## 《软件工程统计方法中期报告》

一、项目进度简介

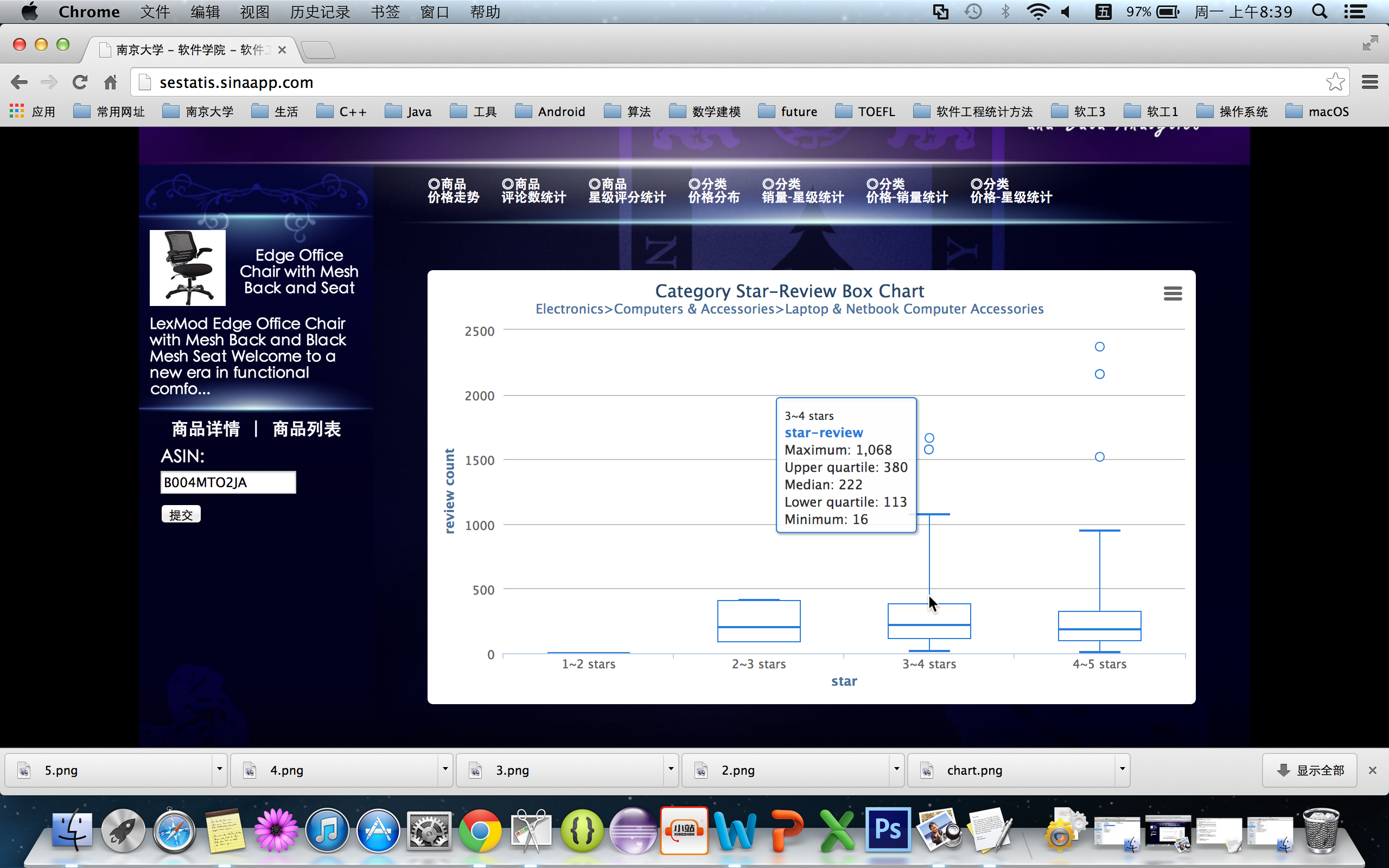
本小组现工程基于trendata.cn提供的开放数据，采用python实现数据分析并已部署在新浪的sae服务器上，URL为：<http://sestatis.sinaapp.com/>，其具体使用情况截图如下：



上图展示了可以选择所有商品种类，并可以选中所有的商品，进而查看每个商品具体的信息。



上图是查看一个具体商品的价格趋势折线图，ASIN为B009CMJCGA，曲线可以动态画出，可以使用鼠标选择来查看每一个数据点的具体值。



上图展示了一个分类下，商品的平均星级和评论总数的盒状图，鼠标选中，可以展示这个星级区间下评论数的最大和最小值，以及第一个四分位数，中位数，第三个四分位数。

二、数据源说明

本次统计数据来源于trendata.cn提供的amazon的商品信息和销售数据。

三、统计分析方法

本次统计活动，我们小组主要采用了如下图表：

1. 折线图：折线图是用直线段将各数据点连接起来而组成的图形，以折线方式显示数据的变化趋势。适用于显示在相等时间间隔下数据的趋势，也可用来分析多组数据随时间变化的相互作用和影响。
2. 柱状图：柱状图用等宽的垂直图条来显示按类别分组的数据，然后利用柱状高度来表示描述统计量的值。适用于对一段时间的数据变化或说明各项之间的比较来说十分有用。
3. 累积频率分布图：将各类别下的频率累加起来，用来展示各分类不同元素的组合比重。
4. 散点图：散点图利用坐标系来表示值，将序列显示为一组点，值由点在图表中的位置表示，类别可以由图表中的不同标记表示。适用于数据集点非常多，处理值的分布问题。
5. 盒状图：盒状图是一种很多的用来表示数据集的位置和变异信息的工具，特别是用于发现不事数据组之间的位置和变异信息的变化。适合用来表示多组数据的分布。

四、完成功能点

1. 单个商品价格走势分析
2. 单个商品评论数统计
3. 单个商品星级评分统计
4. 分类下商品价格分布
5. 分类下评论数与星级评分关系分析
6. 分类下价格与评论数的关系统计
7. 分类下价格与星级的关系分析

