

我们在网上购物的时候，除了关注商品的价格和性能之外，往往会更加关注该类商品的评论内容及好评度如何，这些都是购买商品的重要指标。

本文从 kaggle 平台上下载了一份以前关于 Kindle 产品的消费者评论数据，想通过这份数据分析两个方面：

- 用户对 Kindle 相关产品的好评率如何？
- 对 Kindle 相关产品备受好评的原因是什么？

下面是分析过程。

## 一、数据分析的目的

---

### 提出问题：

- (1) Kindle 相关产品的评论数排名和平均评分能反映哪些问题？
- (2) Kindle 相关产品的好评率如何？推荐评论占比怎样？
- (3) Kindle 相关产品随时间变化其评论数的变化情况是怎样的？
- (4) Kindle 相关产品备受好评的原因主要是什么？

## 二、理解数据

### 数据来源：

<https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products/home>

数据集中有用的字段如下：

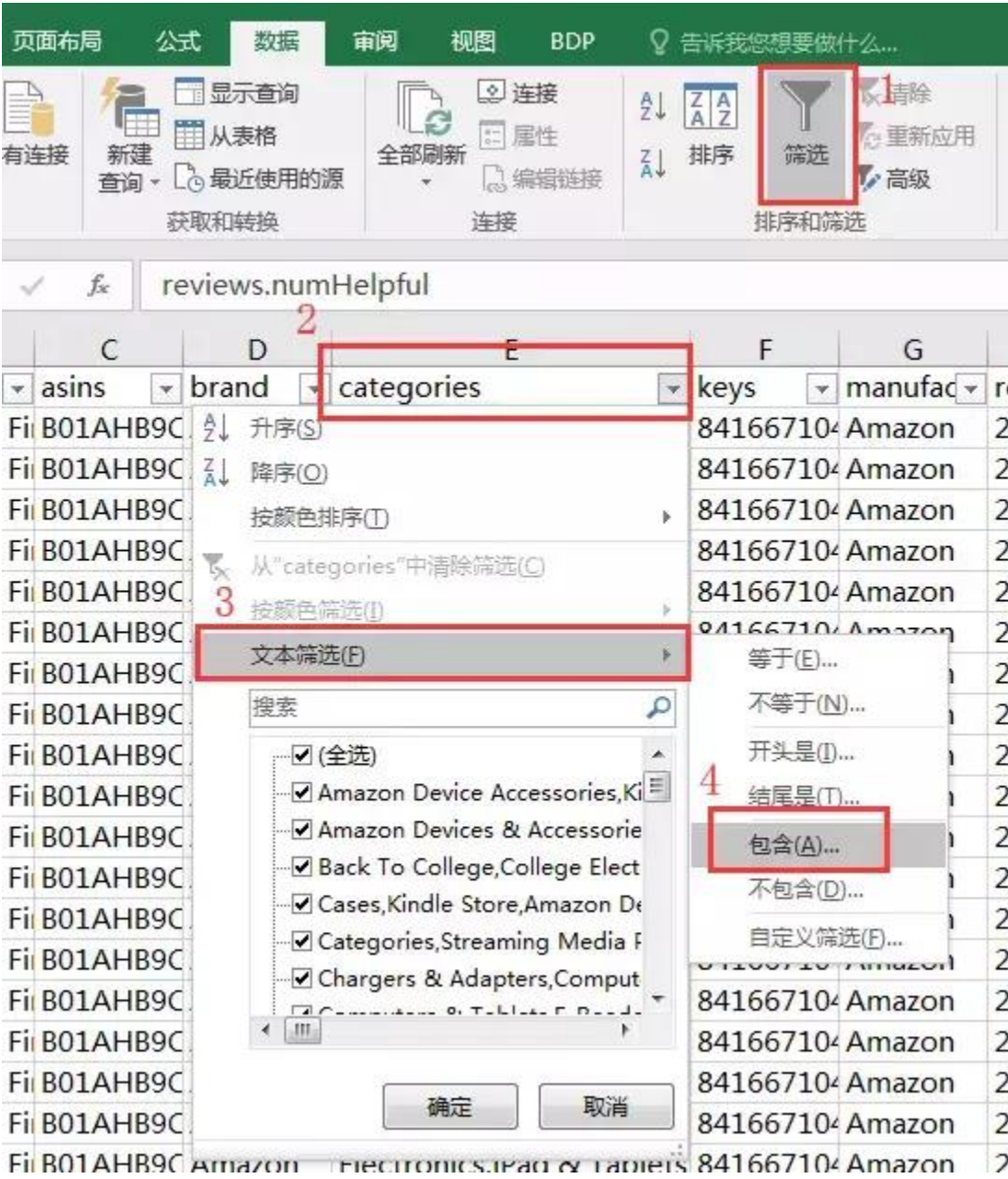
- id-用户编号
- name-产品名称
- asins-产品编号，每一个产品编号对应一个产品，可用它进行产品分组
- brand-品牌
- categories-产品类别
- keys-类别关键字
- manufacturer-制造商
- reviews.date-评论时间，可以分析各时间段评论数
- reviews.dateAdded-追评时间
- reviews.dateSeen-评论可见时间
- reviews.doRecommend-评论是否被推荐
- reviews.numHelpful-帮助性分子数
- reviews.rating-评分，可进行求平均值等描述统计分析
- reviews.sourceURLs-评论链接
- reviews.text-评论文字内容，可以进行抽取关键字，分析评分原因
- reviews.title-评论标题
- reviews.username-用户名

