

基于Python的消费者服装购买数据分析研究

刘旭婷¹, 李春青¹, 荆妙蕾², 冯 堃¹

¹天津工业大学计算机科学与技术学院, 天津

²天津工业大学纺织科学与工程学院, 天津

Email: 951521676@qq.com, franklcq@163.com, jingmiaolei@126.com, fengkun_mail@163.com

收稿日期: 2020年12月9日; 录用日期: 2021年1月1日; 发布日期: 2021年1月8日

摘 要

信息化时代的到来, 打破了传统的服装销售模式, 精准营销成为一种高效的销售模式。为了深入了解消费者的需求, 本研究通过问卷调查采集消费者购买服装产品相关数据, 包括品牌、价格、款式、颜色、面料、质量、舒适度和实用性八个方面, 将采集到的数据存入数据库中, 使用Python语言编程对数据库中的数据进行综合分析处理, 结合年龄、收入等基础条件, 分析影响消费者购买服装时的考虑因素, 使用Python的Matplotlib和Bokeh可视化库将数据分析结果用柱状图和饼状图的形式直观地展示出来, 该数据分析结果可以给服装生产厂家与服装销售商提供参考, 为厂家确定生产规模, 为商家进行有针对性的精准营销提供理论依据和指导。本文的分析结论对服装的生产与销售具有重要意义。

关键词

Python, 数据分析, 精准营销, 可视化

Research on Consumer Clothing Purchase Data Analysis Based on Python

Xuting Liu¹, Chunqing Li¹, Miaolei Jing², Kun Feng¹

¹School of Computer Science and Technology, Tianjin Polytechnic University, Tianjin

²School of Textile Science and Engineering, Tianjin Polytechnic University, Tianjin

Email: 951521676@qq.com, franklcq@163.com, jingmiaolei@126.com, fengkun_mail@163.com

Received: Dec. 9th, 2020; accepted: Jan. 1st, 2021; published: Jan. 8th, 2021

Abstract

The advent of the information age has broken the traditional clothing sales model, and precision

marketing has become an efficient sales model. In order to understand the needs of consumers in depth, this study collected data related to consumer purchases of clothing products through questionnaires, including eight aspects of brand, price, style, color, fabric, quality, comfort and practicality, and stored the collected data. Into the database, use Python language programming to comprehensively analyze and process the data in the database; combined with basic conditions such as age and income, analyze the factors that affect consumers when buying clothing; using Python's Matplotlib and Bokeh visualization libraries, present the data analysis results visually in the form of bar chart and pie chart. The data analysis results can provide references for clothing manufacturers and clothing sellers, determine the production scale for manufacturers, and provide theoretical basis and guidance for merchants to carry out targeted precision marketing. The analysis conclusion of this article is of great significance to the production and sales of clothing.

Keywords

Python, Data Analysis, Precision Marketing, Visualization

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在以往的网络营销中,服装销售商通常是采用全面撒网的方式开展产品营销,尽管这种营销方式能够取得一定的效果,但是在现代技术水平不断发展的今天,想要依靠传统营销方式给企业带来良好的经济效益已经不太可能[1]。在这样的背景下,精准营销作为一种新型高效的营销方式应运而生,不断融入先进的营销理念,通过对信息高度整合从而帮助企业进行准确客户定位,实现精准营销,这样不仅可以降低企业营销成本,还能实现企业经济效益的增长[2]。目前数据分析商用软件有 Matlab、SPSS 和 SAS,然而这些软件费用高昂,安装包较大,安装过程复杂[3],相比于这些商用软件,Python 语言具有庞大的标准库与第三方库,能够处理更大的数据集和更复杂的数据逻辑,更容易实现自动化分析,Python 在数据分析领域的优势逐渐凸显出来,逐渐成为了数据分析领域的首选工具。为了帮助服装厂商和销售商了解消费者的真正需求进而进行精准营销,本研究对服装消费者发放网上问卷调查,将问卷调查得到的数据通过 Python 编程从品牌、价格、款式、颜色、面料、质量、舒适度和实用性这八个角度来分析不同年龄阶层,不同收入阶层的人们在购买服装时的考虑因素。研究结果为服装生产厂家与服装销售商提出咨询建议,为厂家确定生产规模,为商家制定有针对性的精准营销策略提供理论依据和指导。实验的结果显示:相比于 SPSS 等常用的数据分析软件,Python 可以从多个角度多方位地分析数据并使用不同的图表最大程度地展现数据的特征,在数据分析领域更加灵活,在数据分析和交互、探索性计算以及数据可视化等方面都显得比较活跃,因此基于 Python 的消费者服装购买数据分析研究比以往的商用数据分析软件更加快捷、方便和准确,更好地满足了新时代的数据处理要求。

