Capstone 3 Proposal by Zohreh Soltani

How people feel about a product in Musical Instrument Department on Amazon?

To answer this question the sentiment analysis using NLP will be utilized. The below link is the address to this data set which includes product review and metadata from Amazon and has 142.8 million reviews from May 1996-July 2014. For this project the 5-core data for the Musical Instruments will be analyzed. 5 core data is the subset of data in which all users and items have at least 5 reviews.

5 core dataset of the musical instrument includes 10,261 reviews.

Link to the dataset:

http://jmcauley.ucsd.edu/data/amazon/

below is a sample review:

{"reviewerID": "A2SUAM1J3GNN3B", "asin": "0000013714", "reviewerName": "J. McDonald", "helpful": [2, 3], "reviewText": "I bought this for my husband who plays the piano. He is having a wonderful time playing these old hymns. The music is at times hard to read because we think the book was published for singing from more than playing from. Great purchase though!", "overall": 5.0, "summary": "Heavenly Highway Hymns", "unixReviewTime": 1252800000, "reviewTime": "09 13, 2009"}

Solving this problem will help the vendors to quickly understand the customers' opinions about the product and to boost the sales. It also will help the Amazon's executives to informedly decide about the renewal of the contracts with specific vendors, prices and continuation of product supply. This analysis also will save customer's time and provide them with a summarized feedback about the product written by the other users.

In this exercise the reviews will be categorized into multiple groups including Positive, Negative, Neutral or angry, sad, happy, etc using Automatic approach of Natural Language Processing.

The deliverables of this capstone would be a report, a Jupyter Notebook file and a slide deck.

Also, as an extension to this capstone I will work on the visualizing the results to gain hands on experience in visualizing as well as analyzing using NLP.