### 三、数据清洗

---

#### 1、选取子集

1) 在产品类别(categories)中筛选含 Kindle 类别的产品，并将筛选的数据复制到新的文件中：



自定义自动筛选方式

显示行:

categories

包含

☒ 与(A) ☐ 或(O)

可用? 代表单个字符  
用\* 代表任意多个字符

2) 隐藏与分析问题无关的字段，保留以下字段：name-产品名称、asins-产品编号、categories-产品类别、reviews.date-评论时间、reviews.doRecommend-评论是否被推荐、reviews.rating-评分、reviews.text-评论文字内容、reviews.title-评论标题

	B	C	E	H	L	O	Q	R	V
1	name	asins	categories	reviews.da	reviews.dc	reviews.rat	reviews.te	reviews.title	
2	Kindle Oas	B00VINDB	eBook Rea	2017-06-3	TRUE	5	Very lightv	Works Great	
3	Kindle Oas	B00VINDB	eBook Rea	2017-06-3	TRUE	5	I like this s	In love	
4	Kindle Oas	B00VINDB	eBook Rea	2017-06-2	TRUE	5	Replacing	Great size	
5	Kindle Oas	B00VINDB	eBook Rea	2017-06-1	TRUE	4	This is my	Very light	
6	Kindle Oas	B00VINDB	eBook Rea	2017-06-1	TRUE	4	This kindle	Great kindle	
7	Kindle Oas	B00VINDB	eBook Rea	2017-06-0	TRUE	5	The new K	Great entertainment	
8	Kindle Oas	B00VINDB	eBook Rea	2017-05-1	TRUE	5	Met all my	Great Product	
9	Kindle Oas	B00VINDB	eBook Rea	2017-05-1	TRUE	5	I thought I	Best Kindle Ever	
10	Kindle Oas	B00VINDB	eBook Rea	2017-04-2	TRUE	5	I own a Kin	Awesome! I love it!	
11	Kindle Oas	B00VINDB	eBook Rea	2017-04-0	TRUE	4	Got me re	Excellent product.	
12	Kindle Oas	B00VINDB	eBook Rea	2017-04-0	TRUE	5	easy to rea	Very please with our r	

## 2、列重命名

将数据集中字段的名称更改为中文字段

	B	C	E	H	L	O	Q	R	V
1	产品名称	产品编号	产品类别	评论时间	评论是否被	评分	评论文字内	评论标题	
2	Kindle Oas	B00VINDB	eBook Rea	2017-06-3	TRUE		5 Very lightv	Works Great	
3	Kindle Oas	B00VINDB	eBook Rea	2017-06-3	TRUE		5 I like this s	In love	
4	Kindle Oas	B00VINDB	eBook Rea	2017-06-2	TRUE		5 Replacing	Great size	
5	Kindle Oas	B00VINDB	eBook Rea	2017-06-1	TRUE		4 This is my	Very light	
6	Kindle Oas	B00VINDB	eBook Rea	2017-06-1	TRUE		4 This kindle	Great kindle	
7	Kindle Oas	B00VINDB	eBook Rea	2017-06-0	TRUE		5 The new K	Great entertainment	
8	Kindle Oas	B00VINDB	eBook Rea	2017-05-1	TRUE		5 Met all my	Great Product	
9	Kindle Oas	B00VINDB	eBook Rea	2017-05-1	TRUE		5 I thought I	Best Kindle Ever	
10	Kindle Oas	B00VINDB	eBook Rea	2017-04-2	TRUE		5 I own a Kin	Awesome! I love it!	
11	Kindle Oas	B00VINDB	eBook Rea	2017-04-0	TRUE		4 Got me rea	Excellent product.	
12	Kindle Oas	B00VINDB	eBook Rea	2017-04-0	TRUE		5 Great rea	Very close with over	

