# 《手机参数采集分析系统》需求分析

# 1、任务概述

**1.1、目标**

**①数据采集 服务器端定期更新数据库，利用爬虫获取中关村在线各种品牌近五年的手机的价格和各种参数，比如CPU、前置摄像头像素、后置摄像头像素、内存、续航、屏幕尺寸、分辨率等等。**

**②数据存储 将采集到的数据存储到数据库中**

**③数据分析 对各种参数进行数据分析，做出饼状图直方图等用多种形式呈现出来。**

****

****

**1.2、系统（或用户）的特点**

**用户群面向有购买手机或了解手机行业参数信息意图的人群，通过参考软件提供的各种信息，能买到最符合自己需求、对于自己性价比最高的一款。由于利用了python爬虫，本软件包含的数据将较为全面，统计出的结果将会较为精确，实用性强。数据分析将会呈现出手机行业各参数的大致分布和变化趋势以便用户进行参考。与此同时，很多手机广告营销做得非常精致，但是金玉其外败絮其中，通过本软件的分析，用户可以清楚的看到并排除这类招摇撞骗的产品。**

# 2、功能说明

# 1.数据采集 （服务器端）

**服务器端定期（每隔6h）利用爬虫获取手机各种参数如下：**

**手机名（比如中兴）、上市时间、屏幕尺寸、商家指导价、运行内存、存储内存、内核数、CPU频率、屏幕尺寸，屏占比、主屏分辨率、像素（前置摄像头、后置摄像头）、电池容量，重量、用户评分**

**由于手机行业更新换代较为频繁，新产品的出现伴随着旧产品的淘汰，因此对于大众需求不高的旧产品不在我们的统计分析范围内。于是我们对采集的信息进行筛选，仅考虑上市时间为近五年之内的产品。**

**2.数据统计分析（服务器端）**

**通过饼图和直方图了解手机行业内存等数据的大致分布，通过折线图了解某项手机参数的变化趋势，便于加深用户对手机行业的宏观理解，在手机选购上做出更准确的判断和选择。**

1. **饼图：**

**饼图针对较为离散的数据，显示各种类型所占的百分比，便于用户了解行业中该手机参数的大致分布。以饼图方式呈现的数据如下：**

**1.运行内存**

包括2G,4G,6G,8G,10G,12G,显示对应百分比

**2.存储内存**

包括32G,64G,128G,256G,512G,显示对应百分比

**3.屏幕类型**

包括非全面屏，刘海屏，水滴屏，极点屏，常规全面屏，显示对应百分比

**4.解锁方式**

包括面部识别，虹膜识别，屏幕指纹，前置指纹，后置指纹，显示对应百分比

**5.主屏分辨率**

包括4K 2K 1080p 720p，显示对应百分比

1. **直方图 （包括价格，主屏尺寸 屏占比 质量 电池容量（续航） 像素 主屏分辨率**

**直方图针对不太离散的数据，由于数据类型较多，不适合饼图统计，于是按照一定区间分段统计，这样便于用户了解行业中该手机参数的大致分布。以直方图方式呈现的数据如下：**

**1.价格**

按照500元以下 500-1000元 [1000-1500元](http://detail.zol.com.cn/cell_phone_index/subcate57_0_list_1000_s8047_1_2_0_1.html) [1500-2000元](http://detail.zol.com.cn/cell_phone_index/subcate57_0_list_1500_s8047_1_2_0_1.html) [2000-3000元](http://detail.zol.com.cn/cell_phone_index/subcate57_0_list_2000_s8047_1_2_0_1.html) [3000-4000元](http://detail.zol.com.cn/cell_phone_index/subcate57_0_list_3000_s8047_1_2_0_1.html) [4000-5000元](http://detail.zol.com.cn/cell_phone_index/subcate57_0_list_4000_s8047_1_2_0_1.html) [5000元以上](http://detail.zol.com.cn/cell_phone_index/subcate57_0_list_5000_s8047_1_2_0_1.html)进行划分，统计频数并构造直方图

**2.主屏尺寸**

按照[6.5英寸以上](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s5371-s8047_1.html) [6.1-6.4英寸](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s7947-s8047_1.html) [5.6-6.0英寸](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s6258-s8047_1.html) 5.5英寸 5.1-5.4英寸 5.0英寸 [4.5-4.9英寸](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s4218-s8047_1.html) 4.4英寸以下进行划分，统计频数并构造直方图

**3.屏占比**

根据具体情况进行划分，统计频数并构造直方图

**4.质量**

根据具体情况进行划分，统计频数并构造直方图

**5.电池容量**

按照1500mmAh以下，1500-2000mmAh，2000-2500mmAh，2500-3000mmAh，3000-3500mmAh，3500-4000mmAh，4000-4500mmAh，4500-5000mmAh，5000mmAh以上进行划分，统计频数并构造直方图

