

淘宝网手机类商品 数据分析报告

报告人：瞿晨
日期：2018/03/29

目录

- 一、 分析背景与目的3
- 二、 数据来源3
 - 2.1 数据获取3
 - 2.2 数据预处理.....3
- 三、 图表分析4
 - 3.1 数据分布式探索4
 - 3.2 品牌价值分析5
 - 3.3 手机型号价值分析.....8
 - 3.4 品牌市场分析9
 - 3.5 手机主要参数分析.....10
 - 3.6 手机特点/主打卖点词云13
- 四、 结论与不足.....13
 - 4.1 结论13
 - 4.2 不足14
- 五、 附录14
 - 5.1 分析工具及方法14
 - 5.2 原始数据汇总.....14

一、 分析背景与目的

手机已经逐渐成为人们生活中不可缺少的物品，一个人出门什么都可以不带，唯独需要的手机和钥匙，而相信为了智能家居的普及后，每个人出门仅仅需要一部手机。而随着人们对手机的需求的提高，不同手机厂商也纷纷推出产品抢占市场。

本篇报告旨在分析当今手机市场，对不同价格、不同品牌、不同手机性能等相关数据进行汇总，通过图表将其更加清晰的展现出来。

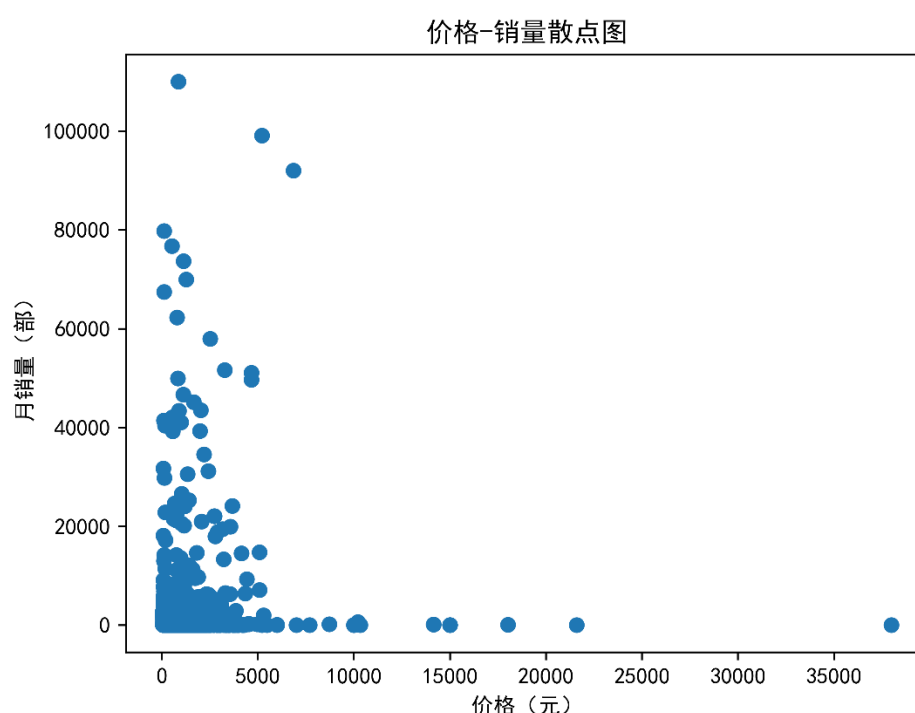
二、 数据来源

2.1 数据获取

本次数据全部获取与淘宝网，使用 Python Urllib 库编写爬虫程序，获取淘宝网中手机类商品的品牌、价格、月销量、商家数及手机本身相关参数信息，爬取手机分类全部共 73 页信息，原始数据 3475 项。

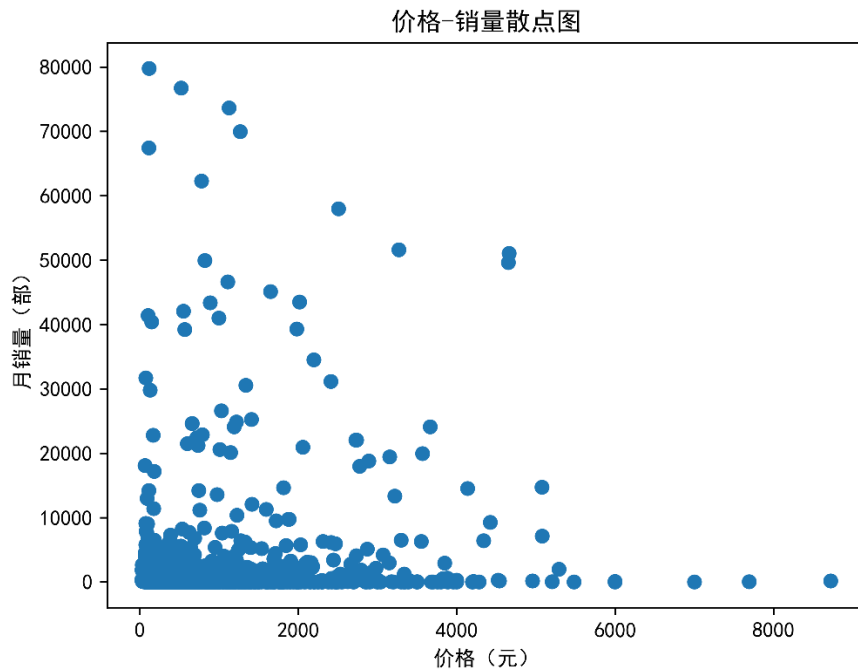
2.2 数据预处理

因代理服务器与爬虫程序的不完善，原始数据中有部分未获取或为 0 的价格和月销量数据，利用 Python 编程清除缺失值 713 条，去除缺失值后数据为 2762 条。



