****问题C: 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产品进行评分和评级的机会。个人评级-称为“星级”，购买者可以使用1星（星级低，满意程度低）到5星（星级高，满意程度高）的等级来表达他们对产品的满意度。此外，客户还可以提交号称“评论”的短信对产品发表进一步的意见和信息。其他客户可以在这些评论中提交评级，确定是否对自己有帮助（称为“帮助性评级”），以协助自己制定产品购买决策。公司使用这些数据来深入了解其参与的市场、参与的时间以及产品设计功能选择的潜在成功。

阳光公司计划在线上市场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，通过顾客过去对其他竞争产品提供的相关评级和评论，识别关键模式、关系、措施和参数，来：1）告知其在线销售策略； 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来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。阳光公司对这些数据中以时间为基准的模式以及它们是否以有助于该公司制作成功产品的互动方式特别感兴趣。

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购买微波炉、婴儿奶嘴和吹风机的客户提供的评级和评论。他们还提供了数据标签定义的词汇表。提供的数据文件包含了您应当使用的唯一数据。

****要求****

1. 用数学证据、有意义的定量或定性模式、关系、度量和参数分析来识别、描述和支持所提供的三个产品数据集。这些在星级评论和帮助等级之内和之间的参数将帮助阳光公司在他们的三个新的线上市场产品销售中取得成功。

2.用您的分析来解决阳光公司市场总监的以下具体问题和要求：

a.确定评级和评论的数据度量，一旦他们的三种产品在线上市场销售，这些数据对阳光公司来说是最有用的信息。

b.识别和讨论每个数据集当中基于时间的度量和模式，这些度量和模式可能表明一个产品的声誉在线上市场上正在上升或下降。

c.确定基于文本的度量和基于评级的度量的组合，来最好地表示潜在成功或失败的产品。

d.具体的星级会引起更多评论吗？ 例如，在看到一系列低星评级之后，顾客是否更有可能发表某种类型的评论？

2. e.文本评论的具体质量描述（如“热情”，“失望”等）与评分水平是否密切相关？

3. 给阳光公司的市场总监写一封一到两页的信，总结您团队的分析和结果，包括您的团队最有信心推荐给市场总监的结果的具体理由。

你提交的内容应包括：

·一页摘要表

·目录

·一封一到两页的信函

你的解决方案不得超过20页，加上摘要、目录和两页信函最多不得超过24页。

注意：参考列表和任何附录不计入页数限制，应在完成解决方案之后出现。你不应使用未经授权的图像和材料。受版权法的限制，确保你引用的想法和报告中使用的材料的来源。

****词汇****

****帮助等级：****指在决定是否购买该产品时特定产品评论的价值。

****奶嘴：****一种橡胶或塑料舒缓装置，通常是乳头状的，给婴儿吮吸或咬。

****评论****：对产品的书面评价。

****星级****：在一个系统中给出的分数，允许人们对一个产品进行多个星级的评分。

****附件：问题数据集****

Problem\_C\_Data.zip

所提供的三个数据集包含产品用户评分和通过Amazon Simple Storage Service（Amazon S3）从亚马逊顾客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（string）：撰写评论的市场的2个字母的国家/地区代码。

●customer\_id（string）：随机标识符，可用于汇总单个作者撰写的评论。

●review\_id（string）：评论的唯一ID。

●product\_id（string）：评论所属的唯一产品ID。

●product\_parent（string）：随机标识符，可用于汇总同一产品的评论。

●product\_title（string）：产品的标题。

●product\_category（string）：产品的主要消费者类别。

●star\_rating（int）：评论的1-5星级。

●helpful\_votes（int）：有帮助的投票数。

●total\_votes（int）：评论获得的总票数。

●vine（string）：基于客户在Amazon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。 Amazon不会影响Amazon Vine成员的意见，也不会修改或编辑评论。

●verified\_purchase（string）：“ 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，并且没有以大幅度折扣购买该产品。

●review\_headline（string）：评论的标题。

●review\_body（string）：评论文本。

●review\_date（bigint）：撰写评论的日期。