Boston Crime Data Analysis (2017-2022)

Individual Project COMP-3125

Isabella Crowe

***Abstract*—This is the last part to be filled out**

***Index Terms*—keywords, temperature, xxxx equation, etc.**

I. Introduction

AWhile this essay focuses on crime, it is part of a larger, more complex issue of the student housing crisis that occurs year after year in Boston and Cambridge, Massachusetts. A significant challenge for college students searching for housing is where to begin. The first issue is that many colleges in Boston do not have a dedicated housing assistance center for students who wish to move off campus. The second most important factor that students take into consideration when finding housing is safety. Students have encountered issues with scams, inconsistent listing quality, misleading apartment descriptions, and unverified information about landlords. Students are willing to navigate these challenges of false listings and inaccurate information at the cost of their safety. Finding apartments in metropolitan areas is challenging, and Boston is one of the most expensive cities to reside in. For the general public, 2,400 seems high but not outrageous. But how does this affect college students? To comfortably adhere to the 30 rule, a student must earn 8,000 monthly to support themselves in the metropolitan area. This arrangement is often unmanageable for most students, particularly those balancing a full course load and extracurricular activities. Students in the metropolitan area spend over 50 of their total income on rent, which is considered a severe rent burden. This causes extreme unnecessary stress on top of the everyday pressure that college students endure. Not only does this create an extreme rent burden, but it also causes students to look for other housing options. Students will justify trade-offs, such as safety when seeking affordable housing, more than the general public might believe. This issue becomes significant when many Boston students are unaware of where to find the best and safest housing options while searching for an apartment. In the past 25 years, Boston has been one of the safest megalopolises in the United States. In 2023, the nationwide rate of property crimes (which includes but is not limited to shoplifting, motor vehicle theft, embezzlement, burglary, and forgery) in the United States was 1,917 cases per 100,0000, equating to nearly 1.9. However, in Boston, property crime was below the average, coming in at 1,820 cases per 100,000 or 1.8 of the population per year, lower than the national average. Boston does have a higher-than-average violent crime rate. Violent crime encompasses offenses involving force or the threat of force, including murder, rape, robbery, and aggravated assault. The FBI recorded 364 violent crimes per 100,000 in 2023. Boston has nearly double the amount of violent crime rate than the nationwide average, coming in at 630 cases per 100,000. But these numbers mean very little when considering Boston as a whole. To accurately analyze the crime rate in Boston, this paper will look into 13 different neighborhoods. This will include Allston/Brighton, Charlestown, Dorchester, Downtown, East Boston, Fenway/Lower Roxbury, Jamaica Plain, Mattapan, Roxbury, and West Robury/Roslindale.

II. Datasets

A. Source of Dataset

This article will reference a data set from Kaggle about Boston Crime from 2017 to 2022. Kaggle is a platform for data scientists and machine learning practitioners created and monitored by Google LLC. Kaggle is a platform where individuals can publish datasets, build models collaboratively, and participate in competitions to solve data science challenges. Kaggle is widely regarded as a source of reliable datasets, especially in professional settings. Over the years, there has been suspicion that the datasets are too clean, lacking inconsistencies, duplicates, or missing values. These issues are common in datasets, but Kaggle's do not often exhibit these errors.

B. Character of the Datasets

The dataset only contains crime-related records from Boston with specific attributes such as offense codes, locations of the crime, and dates. The CSV file on Boston Crime from 2017 - 2022 from Kaggle is divided into four categories: Column Name description, unit, and data type. The Column Name is the overarching nature of the crime (Ex. OFFENSE-DESCRIPTION could be burglary - commercial, towed motor vehicle, death investigation). The description is the specific events that happened during the arrest. The Unit is defined as the measurement unit for each column if applicable. The measurements in this CSV file are either N/A, numeric, categorical, Date/Time, or geographic. The data type is the data stored in the columns, whether it is an integer, float, string, or datetime. Each row represents an individual crime record identified by a unique Incident Number. The dataset includes both timestamps (OCCURRED-ON-DATE) and separate columns for YEAR, MONTH, DAY-OF-WEEK, and HOUR, facilitating time-based analysis. Other aspects of this data set include geographic location as a tuple of latitude and longitude. There were a few missing values; some of the rows had 0 values for the latitude and longitude of missing offense codes. These incidents may not be tied to a specific location or may involve improperly recorded data. This data set does not show evidence of merging multiple data sets. Each row corresponds to a separate incident record contained within the CSV file. The Kaggle CSV file on Boston Crime from 2017 -

2022 has 31 rows representing individual incident reports and nearly 550 thousand columns.

III. Materials and Methods

The methods utilized throughout this project focus on analyzing and visualizing crime data in Boston, Massachusetts. The main goal of this analysis is to identify the most significant types of offenses and the most frequent areas where crimes occur, illustrated by pie charts, bar graphs, and maps. The process involves reading data from the CSV file, filtering specific districts, cleaning the data, sorting the data, and sampling to create visual presentations of the plots and map of Boston's crime. Data cleaning is one of the most essential parts of this project and ensuring that data are accurate, reliable, and useful to the analysis. We can assume that the data contains records specific to East Boston, Dorchester, Roxbury, and Charlestown, and filter the data according to to the district. We can assume that the data do not contain irrelevant information, as they have been cleaned. To manage the size of the CSV file, the third row is sampled; this approach ensures that viewers can meaningfully visualize the diagrams of the data.