以价格为 X 轴，月销量为 Y 轴，绘制价格-销量散点图，可以看到月销量集中在 80000 部以内，价格在集中于 9000 元以下，因此对范围外的数据进行处理。

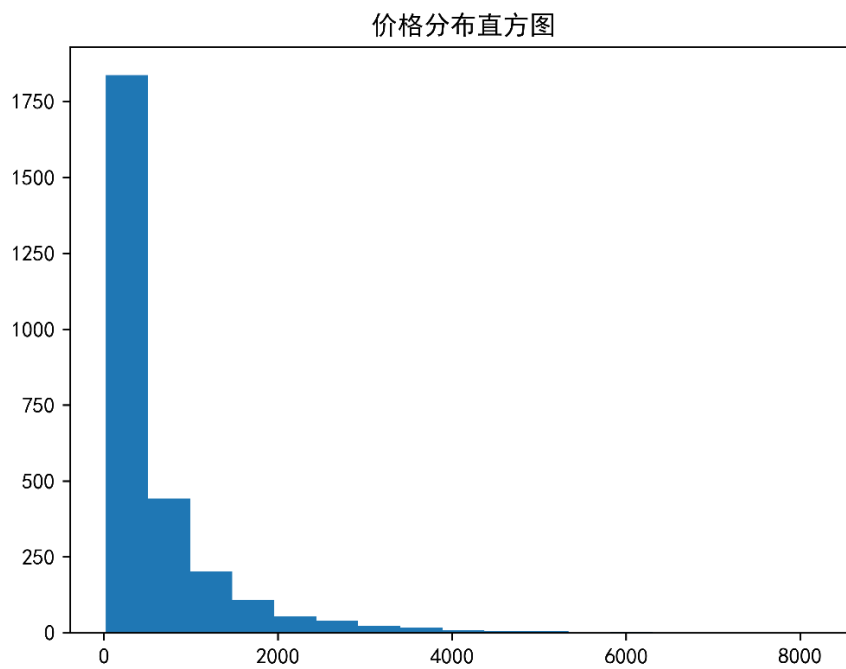
删除价格高于 9000 元及月销量高于 80000 部的数据，得到如下新的价格-销量散点图。

