**Team Name:**

Shopaholics Anonymous (SA)

**Team Members**

Saleha Siddique

**Project Proposal**

**Text Analytics & Visualization of Amazon Customer Reviews**

Customer reviews play an important role in influencing purchase decisions in e-commerce. Many products have 100s of reviews though, and going through all of them takes too much time. I would like to analyse the reviews of a product to count the most frequently occurring words, and display them in a chart (similar to how we did it for the course scraping exercise).

Another chart could display the most common words in positive reviews, and the most common words in negative reviews.

Nice to have functionality could include analysing the sentiment of particular phrases and visualizing that.

**Modules/Technologies that I plan to use:**

I have broken these down by the problems that need to be solved.

1. *How to get the customer reviews* – ideally I want this to be dynamic – i.e you enter a a product key and it pulls out the reviews for you.

**Ideal solution:** Get them from Amazon’s Product Advertising API. Github has a Node.js solution to access the Product Advertising API which can pull reviews:

<https://github.com/livelycode/aws-lib>

**Secondary solution:** Use curl to scrape them

**If everything else fails**: Use a publically available dataset e.g. <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>

1. *How to analyse the reviews:*

Node.js + I’d need to shortlist from NLP libraries available on NPM:

<https://www.npmjs.com/browse/keyword/nlp>

Other resources could be e.g. Watson API <https://www.ibm.com/watson/developercloud/alchemy-language.html>

1. *How to visualize the analysis:*

* d3