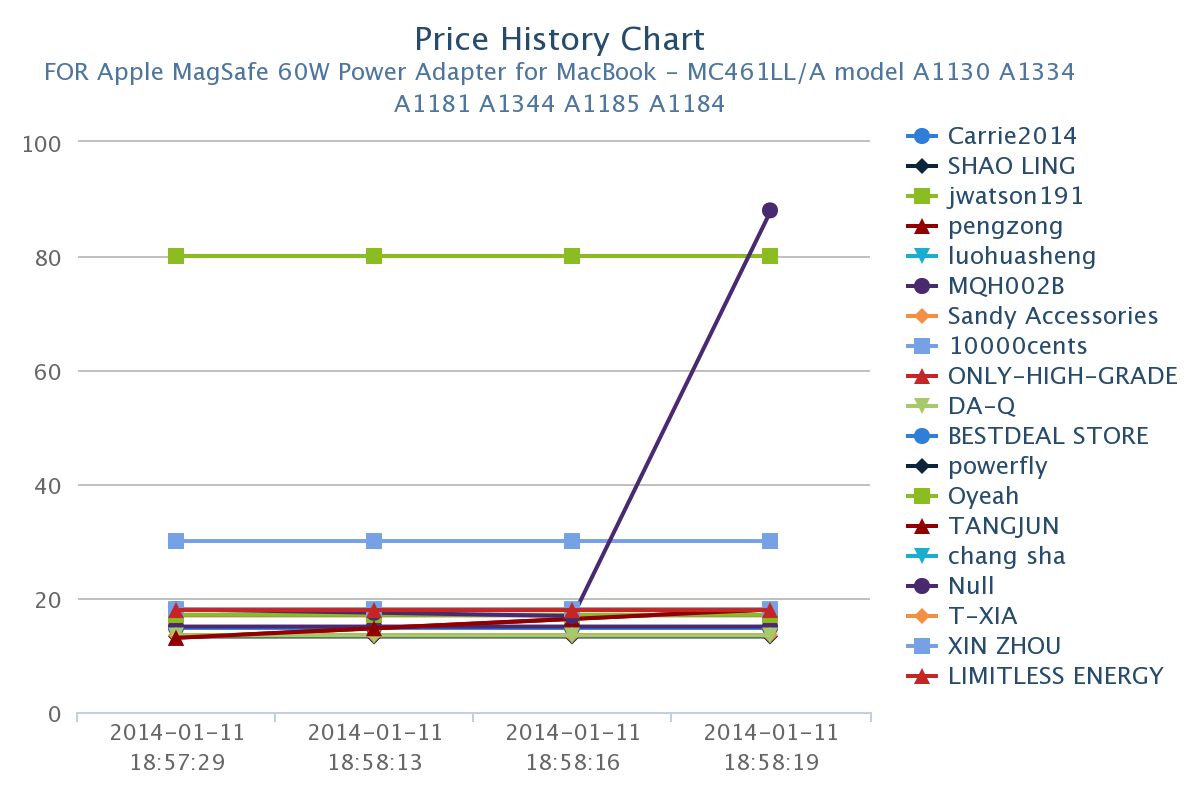
五、结果图表展示及解释说明

接下来将挑选几个分类下的若干种商品进行结果展示：

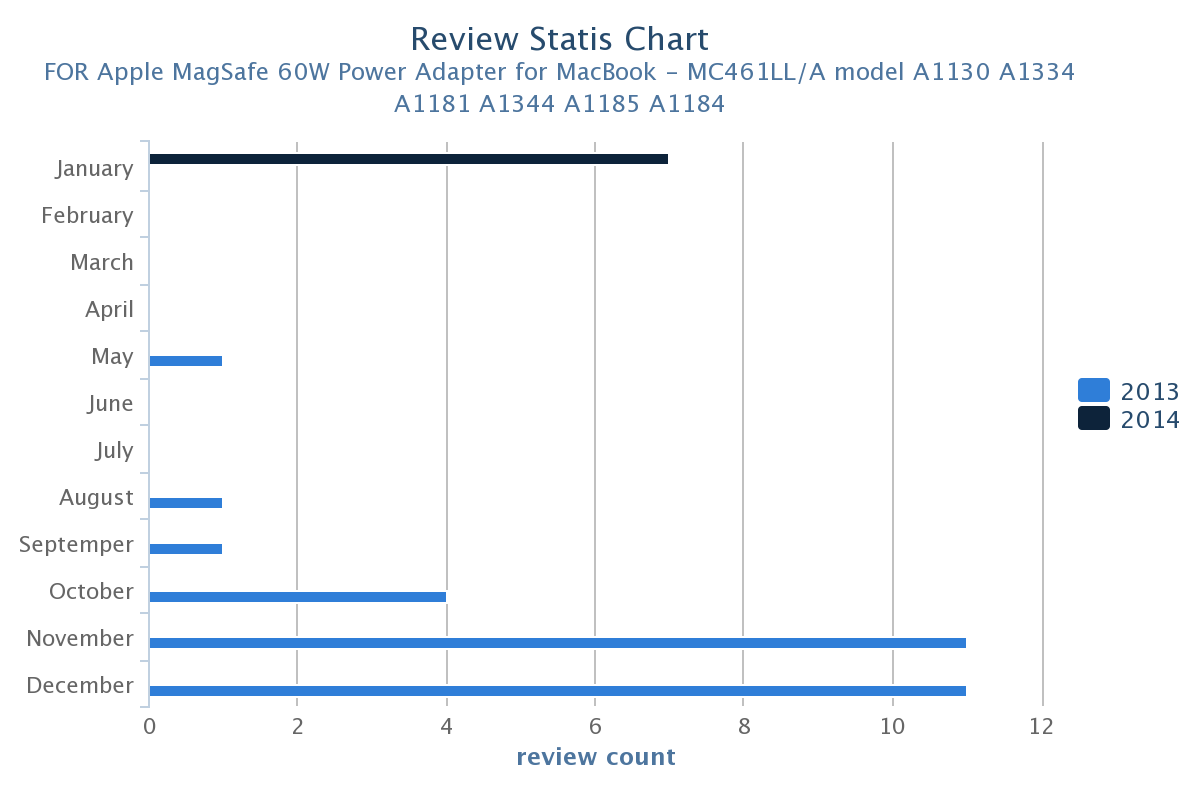
1. 分类:Electronics>Computers & Accessories>laptop & Netbook Computer Accessories