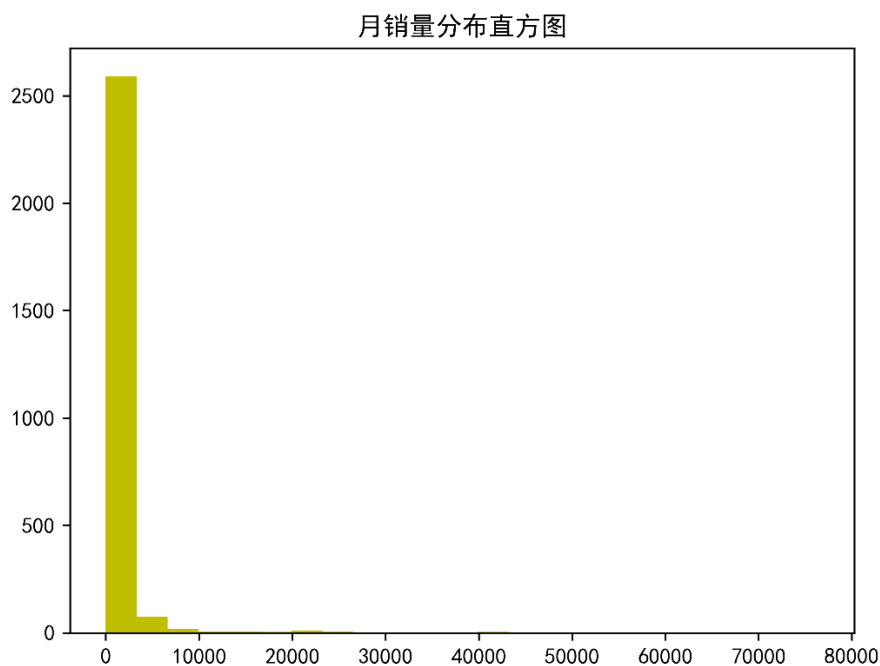


通过处理异常值，得到的新散点中可以看出，数据显示，月销量更集中在低于 10000 部，而价格更集中与低于 4000 元。

三、 图表分析

3.1 数据分布式探索

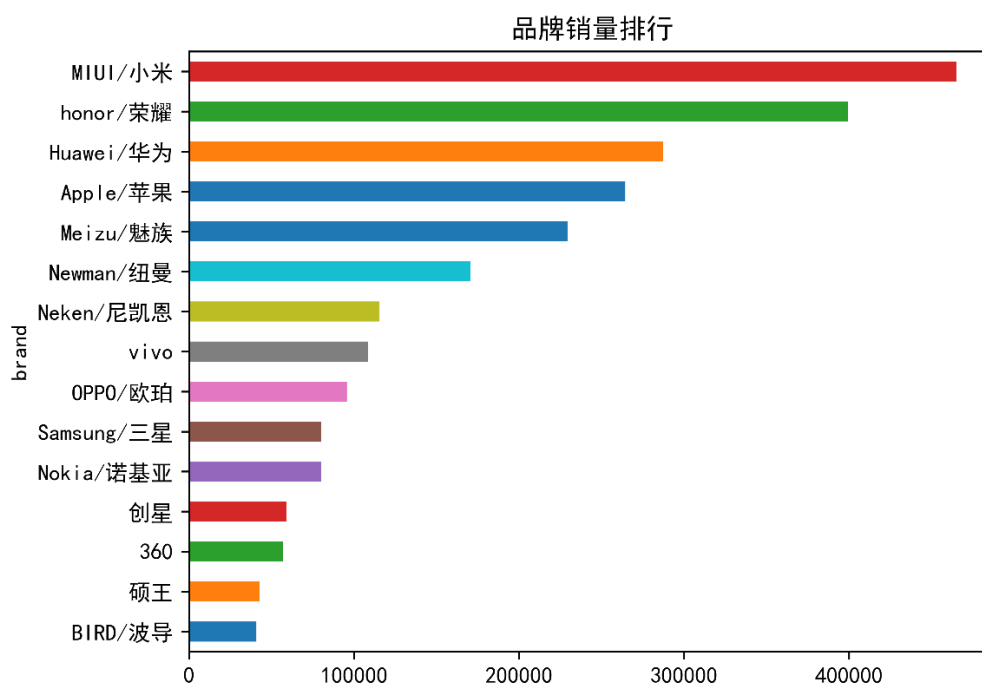




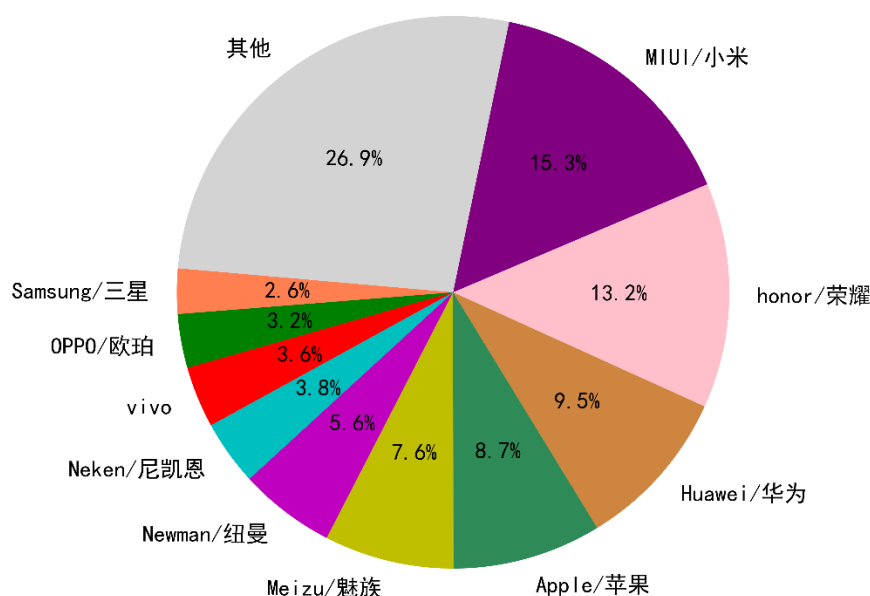
价格分布直方图与月销量分布直方图显示，价格与月销量均呈现长尾分布，这与长尾效应的描述一致，大多数的需求会集中于头部，也就是说相对来说人们接收的，市场流行的更多是不算高的销量和较低的手机价格。

