DSC 550 Final Project Proposal

Name: Katie Briggs   
Title: Mining for Recalls

**Proposal**

Recalls on products happen when a large quantity of items are no longer safe for use. Items that are frequently recalled are vehicles, medications, foods, and products (i.e. small choking hazards for infants). There are a various amount of information on recalls located at data.gov and <https://www.recalls.gov/> that I would like to mine and find the most frequent recalls on food. I would also like to know which class of medications are the most recalled. Along with frequent itemset mining, I would also conduct rare itemset mining. Hopefully, I can find patterns that are rarely seen.

**Research**

The data that I would like to research involves:

1) What food items are most frequently recalled and what cause?

2) What part of a vehicle is most frequently recalled?

3) What class of prescription medication is most frequently recalled?

4) What consumer product is rarely recalled? Which category is most recalled?

**Approach**

For this project I may use all data techniques that include, shingling, hashing, locating candidate pairs and placing those candidate pairs into bands. I will band similar items together. Hopefully, I will learn the best approach on finding rare itemsets.

Reference/Resource:

<https://www.recalls.gov/cpsc.html>

<https://catalog.data.gov/dataset?q=recalls&sort=score+desc%2C+name+asc&as_sfid=AAAAAAUUUKH5QIclpY9sqIShUbfkvdUO-1lRZJUWu6S6UjWX39lacUd9hKxSbrgyKtiujDdiCOE88cYUBJphxmKgIPyIhQMENRPwMPVWjlyHLLimxdmw0GzQ03nBUVsp7-u7p9M%3D&as_fid=ad80cc5649bb6a7b240fef3e9c0e7f1c45717c70>