* 商品FOR Apple MagSafe 60W Power Adapter for MacBook，

ASIN: B009CMJCGA



上图为折线图，展示的是这个商品价格与时间的关系，每个供应商对应一条曲线。从上图可以看出，不同的供应商，对这个商品的价格相差比较悬殊，但一般而言，一个供应商的定价较为稳定，但其中一个名为null的供应商，有个时间点突然提高了价格。



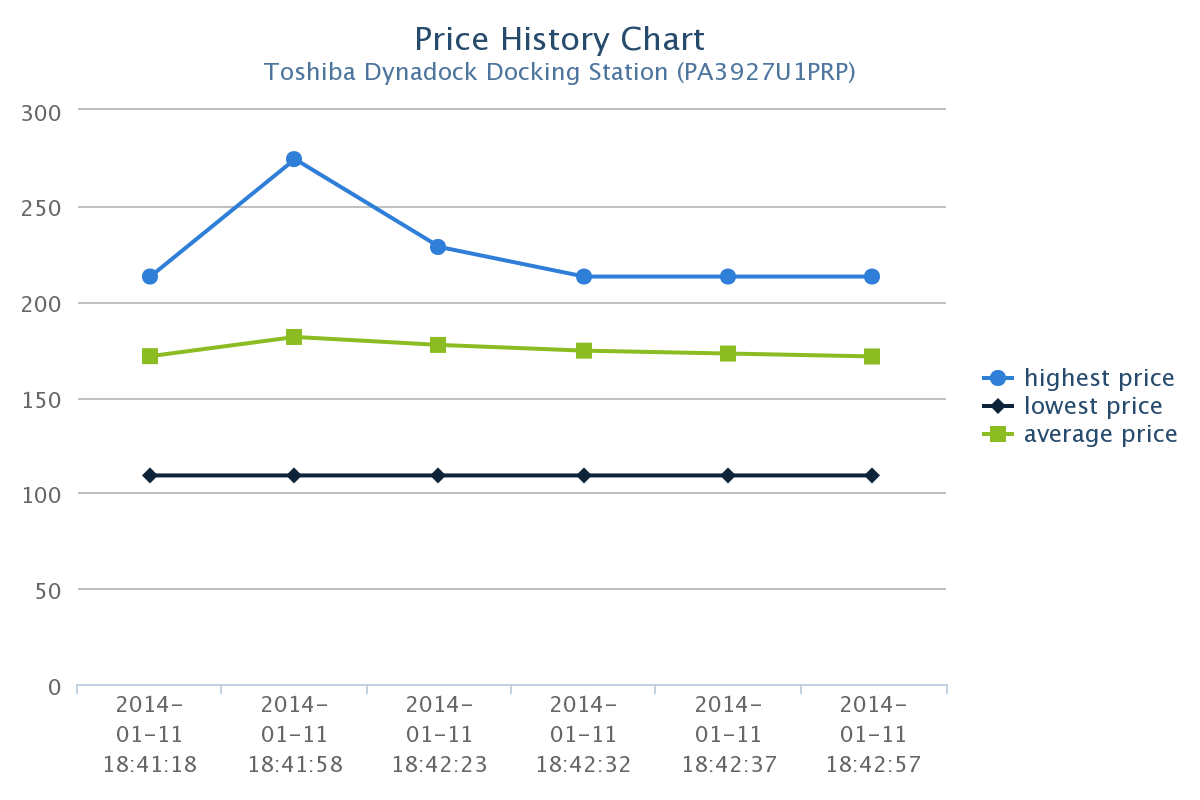
上图为这个商品不同年份，各个月份的评论数的柱状图，可以看出不同月份该商品的销量差距较大，其中在2013年末的销量最好。



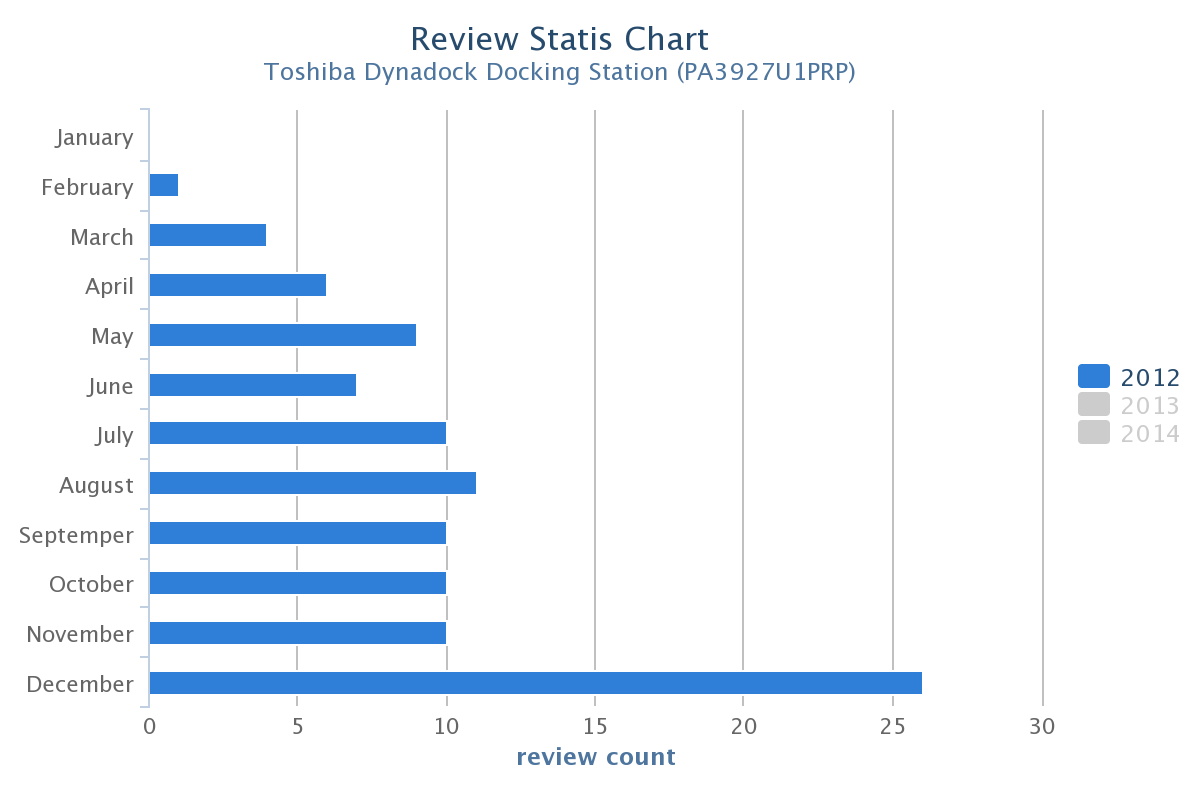
上图展示了该商品星级分布的柱状图，发现一星评论和五星评论占据最多，说明消费者对该商品的评价差距较大，且评分方式容易极端化。