3.2 品牌价值分析

统计不同品牌的销量总数、品牌型号数与销售商数总数进行排序，绘制如下品牌销量排行图、品牌型号数排行图与品牌商家数排行图，同时展示品牌的市场份额饼图。

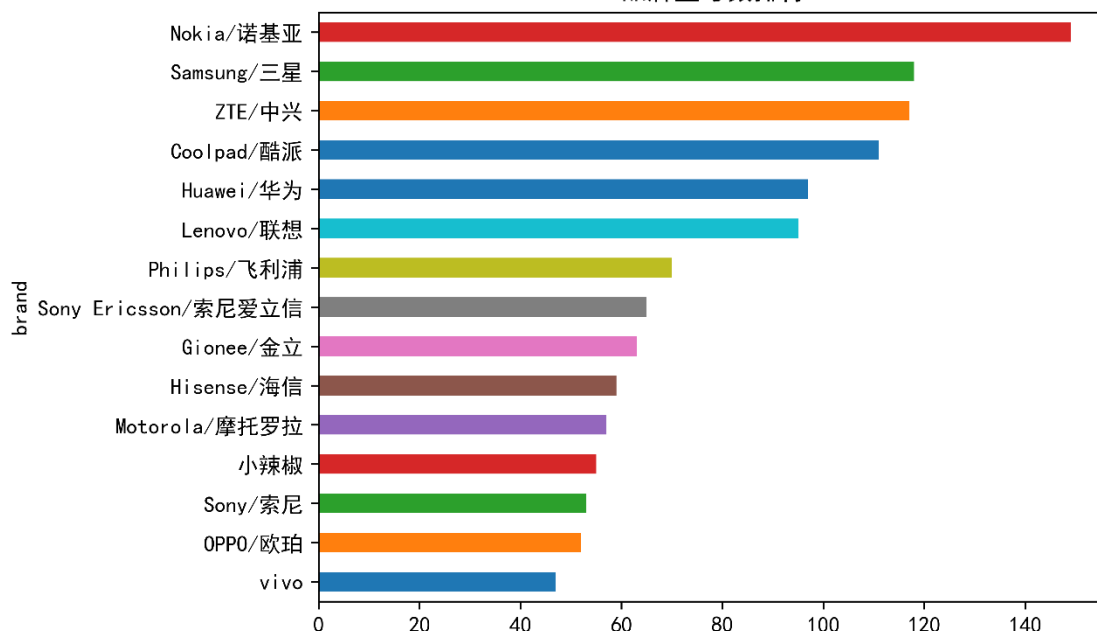


十大品牌的市场占比



若将荣耀（华为的线上品牌）并入华为中，销量排行的前三名分别是华为、小米和苹果，这是符合预期的，这也表示人们在选购手机时已经更加倾向于购买国产手机，而非一味崇拜苹果、三星等国际大牌。然而 vivo、OPPO 以及三星虽然看起来粉丝众多，广告铺天盖地，但是实际销量却仅仅刚进入前十。

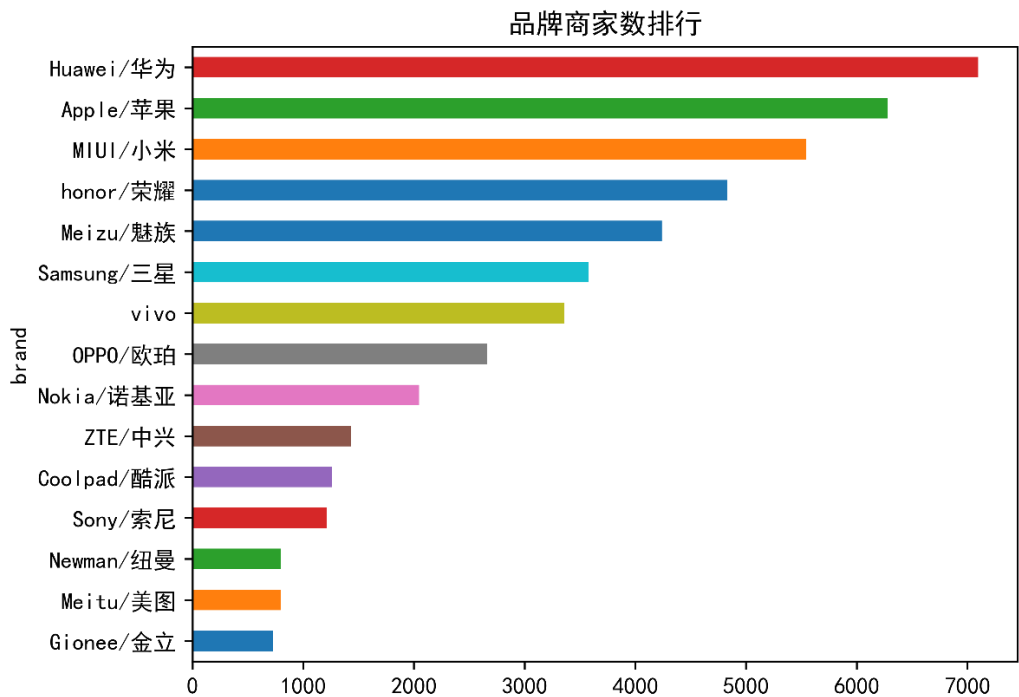
品牌型号数排行



而从品牌型号数排行中，型号最多的是老牌手机诺基亚，诺基亚高达 140+ 的型号总数更多的是安卓系统飞速发展之前，借助没有对手的塞班系统，诺基亚推出了无数手机，也获得了极高的销量，可惜在智能手机飞速发展的今天已经有些没落。

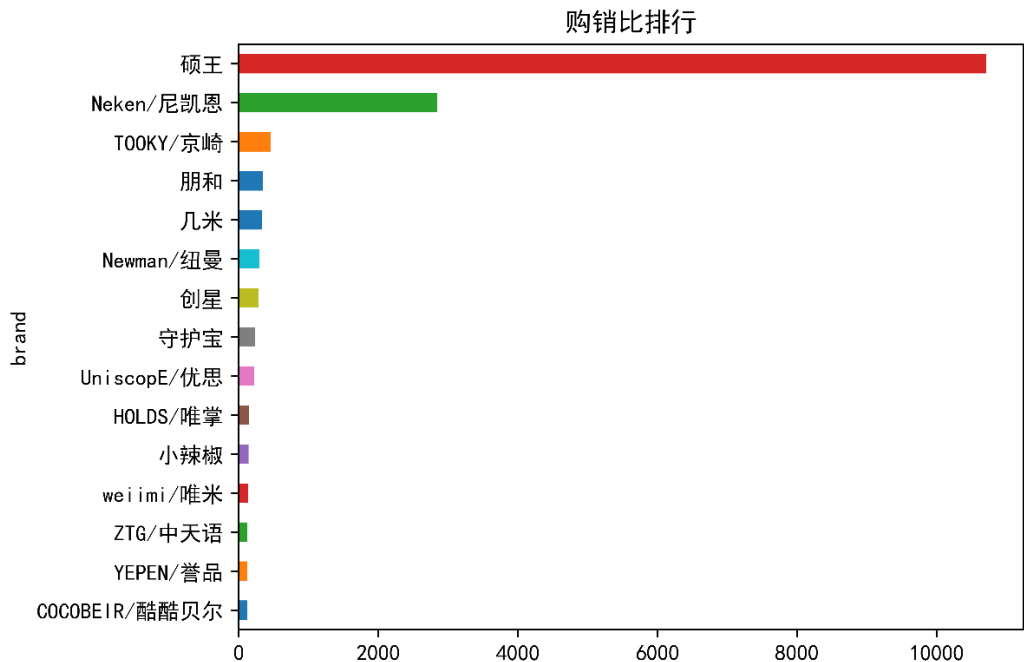
而除了诺基亚之外，三星、中兴、酷派、华为和联想以相差不多的数量领先于其他手机品牌。销量极高的小米和苹果生产线则相对较少，依靠不多的产品获

