**PROJECT PROPOSAL**

**DATA SOURCE AND DESCRIPTION**

The datasets for this project were obtained from <https://www.kaggle.com/cynthiarempel/amazon-us-customer-reviews-dataset> which are Amazon Customer Reviews from 1995 to 2015 with several datasets per category of products and they are listed below:

1. <https://www.kaggle.com/cynthiarempel/amazon-us-customer-reviews-dataset?select=amazon_reviews_us_Baby_v1_00.tsv>
2. <https://www.kaggle.com/cynthiarempel/amazon-us-customer-reviews-dataset?select=amazon_reviews_us_Electronics_v1_00.tsv>
3. <https://www.kaggle.com/cynthiarempel/amazon-us-customer-reviews-dataset?select=amazon_reviews_us_Office_Products_v1_00.tsv>
4. <https://www.kaggle.com/cynthiarempel/amazon-us-customer-reviews-dataset?select=amazon_reviews_us_Toys_v1_00.tsv>

I would be performing an exploratory data analysis (EDA) on the above datasets while reporting any findings, the datastore used is the MongoDB and Sentiment analysis using logistic regression is the supervised learning model adopted for this project.