2. 相关技术

2.1. Python 语言

Python 是一种结合了解释性、编译性和面向对象的计算机程序设计语言[4]。Python 语言由于简洁易读、易维护和可扩展性,近年来已经发展成为最受欢迎的程序设计语言之一,并在数据科学与数据分析

领域逐渐凸显其优势[5]。

2.2. Matplotlib 可视化库

Python 的 2D 绘图库 Matplotlib 功能非常强大，是进行 Python 可视化的常用库。Matplotlib 的 plot 工具包提供了 MATLAB 类似的绘图 API，用户只需调用 pyplot 模块中的函数[6]就可以用简洁高效的代码画出直方图、散点图和条形图等。

2.3. Bokeh 可视化库

Bokeh 是一个支持 web 浏览器的交互式可视化库，允许输出为 JSON，HTML 等网络格式，也支持数据流和实时数据[7]。Bokeh 可以为流式数据集或超大型数据集提供高性能交互的多功能图形。BokehJS 是 Bokeh 绘图的独特之处之一，BokehJS 是一个一流 JavaScript 绘图库，Bokeh 通过捆绑的多种语言可以生成一个 JSON 文件，此文件会作为 BokehJS 的一个输入，从而实现在浏览器中的动态和交互式可视化。

3. 数据来源

采用问卷调查的方式，问卷对不同地区，不同年龄段，不同收入段的服装消费者发放网上问卷，收回有效问卷 353 份。问卷主要调查内容分为两部分，一部分是被调查者的个人信息，包括性别、年龄、收入；另一部分是被调查者购买服装的相关信息，包括每季购买服装数量，习惯购买服装的价格、面料、方式、考虑因素和是否受品牌影响。

4. 数据分析过程

将问卷调查的结果录入 mysql 数据库中，开发工具为 Visual Studio Code，使用 Python 语言从数据库中读取数据，通过多种角度统计与分析，使用 Python 的 Bokeh 和 Matplotlib 绘图库将数据分析结果以柱状图、饼状图的形式来显现出来，便于人们更直观的了解不同人群在购买服装时的考虑因素。

4.1. 数据库设计

根据问卷调查的结果来设计数据库表，采用 mysql 数据库录入问卷调查的信息。数据库表结构如图 1。


名	类型	长度	小数点	不是null	
序号	int	255	0	<input checked="" type="checkbox"/>	 1
性别	varchar	255	0	<input checked="" type="checkbox"/>	
年龄	varchar	255	0	<input checked="" type="checkbox"/>	
收入	varchar	255	0	<input checked="" type="checkbox"/>	
春季购买服装数量	varchar	255	0	<input checked="" type="checkbox"/>	
夏季购买服装数量	varchar	255	0	<input checked="" type="checkbox"/>	
秋季购买服装数量	varchar	255	0	<input checked="" type="checkbox"/>	
冬季购买服装数量	varchar	255	0	<input checked="" type="checkbox"/>	
平时购买衣服单价	varchar	255	0	<input checked="" type="checkbox"/>	
平时购买衣服面料	varchar	255	0	<input checked="" type="checkbox"/>	
平时购买衣服方式	varchar	255	0	<input checked="" type="checkbox"/>	
购买衣服比较看重	varchar	255	0	<input checked="" type="checkbox"/>	
是否受品牌影响	varchar	255		<input checked="" type="checkbox"/>	

