**Project Proposal**

**Project Name:** Cheap & Best

**Group Members:** Utkarsh Srivastava,Ashwin Jayadev,Aakriti Arya, Divyam Malik

**Objective:** Return the product that are cheap and have great reviews, based on a query entered by user.

**Description:**

E-commerce websites have thousands of products in their database,whenever we enter a query i.e the name of the product.They return the product list and we manually chose the products, here the concept of specific advertising as well as popularity of the product works.But a user basically is willing to buy a product that doesn’t cost much as well is durable and worthy of the price and time he spends.

The quality of the product is sometimes expressed in terms of star rating ,and also the reviews given by several users who have bought the product .

**Our View Point:**

Data Extraction:

A very basic need was for data extraction,and hence we came up with a large bunch of data that we have include in a single file.

We have extracted a dataset that has the details of the product in the following format.

(Name of the product,Tags of the product,Price,Reviews)

And we will work on this dataset.And try finding the result based on the query by reviewing this dataset.

**How we are willing to proceed??**

We divided our methodology into three phases:

* Phase 1: Query Segmentation
* Phase 2: Product Search
* Phase 3: Sentiment Analysis of the review
* Phase 4: Implement an algorithm to sort the products
* Phase 5:return the list as per the methodology

Phase 1: The entered Query will be subdivided into keywords, these keywords will lead us to the required product

Phase 2:The segmented query will help us give the list of the most probable products that we will work upon from the dataset.

Phase 3: For all the products that we received ,we perform a sentiment analysis in order to get insight of the review of the product.

Phase 4:A proper algorithm will be generated which will sort the products

**Limitation:**

* Phase 1:The query can only be entered in English
* Phase 2:Products are only limited to the ones present in the dataset
* Phase 3:Won’t be able to predict sarcasm and null reviews
* Phase 4:The algorithm will generate a particular id : that will be used for further Comparison