A. HAVE NOT COMPLETED ANYTHING FROM HERE ON

IV. Results

Show plots of any data collected and describe with words what your plots are showing. Describe the relationship between variables and time. Remember to number all your figures. This is the most critical part affect the technical achievement.

No picture, table, schematic, or graph should appear without a name (generally of the form Fig.1 o Table 1). None should appear without a reference to them by name in the main body of the writing. All figures and tables must be discussed in the text, including what it is, significant observations, and analysis.

Capitalize "Table" and "Fig." any time they are accompanied by specific table or figure numbers. Examples: "The measured data are plotted in Fig. 2. The figure shows a linear relationship in....". "The table shows ..." vs. "The data of Table 3..."

Student	Max Temperature
aabbbccc	35°
eeeddd	54°
eeeddd	54°

TABLE I
TEMPERATURE MEASUREMENTS PERFORMED FOR SESSION 1.

Use your word processor to make "real tables" (i.e., boxed in, etc.). Center all tables and include a heading and caption with the appropriate table number below each table. For example, "Table 1: Temperature measurements performed for session 1."

Figures must be centered, and the figure number and caption is centered beneath the figure. For example, "Figure 1".

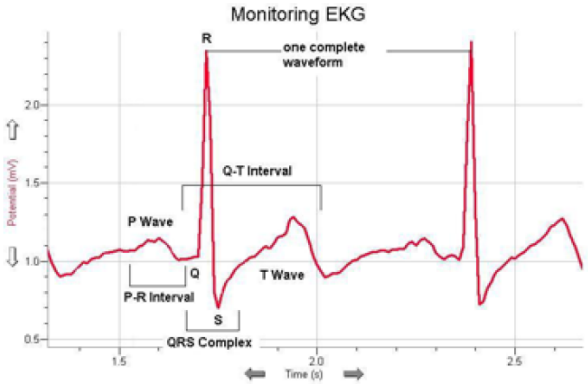

Fig. 1. Illustrations, graphs, and photographs may fit across both columns, if necessary. Your artwork must be in place in the article.

Always spell out table or Table. Give abbreviation of Figure, i.e., Fig., when used in the middle to end of sentence, but spell it out when used at the very start of the sentence.

All graphs must be done with a computer (i.e., spreadsheet software such as Microsoft Excel or even Matlab.). Do not include hand drawn graphs unless specifically instructed to do so.

Include a leading zero when a number's magnitude is less than 1 (use 0.83 instead of writing .83).

Use your word processor for Greek symbols for common engineering quantities as β, π, γ ,Ω .

V. Discussion and Summary

Discuss any interesting result related to the materials used or to any claim from the introduction. Discuss your measurements using engineering terms (accuracy, precision, resolution, etc). Give technical conclusions. Restate the main objectives and how or to what degree they were achieved. What principles, laws and/or theory were validated by the experiment? Describe some applications of your results and comment any possible recommended future work.

Appendix A
Hand calculations (or name your title for appendix subtitle)

List any extra evidence such as photos of the session, that may help you support your claims. You can include all hand calculations, extra graphs and plots, simulation results, etc.

Acknowledgment

The authors would like to thank...

Examples of references:

Example of data book:

[2] National Operational Amplifiers Databook. Santa Clara: National Semiconductor Corporation, 1995 Edition, p. I-54.

Example of textbook:

[3]M. Young, The Technical Writer's Handbook. Mill Valley,

CA: University Science, 1989.

Example of scientific journal paper:

[4] J.W. Smith, L.S. Alans and D.K. Jones, “An operational amplifier approach to active cable modeling”, IEEE Transactions on Modeling, vol. 4, no. 2, 1996, pp. 128-132.

Example of conference paper proceedings:

[5] J.W. Smith, L.S. Alans and D.K. Jones, “Active cable models for lossy transmission line circuits”, in Proc. 1995 IEEE Modeling Symposium, 1996, pp. 1086-89.

Example of Internet web page:

[6] Approximate material properties in isotropic materials. Milpitas, CA: Specialty Engineering Associates, Inc. web site: www.ultrasonic.com, downloaded Aug. 20, 2001.

List and number all bibliographical references at the end of your paper in **9 or 10 point** Times, with 10-point interline spacing. When referenced within the text, enclose the citation number in square brackets, for example [1].
Use IEEE format. Cite any external work that you used (data sheets, text books, Wikipedia articles, . . .). If you get a formula from a Wikipedia article, you must cite the article, giving the title, the URL, and the data you accessed the article as a minimum. If you copy a figure, not only must you cite the article you copied from, but you must give explicit figure credit in the caption for the figure: This image copied from . . . . If you modify a figure or base your figure on one that has been published elsewhere, you still need to give credit in the caption: This image adapted from . . . .