Figure 1. Prediction model of sludge age

图 1. 数据库表结构

4.2. 数据分析与绘图

使用 Python 语言连接 mysql 数据库, 读取数据, 从多个角度分析数据, 将数据分析的规律使用 Python 的 Matplotlib 和 Bokeh 绘图库用柱状图与饼状图表现出来。由于本文绘制图表较多, 节选以下两个图的生成过程展示, 其余图绘制方式类似。

4.2.1. 不同年龄段人群购买服装考虑因素多系列柱状图

对不同年龄段人群在购买服装时关注因素占比生成多系列柱状图。具体操作如下:

- 1) 连接 mysql 数据库读取数据;
- 2) 创建四个列表存放数据, 四个列表包含的年龄段的值分别为 18~25 岁、26~40 岁、41~55 岁和 56 岁以上;
- 3) for 循环依次遍历四个列表, 获取每个列表中购买服装关注点为品牌、价格、款式、颜色、面料、质量、舒适度和实用性的数据, 然后使用 count()方法分别计算其出现次数;
- 4) key()方法获取属性, value()方法获取对应的值, vbar()函数绘制柱状图。

4.2.2. 被调查者平时购买衣服方式占比饼状图

对所有被调查者平时购买衣服方式占比生成饼状图。具体操作如下:

- 1) 连接 mysql 数据库读取数据;
- 2) 创建列表存储“购买方式”列信息;
- 3) for 循环遍历列表获取不同购买方式值及 count()方法计算其出现次数;
- 4) key()方法获取购买方式属性, value()方法获取对应的值, plt.pie()绘制饼状图。

5. 服装购买因素结果分析

有效数据量为 353 份, 其中男性数量为 154, 占总样本量的 43.63%, 女性数量为 199, 占总样本量的 56.37%。

5.1. 消费者在购买服装时的考虑因素分析

为调查消费者购买衣服时的考虑因素列出 8 个选项, 分别为品牌、价格、款式、颜色、面料、质量、舒适度和实用性, 消费者对于这 8 个考虑因素选项占比如图 2 所示。

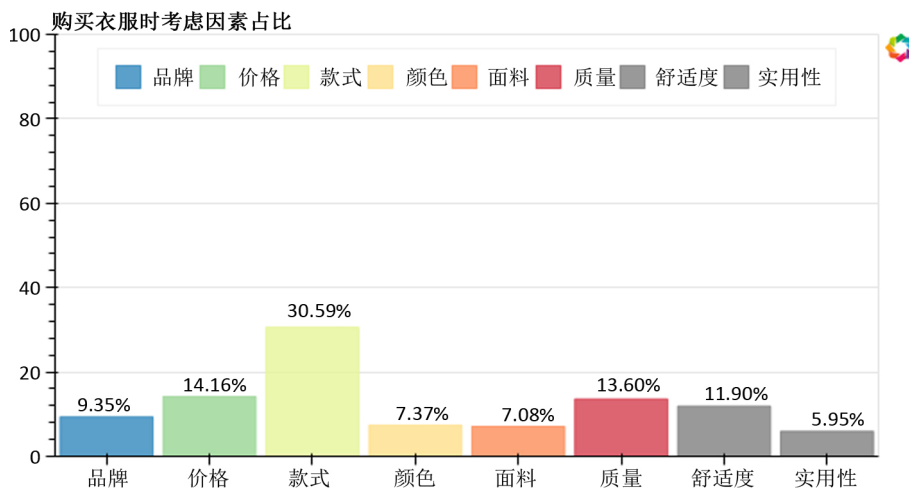


Figure 2. Percentage of factors considered by consumers when buying clothes

图 2. 消费者购买衣服时考虑因素占比

从图 2 可以看出,款式占比最高,为 30.59%,其次是价格占 14.16%,由此可知消费者在购买衣服时比较看重款式与价格,说明款式新颖并且价格合适的服装更能吸引消费者。

