**问题C:大量的数据**

在其创建的在线市场中，亚马逊为客户提供了一个评价和审查购买的机会。

个人评级——称为“星级评级”——允许购买者使用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度。

客户可以提交基于文本的消息—称为“评论”—来表达关于产品的进一步意见和信息。

其他客户可以在这些评论中提交对他们自己的产品购买决策是否有帮助的评级——称为“帮助度评级”。

公司利用这些数据来洞察他们所参与的市场、参与的时机以及产品设计特性选择的潜在成功。

阳光公司计划在网上市场推出并销售三种新产品:微波炉、婴儿奶嘴和电吹风。他们聘请你的团队作为顾问，以确定过去客户提供的与其他竞争产品相关的评级和评论中的关键模式、关系、度量和参数，以1)告知他们的在线销售策略，2)确定可能提高产品吸引力的重要设计特性。阳光公司过去曾使用数据来告知销售策略，但他们以前从未使用过这种特定的组合和类型的数据。Sunshine公司特别感兴趣的是这些数据中的基于时间的模式，以及它们之间的交互方式是否有助于该公司打造成功的产品。

为了帮助您，Sunshine的数据中心为您提供了三个数据文件:hair\_dryer.**tsv,microwave.tsv, pacifier.tsv。**这些数据代表了在数据中显示的时间段内亚马逊市场上销售的微波炉、婴儿奶嘴和吹风机的客户提供的评级和评论。还提供了数据标签定义的术语表。提供的数据文件包含您应该用于此问题的唯一数据。

**需求**

1. 分析所提供的三个产品数据集，利用数学依据/有意义的定量、定性模式/关系/措施/星级评级、评论和有意义的帮助度评级中的参数来识别，描述和支持，帮助阳光公司在他们的三个新的在线市场产品提供成功。

2.使用您的分析来解决以下具体问题和阳光公司营销总监的要求:

a.一旦阳光公司的三种产品在网上市场销售，根据评级和评论确定最能提供信息的数据度量，以便阳光公司跟踪。

b.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中的声誉正在增加或减少。

c.确定基于文本的测量方法和基于评级的测量方法的组合，以最好地指示潜在的成功或失败的产品。

d.特定的明星评级是否会引发更多的评论?例如，在看到一系列的低星级评价后，客户更有可能写一些评论吗?

e.基于文本的评论的特定质量描述符，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1到2页的信，总结你的团队的分析和结果。包括你的团队最自信地向市场总监推荐的结果的具体理由。

提交应包括:

 1页总结表

 目录

 1-2页的信

 解决方案不超过20页，带目录和信不超过24页

注意:参考文献和附录不计入页面限制，应在完成解决方案后显示。你不应该使用未经授权的图像和材料，其使用受到版权法的限制。确保你在报告中引用了你的观点和材料。

**术语表**

**帮助度评价:在决定是否购买某个产品时，对某个产品的评价是否有价值的一个指标。**

**奶嘴:一种橡胶或塑料的安抚装置，通常是乳头状的，给婴儿吮吸或咬咬。**

**评审:对产品的书面评价。**

**星级评定:在一个允许人们用若干颗星来评价一个产品的系统中给出的分数。**

**附件:问题数据集**

**Problem\_C\_Data.zip**

所提供的三个数据集包含从Amazon Customer reviews数据集中提取的产品用户评分和评论，数据通过Amazon Simple Storage Service (Amazon S3)存储。

hair\_dryer.tsv

microwave.tsv

pacifier.tsv

**数据集定义:每一行表示划分为以下列的数据。**

●市场(字符串):撰写评论的市场的2个字母国家代码。

●customer\_id(字符串):可用于聚合单个作者所写评论的随机标识符。

●review\_id(字符串):评论的唯一ID。

●product\_id (string):本文所涉及的唯一产品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(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):有用投票的数量。

●total\_votes (int):收到的总票数。

●vine (string):用户被邀请成为Amazon vine上的声音，这是基于他们在Amazon社区中赢得的信任，因为他们写的评论准确而有见地。Amazon向Amazon Vine的会员提供免费的产品拷贝，这些产品是由供应商提交给该计划的。亚马逊不会影响会员的意见，也不会修改或编辑评论。

●verified\_purchase(字符串):一个“Y”表示Amazon验证了写评论的人在Amazon上购买了该产品，并且没有收到折扣优惠的产品。

●review\_heading(字符串):评论的标题。

●review\_body(字符串):查看文本。

●review\_date (bigint):写评论的日期。