### 3、删除重复项

这里需要对评论文字内容列进行删除重复项，此操作删除 0 个重复项。

### 4、处理缺失值

缺失值处理方法有四种：

- 1) 对比缺失内容，通过手动输入内容补全
- 2) 删除缺失的数据
- 3) 用平均值代替缺失值
- 4) 用统计模型计算出的值去代替缺失值

通过筛选查看各字段内容，发现产品名称(name)、产品品牌(asins)、评分(rating)、评论时间(date)和是否被推荐(doRecommend)字段中均存在缺失值。

a. 产品名称(name)、产品编号(asins)、评分(rating)、评论时间(date)：对这 4 个字段中存在的缺失值，为了避免输入错误信息或删除有用的信息，这里选择直接删除空白项对应的行

操作：选择对应字段 - 定位条件 - 空值 - 删除行



b. 是否被推荐(doRecommend)：用 FALSE 手动补全评论是否被推荐  
(doRecommend)字段中的缺失值

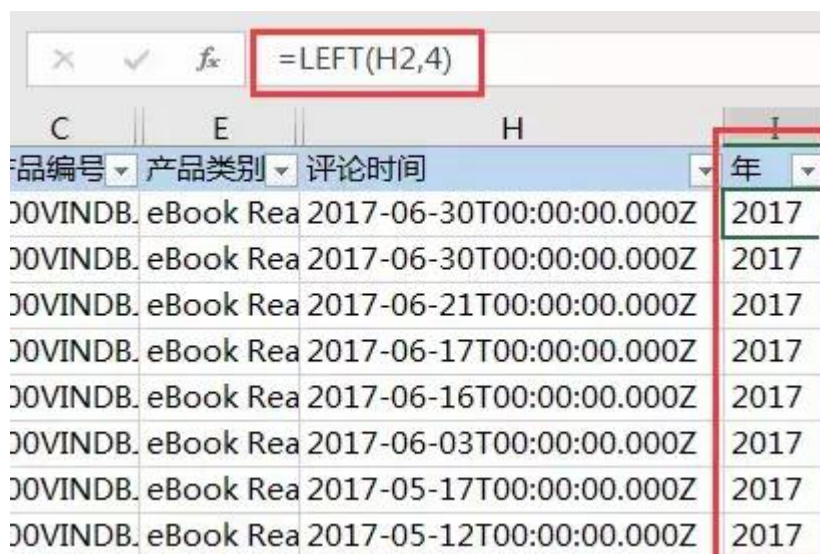
操作：选择对应字段 - 定位条件 - 空值 - 输入 FALSE - ctrl+Enter

## 5、一致化处理

### 1) 日期格式转换

评论时间字段中格式包含日期+时间，分析中并未使用到具体时间段，这里把评论日期拆分为三列，分别为年、月、日

年：插入一列年，使用 LEFT()函数



The screenshot shows an Excel spreadsheet with columns C, E, H, and I. Column C is labeled '品编号' (Product ID), column E is '产品类别' (Product Category), and column H is '评论时间' (Review Time). A new column I, labeled '年' (Year), has been added. The formula bar shows the formula =LEFT(H2,4) for cell I2. The data in column I shows the year 2017 for all rows.

品编号	产品类别	评论时间	年
00VINDB.	eBook Rea	2017-06-30T00:00:00.000Z	2017
00VINDB.	eBook Rea	2017-06-30T00:00:00.000Z	2017
00VINDB.	eBook Rea	2017-06-21T00:00:00.000Z	2017
00VINDB.	eBook Rea	2017-06-17T00:00:00.000Z	2017
00VINDB.	eBook Rea	2017-06-16T00:00:00.000Z	2017
00VINDB.	eBook Rea	2017-06-03T00:00:00.000Z	2017
00VINDB.	eBook Rea	2017-05-17T00:00:00.000Z	2017
00VINDB.	eBook Rea	2017-05-12T00:00:00.000Z	2017

月：插入一列月，使用 MID()函数和 FIND()函数

=MID(H2,FIND("-",H2,1)+1,2)						
B	C	E	H	I	J	
产品名称	产品编号	产品类别	评论时间	年	月	
Kindle Oas	B00VINDB.	eBook Rea	2017-06-30T00:00:00.000Z	2017	06	
Kindle Oas	B00VINDB.	eBook Rea	2017-06-30T00:00:00.000Z	2017	06	
Kindle Oas	B00VINDB.	eBook Rea	2017-06-21T00:00:00.000Z	2017	06	
Kindle Oas	B00VINDB.	eBook Rea	2017-06-17T00:00:00.000Z	2017	06	
Kindle Oas	B00VINDB.	eBook Rea	2017-06-16T00:00:00.000Z	2017	06	
Kindle Oas	B00VINDB.	eBook Rea	2017-06-03T00:00:00.000Z	2017	06	
Kindle Oas	B00VINDB.	eBook Rea	2017-05-17T00:00:00.000Z	2017	05	
Kindle Oas	B00VINDB.	eBook Rea	2017-05-12T00:00:00.000Z	2017	05	