5.2. 不同年龄段人群购买服装考虑因素分析

调查样本中,18~25 岁的人有 112,占总样本量的 31.73%,26~40 岁的人有 182,占总样本量的 51.56%,41~55 岁的人有 54 人,占总样本量的 15.30%,56 岁以上的有 5 人,占总样本量的 1.42%。对各年龄段的人群购买服装时的考虑因素分析,结果如图 3。

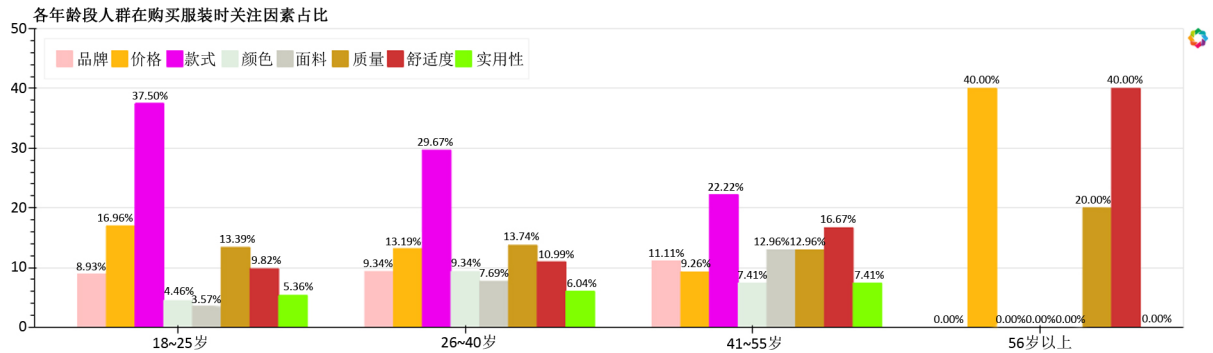


Figure 3. Proportion of factors that people of all ages pay attention to when buying clothing
图 3. 各年龄段人群在购买服装时关注因素占比

图 3 显示,年龄在 18~25 岁的人群中,在购买衣服时看重款式的人占比最高,看重舒适度的人占比较低,随着年龄段的的增长,看重款式的人占比逐渐减少,而看重舒适度的人占比明显提高。说明年轻人购买衣服时比较关注款式,随着年龄的增加,消费者更加关注舒适度。

5.3. 不同收入段人群购买服装考虑因素分析

调查样本中,收入在 3000 元以下的有 95 人,占总样本量的 26.91%。收入在 3001~6000 元的有 127 人,占总样本量的 35.98%。收入在 6001~10,000 元的有 108 人,占总样本量的 30.59%。

不同收入段人群购买服装考虑因素分析,结果如图 4 所示。

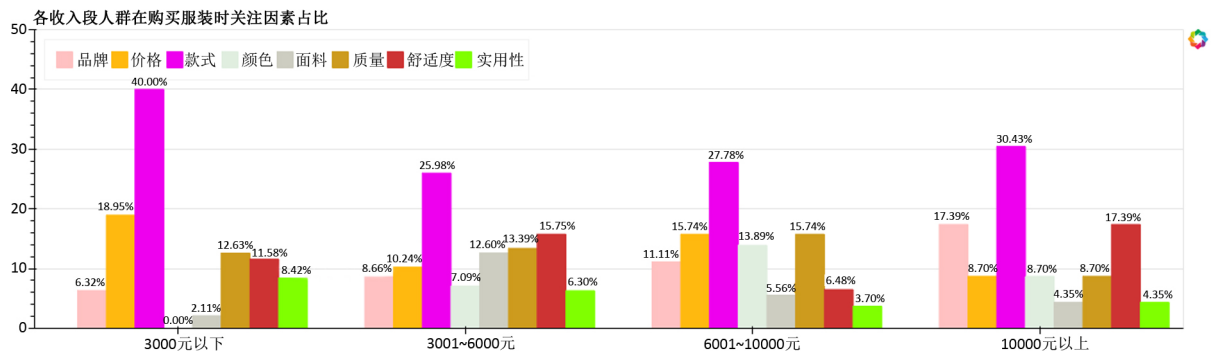


Figure 4. Proportion of factors of concern when buying clothing for all income groups
图 4. 各收入段人群在购买服装时关注因素占比