得了领先的销量和销售商。

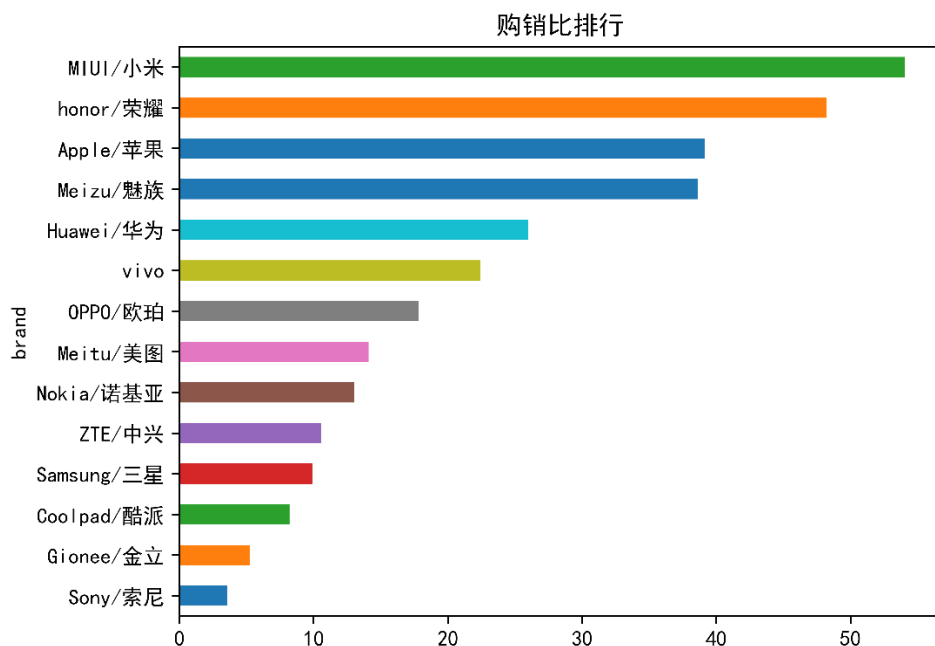


相比于品牌销量来说，销售商的数量也整体相似，仅仅小范围内名次有轻微变动，这一方面可能有销售商数提高带来销量的影响，但更多应当是由于相关品牌的大卖导致销售商随之增加。

因此，构建新参数购销比=品牌销量/品牌商家，来观察不同品牌的销售转换率情况，同样统计前五名如下图。



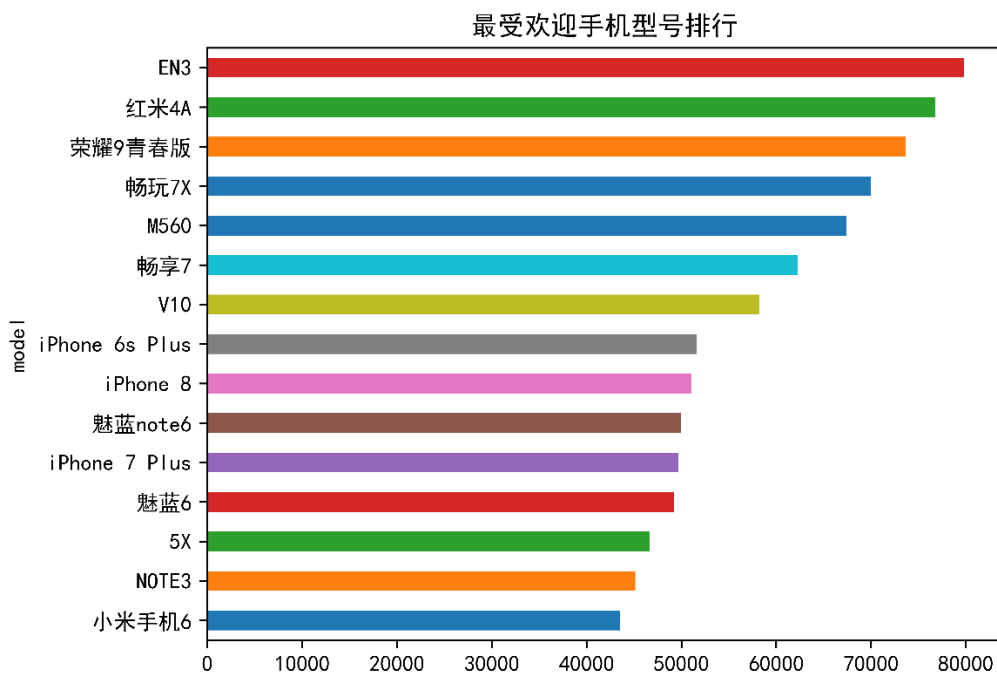
由于店铺数量较少的缘故，排行榜前面几乎完全被一些小众品牌货老人机占据，表中未体现的小米也仅仅进入前三十名，因此去掉小众品牌和老人机品牌，我只对比销量和店铺数量较多的流行品牌，处理后如下图。