日：插入一列日，使用 MID ()函数和 FIND()函数

=MID(H2,FIND("T",H2,1)-2,2)						
B	C	E	H	I	J	K
产品名称	产品编号	产品类别	评论时间	年	月	日
Kindle Oas	B00VINDB.	eBook Rea	2017-06-30T00:00:00.000Z	2017	06	30
Kindle Oas	B00VINDB.	eBook Rea	2017-06-30T00:00:00.000Z	2017	06	30
Kindle Oas	B00VINDB.	eBook Rea	2017-06-21T00:00:00.000Z	2017	06	21
Kindle Oas	B00VINDB.	eBook Rea	2017-06-17T00:00:00.000Z	2017	06	17
Kindle Oas	B00VINDB.	eBook Rea	2017-06-16T00:00:00.000Z	2017	06	16
Kindle Oas	B00VINDB.	eBook Rea	2017-06-03T00:00:00.000Z	2017	06	03
Kindle Oas	B00VINDB.	eBook Rea	2017-05-17T00:00:00.000Z	2017	05	17
Kindle Oas	B00VINDB.	eBook Rea	2017-05-12T00:00:00.000Z	2017	05	12
Kindle Oas	B00VINDB.	eBook Rea	2017-04-23T00:00:00.000Z	2017	04	23

## 2) 分组

在评分字段后面插入一列评价等级，对评分进行分组，03 分为差评，45 分为好评

使用 IF 函数实现：=IF(P2<=3,"差评","好评")

阈值	评价等级	取值范围
0	差评	$0 < x \leq 3$
4	好评	$3 < x \leq 5$

S2	=IF(R2<=3,"差评","好评")									
	B	C	E	H	I	J	K	O	R	S
1	产品名称	产品编号	产品类别	评论时间	年	月	日	评论是否	评分	评价等级
2	Kindle Oas	B00VINDB	eBook Rea	2017-06-30T00:00:00.000Z	2017	06	30	TRUE	5	好评
3	Kindle Oas	B00VINDB	eBook Rea	2017-06-30T00:00:00.000Z	2017	06	30	TRUE	5	好评
4	Kindle Oas	B00VINDB	eBook Rea	2017-06-21T00:00:00.000Z	2017	06	21	TRUE	5	好评
5	Kindle Oas	B00VINDB	eBook Rea	2017-06-17T00:00:00.000Z	2017	06	17	TRUE	4	好评
6	Kindle Oas	B00VINDB	eBook Rea	2017-06-16T00:00:00.000Z	2017	06	16	TRUE	4	好评
7	Kindle Oas	B00VINDB	eBook Rea	2017-06-03T00:00:00.000Z	2017	06	03	TRUE	5	好评
8	Kindle Oas	B00VINDB	eBook Rea	2017-05-17T00:00:00.000Z	2017	05	17	TRUE	5	好评
9	Kindle Oas	B00VINDB	eBook Rea	2017-05-12T00:00:00.000Z	2017	05	12	TRUE	5	好评
10	Kindle Oas	B00VINDB	eBook Rea	2017-04-23T00:00:00.000Z	2017	04	23	TRUE	5	好评
11	Kindle Oas	B00VINDB	eBook Rea	2017-04-07T00:00:00.000Z	2017	04	07	TRUE	4	好评