* 商品Toshiba Dynadock Docking Station

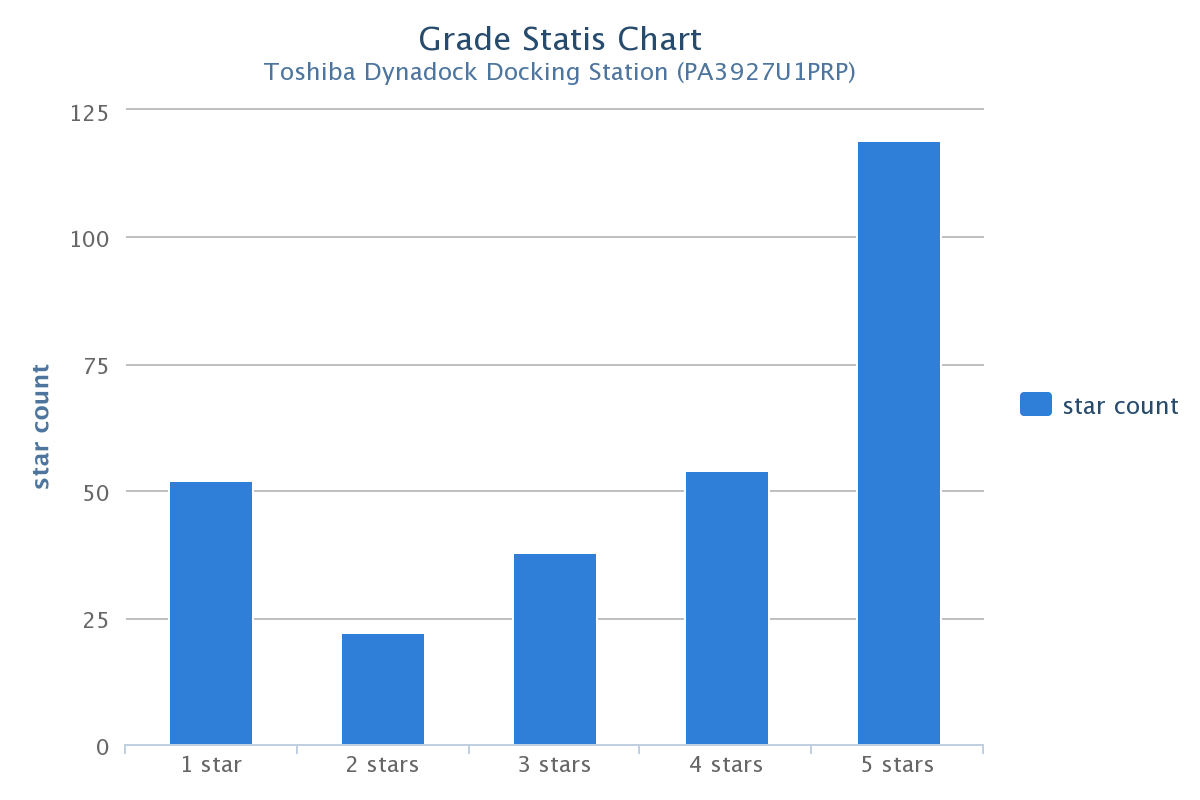
ASIN: B006ZGWJU2



上图展示的是这个商品的价格与时间的折线图，由最高、最低和平均价格组成，由上图可以看出，该商品的最高价格随时间变化较大，而最低价格和平均价格都比较稳定。



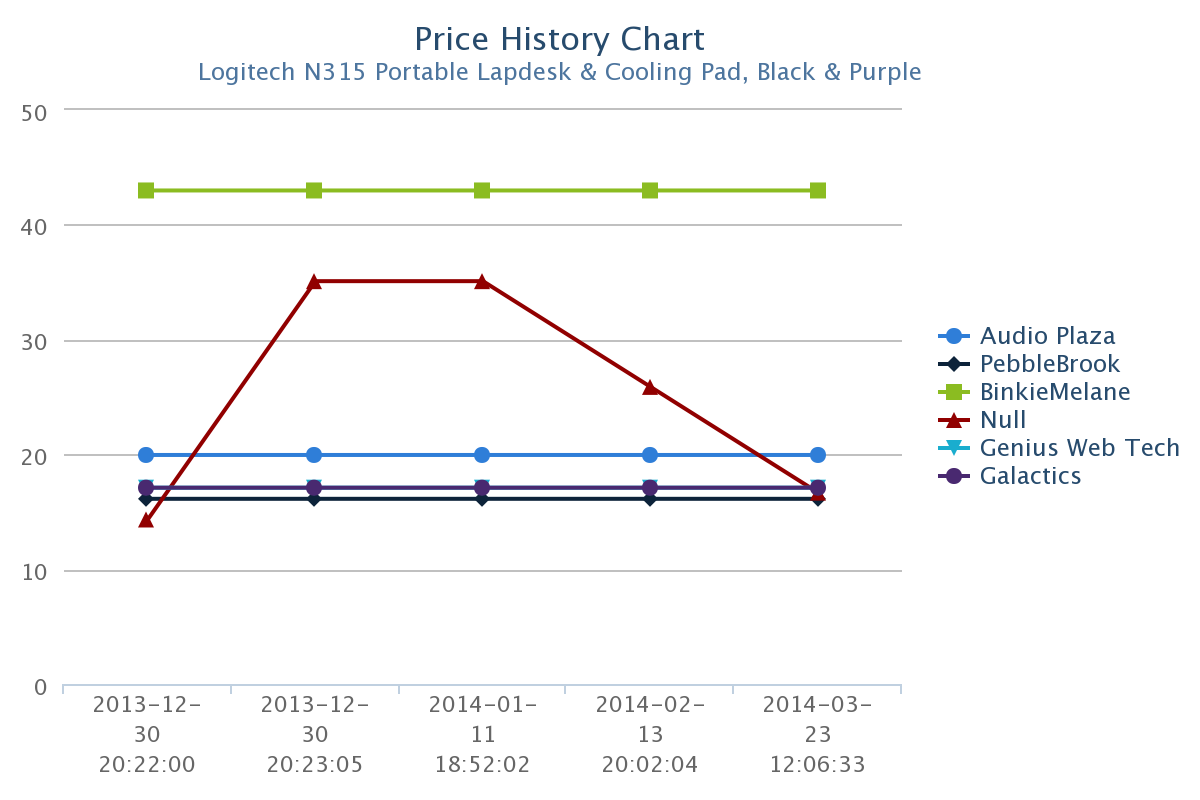
上图选择展示了该商品评论数在2012年各月份的柱状图，可以看出，除个别月份外，该商品在各月份的销量都基本稳定，而在12月份，销量突然升高，可能是由于西方圣诞节的原因。



