2020 MCM Weekend 2
Problem C: A Wealth of Data

In the online marketplace it created, Amazon provides customers with an opportunity to rate and review purchases. Individual ratings - called "**star ratings**" – allow purchasers to express their level of satisfaction with a product using a scale of 1 (low rated, low satisfaction) to 5 (highly rated, high satisfaction). Additionally, customers can submit text-based messages – called "**reviews**" – that express further opinions and information about the product. Other customers can submit ratings on these reviews as being helpful or not – called a "**helpfulness rating**" – towards assisting their own product purchasing decision. Companies use these data to gain insights into the markets in which they participate, the timing of that participation, and the potential success of product design feature choices.

Sunshine Company is planning to introduce and sell three new products in the online marketplace: a microwave oven, a baby **pacifier**, and a hair dryer. They have hired your team as consultants to identify key patterns, relationships, measures, and parameters in past customer-supplied ratings and reviews associated with other competing products to 1) inform their online sales strategy and 2) identify potentially important design features that would enhance product desirability. Sunshine Company has used data to inform sales strategies in the past, but they have not previously used this particular combination and type of data. Of particular interest to Sunshine Company are time-based patterns in these data, and whether they interact in ways that will help the company craft successful products.

To assist you, Sunshine's data center has provided you with three data files for this project: **hair_dryer.tsv**, **microwave.tsv**, and **pacifier.tsv**. These data represent customer-supplied ratings and reviews for microwave ovens, baby pacifiers, and hair dryers sold in the Amazon marketplace over the time period(s) indicated in the data. A glossary of data label definitions is provided as well. THE DATA FILES PROVIDED CONTAIN THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM.

Requirements

1. Analyze the three product data sets provided to identify, describe, and support with mathematical evidence, meaningful quantitative and/or qualitative patterns, relationships, measures, and parameters within and between star ratings, reviews, and helpfulness ratings that will help Sunshine Company succeed in their three new online marketplace product offerings.

2. Use your analysis to address the following specific questions and requests from the Sunshine Company Marketing Director:
   a. Identify data measures based on ratings and reviews that are most informative for Sunshine Company to track, once their three products are placed on sale in the online marketplace.
   b. Identify and discuss time-based measures and patterns within each data set that might suggest that a product's reputation is increasing or decreasing in the online marketplace.
   c. Determine combinations of text-based measure(s) and ratings-based measures that best indicate a potentially successful or failing product.

d. Do specific star ratings incite more reviews? For example, are customers more likely to write some type of review after seeing a series of low star ratings?

e. Are specific quality descriptors of text-based reviews such as 'enthusiastic', 'disappointed', and others, strongly associated with rating levels?

3. Write a one- to two-page letter to the Marketing Director of Sunshine Company summarizing your team's analysis and results. Include specific justification(s) for the result that your team most confidently recommends to the Marketing Director.

Your submission should consist of:
- One-page Summary Sheet
- Table of Contents
- One- to Two-page Letter
- Your solution of no more than 20 pages, for a maximum of 24 pages with your summary sheet, table of contents, and two-page letter.

Note: Reference List and any appendices do not count toward the page limit and should appear after your completed solution. You should not make use of unauthorized images and materials whose use is restricted by copyright laws. Ensure you cite the sources for your ideas and the materials used in your report.

Glossary

Helpfulness Rating: an indication of how valuable a particular product review is when making a decision whether or not to purchase that product.

Pacifier: a rubber or plastic soothing device, often nipple shaped, given to a baby to suck or bite on.

Review: a written evaluation of a product.

Star Rating: a score given in a system that allows people to rate a product with a number of stars.

Attachments: The Problem Datasets

Problem_C_Data.zip
The three data sets provided contain product user ratings and reviews extracted from the Amazon Customer Reviews Dataset thru Amazon Simple Storage Service (Amazon S3).
    hair_dryer.tsv
    microwave.tsv
    pacifier.tsv

Data Set Definitions: Each row represents data partitioned into the following columns.

- marketplace (string): 2 letter country code of the marketplace where the review was written.

- customer_id (string): Random identifier that can be used to aggregate reviews written by a single author.

- review_id (string): The unique ID of the review.

- product_id (string): The unique Product ID the review pertains to.

- product_parent (string): Random identifier that can be used to aggregate reviews for the same product.

- product_title (string): Title of the product.

- product_category (string): The major consumer category for the product.

- star_rating (int): The 1-5 star rating of the review.

- helpful_votes (int): Number of helpful votes.

- total_votes (int): Number of total votes the review received.

- vine (string): Customers are invited to become Amazon Vine Voices based on the trust that they have earned in the Amazon community for writing accurate and insightful reviews. Amazon provides Amazon Vine members with free copies of products that have been submitted to the program by vendors. Amazon doesn't influence the opinions of Amazon Vine members, nor do they modify or edit reviews.

- verified_purchase (string): A "Y" indicates Amazon verified that the person writing the review purchased the product at Amazon and didn't receive the product at a deep discount.

- review_headline (string): The title of the review.

- review_body (string): The review text.

- review_date (bigint): The date the review was written.