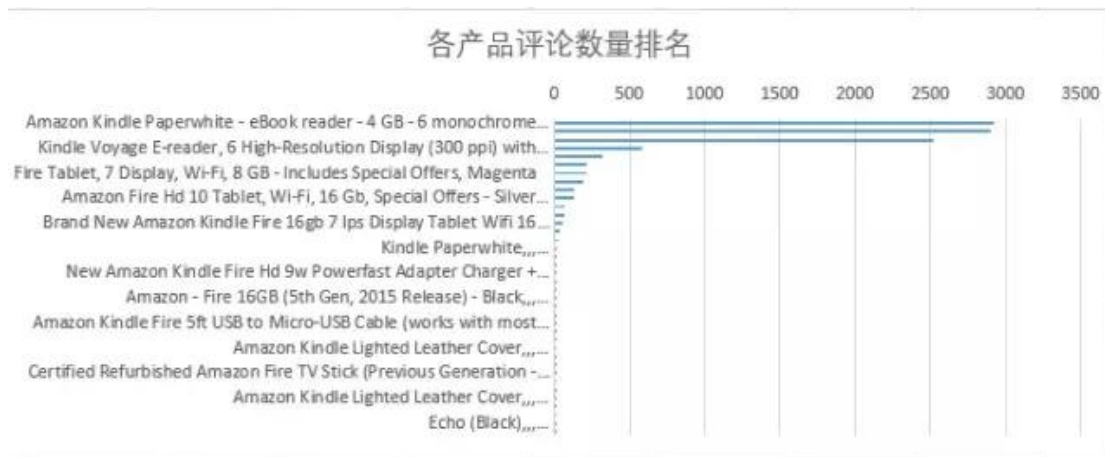
## 四、构建模型

### 1、Kindle 相关产品的评论数排名和平均评分能反映哪些问题？

利用数据透视表对每个产品进行统计评论数，并按高低排名，通过产品评论数，可以大概了解到各产品的销售情况，评论数越多，表示该产品销量越多。

行标签	计数项:评论文字内容
Amazon Kindle Paperwhite - eBook reader - 4 GB - 6 monochrome Paperwhite - touchscreen - Wi-Fi - black...	2927
Echo (White)... Echo (White)...	2898
Amazon Fire TV... Amazon Fire TV...	2527
Kindle Voyage E-reader, 6 High-Resolution Display (300 ppi) with Adaptive Built-in Light, PagePress Sensors, Wi-Fi - Includes Special Offers...	580
Amazon - Amazon Tap Portable Bluetooth and Wi-Fi Speaker - Black... Amazon - Amazon Tap Portable Bluetooth and Wi-Fi Speaker - Black...	318
All-New Kindle E-reader - Black, 6 Glare-Free Touchscreen Display, Wi-Fi - Includes Special Offers...	212
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Magenta	211
Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders... Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets ar	195
Amazon Fire Hd 8 Bin Tablet 16gb Black 8018ast3bk 6th Gen (2016) Android... Amazon Fire Hd 8 Bin Tablet 16gb Black 8018ast3bk 6th Gen (2016) Android...	125
Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special Offers - Silver Aluminum... Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special Offers - Silver Aluminum...	128
Kindle Oasis E-reader with Leather Charging Cover - Merlot, 8 High-Resolution Display (300 ppi), Wi-Fi - Includes Special Offers...	67
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Black	64
Brand New Amazon Kindle Fire 16gb 7 Ips Display Tablet Wifi 16 Gb Blue...	57
Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders... Amazon 9W PowerFast Official OEM USB Charger and Power Adeq	36
Kindle Paperwhite E-reader - White, 6 High-Resolution Display (300 ppi) with Built-in Light, Wi-Fi - Includes Special Offers...	30
Kindle Paperwhite... Kindle Paperwhite...	9
Kindle Dx Leather Cover, Black (fits 9.7 Display, Latest and 2nd Generation Kindle Dxs)...	9
Amazon Echo and Fire TV Power Adapter... Amazon Echo and Fire TV Power Adapter...	8
New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable... New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable...	6
Amazon Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 GB, Blue Kid-Proof Case - Blue	6
Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red... Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Caye	6
Amazon - Fire 16GB (5th Gen, 2015 Release) - Black... Amazon - Fire 16GB (5th Gen, 2015 Release) - Black...	5
Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 GB, Green Kid-Proof Case	4
Certified Refurbished Amazon Fire TV with Alexa Voice Remote... Certified Refurbished Amazon Fire TV with Alexa Voice Remote...	4
Amazon Kindle Fire 5ft USB to Micro-USB Cable (works with most Micro-USB Tablets)... Amazon Kindle Fire 5ft USB to Micro-USB Cable (works with most Micro-USB Tablets)...	4
Echo (Black)... Echo (Black)...	3
Amazon Kindle Fire Hd (3rd Generation) 8gb... Amazon Kindle Fire Hd (3rd Generation) 8gb...	2
Amazon Kindle Lighted Leather Cover... Amazon Kindle Lighted Leather Cover...	2
Amazon Fire TV... Kindle Dx Leather Cover, Black (fits 9.7 Display, Latest and 2nd Generation Kindle Dxs)...	1
Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red... Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kir	1
Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st)... Kindle Paperwhite...	1
Echo (White)... Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Tangerine	1
Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st)... Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st)...	1
Amazon Kindle Lighted Leather Cover... Kindle Keyboard...	1
New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable...	1
Kindle Keyboard... Kindle Keyboard...	1
Echo (Black)... Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders...	1
Coconut Water Red Tea 16.5 Oz (pack of 12)... Amazon Fire Tv...	1
总计	10463