上图中可以看到，前几名几乎与销量和销售商数排行几乎没有太多的变化，而图中我们可以看到三星这个安卓手机中的招牌，平均店铺的销量已经与落后于大部队。

3.3 手机型号价值分析

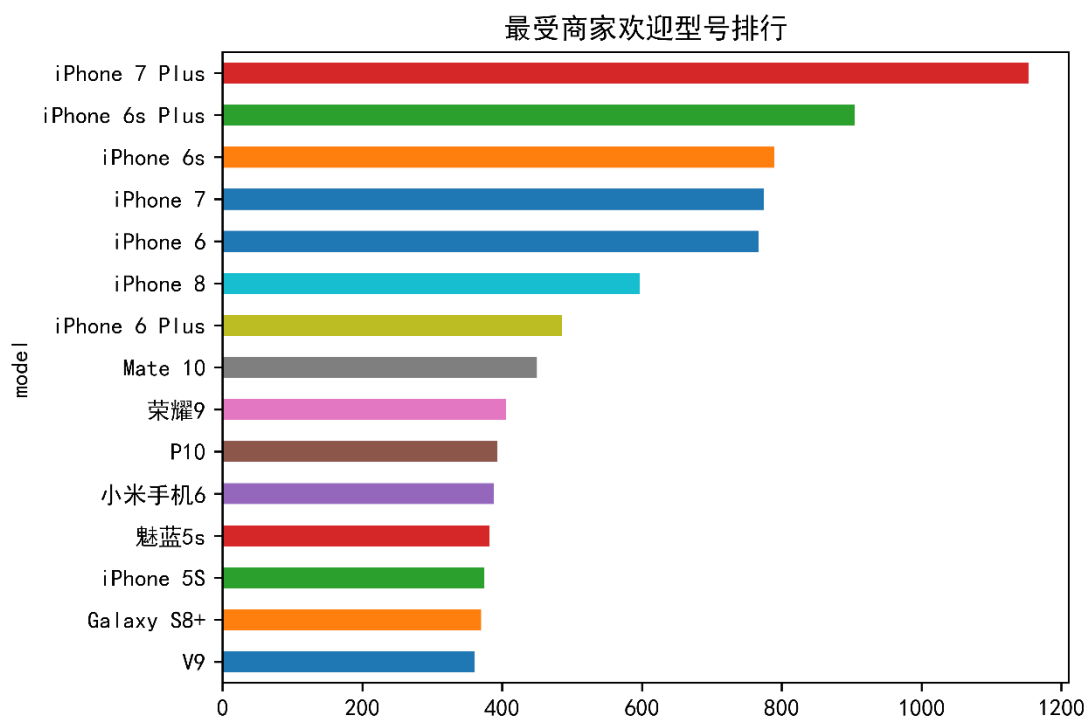
相较于对品牌价值的分析，本节将研究对象具体到每一部具体手机型号，统计最受消费者欢迎和最受商家青睐的手机型号，并对其排序。



上图展示的最受消费者欢迎的 15 个手机型号，其中 EN3 是一部老人机，就是它。



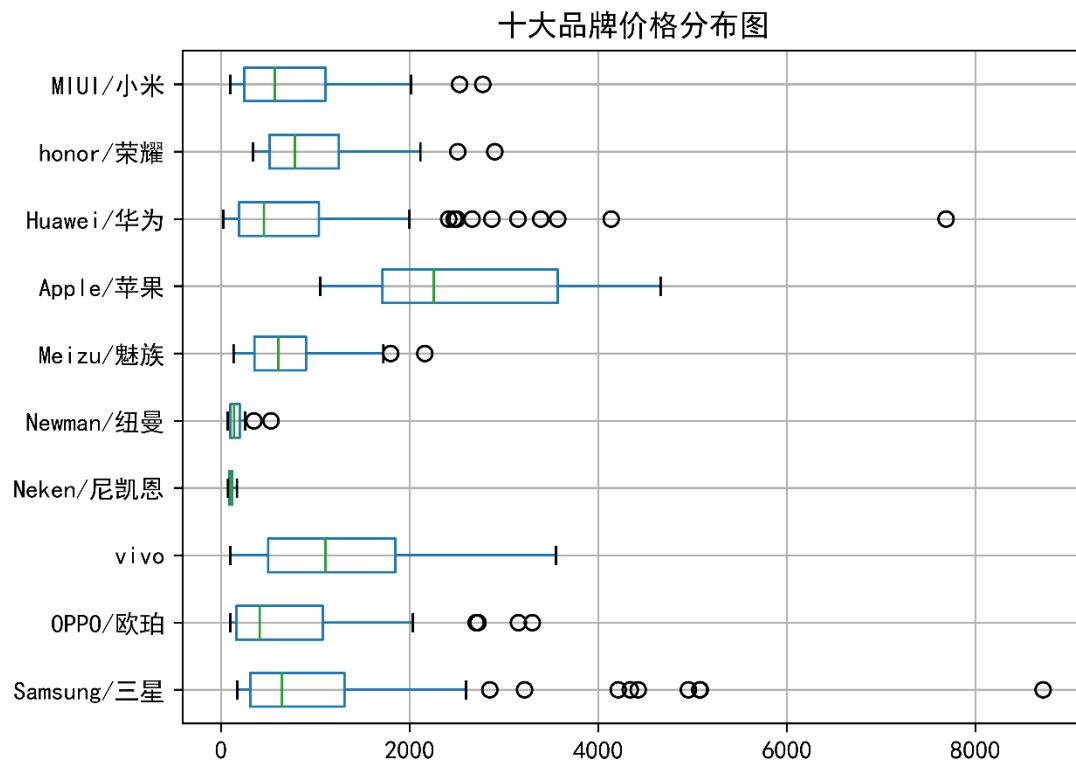
除此以外，小米、华为和苹果的手机占据了大部分江山，而且除了苹果手机以外，这些高销量的手机型号多为千元级别，可以看出消费者相比于昂贵的旗舰机，更倾向于性价比高的低端手机。



