Everest Kehmeier

Professor Blanche Cohen

Computer Science Principles-CS-1030-001

1 May 2023

Textual Information on Analyzed Dataset

To begin gathering an understanding for this final project, I went through the recommended Khan Academy Unit on “Data Analysis.” Within this unit, there where three different lessons to get the student familiar with the ideas of how data is stored, big data sources, and the bias in machine learning. The more important lessons for this project where: “Data Tools,” and “Big Data”. With data tools I learn how to look at a dataset and analyze smaller portions for patterns or an explanation to a particular question. In big data, I figured out how to look at a dataset and find the size of the file, so I can determine if it’s manageable. All my project information and results are uploaded to my GitHub account, which will be provided on the submission on Canvas.

After this step was locked down, I began my search for a dataset that seemed interesting and that I could actually analyze. Instructor Cohen provided us with four different pre-approved data sources where we could collect a set. The source that I chose was: Data.gov, which is open to the public and provides gathered data from agencies across the federal government. When the final project was first introduced, I knew that I wanted to look at crime data. It didn’t take me long to find a dataset that I could connect with and find value in analyzing. This connection came in the form of a trip I took to Honolulu HI in the summer of 2022. It just so happened that one of the first organization contributing to Data.gov was the City of Honolulu and the crime incidents documented by the Honolulu Police Department. This data was simple to view, I was able to download it in comma-separated value (CSV) format and open it in Microsoft Excel. It’s a fairly large data set with a file size of 2.12 MB and contains 28,161 total reported crimes from the year 2015-2016. The reported crimes consist of burglary, motor vehicle theft, theft/larceny, vandalism, and vehicle break-in/theft. Unfortunately, it’s actually quit a disorganized dataset. This is where it was then time for me to analyze the data in a way that would produce useable results.

The main problem that I wanted to address was: what type of reported and documented crime should a tourist to Honolulu be most aware of? I started by getting a count of how many times each crime type took place during the reporting year. A lot of crime took place that year and trying to get a count of which ones were committed, by just looking at the dataset would be difficult. To help with this, I used the *COUNTIF* function in Excel. I then put that count type into a “PivotTable/Chat,” pie chart, to determine the percentage that each crime was happening at to compare them. The *SUM* function to get a count for the total amount of all crime was also initiated. The result of this found that theft/larceny took place 51% of the time, totaling 14,261 incidents in a years’ time. I then took the totals for each crime type and put them into a bar graph, so a major distinction could be made between the differences in occurrence. With this analysis I wanted to understand why theft/larceny took place 13,867 more times than vandalism, which had the lowest commitment at 394 incident in one year. With this analysis I was able to see that there is a downward trend in the commitment of crimes with a more severe punishment.

In order to complete this project, I used and visited multiple different sources. The first was Khan Academy. On this source I followed along with the “Data Analysis” unit to get an understanding/training in what big data is and how to analyze it. My progress with this is all documented in a separate Microsoft Word document. The second source was Data.gov. Here I was able to search for and find a dataset that I found interesting and could work with. Like I mentioned above, this dataset was a report of multiple different crimes that took place in Honolulu HI during the time of 2015 to 2016. From this website, I was able to download a copy of the data in the format of comma-separated value (CSV) to my computer. The file is large and required some organization to really grasp what the crime types where and how many times they happened. Microsoft Excel was the third source. Here I was able to group the reported crime types into four distinct categories of: burglary, motor vehicle theft, theft/larceny, vandalism, and vehicle break-in/theft. I then was able to get a count of how many times each crime occurred during the year of reporting. This helped me get a percentage comparison and create a bar graph to find a trend. The fourth source that helped me, was Spaces.w3schools.com. On this website I was able to create an HTML file that documented textual information in the HTML format. This is where I answered the question and explained the hypothesis that I produced based on the crime incidents dataset. The fifth source was shouselaw.com. I used this website to find the punishment ranges for vandalism. The sixth and final source was capitol.hawaii.gov. This is where I found the punishment for theft/larceny. I used both websites to help explain my hypothesis.

In the end, this final project was a challenge that required me to think on a more meticulous level. I needed this so that I could analyze a dataset and produce results to a problem, that the numbers could answer. I appreciated that the use of some HTML format was a must, since I got to learn something new and see the basics of writing a few paragraphs in this format. Now I have a better understanding of where to find datasets and how to analyze them when a particular question becomes apparent.

 Associated Sources

Khan Academy. (2023). *Storing data sets | AP CSP (article)*. Khan Academy. Retrieved April 26, 2023, from https://www.khanacademy.org/computing/ap-computer-science-principles/data-analysis-101/data-tools/a/storing-data-sets

Thompson, M. (2020, November 10). *Crime incidents*. Catalog. Retrieved April 28, 2023, from https://catalog.data.gov/dataset/crime-incidents

Microsoft Excel

W3Schools. (1999). *HTML Tutorial*. W3Schools. Retrieved May 2, 2023, from https://www.w3schools.com/html/

T., J. (2022, November 9). *Vandalism laws in Colorado - definition, penalties, legal defenses*. Shouse Law Group. Retrieved May 2, 2023, from https://www.shouselaw.com/co/defense/laws/vandalism/

capitol.hawaii.gov. (n.d.). *Theft in the Fourth Degree*. House Bill. Retrieved May 2, 2023, from https://www.capitol.hawaii.gov/hrscurrent/vol14\_ch0701-0853/HRS0708/HRS\_0708-0833.htm#:~:text=%C2%A7708%2D833%20Theft%20in,not%20in%20excess%20of%20%24250.&text=Act%20231%2C%20Session%20Laws%202016,monetary%20thresholds%20for%20the%20offenses.