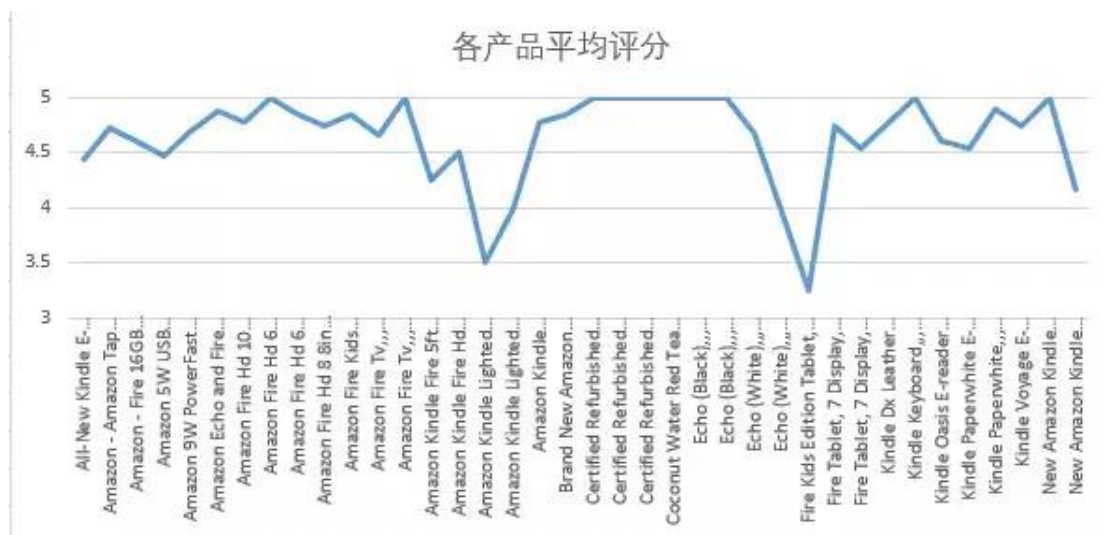




由图上可以看出，产品(Amazon Kindle Paperwhite - eBook reader - 4 GB - 6 monochrome Paperwhite - touchscreen - Wi-Fi - black) 的评论数最多，前 3 名的评论数差别不大，从第 4 名起评论数明显减少，相比第 1 名减少 5 倍多，这说明 Kindle 类产品容易出现聚集效应，绝大部分用户只喜欢特定的两三种产品。

产品的平均评分表示消费者对该产品的评价。

下图为各产品平均评分的折线图：



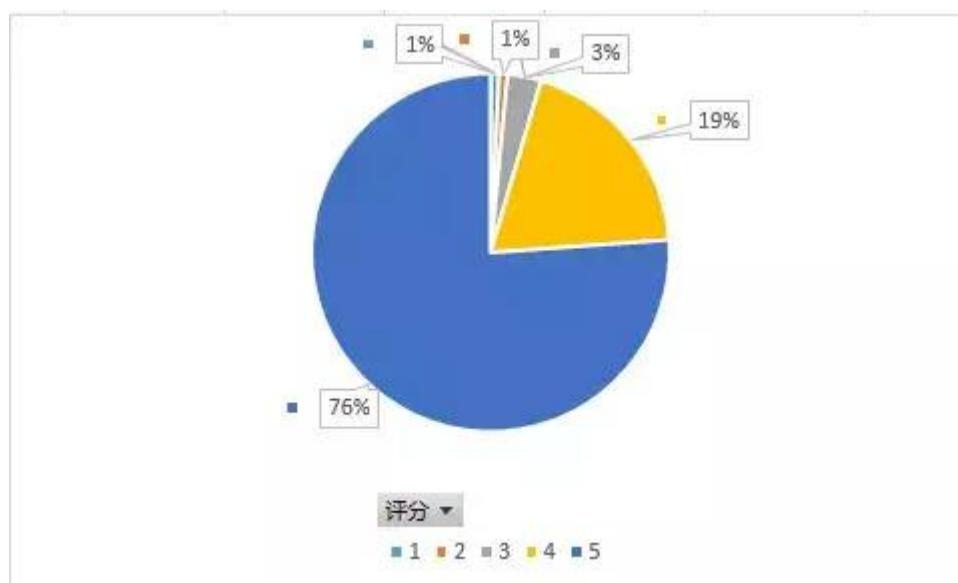
由图上可知，大多数产品的平均评分都在 4.5 分或以上，这说明 Kindle 类产品的评价都较高。但有些产品的评论数很少，考虑是知名度较低，或者相对其他产品有明显的缺点。