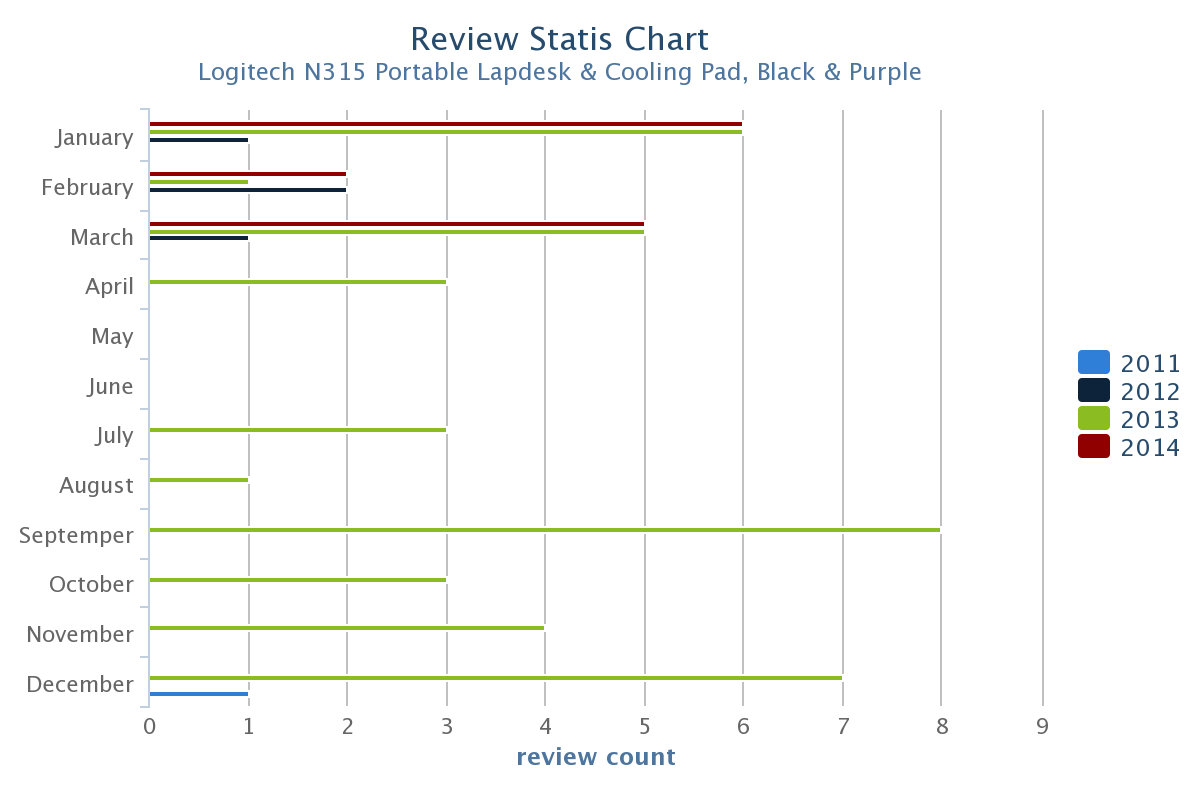
上图为该商品星级评分的柱状图，看出其5星评分占较大比例，因此该商品比较受到消费者的好评。

* 商品名Logitech N315 Portable Lap desk & Cooling Pad, Blac...

