

C 题翻译：数据财富

在亚马逊创建的在线市场中，亚马逊为客户提供了对购买进行评级和审查的机会。个人评级-称为“**星级评级**”-允许购买者使用从 1(低评级，低满意度)到 5(高评级，高满意度)的等级来表达他们对产品的满意程度。此外，客户还可以提交基于文本的消息-称为“**评论**”-表达对产品的进一步意见和信息。其他客户可以在这些评论上提交评级，认为它们对帮助他们自己的产品购买决策有帮助或没有帮助-被称为“**有帮助的评级**”。公司使用这些数据来洞察他们参与的市场，参与的时机，以及产品设计功能选择的潜在成功。

阳光公司正计划在网上市场推出并销售三种新产品：微波炉、婴儿奶嘴和吹风机。他们已聘请您的团队作为顾问，以确定过去客户提供的与其他竞争产品相关的评分和评论中的关键模式、关系、度量和参数，以 1)为其在线销售战略提供信息，2)确定潜在的重要设计功能，以增强产品的可取性。阳光公司过去曾使用数据为销售战略提供信息，但他们以前没有使用过这种特定的组合和类型的数据。阳光公司特别感兴趣的是这些数据中基于时间的模式，以及它们是否以有助于公司制作成功产品的方式进行交互。

为了帮助您，阳光数据中心为您提供了该项目的三个数据文件：Hair_dryer.tsv、microwave e.tsv 和 pacifier.tsv。这些数据代表客户对亚马逊市场上销售的微波炉、婴儿安抚奶嘴和吹风机在数据显示的时间段内提供的评级和评论。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该用于此问题的唯一数据。

要求

(1) 使用数学证据、有意义的定量和/或定性模式、星级评级、评论和帮助评级之间的关系、衡量标准和参数，分析提供的三个产品数据集来识别、描述和支持，这些数据集将帮助阳光公司在其三个新的在线市场产品中取得成功。建造在与同一海滩上的与水的距离大致相同的地方；

(2) 使用您的分析来解决阳光公司营销总监提出的以下具体问题和要求：

- 一旦阳光公司的三种产品在网上市场上销售，根据评级和评论确定最具信息量的数据衡量标准，以供阳光公司跟踪。
- 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在线市场上的声誉在上升或下降。

-
- 确定基于文本的衡量标准和基于评级的衡量标准的组合，以最好地指示潜在的成功或失败的产品。
 - 具体的星级评分是否会引发更多评论？例如，客户是否更有可能在看到一系列低星评级后撰写某种类型的评论？
 - 基于文本的评论的具体质量描述，如“热情”、“失望”等，是否与评级水平密切相关？

(3) 给阳光公司的市场总监写一封一到两页的信，总结你的团队的分析和结果。对于您的团队非常自信地向市场营销总监推荐的结果，请提供具体的理由。

你的意见应包括：

- 一页摘要
- 目录
- 1-2 页的信
- 您的解决方案不超过 20 页，最多 24 页包含您的摘要、目录和两页的信。

注意：参考文献和任何附录不计入页面限制，应在完成解决方案后显示。您不应使用未经授权的图片和材料，其使用受到版权法的限制。确保你引用了你的观点的来源和你报告中使用的材料。

专业词汇

有益性评级：在做出是否购买该产品的决定时，表明某一特定产品评论的价值有多高。

奶嘴：一种橡胶或塑料的抚慰装置，通常为乳头状，供婴儿吮吸或咬住。

评论：对产品的书面评价。

星级：在一个系统中给出的分数，该系统允许人们用多个星级对一种产品进行评级。

C 题数据解释

提供的三个数据集包含通过 Amazon Simple Storage Service(Amazon S3)从 Amazon Customer Reviews 数据集中提取的产品用户评分和评论。

数据集定义：每行表示分区到以下列中的数据。

Marketplace(String)：撰写评论的市场的 2 个字母的国家/地区代码。

Customer_id(String)：可用于聚合单个作者撰写的评论的随机标识符。

Review_id(String)：评论的唯一 ID。

product_id(String)：审核所属的唯一产品 ID。

PRODUCT_PARENT(String)：可用于聚合对同一产品的评论的随机标识符。

PRODUCT_TITLE(String)：产品的标题。

PRODUCT_CATEGORY(String)：产品的主要消费类别。

STAR_RATING(INT)：点评的 1-5 星级。

HELPUV_VOTES(INT)：有帮助的投票数。

TOTAL_VOTES(INT)：评论收到的总票数。

Vine(String)：邀请客户成为 Amazon Vine Voices 是基于他们在 Amazon 社区中因撰写准确且有洞察力的评论而赢得的信任。亚马逊为 Amazon Vine 会员提供供应商提交给该计划的产品的免费副本。亚马逊不会影响 Amazon Vine 会员的意见，他们也不会修改或编辑评论。

VERIFIED_PURCHASE(String)：“Y”表示亚马逊验证了撰写评论的人在亚马逊购买了该产品，并且没有以很大的折扣收到该产品。

REVIEW_HEADLINE(String)：评论的标题。

REVIEW_BODY(String)：审核文本。

REVIEW_DATE(Bigint)：编写评论的日期。