**6.前置摄像头像素**

按照[2000万像素以上](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s8047-s7348_1.html) [1600万像素](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s8047-s7349_1.html) [1300万像素](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s8047-s7321_1.html) [800万像素](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s8047-s5747_1.html) 500万像素 400万像素以下进行划分，统计频数并构造直方图

**7.后置摄像头像素**

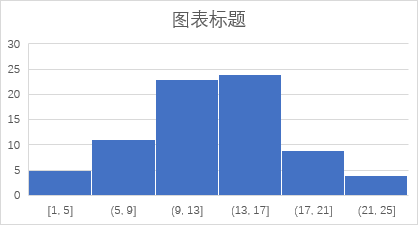
按照[2000万像素以上](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s8047-s6270_1.html) [1600万像素](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s8047-s5536_1.html) 1300万像素 [1200万像素](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s8047-s7572_1.html) 800万像素 700万像素以下进行划分，统计频数并构造直方图

**8.屏幕像素密度**

按照[500ppi以上](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s6192_1.html) [400ppi至500ppi](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s4691_1.html) [350ppi至399ppi](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s4690_1.html) [300ppi至349ppi](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s4689_1.html) [300ppi以下](http://detail.zol.com.cn/cell_phone_index/subcate57_list_s4688_1.html)进行划分，统计频数并构造直方图