ASIN：B005G40MVW

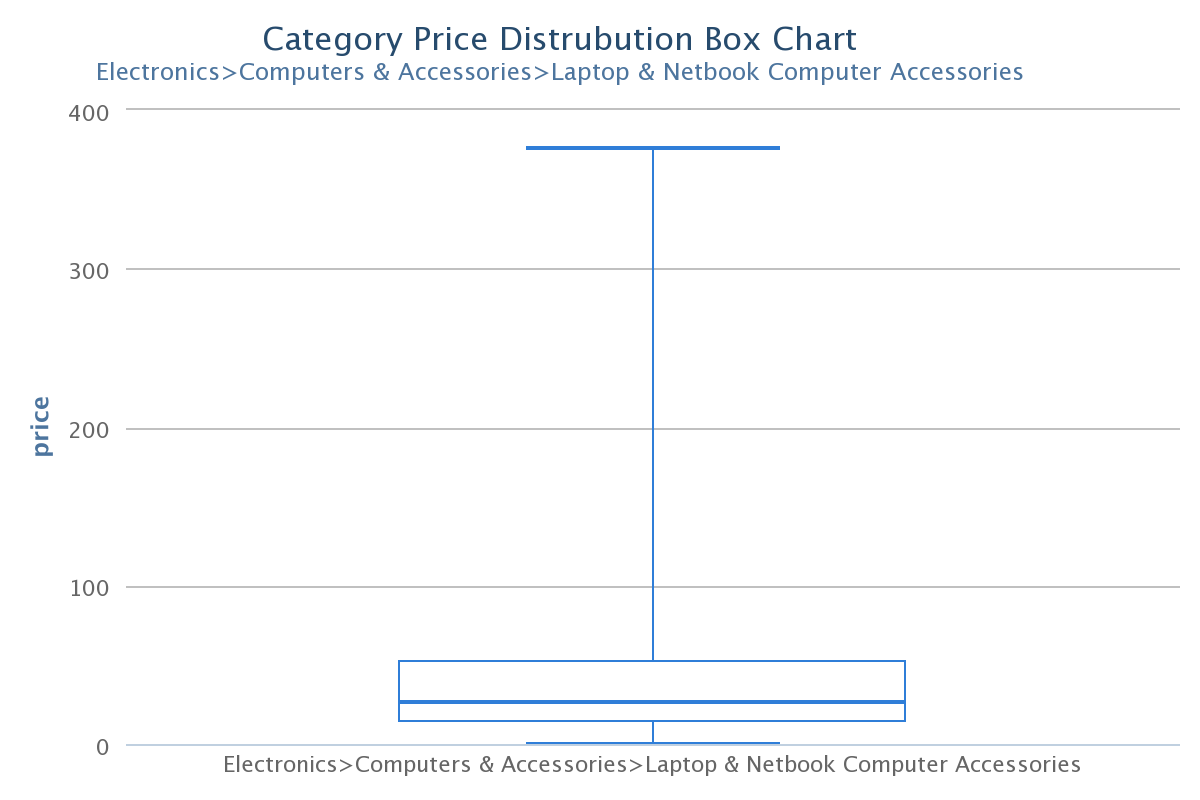


上图为该商品价格时间的折线图，每个供应商对应一条折线，可以看出的供应商，对这个商品的价格相差比较悬殊，但一般而言，一个供应商的定价较为稳定，但其中一个名为null的供应商，不同时间价格变化较大。

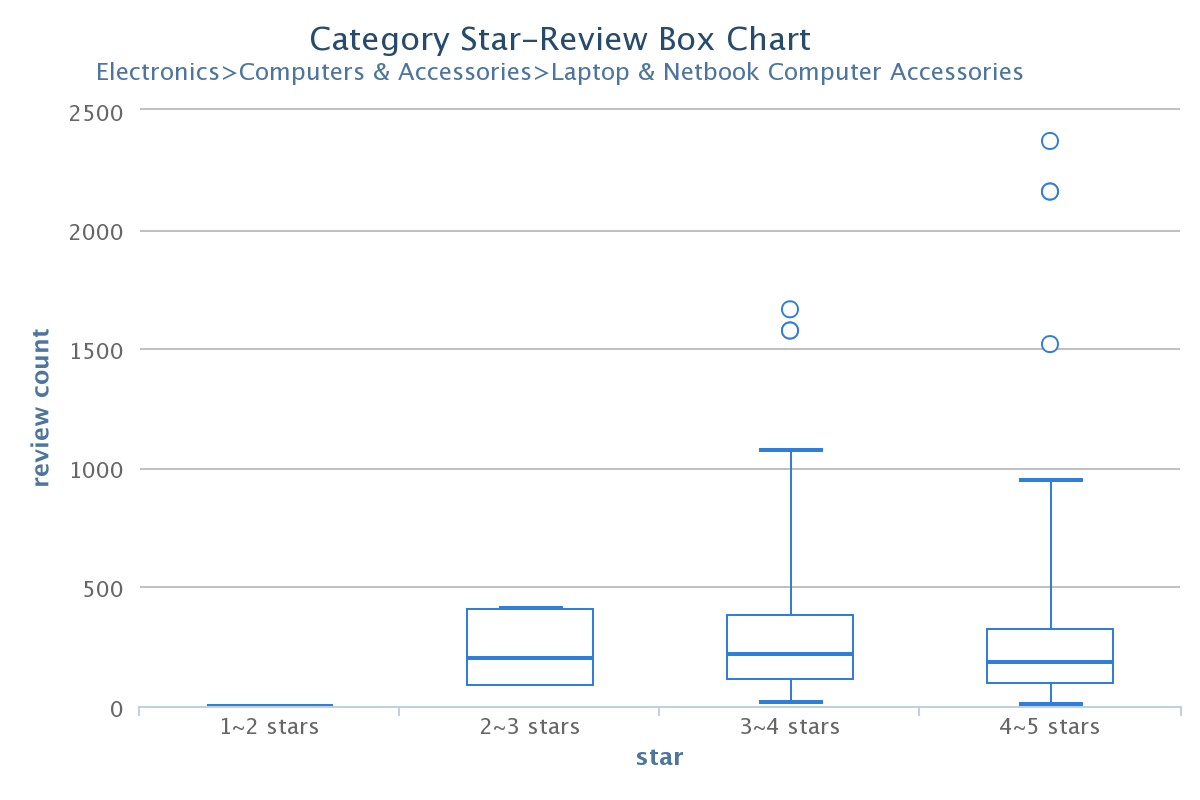


上图为该商品在不同年份的各月的销量情况柱状图，从现有的数据可以看出，不同年份，在相同的月，其销量基本保持稳定。

* 分类下统计分析



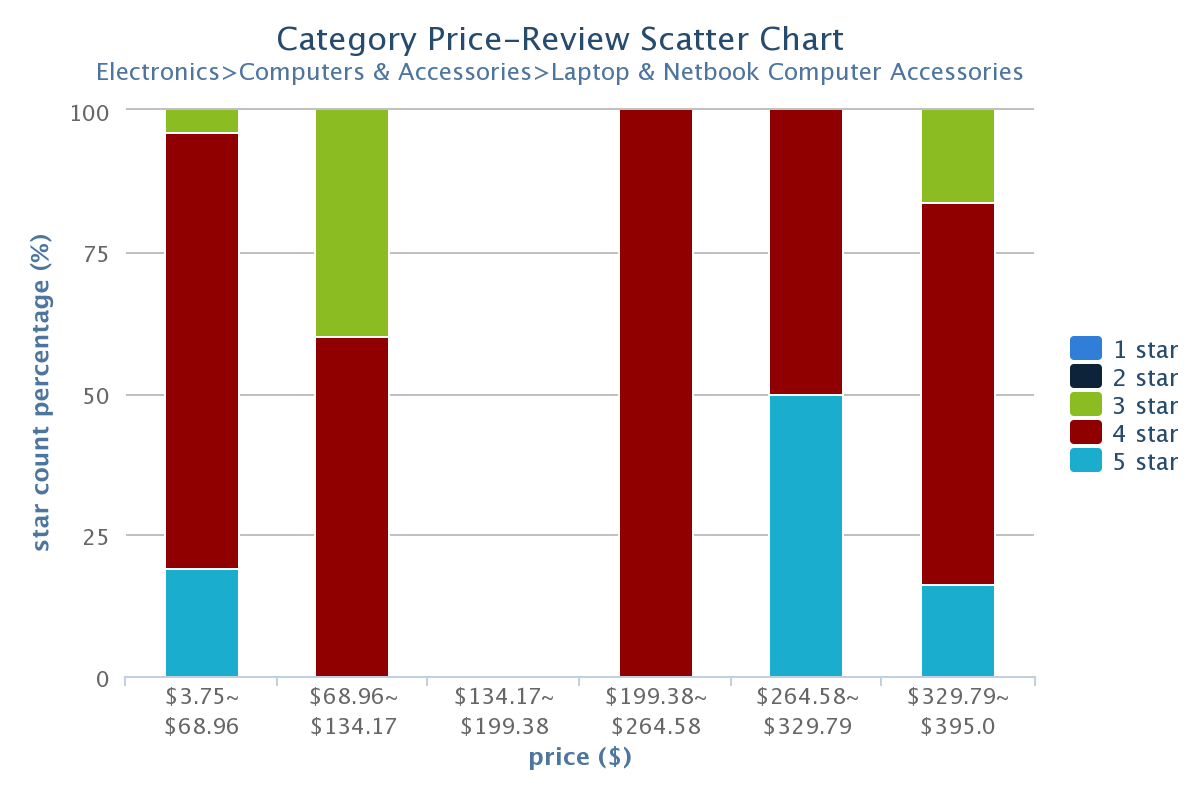
上图为这个分类下价格分布的盒状图，从上图可以看出，该分类下商品价格分布较集中，且主要集中在15～50美元，但也存在极端值，如最高价格就和主要价格区间相差较大。



