

Project 3: Association Rule Mining

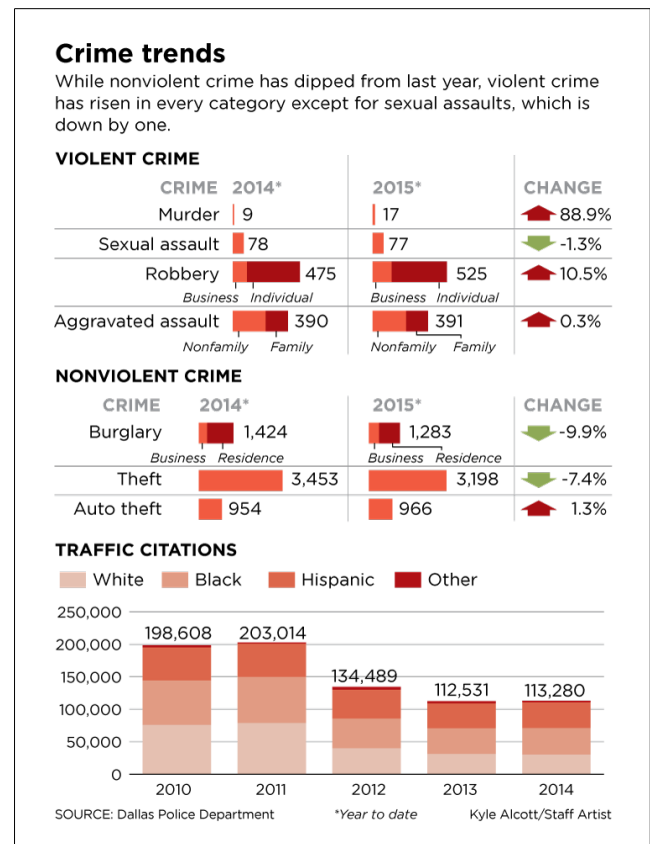
Assigned: 3/17/2016
 Due: 4/6/2016
 (via Canvas)
 Points: 100

Please submit your report in **PDF format**.

Introduction

You will work again with crime report data for Dallas published by Dallas Open Data. You can use the cleaned data from Projects 1 and 2, and any additional data you find.

Write a report covering in detail all steps of the project. The results have to be **reproducible** using your report. Carefully describe every **assumption** and every step in your report. Also mention (not include) any program/code other than R that you are using for your analysis.



Follow the CRISP-DM framework (Steps 3-6)

3. Data Preparation [30 points]

- Select and describe the features you will use. Explain how you preprocess the data. [20 points]
 For example:
 - Create new binary features to indicate the presence/absence of a certain fact.
 - Discretize continuous features. What method do you use? Explain why? Do you use different discretization methods for different features?
- Construct at least two transaction data sets using different subsets of a single data set. You may choose how to select the subsets (e.g., different location, different time, different complainant demographics). Show and discuss some simple statistics about the transaction sets. [10 points]

4. Modeling [40 points]

- Create sets of frequent, closed and maximal itemsets. [5 points]
- Create sets of association rules. [5 points]
- Use filtering, sorting, tables, and visualization to discuss the found patterns. What do the patterns mean and how are they useful? [20 points]
- Compare the itemsets/rules found in the different transaction data sets. Discuss what the differences mean. [10 points]

5. Evaluation [20 points]

- What findings are the most interesting? Do they support what you have found in the other two projects? [10 points]
- What recommendations do you have for different stakeholders based on your findings? [5 points]
- How useful are your results for different stakeholders? [5 points]

6. Deployment [0 points]

- Write a report covering in detail all steps of the project. The results have to be reproducible using this documentation.

Exceptional Work [10 points]

Examples include:

- Using the arrest data (found here <https://www.dallasopendata.com/Police/Dallas-Police-Public-Data-RMS-Arrest/r4wm-ig9m>).
- Comparing more subsets of the data to reveal interesting relationships.
- Exceptional postprocessing and representation of the found rules.
- Use different quality measures for rules, explain what they do and why they produce meaningful insights.