**9.窄边框**

按照2-3mm 3-4mm 4-5mm 5-6mm 6mm以上进行划分，统计频数并构造直方图

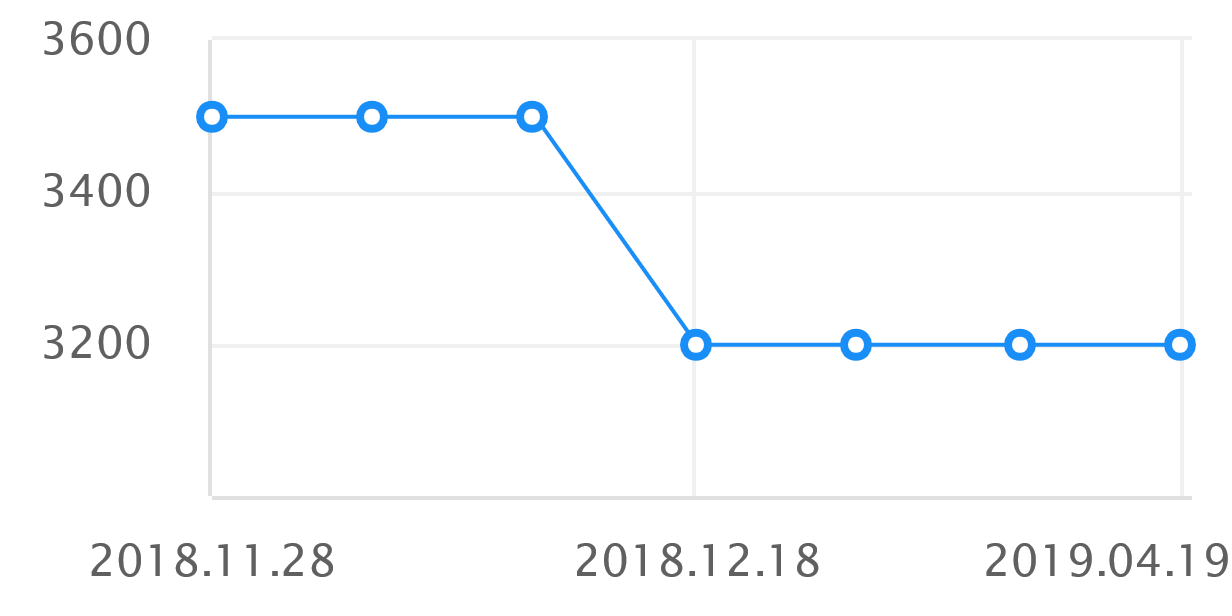
****

1. **折线图（内存，价格等）**

**通过折线图，以时间（或上市时间）作为横轴，展示某些手机参数的变化趋势，计该月内上市手机的某项参数的均值，并用折线图展示出来，方便用户了解该数据的变化趋势**

**1. 价格**

显示近期同款手机价格走势，实例如下图所示



**2.运行内存**

**显示一段时间内，不同类型的运行内存（**包括2G,4G,6G,8G,10G,12G）**的占比走势（共六条折线），构造折线图**

**3.存储内存**

**显示一段时间内，不同类型的存储内存（**包括32G,64G,128G,256G,512G）**的占比走势（共六条折线），构造折线图**

**4.屏幕类型**

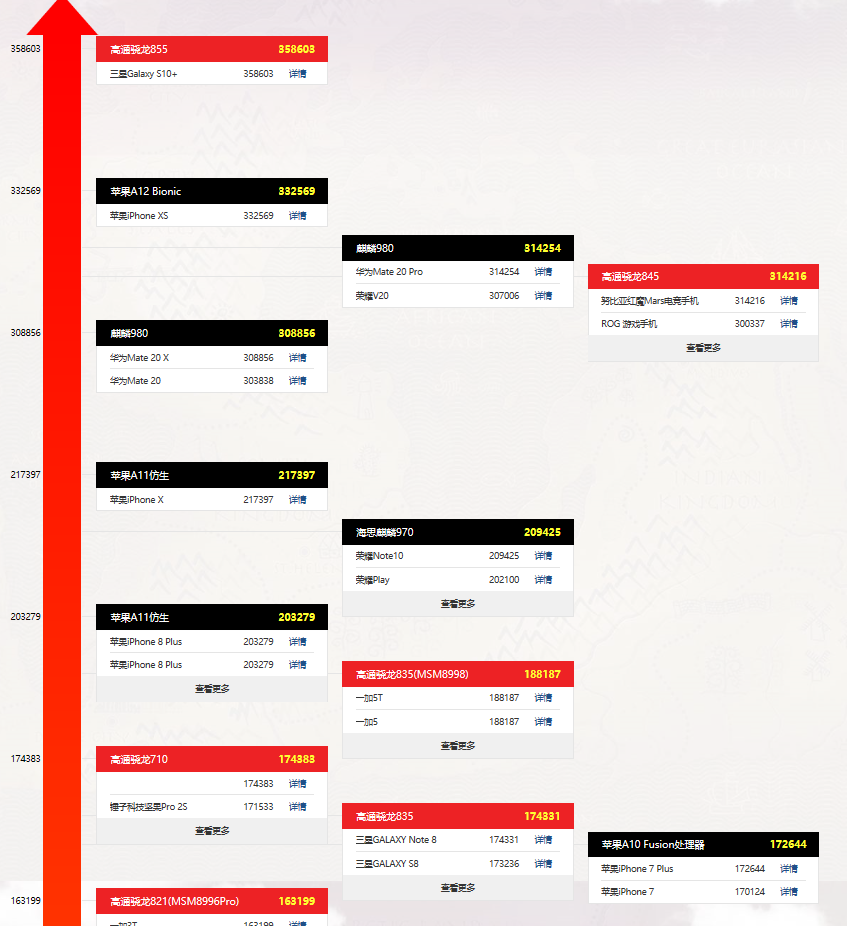
**显示一段时间内，不同类型的屏幕类型（**包括非全面屏，刘海屏，水滴屏，极点屏，常规全面屏）**的占比走势（共五条折线），构造折线图**

**5.解锁方式**

**显示一段时间内，不同类型的解锁方式（**包括面部识别，虹膜识别，屏幕指纹，前置指纹，后置指纹）**的占比走势（共五条折线），构造折线图**

1. **天梯图**

**手机处理器不宜用饼图直方图折线图的方式呈现，于是我们提供对于处理器性能评分的天梯图供用户参考，同时我们还提供中关村在线上评测的各品牌手机的热度，性能排行榜。**

****

****

**其中排行榜功能包括手机总排行，热门手机排行，上升最快手机排行以及按品牌的排行榜（OPPO、VIVO、华为、三星、苹果、荣耀、小米、魅族、诺基亚），如下图所示：**



# 3、技术实现

**基于python和数据库mongodb or mysql**

**1.数据采集 request请求网页（selenium），beautifulsoup or lxml解析网页**

**，之后将获取的数据写入到数据库中。**

**2.数据分析 numpy scipy pandas 对获得的数据进行一些必要的处理。**

**3.数据可视化 matplotlib or seaborn 通过图表可视化数据，更为直观清晰。**

# 4、备选功能\*

**性价比评估**

**该评估方式基于经济学里的特征价格法。将手机的价格分解，以显现出其各项特征的隐含价格，在保持手机的特征不变的情况下，将手机价格变动中的特征因素分解，从价格的总变动中逐项剔除特征变动的影响，剩下的便是纯粹由供求关系引起的价格变动。**

**因为手机的价格很大程度上是由手机的各项参数（cpu，rom，ram，像素，分辨率等）决定的。我们以手机各项性能参数作为特征，以价格作为输出，通过机器学习算法（linear regression）计算出手机的价格模型(hypothesis).这样对于一个新的样本，通过hypothesis计算出模型价格，和该手机的实际价格对比，就可以评估该手机的性价比。**