上图为在该分类下，对各商品计算其平均星级，统计在每个平均星级区间内各商品评论总数分布的盒状图，从而分析商品星级与销量的关系。由上图可以看出，平均星级过低的商品，如只有1～2星，其销量也较低，而在其他星级中，销量比较接近，可以理解成用户购买一些商品，总会根据个人主观的一些看法作出评分，导致星级分布的差异。而值得注意的是，在星级较高的分类中，如4～5星的商品，存在一些离群的极值点，可能是因为一些物美价廉的商品受到大家的喜爱，因些销量特别好。

这里放散点图，去掉点

上图为该分类中价格与评论总数的散点图，可以看出，数据点主要集中在价格较低的区间上，我们可以得出对于一些价格较低的商品，可能获得较高的销量，因为低价商品能使顾客的购买负担更小，更能吸引顾客；而对于价格过高的商品，则少有人问津。

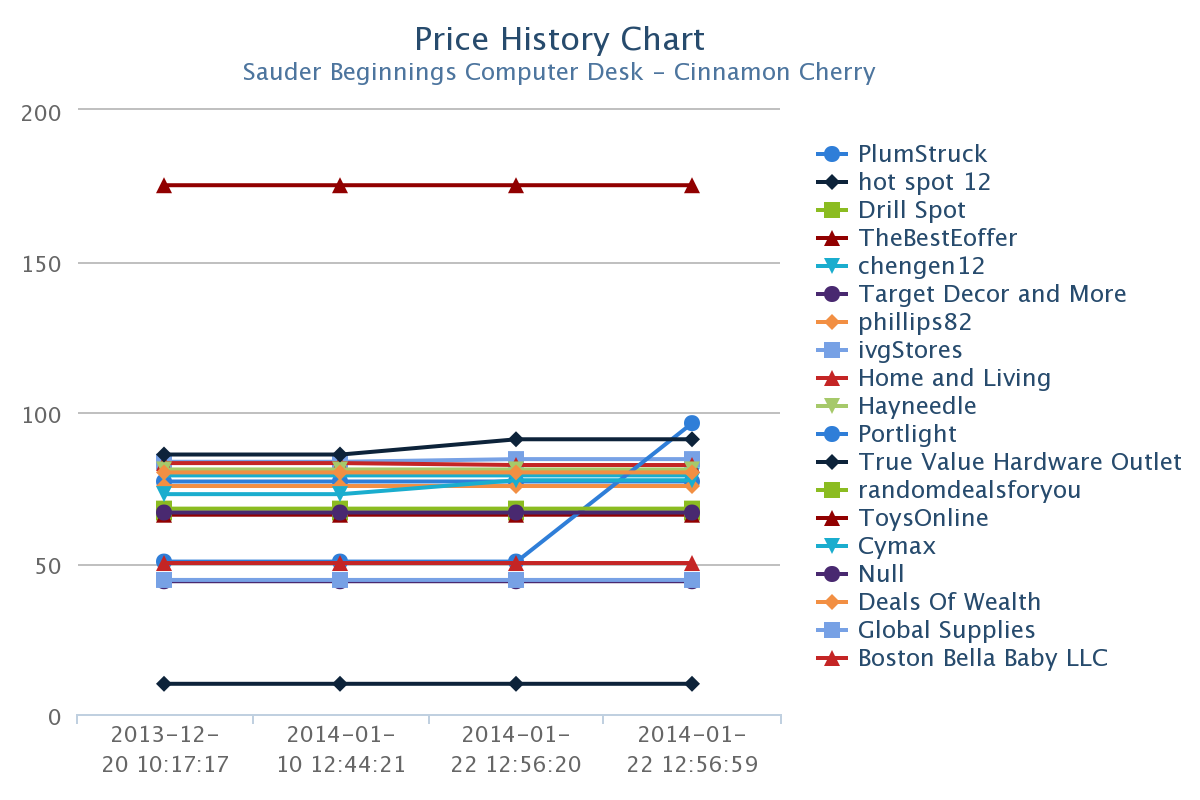


上图为该分类下，不同价格区间内的商品的星级评分的频数分布图，可以看出来，即使是廉价的商品，其好评（4～5星）的比例也不会低于高价的商品。由于在某一个价格区间没有商品，因此存在空白。