图 4 显示收入在 3000 元以下的人群在购买衣服时的考虑因素中品牌只占 6.32%,而收入在 10,000 元以上的人群中,在购买衣服看重品牌的人数增加到了 17.39%,由此可以看出随着收入的增加,消费者在购买衣服的时候对品牌的关注度提高。

5.4. 不同收入段人群购买服装的价格分析

对不同收入段的人群习惯购买的衣服价格分析，如图 5 所示。

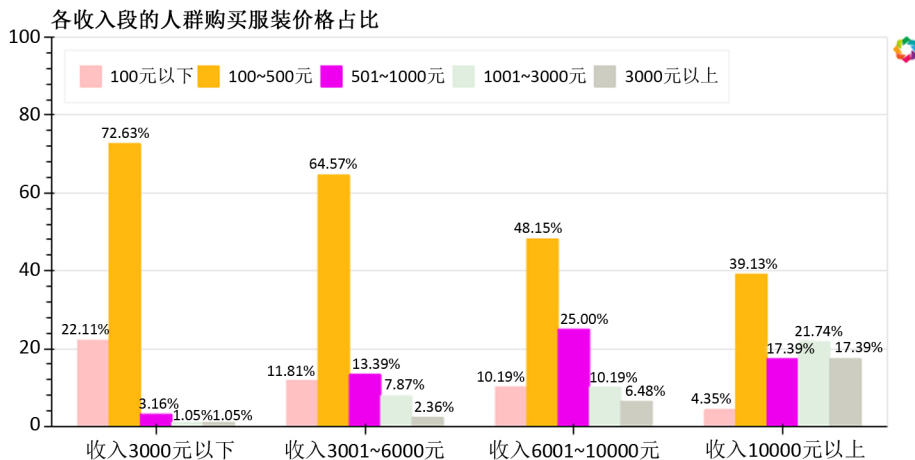


Figure 5. Proportion of the price of clothing purchased by the groups of people
图 5. 各收入段的人群购买服装价格占比

图 5 显示，收入在 3000 元以下的人群中，购买衣服价格在 100~500 元区间内的人占比较高，为 72.63%，购买衣服价格区间在 3000 元以上的人仅占比 1.05%。随着收入的增加，购买低价衣服的人在减少，购买高价衣服的人在增多，在收入为 10,000 元以上的人群中，购买衣服价格在 100~500 元区间内的人占比降到了 39.13%，而购买衣服价格在 3000 元以上的人占比增加到了 17.39%。由此可以得出收入高的人更倾向于购买高价的服装。

5.5. 被调查者平时购买衣服的方式分析

本次调查设置淘宝店，品牌类官网，实体店，小红书，蘑菇街，微商和其他七个选项，被调查者选项占比如图 6 所示。

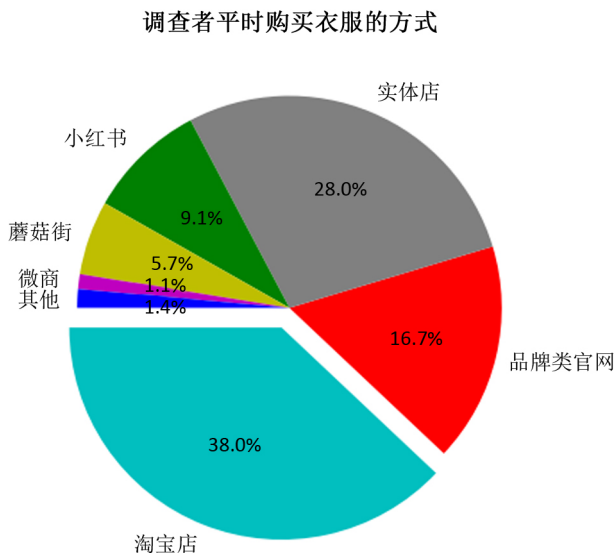


Figure 6. The proportion of respondents who usually buy clothes
图 6. 被调查者平时购买衣服方式占比

