

2018 年"泰迪杯"数据分析职业技能大赛

A 题

电商平台手机销售数据采集与分析

一、背景

近年来随着手机行业的竞争日趋激烈，手机厂商也越来越关注用户体验。但是，关于手机的数据（如销售数据、评价数据等）并不被手机厂商直接掌握，而是分布在不同的电商平台中。因此，从电商平台上采集和分析相关数据成为手机厂商了解用户需求痛点、掌握经营方向的重要手段，也对优化用户体验、推动手机行业的发展有着非常重要的意义。

二、目标

1. 从目标网站采集数据，并对数据进行预处理，转化为便于分析的结构化数据。
2. 依据这些数据，对商品销售及市场需求进行分析，并将分析结果进行可视化展示。

三、任务

选择合适的方式从目标网站上采集手机销售的相关数据，分析价格、用户评价等因素对手机销量带来的影响，并对手机生产商或电商商家的经营决策给出建议（例如：定价、手机外观、手机功能等）。采集软件可以使用八爪鱼、可视化工具可以使用微图。

任务 1 数据采集与预处理

任务 1.1 选择合适的软件导入目标网站，或者使用 python 查看网站源代码。目标网站地址：<http://139.196.113.162:8090/>（建议使用谷歌浏览器）。
请保存任务截图。

任务 1.2 根据任务 2、任务 3 的需求确定所需采集的字段。

任务 1.3 确定、配置采集规则或编写爬虫代码。请保存规则截图或爬虫代码（代码请以 txt 格式保存）。

任务 1.4 抓取数据。请保存数据采集过程的截图。

任务 1.5 保存数据。请以 excel 格式保存采集到的数据。

任务 1.6 数据预处理（探查数据质量并进行必要的预处理，如：数据清洗、去噪等）。请以 excel 格式保存处理后的数据。

任务 2 数据可视化

任务 2.1 将任务 1 的结果数据导入相关可视化工具。请保存任务截图。

任务 2.2 选择合适的图形从不同角度展示不同品牌的商品销量情况。请保存图形并在任务四的报告中呈现、分析。

注：将累计评论数视为商品销量，下同。

任务 2.3 分析销量与价格之间的相关性，选择合适的图形绘制两者之间的关系。请保存图形并在报告中呈现、分析。

任务 2.4 选择两到三个品牌，根据客户评论内容文本，分析客户对商品的评价的情感偏向，并选择合适的图形绘制情感倾向图。请保存图形并在任务四的报告中呈现、分析。

任务 2.5 选择两到三个品牌，根据客户评论内容文本，提取用户典型意见，并选择合适的图形展示出来。请保存图形并在任务四的报告中呈现、分析。

任务 3 购买推荐

小明平时喜欢用手机看视频学习，也喜欢用手机听歌玩游戏。但是他的手机最近在打球的时候不小心摔坏了，小明想要重新购买一款手机。请你依据所抓取的数据，帮小明推荐一款手机并给出推荐的理由。请在任务四的报告中呈现。

任务 4 撰写报告

完成上述任务后，以报告形式完整地呈现数据分析的背景、过程、结果及建议，内容包括：

- (1) 数据采集过程、数据结果的说明。
- (2) 任务 2.2 至 2.5 中数据结果的呈现，并进行分析说明。
- (3) 帮小明推荐一款手机并阐述理由。
- (4) 归纳总结数据分析结果，并根据分析结果对手机厂商或电商商家给出合理的建议。

四、关于竞赛成果提交的说明

1. 关于提交方式

请使用**队员 1**的账号登录数睿思，进入第一届技能大赛页面。为保证成功提交，请使用谷歌浏览器。

2. 关于作品提交

将任务 4 所要求的报告转成 PDF 文件作为作品提交，文件名为“**report.pdf**”。

3. 关于附件提交

请将任务 1 至任务 3 每个子任务的结果（包括任务截图、数据文件、图表）进行保存，并按照下述命名规则命名，与报告的 Word 版一起打包成压缩包作为“附件”上传，文件名为“**appendix.zip**”。

3.1 命名格式统一为“任务 x. x. 扩展名”，如单个任务出现多个结果，则命名为“任务 x. x. x. 扩展名”。如：任务 1. 1. jpg、任务 1. 1. 2. xlsx。所有图片文件都保存为. jpg 格式。

3.2 任务输出结果请参考任务说明。

4. 提交界面

4.1 在依次上传完“竞赛承诺书”、“作品”、“附件”后，点击“提交”。

竞赛资讯	*对应题目:	A
赛制介绍	*上传竞赛承诺书(PDF):	选择文件: 竞赛承诺书.pdf 注: 1、承诺书模板下载: 竞赛承诺书模板。 2、竞赛承诺书打印后签名, 并扫描生成pdf文件。
报名缴费	*上传作品(报告PDF):	选择文件: report.pdf
提交A题	*上传附件(zip文件):	选择文件: appendix.zip 注: 1、若选择附件错误, 请按F5刷新页面, 重新选择附件。 2、附件总大小不能超过150MB。
提交B题	举报:	是否发现违反竞赛章程的现象: <input checked="" type="radio"/> 是 <input type="radio"/> 否 选择文件: 未选择任何文件 注: 若发现有违反竞赛章程的现象, 请以Word描述该现象并收集相关证据(图片, 音频, 视频等)以压缩包的形式提交。
我的成绩	作品论文内容及命名请勿出现学校、学院、队号、队员以及指导老师相关任何信息, 否则该作品视为无效作品。 提交	

4.2 请待页面弹出“上传成功”对话框，点击“确定”，在相应位置可以看到“已上传 XXX”字样，表示相关文件提交成功。

竞赛资讯	*对应题目:	A
赛制介绍	*上传竞赛承诺书(PDF):	<div>选择文件 未选择任何文件 已上传承诺书!</div> <div>注: 1、承诺书模板下载: 竞赛承诺书模板。 2、竞赛承诺书打印后签名, 并扫描生成pdf文件。</div>
报名缴费	*上传作品(报告PDF):	<div>选择文件 未选择任何文件 已上传作品!</div>
提交A题	*上传附件(zip文件):	<div>选择文件 未选择任何文件 已上传附件!</div> <div>注: 1、若选择附件错误, 请按F5刷新页面, 重新选择附件。 2、附件总大小不能超过150MB。</div>
提交B题	举报:	<div>是否发现违反竞赛章程的现象: <input checked="" type="radio"/> 是 <input type="radio"/> 否 选择文件 未选择任何文件</div> <div>注: 若发现有违反竞赛章程的现象, 请以Word描述该现象并收集相关证据 (图片, 音频, 视频等) 以压缩包的形式提交。</div>
我的成绩		

作品论文内容及命名请勿出现学校、学院、队号、队员以及指导老师相关信息, 否则该作品视为无效作品。

4.3 在比赛当天 20:00:00 竞赛成果截止提交之前, 可多次上传相关文件, 系统默认以最后上传的文件为准。