## 2、Kindle 相关产品评论总数中各评分的占比？推荐评论中好评数量占比怎样？

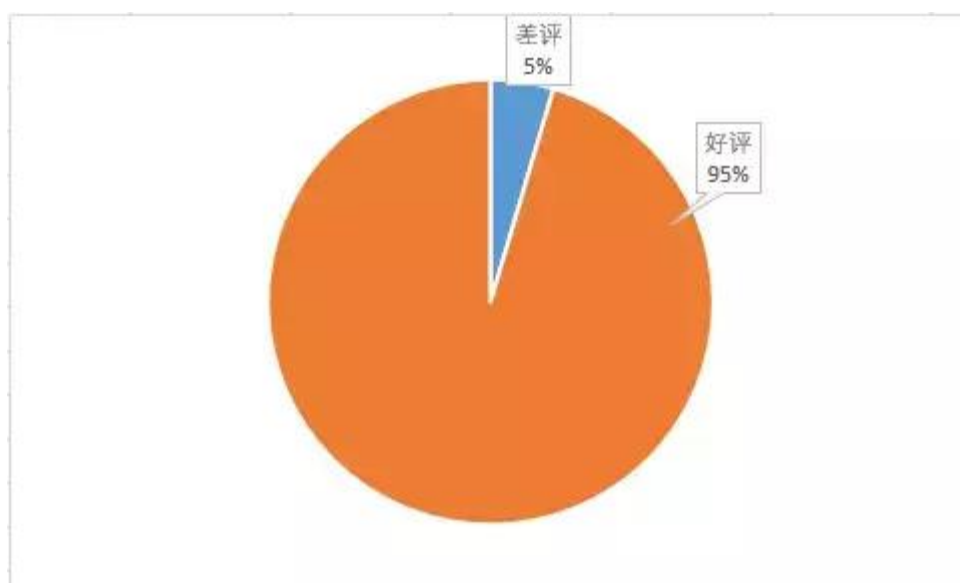
首先，统计评论中每个评分(1~5 分)占有多少数量。使用数据透视表，将评分字段拉入行和值选框，得到如下数据：

行标签	计数项:评分
1	80
2	80
3	318
4	2019
5	7967
总计	10464

插入饼图，了解各评分段的占比：



通过数据透视表对评价等级中统计好评数量与差评的占比，并插入饼图：

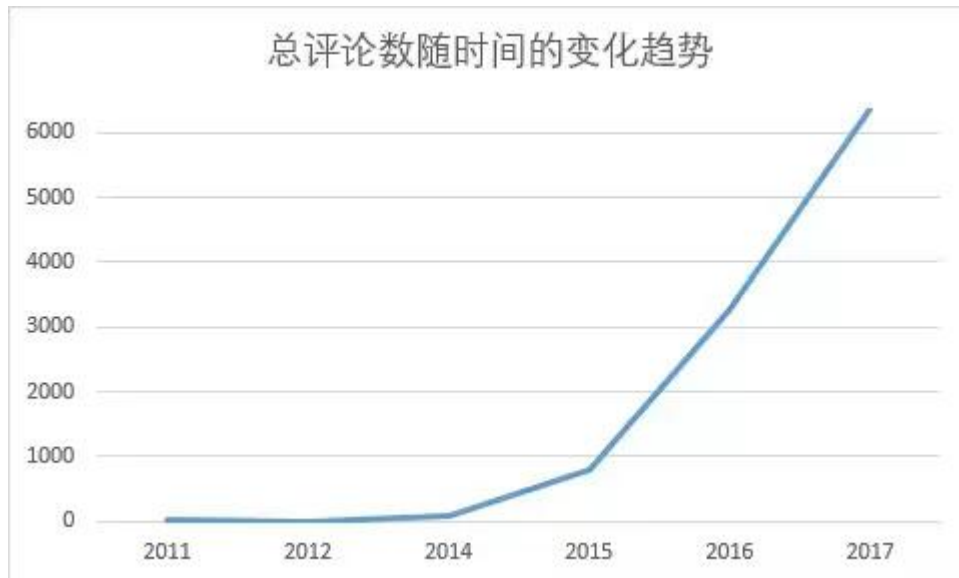


由图上可以看出,评分中 3 分及 3 分以下的占比 5%, 4 分及以上的占比 95%, 由此可见该类好评率高达 95%, 虽然不能保证所有评论评分都值得相信, 但是仍然反映用户对 Kindle 相关产品的使用普遍感到满意。

### 3、Kindle 相关产品随时间变化其评论数的变化情况是怎样的？

使用数据透视表, 将年拉入行标签处, 对评论文字内容进行计数统计, 并插入折线图, 分析总评论数随时间的变化趋势:

行标签 ▼	计数项:评论文字内容
2011	6
2012	1
2014	66
2015	779
2016	3256
2017	6355
<b>总计</b>	<b>10463</b>



使用数据透视表，将年拉入行标签处，将产品名称拉入列标签处，对评论文字内容进行计数统计：、

在以下区域间拖动字段:

<p>▼ 筛选器</p> <div></div>	<p>    列</p> <div>产品名称 ▼</div>
<p>≡ 行</p> <div>年 ▼</div>	<p>Σ 值</p> <div>计数项:评论文... ▼</div>