相比于销量的平分天下，在销售商的对比上，苹果让其他品牌望尘莫及，最受欢迎的十五个手机型号中有 8 个是苹果的产品，前十中更是高达 7 个，其余产品则是分别华为、小米和三星三个品牌。

3.4 品牌市场分析

为了了解这些畅销流行品牌的分别进军的市场以及其所占据的市场份额，绘制十大流行品牌的价格分布箱型图和各品牌销售量占全部市场的份额占比图。



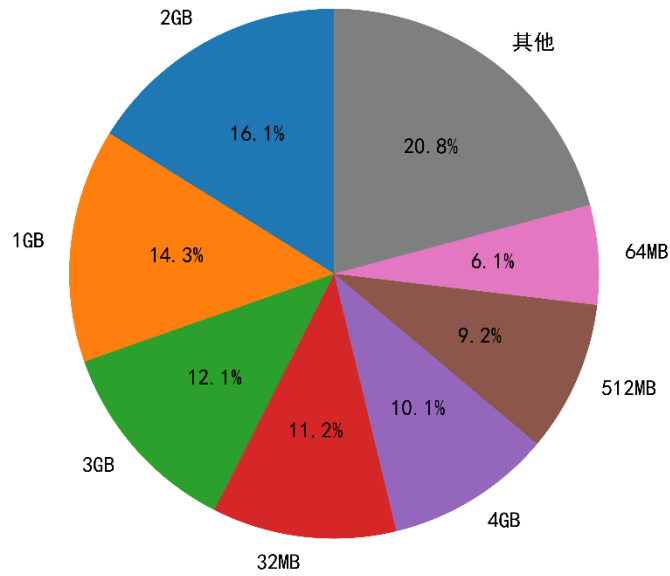
从上图可以看到，其实除了苹果以为各大手机品牌其实从未放弃过低端机市场，而从之前对市场价格和最受消费者欢迎的品牌分析中也可以看出，大品牌对市场情况的把握是十分准确的。

这张数据图中令我没有想到的一个是 vivo 的产品跨度，本以为是一个走中端机市场的品牌，其实横跨低中高市场，第二个则是三星，图中可以看出三星除了每年的两部旗舰机，更多的目光其实的放在了低端机市场。

3.5 手机主要参数分析

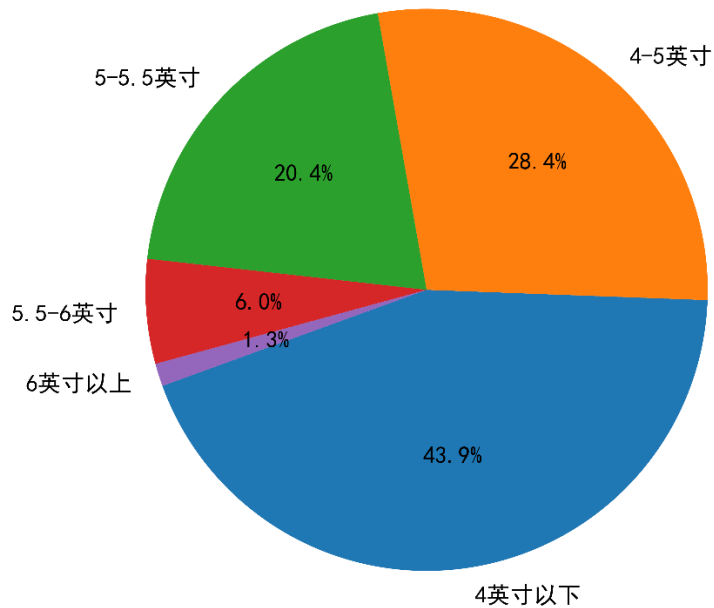
以手机型号数量为依据，对不同内存、不同屏幕尺寸、不同摄像头像素和不同核心数的手机绘制饼图，分析不同参数所占据的市场。