图 6 显示, 被调查者平常购买衣服的方式中, 淘宝店占比最高, 为 38.0%, 其次是实体店与品牌类官网, 占比分别为 28.0%和 16.7%。由此可见, 线上销售具有很大市场, 服装销售商应该在原有销售模式上拓宽自己的销售渠道, 采用线上线下相结合的服装销售模式。

6. 结论与展望

本研究使用 Python 实现了数据向图表的转化, 相比于之前碎片化的信息, 以图表的形式呈现出来的数据更有价值, 使繁多的数据变得直观, 研究从多个角度分析数据, 使用不同的图表最大程度地展现了数据的特征, 使得服装销售商家和厂商更能够容易发掘隐藏的数据关联, 发现消费者潜在的需求, 筛选出消费者真正关注的信息[8]。本文数据分析结果对服装生产与销售具有一定的参考价值, 对服装生产商和服装销售商的市场行为具有一定的指导意义。

6.1. 厂家生产服装应着重考虑款式与价格

本文数据分析的结果显示, 大部分消费者在购买服装时比较看重款式与价格, 款式新颖、价格在 100~500 元的服装销量最高。服装厂家应关注服装流行趋势, 注重设计感和时尚感, 适当控制服装生产成本, 加大生产款式新颖、中低档价位的服装。

6.2. 商家应该对不同人群精准营销

本文数据分析的结果显示, 不同的年龄段与收入段的人群在购买服装的时候关注点是不同的, 例如年轻人比较看重款式, 然而年龄大的人关注的是舒适度。收入高的人在购买衣服时更关注品牌而不是价格。服装销售商户应分析不同年龄阶层, 不同收入阶层的人们的喜好, 根据分析结果制定有针对性的精准营销策略, 根据每位消费者的需求推荐适合他们的产品, 这样既可以增加销量, 又能节省广告与人力成本。

6.3. 实体店应该扩展网上销售渠道

本文数据分析的结果显示, 一半以上的被调查者习惯在淘宝店购买衣服, 淘宝店购物方便快捷, 可以货比三家, 并且可以自由退换货, 随着人们生活节奏的加快, 越来越多的人选择在网上购物。网络购物的兴起对实体店产生了一定的影响。为了提升销量, 服装销售实体店也应该扩展自己的网上销售渠道, 实行线上线下相结合的服装销售模式。

参考文献

- [1] 廖衡. 大数据背景下网络精准营销[J]. 营销界, 2020(44): 197-198.
- [2] 陈旭. 基于大数据环境下电商精准营销策略探析[J]. 电子商务, 2020(5): 51-52.
- [3] 宋永生, 黄蓉美, 王军. 基于 Python 的数据分析与可视化平台研究[J]. 现代信息科技, 2019, 3(21): 7-9.
- [4] 刘艳玲, 姚建盛. Python 在数据可视化中的应用[J]. 福建电脑, 2020, 36(3): 68-70.
- [5] 罗博炜. 基于 Python 的数据可视化[J]. 信息记录材料, 2019, 20(12): 72-74.
- [6] 冯艳茹. Python 语言在大数据分析中的应用[J]. 电脑知识与技术: 学术版, 2020, 16(24): 72-73.
- [7] 胡非, 韩永辉, 许超, 周巧妮. 基于 Python 的可视化工具研究与应用[J]. 工程技术研究, 2018(13): 25-26.
- [8] 刘娟, 管希东. 基于 Python 爬虫的职位信息数据分析和可视化系统实现[J]. 软件工程与应用, 2020(4): 317-325.