

Performance Assessment: Fundamentals of Data Analytics

Emergency 911 Call Log Analysis

Savahnna L. Cunningham

Western Governors University

Part 1a: Data preparation

A. Address Quality & Tidiness Issues

Remove unnecessary columns:

- CAD Event Number – unnecessary information for analysis; information can be referenced using the primary key.
- General Offense Number – unnecessary information for analysis; information can be referenced using the primary key.
- Event Clearance Code – unnecessary information for analysis
- Event Clearance Description – duplicate information
- Event Clearance SubGroup – duplicate information
- Hundred Block Location – poor quality data; abbreviations, XX representing numbers, backslash representing a new address line, etc... Data is unnecessary for analysis. If location is needed for analysis, the Latitude and Longitude will be used for geographical reference.
- Zone/Beat – unnecessary information for analysis
- Census Tract – unnecessary information for analysis
- Latitude – unnecessary information for analysis
- Longitude – unnecessary information for analysis
- Incident Location – duplicate information, unnecessary for analysis
- Initial Type Description – unnecessary information for analysis
- Initial Type Subgroup – unnecessary information for analysis.
- Initial Type Group – unnecessary information for analysis.
- At Scene Time – too much missing data; unnecessary for analysis

Timestamp – separate date and time information, already a “DateTime Object” means no need to convert for easier data analysis

Rename columns

- CAD CDW ID → CAD_ID
- District/Sector → District
- OFFICERS_AT_SCENE → Total Officers

Part 1b: Data Analysis: see Excel spreadsheet for information

Part 2a: Monte Carlo Simulation: see Excel spreadsheet for information

Part 2b: Recommendation

F. Determine if the police department currently qualifies for the funding.

The analysis indicates there was an average of 55.1 incidents during the 3-day sample period with an average total of 1.25 officers per incident. Therefore, the department does not currently meet the minimum standard to receive the governor’s funding incentive.

G. Calculate the probability that the department will or will not qualify for the funding in the future.

Avg Officers @ Scene	4.836012014
Incidents with ≥ 2.5 Officers	321
Probability of ≥ 2.5 Officers	32.1%

The table above shows the results of the Monte Carlo simulation. As you can see, the simulation shows the department has a 32.1% probability that an incident will have more than 2.5 officers present.

H. Describe the precautions or behaviors that should be exercised when working with and communicating about the sensitive data in this scenario.

When a Data Analyst is hired to conduct an analysis the person who is hiring the Analyst instinctive enters into a trusting relationship, assuming the Data Analyst will keep their information private and confidential. The University of California-Irvine Office of Research defines privacy as the control over the extent, timing, and circumstances of sharing oneself (physically, behaviorally, or intellectually) with others. While confidentiality pertains to the treatment of information that an individual has disclosed in a relationship of trust and with the expectation that it will not be divulged to others without permission in ways that are inconsistent with the understanding of the original disclosure (Privacy and Confidentiality, 2015). Thus, protocols should be designed and adhered to that provide as much protection to people as possible including keeping the collection of identifiable data to a minimum. Data should always be collected anonymously or identifiers should be removed and replaced with a numbered primary key.

To make the Seattle PD 911 Call Log database more secure the would take the following steps:

1. Make a separate spreadsheet containing the unique identifiers CD CDW ID, CAD Event Number and General Offense Number. Reassign each case an integer from 1 to 1045. Use this new number as the primary key in the data analysis spreadsheet. This separates unique identifiable information from your analysis dataset and if one spreadsheet is breached, the other will still be protected.
2. Remove all location data from the spreadsheet used for the analysis. Although the API has altered some information, it is not private or confidential. If the police

department needs location information for a specific case, they can access that information through their internal database using the CAD ID information.

References

In-text: (Fairclough, 2016)

Fairclough, B. (2016). Privacy Piracy: The Shortcomings of the United States' Data Privacy Regime and How to Fix. *Journal Of Corporation Law*, 42(2), 461-480.

In-text: (Pardoe et al., 2016)

Pardoe, I., Sturdivant, R., Berrier, J., Nestler, S., Watts, K., Vahid, F., & Chan, C. (2016). zyBooks. My.zybooks.com. Retrieved 23 April 2018, from <https://my.zybooks.com/#/zybook/WGUFundamentalsOfDataAnalytics/tableofcontents>

In-text: (Privacy and Confidentiality, 2015)

University of California Irvine, Office of Research. (2015) *Privacy and Confidentiality*. Retrieved 23 April 2018, from <http://www.research.uci.edu/compliance/human-research-protections/researchers/privacy-and-confidentiality.html>

Silverstein, S. (2015, January 27). *Monte Carlo Simulations: Run 10,000 Simulations at Once* [Video file]. Retrieved from <https://www.youtube.com/watch?v=wKdmEXCvo9s>

NIST/SEMATECH e-Handbook of Statistical Methods. (2013). *What are outliers in the Data?*. Retrieved 23 April 2018 from <https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm>.