# 2020 年美国大学生数学建模竞赛（MCM/ICM）C 题 中英版



## 翻译：数据的财富

亚马逊在其创建的在线市场中，为顾客提供了对购买进行评分和评价的机会。个人评级（称为"**星级**"）使购买者可以使用 1（低评级，低满意度）到 5（高评级，高满意度）的等级来表示他们对产品的满意度。此外，客户可以提交文本消息（称为"**评论**"），以表达有关产品的更多意见和信息。其他客户可以在这些评论中提交有帮助或无帮助的评级（称为"**帮助评级**"），以协助他们自己的产品购买决策。公司使用这些数据来深入了解客户参与的市场，参与的时间以及产品设计功能选择的潜在成功。

阳光公司计划在网络上推出和销售三种新产品：微波炉、婴儿奶嘴和吹风机。他们已聘请您的团队作为顾问，识别与其他竞争产品相关的客户提供的过去评级和评论中的关键模式、关系、度量和参数，用于 1）告知其在线销售策略；2）识别潜在的重要设计功能，提高产品的吸引力。阳光公司过去曾使用数据为销售策略提供信息，但他们以前从未使用过这种特殊的组合和数据类型。阳光公司特别感兴趣的是这些数据中的基于时间的模式，以及顾客的互动方式是否有助于公司打造成功的产品。

为了帮助您，阳光公司的数据中心为您提供了该项目的三个数据文件：**hair_dryer.tsv**，**microwave.tsv** 和 **pacifier.tsv**。这些数据代表在数据指示的时间段内在亚马逊市场上出售的微波炉、婴儿奶嘴和吹风机的客户提供的评级和评论。 并且，这里还提供了数据标签定义的词汇表，提供的数据文件包含您应用于此问题的唯一数据。

## 要求

1. 分析提供的三个产品数据集，以使用数学证据来识别、描述和支持有意义的定量和/或定性模式、关系、量度和参数，这些数据在星级、评论和帮助评分之内和之间进行，这将有助于阳光公司在他们三个新的在线市场产品中都取得成功。

2. 使用您的分析来解决阳光公司市场总监的以下特定问题和要求：

   - 一旦阳光公司的三款产品在网络市场上销售，根据对其信息量最大的评级和评论确定数据衡量标准。

   - 在每个数据集中识别并讨论基于时间的度量和模式，这些度量和模式可能表明产品在网络市场中的声誉在上升或下降。

   - 确定最能表明潜在成功或失败产品的基于文本的度量和基于评级的度量的组合。

   - 特定星级会引起更多评论吗？例如，在看到一系列低星级评级之后，客户是否更有可能撰写某种类型的评论？

   - 诸如"热情"，"失望"之类的基于文本的评论的特定质量描述符是否与评级水平密切相关。

3. 写一两页的信给阳光公司市场总监，总结您团队的分析和结果。包括针对您的团队最有信心推荐给市场总监的结果的具体理由。

您的提交应包括：

- 一页摘要表

- 目录

- 一到两页的信函

- 您的解决方案不超过20页，包含摘要，目录和信函时最多24页。

**注意**：参考列表和任何附录均不计入页数限制，应在完成解决方案后显示。您不应使用受版权法限制使用的未经授权的图像和材料。要确保您引用的想法的来源和报告中使用的材料。

## 词汇表

- **帮助等级**：表示在决定是否购买该产品时特定产品评论的价值。
- **奶嘴**：一种橡胶或塑料的舒缓装置，通常为乳头状，提供给婴儿吸吮或咬。
- **评论**：对产品的书面评估。
- **星级**：在系统中给出的分数，该分数使人们可以对具有多个星级的产品进行评分。

## 数据

Problem_C_Data.zip

提供的三个数据集包含产品用户评分和通过亚马逊简单存储服务（Amazon S3）从亚马逊客户评论数据集提取的评论。

hair_dryer.tsv

microwave.tsv

pacifier.tsv

**数据集定义**：每行代表划分为以下几列的数据。

- marketplace（字符串）：撰写评论的市场的 2 个字母的国家/地区代码。

- customer_id（字符串）：随机标识符，可用于汇总单个作者撰写的评论。

- review_id（字符串）：评论的唯一 ID。

- product_id（字符串）：该评论所属的唯一产品 ID。

- product_parent（字符串）：随机标识符，可用于汇总同一产品的评论。

- product_title（字符串）：产品的标题。

- product_category（字符串）：产品的主要消费者类别。

- star_rating（int）：评论的 1-5 星评级。

- helpful_votes（int）：有用的投票数。

- total_votes（int）：评论收到的总票数。

- vine（字符串）：根据客户在亚马逊社区中撰写准确而有见地的评论所获得的信任，邀请他们成为 Amazon Vine Voices（亚马逊藤之声）。亚马逊为 Amazon Vine 成员提供了供应商已提交给该程序的产品的免费副本。亚马逊不会影响 Amazon Vine 成员的意见，也不会修改或编辑评论。

- verify_purchase（字符串）："Y" 表示亚马逊已验证撰写评论的人在亚马逊上购买了该产品，并且没有以大幅折扣收到该产品。

- review_headline（字符串）：评论的标题。

- review_body（字符串）：评论文本。

- review_date（bigint）：撰写评论的日期。

## 原文：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:

   - 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.

   - 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.

   - Determine combinations of text-based measure(s) and ratings-based measures that best indicate a potentially successful or failing product.

- 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?

- 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.

## 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.