Proposal: RZ_capstone

DATA 450 Capstone

Robert Zhong

February 8, 2024

1 Introduction

EMS is an essential part of the healthcare system. It plays a role in delivering direct medical care and transportation to patients requiring attention. This analysis explores the different factors that affect EMS response times and examine how they impact patient outcomes in relation to the event type and location to the efficiency of the overall system. There will also be a deeper analysis on the MPDS codes. Objectives:

- Assess the impact of EMS response times on patient outcomes.
- Examine the temporal changes in EMS response times over a specified period.
- Identify factors contributing to changes in response times.
- Compare response times and outcomes between different locations.
- Evaluate the effect of using lights and sirens on response times and safety.
- Analyze the influence of MPDS prioritization on response efficiency.

2 Dataset

- Call year: The year the call was received
- Community type: The community type
- Call location: The location (municipality) —partial — with at least 10,000 total calls per year
- MPDS Code: The intital/Final MPDS code used for the call (i.e. “Code 2” or “Code 3”)
- Description: Description on the event
- Qualification: Required qualification for service provider
- MPDS Determinant code: Initial/Final “MPDS Determinant” code assigned to the call
- Response mode: The code used to respond to a call
- Event Cancelled: Whether the event was cancelled

- Response time: The time in minutes from when the call was received to when the first ambulance was on scene
- To Hospital time: The time in minutes from when the call was received to when the first ambulance arrived at hospital

[This data was obtained from the Provincial Health Services Authority through a Freedom of Information request. Chad Skelton, a journalist for the Vancover Suns made it available.] [Provincial Health Services Authority. (2015). B.C. Ambulance Raw Data. Chad Skelton] # Data Acquisition and Processing

[In this section, if applicable, describe how you will obtain the data (if it's anything more complicated than a simple download). Discuss what data processing steps will be needed, such as recoding variables, data cleaning, data tidying, imputing missing values, etc. See sections 1c, 1d, 1e in the "Good Enough Practices" paper.] * First off, I will have to make sure the datatypes a correct. * Then I might fill missing vlaues, but I will most likely get rid of it when Im doing modeling. * Save raw data for recovery and experimentation # Research Questions and Methodology

[In this section, list each of the questions you will explore. Following each question, provide a detailed and specific plan for how you plan to answer the question. Include the specific steps you will take, what form the answer will take (a number? table? visualization? model? Give all the specifics), and estimate how many hours each question will take to complete.]

Questions:

1. How does the response time of ambulances vary across different types of communities? Filter data by community type and calculate response times. Then using descriptive statistics to summarize response times for each community type. Then make set of boxplots and a table summarizing key statistics mean, median, range for each community type. Perform ANOVA or Kruskal-Wallis test to determine if there are statistically significant differences in response times between community types

2. What is the impact of different MPDS codes on the response time of ambulances? Group data by MPDS codes, calculate average response times for each MPDS code, create a bar chart and a table listing average response times for each MPDS code. ccess using correlation and regression modeling.

3. How does the change in initial and final codes affect the response time and outcomes of emergency calls? Compare response times and outcomes before and after code changes to identify calls with changes in initial and final codes. Make paired scatter plots and a summary of statistical tests assessing the impact of code changes.

4. What is the correlation between the cancellation of events and the response time of ambulances? Perform correlation analysis between cancellation status and response times.Scatterplot with a trend line to show the relationship. Get correlation coefficient indicating the strength and direction of the relationship. Regression

5. How does the time from when the call was received to when the ambulance arrived at the hospital vary across different types of emergencies? Group data by eevent type, then calculate average times from call receipt to hospital arrival for each emergency type. Make bar chart and a table listing average times for each event type. Conduct different tests.
6. What factors significantly affect ambulance response times? Use multiple and multivariate regression analysis to identify significant predictors of response times with summary including coefficients, significance levels, and model fit statistics
7. How have response times changed over the years and what factors might be contributing to these changes? Organize data by year, then perform a time series analysis to identify trends and fluctuations in response times over the years. I will make a line graph and a summary of the time series analysis. Regression
8. How does the use of lights and sirens affect the response time? Compare response times with and without the use of lights and sirens.Use histograms and a summary of statistical tests to compare the two cases. t-tests or non-parametric tests to assess the significance of the differences.
9. Does the Response code and MPDS determination code (and maybe other factors) affect the probability that a call could be cancelled? Use logistic regression to assess the impact of codes on the likelihood of cancellation.
10. Was there a large difference between the intitial and final values? Calculate the difference between initial and final codes and response times. Use descriptive statistics to summarize the differences, then analyze the significances. [Correlation (ex. Heatmap)]
11. What are the most frequent call types? Count the number of event types, MPDS codes, and MPDS determinant codes then rank by frequency

- [NOTE: More EDA and modeling might be used. This is just a basic outline]

3 Work plan

Week 4 (2/12 - 2/18):

- Set up project management tools
- Begin data cleaning, standardizing formats
- Dividing into subsets

Week 5 (2/19 - 2/25):

- Conduct descriptive statistics to understand the distribution of key variables
- Start group by analysis

Week 6 (2/26 - 3/3):

- Correlation analysis to identify potential relationships between variables.
- Time series analysis to examine trends over the years.

Week 7 (3/4 - 3/10):

- Create graphs, charts, and plots
- Presentation prep and practice (4 hours)

Week 8 (3/11 - 3/17):

- Geospacial maps from events that happened in the vicinity from other databases.
- Poster prep (4 hours)
- Presentation peer review (1.5 hours)

Week 9 (3/25 - 3/31): *Final Poster due Sunday 3/31.*

- Peer feedback (3.5 hours)
- Poster revisions (3.5 hours)

Week 10 (4/1 - 4/7):

- Select and train regression and classification algorithms

Week 11 (4/8 - 4/14):

- Apply clustering to identify patterns in call types or community responses.
- Time series forecasting to predict future trends in response times.

Week 12 (4/15 - 4/21):

- Compile and interpret findings

Week 13 (4/22 - 4/28):

- Draft blog post (4 hours).

Week 14 (4/29 - 5/5):

- Peer feedback (3 hours)
- Blog post revisions (4 hours)

Week 15 (5/6 - 5/12):

- Blog post revisions (2 hours)
- Peer feedback (2 hours)