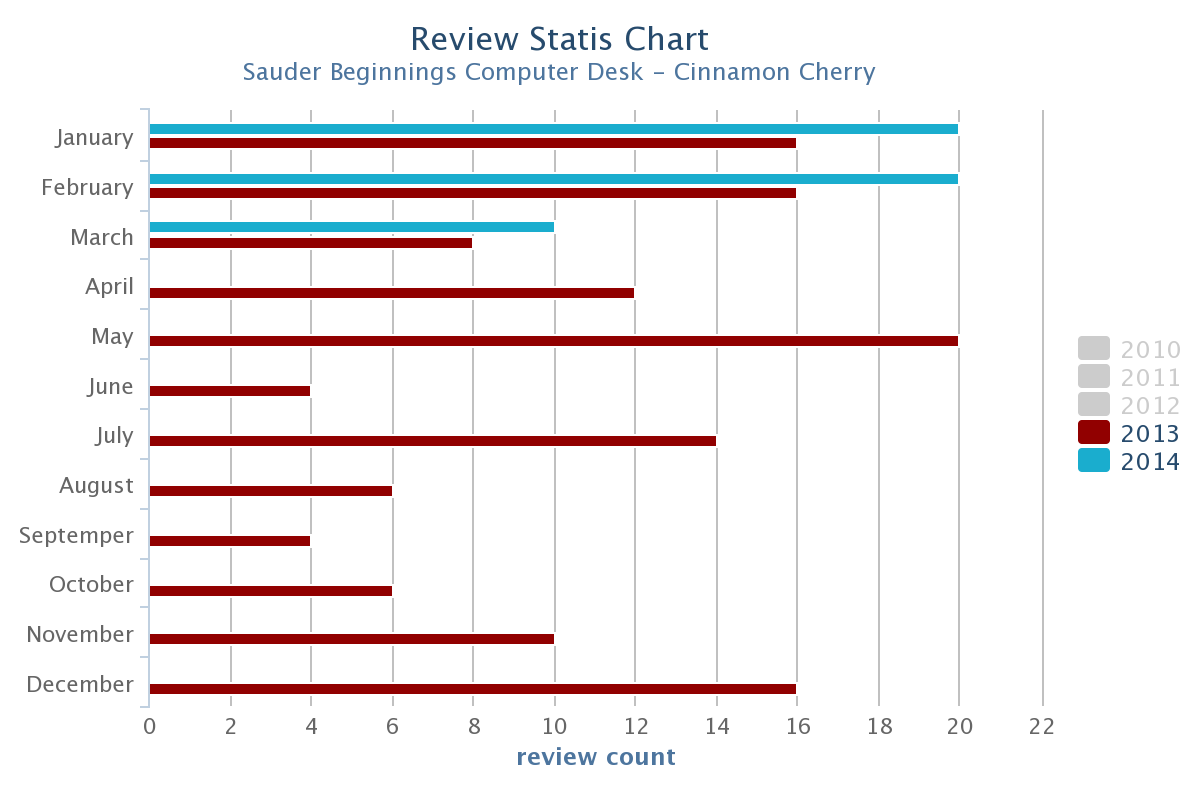
1. 分类：home & kitchen >furniture >Home Office Furniture

* 商品名称Sauder Beginnings Computer Desk - Cinnamon Cherry

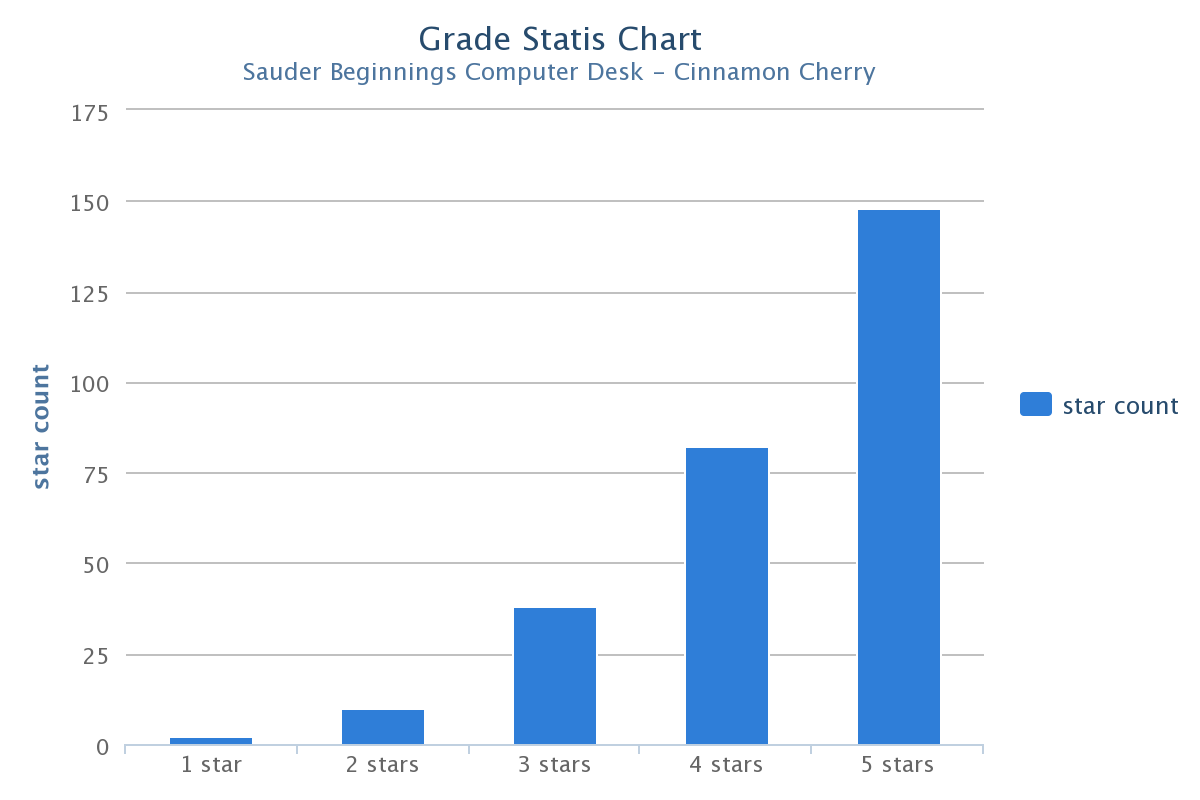
ASIN: B002QGLBI6



上图为折线图，展示的是这个商品价格与时间的关系，每个供应商对应一条曲线。从上图可以看出，大多数供应商对商品的定价都集中在一个很小的区间中，且一个供应商的价格基本保持稳定。



上图为该商品的评论数统计条状图，选择了2013、2014两年的数据，可以明显看出，在2014年的各个月份，同比2013年，评论数均有所上升，说明该商品的受欢迎程度正在上升，市场越来越好。



上图为该商品星级评分条状图，可以看出，该商品好评数占较大比例，说明该商品受到顾客的一致好评。