaver.
- Another such life saving action is to provide the pre-hospital care to any patient present in an ambulance or waiting for one. Therefore, having an accurate arrival time of ambulance could be very helpful.
- This project covers process of building a machine learning model that predicts a response time of any ambulance in relation with the weather on that day.
- This poster mainly covers details about the datasets in focus, what are the steps that are to be followed to get the data ready for any model building or evaluation. It also covers some of the plots which helps identify the relation between variables and finally contains the results and the conclusion drawn out from those results.

Inspiration

- It is crucial to know the estimated arrival time of an Ambulance in case of any emergency.
- This could help provide pre-hospital care to the patient which could save many lives.
- Weather can have impact on Ambulance times.
- This project performs data pre-processing, identifies relation between Weather & EMS data and develops model to predict future response times in any given weather condition.

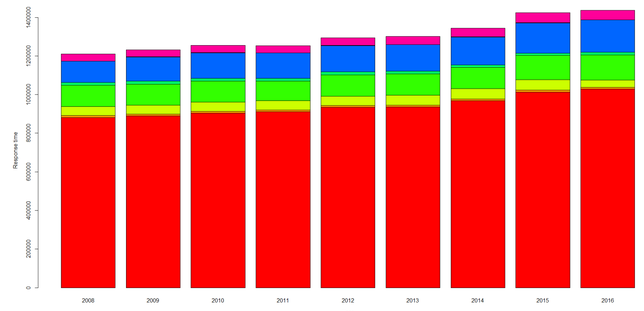
Data Sets

- EMS Dataset
  - Provided by Fire Department of New York City
  - 11863759 number of rows
  - 32 number of columns
  - The target column to be predicted using various machine learning regression models is INCIDENT_RESPONSE_SECONDS_QY
- Weather Dataset
  - Provided by Global Historical Climatology Network
  - 3288 number of rows
  - 34 number of columns
  - Weather on a particular date is used to link these datasets. Datasets are joined using INCIDENT_DT column of EMS and DATE column of Weather dataset

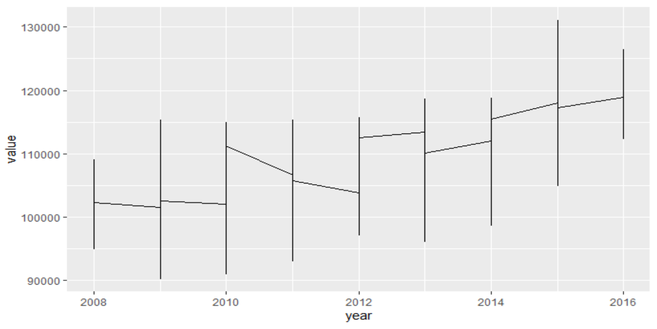
Data Analysis

- Data preparation steps:
  - Quality Assessment
    - ✓ High level checks to determine if the data meet the required quality.
  - Data Cleaning
    - ✓ Fixing incorrect/corrupted entries in dataset
  - Data Munging
    - ✓ Modifying or changing dataset beyond its original state.
  - Exploratory Data Analysis
    - ✓ Performing initial investigation on data to discover patterns and to spot anomalies.
  - Model Preparation
    - ✓ Applying 4 regression models on the final merged datasets.
  - Model Evaluation
    - ✓ Evaluating the accuracy of these models and determining which works best for given set of data.

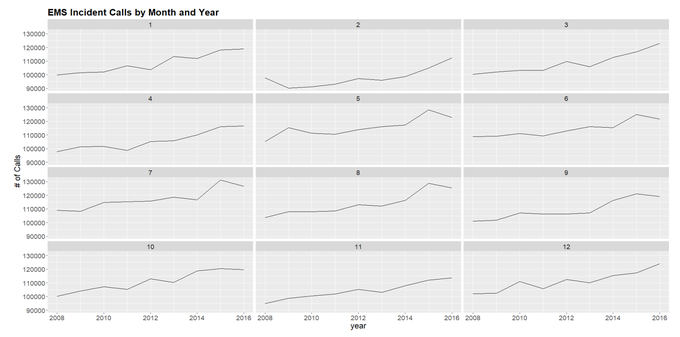
- Bar chart showing response times for different disposition codes. Red color refers to transportation time.



- Plot showing changes in incident counts by year



- Chart describing incident counts by month for every year from 2008-2016



- Following were performed for data cleaning:
  - Removed about 0.22M outliers from the Weather and EMS dataset.
  - Removed about 3.7M rows with N/A values from EMS dataset.
  - Removed 3 columns from Weather dataset as they had majority of N/A values.
  - Converted Date Format & String to numbers.

Model Development

- The final merged dataset obtained after joining the EMS and weather datasets on date column had about 0.36M rows.
- This number is after all data filtering and cleaning were done and dataset was ready for model building.
- Models performed significantly better on datasets without outliers and N/A values.
- Following 4 regression models were built for evaluation:
  - Linear Regression
  - Support Vector Machine
  - Random Forest
  - K-Nearest Neighbors
- Hyper tuning turned out to be an essential tool to optimize model parameters and maximize the performances of all the four models.

Results and Conclusion

- Weather data and EMS response times are not entirely co-related, meaning it does not assert the initial hypothesis I have made that weather has impact on the ambulance response time.
- All the four regression models namely, Linear Regression, SVM, Random Forest, and KNN were built on the same dataset and evaluated basis their accuracy, root mean square error, deviation from the actual expected result, and other evaluation techniques.
- Linear Regression had the highest accuracy and least mean square error amongst other 4 models which means that it is best suited to predict ambulance response times for our problem.
- The aim of the project was to analyze the impact of Weather on EMS response times using the concepts taught by Prof. Thilanka Munasinghe in his Data Analytics class.

Glossary

- EDA: Exploratory Data Analysis
- R: A programming language like Python
- Library(R): Load packages containing useful functions
- EMS: Emergency Medical Services
- Boxplot: Graphical Representation of five num

References

- R Basics: https://towardsdatascience.com/r-basics-everything-you-need-to-know-to-get-started-with-r-10c8e566d7b3
- Scrape-R Package: https://cran.r-project.org/web/packages/scrapeR/scrapeR.pdf
- Read Excel File in R: https://www.datacamp.com/community/tutorials/r-tutorial-read-excel-into-r
- Dealing with missing data in R: https://www.statmethods.net/input/missingdata.html
- Visualization in R: https://www.analyticsvidhya.com/blog/2015/07/guide-data-visualization-r/