不同内存手机的市场占比



可以看到，虽然现在各大手机品牌开始逐渐开发大内存手机，但目前的市场2GB、1GB 的手机数量仍然占据领先，但很显然地，未来市场更多的会是更大内存的手机占据主流。

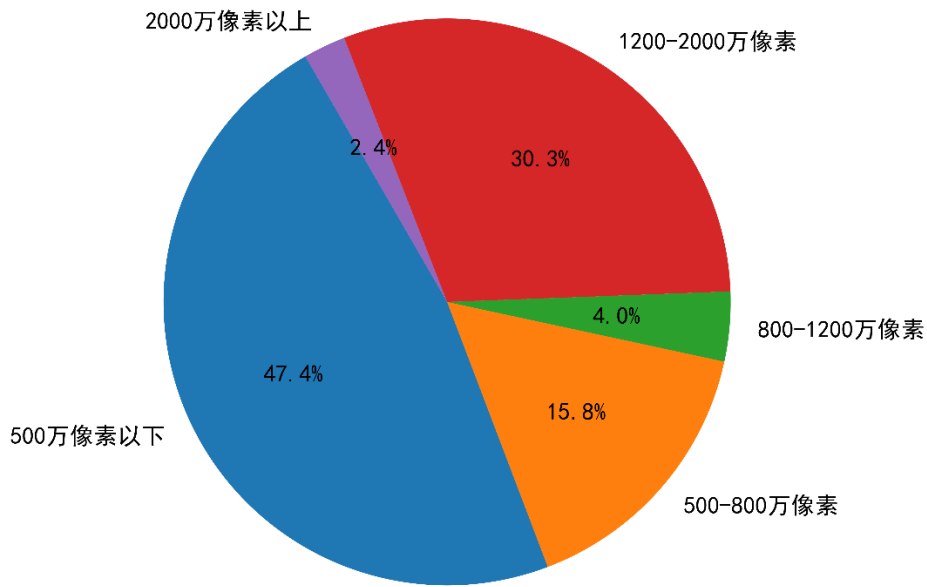
不同尺寸手机的市场占比



和内存一样，一方面由于老人机的数量较多，更一方面之前的小尺寸屏幕也依然未退出市场，4 英寸屏幕以下的手机仍是多数。

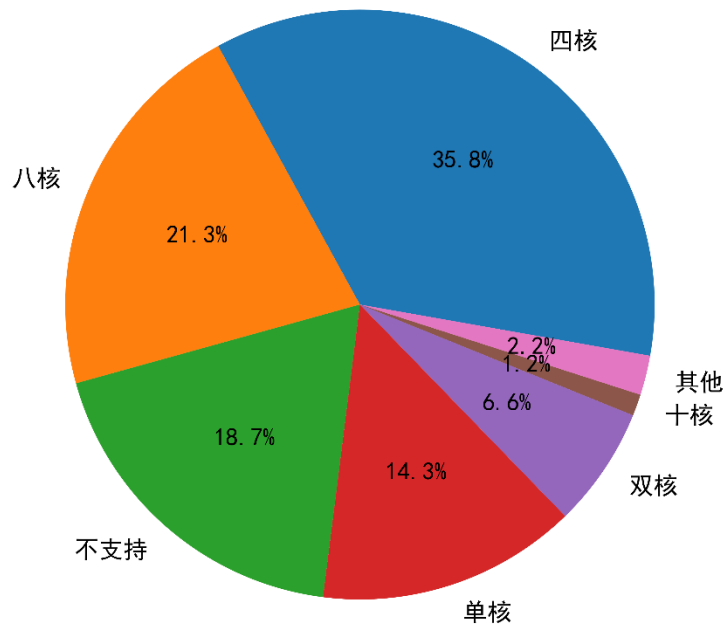
另外，市场的主流仍是 4-5.5 英寸屏幕的手机，5.5-6 英寸的数量并不多，6 英寸以上的超大屏手机更是仅仅占据很少的一部分。

不同相机像素手机的市场占比



抛开未退出市场的古董手机和老人机，我们可以看到市场的主流更多是摄像头 1200-2000 万像素的手机，摄像头低于 1200 万像素和高于 2000 万像素的手机数量同样并不多，特别是 2000 万像素以上的数量仅为全部市场的 2.4%，但不可否认未来市场手机的摄像头像素一定会越来越高。

不同核心数手机的市场占比



在手机核心数的统计中，四核手机与八核手机已经逐渐占据了绝大多数的市场，手机厂商认为较为低端的单/双核已经无法满足消费者，单高于八核的十核手机也数量极少，而这更多的是需要技术的发展来推动。

3.6 手机特点/主打卖点词云

统计手机的特点信息，利用 jieba 进行分词，使用 wordcloud 库绘制词云，如下图。

可以看到手机厂商们主打的卖点更多的是全网、双卡双待、智能和超长待机等。



四、 结论与不足

4.1 结论

1. 淘宝网手机商品的价格集中在 4000 元以下，月销量集中在 10000 部以下，同时市场上大多数手机的销量和价格是不高的；
2. 华为、小米和苹果三个品牌占据销量榜前三，OV 和三星虽然宣传很多但是实际销量并没有很高；
3. 手机型号最多的品牌是老牌厂商诺基亚，而销量前十中只有三星和华为的手机型号较多；
4. 购销比排行中小米和荣耀是领先于其他品牌的，如果要开一家网店，除了老人机，小米和荣耀是最正确的选择，非常不推荐三星；
5. 大部分的店家非常钟爱苹果，销售商数前五的手机型号中有 8 个是 iPhone，前十中更是高达 7 个；
6. 除苹果以为销量前十的品牌几乎将主要市场放在了中低档手机，只有 vivo 的产品线价格要领先于其他品牌；
7. 目前市场的主流仍然是 2GB/1GB 内存、5 英寸以下屏幕、500 万像素和 1200-2000 万摄像头像素和四核/八核的手机；
8. 手机厂商的主打卖点更多是全网、双卡双待、智能以及超长待机。

4.2 不足

1. 本篇报告数据获取自淘宝网，由于手机市场线上出来淘宝以为，依然还有大量其他平台，同时手机市场还有线下渠道，因此数据具有局限性；
2. 本篇报告是笔者学习 Python 数据分析之后，所进行的第二次数据分析尝试，报告中参考了互联网中一些他人的分析思路，不足之处敬请谅解。

五、 附录

5.1 分析工具及方法

1. 数据获取——Python urllib 库爬虫、MySQL 数据库
2. 数据清洗与预处理——Excel、Python
3. 数据整理——Python pandas
4. 数据可视化——Python matplotlib、wordcloud

5.2 原始数据汇总



原始数据汇总.xlsx