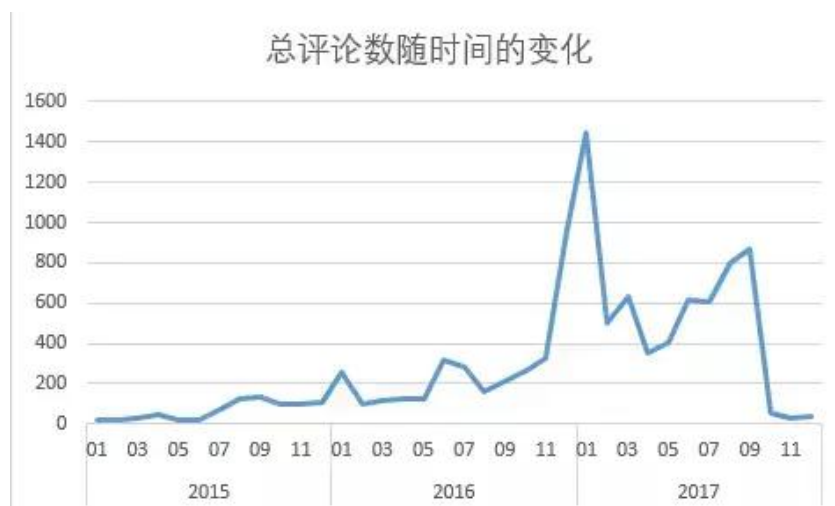
计数项:评	列标													
行标签	All-Ne	Amaz	Amaz	Amaz	Amaz	Amaz	Amaz	Amaz	Amaz	Amaz	Amaz	Amaz	Amaz	Amaz
2011														
2012														
2014														
2015														
2016		113	4	7	36	8	128	1	6	126	6	336		
2017	212	205	1	188						9		2191		
总计	212	318	5	195	36	8	128	1	6	135	6	2527		

插入折线图，分析 Kindle 类各产品评论数随时间的变化趋势：



由图上可知，评论数量是从 2015 年开始快速变化的，从 15 年急剧增加到 6000 多，有关 Kindle 的产品种类也是从 15 年快速上升。

选取 15、16 和 17 年三个主要年份进行逐月分析：





由图上可以看出，评论数量是从 2015 年开始迅速增加，在 2016 年 12 月达到最高峰，用户一年中的消费主要集中在 11 月、12 月和 1 月。

#### 4、Kindle 相关产品广受好评的原因主要是什么？

对评论文字内容字段进行关键字提取，获取产品受欢迎的原因有哪些？

该部分使用到 Python 软件，通过调用第三方工具包，对评论内容生成词云，从而获得关键信息。

Python 代码如下：

```
#导入相关模块
```

```
import numpy as np
```

```
import pandas as pd
```

```
from wordcloud import
```

```
WordCloud,STOPWORDS,ImageColorGenerator
```

```
import matplotlib.pyplot as plt
```

```
from scipy.misc import imread
```

```
# 加载数据
```

```
word =
```

```
pd.read_csv(r'C:\Users\Administrator\Desktop\amazon.csv')
```

```
# for 循环遍历每个单词
```

```
text = ''
```

```
for i in word['评论文字内容']:
```

```
    i = i.lower().replace(".", "")    # lower 将单词改为小写
```

```
    text += i + ' '
```

```
# 生成词云图
```

```
"""
```

```
WordCloud 参数
```

```
collocations: 是否包括两个词的搭配
```

```
width : int (default=400) #输出的画布宽度, 默认为 400 像素
```

```
height : int (default=200) #输出的画布高度, 默认为 200 像素
```

```
background_color: 设置背景颜色
```

```
scale: 按照比例进行放大画布, 如设置为 1.5, 则长和宽都是原来画布的 1.5 倍
```

```
"""
```





## 五、总结月建议

---

### 1、总结

Kindle 及周边各产品评分比较高，但评论数排名前 3 和排名第 4 名起其数目差距比较大，大部分产品只有少数评论数，绝大部分用户只喜欢特定的两三种产品。

Kindle 及周边商品深受消费者喜爱，好评度很高。

Kindle 及周边产品在每年 1 月、11 月和 12 月评论数骤增，应加大销售手段，刺激消费增长销售额。

Kindle 及周边产品广受好评的主要原因是：其产品特点轻便小巧、电池寿命长、运行流畅、价格实惠、封面漂亮。

### 2、建议

选择最受欢迎的三种产品作为主要商品，分别是

- Amazon Kindle Paperwhite - eBook reader - 4 GB - 6 monochrome Paperwhite - touchscreen - Wi-Fi – black
- Echo (White)
- Amazon Fire Tv

利用商品使用、性能及价格的优势作为主要宣传手段。

在黄金时期(11、12、1 月)，加大宣传力度，刺激产品销售，提高销售量。