

2021-2022学年第二学期

**《互联网+项目实战》**

**课程设计报告**

**题 目：**优衣库整体销售情况可视化分析

**姓 名：** 2019级网络工程二班连子蒙

**指导教师：** 贺 萍

**完成时间：** 2022年5月21日

河北经贸大学信息技术学院

目录

[1. 项目介绍 3](#_Toc105787103)

[1.1. 项目背景 3](#_Toc105787104)

[1.2. 项目说明 3](#_Toc105787105)

[1.2.1. 项目目的 3](#_Toc105787106)

[1.2.2. 项目所需技术 3](#_Toc105787107)

[2. 数据处理及可视化 4](#_Toc105787108)

[2.1. 配置项目与读取数据 4](#_Toc105787109)

[2.2. 数据介绍 5](#_Toc105787110)

[2.3. 数据处理 6](#_Toc105787111)

[2.3.1. 删除无用数据 6](#_Toc105787112)

[2.3.2. 修改数据类型 6](#_Toc105787113)

[2.3.3. 增加计算字段 6](#_Toc105787114)

[2.4. 数据分类统计 7](#_Toc105787115)

[2.4.1. 对 city（城市）字段进行分类统计 7](#_Toc105787116)

[2.4.2. 对 channel（购买方式）字段进行分类统计 8](#_Toc105787117)

[2.4.3. 对 gender\_group（性别分类）字段进行分类统计 9](#_Toc105787118)

[2.4.4. 对 age\_group（年龄分类）字段进行分类统计 10](#_Toc105787119)

[2.4.5. 对 revenue（营收）字段进行分类统计 11](#_Toc105787120)

[2.4.6. 对 wkd\_ind（购买时间）字段进行分类统计 12](#_Toc105787121)

[2.4.7. 对 product（产品类型）字段进行分类统计 14](#_Toc105787122)

# 1. 项目介绍

## 1.1. 项目背景

优衣库是日本迅销公司的核心品牌，建立于1984年。最初是一家销售西服的小服装店，现已成为国际知名品牌。

优衣库（Uniqlo）的内在涵义是指通过摒弃不必要装潢装饰的仓储型店铺，采用超市型的自主购物方式，以合理可信的价格提供顾客希望的商品。那么，如何更好地提供顾客所喜欢的商品、如何提上商品的利润、高效利用仓储空间，则是一个非常重要的企业课题。就此，笔者希望通过分析优衣库品牌的商品、用户以及销售数据，来更好地分析用户购买偏好以及不同产品的盈利情况，进而提供国内竞品品牌相关经验。

## 1.2. 项目说明

### 1.2.1. 项目目的

本项目使用 Python 进行数据清洗与可视化，尝试分析以下关键问题：

整体销售情况随着时间的变化是怎样的？

不同产品的销售情况是怎样的？

顾客偏爱哪一种购买方式？

销售额和产品成本之间的关系怎么样？

### 1.2.2. 项目所需技术

#### (1) Numpy

NumPy是Python中科学计算的基础包。 它是一个Python库，提供多维数组对象，各种派生对象（如掩码数组和矩阵），以及用于数组快速操作的各种API，有包括数学、逻辑、形状操作、排序、选择、输入输出、离散傅立叶变换、基本线性代数，基本统计运算和随机模拟等等。

#### (2) Matplotlib

Matplotlib是一个Python的2D绘图库，它以各种硬拷贝格式和跨平台的交互式环境生成出版质量级别的图形。 通过Matplotlib，开发者可以仅需要几行代码，便可以生成绘图。 一般可绘制折线图、散点图、柱状图、饼图、直方图、子图等等。

#### (3) Pandas

Pandas 是 Python 的核心数据分析支持库，提供了快速、灵活、明确的数据结构，旨在简单、直观地处理关系型、标记型数据。 Pandas 的目标是成为Python 数据分析实践与实战的必备高级工具，其长远目标是成为最强大、最灵活、可以支持任何语言的开源数据分析工具。

# 2. 数据处理及可视化

## 2.1. 配置项目与读取数据

项目分析数据条数共**22293** 条。

具体配置如图2-1 配置项目与读取数据所示：

图2-1 配置项目与读取数据

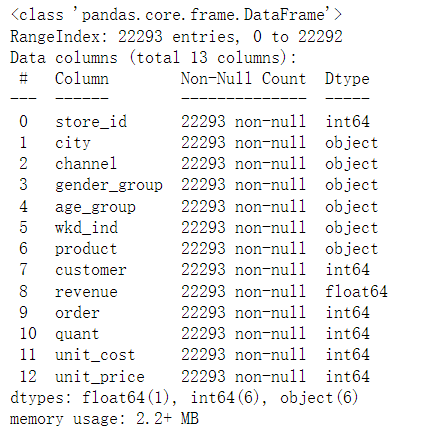
## 2.2. 数据介绍

本次数据如表2-1数据字段所示：

表2-1数据字段

|  |  |
| --- | --- |
| 数据字段 | 字段释义 |
| store\_id |  |
| city | 城市销量 |
| channel | 购买方式（线上/线下） |
| gender\_group | 性别分类（男性/女性） |
| age\_group | 年龄分类 |
| wkd\_ind | 购买时间（工作日/假期） |
| product | 产品类型 |
| customer | 购买人数 |
| revenue | 营收 |
| order | 订单数量 |
| quant | 定量 |
| unit\_const | 单价成本 |
| unit\_price | 单价 |

具体数据类型介绍见图2-2 数据类型介绍。

图2-2 数据类型介绍

## 2.3. 数据处理

### 2.3.1. 删除无用数据

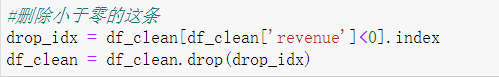
具体如图2-3 删除小于零的数据所示。

图2-3 删除小于零的数据

### 2.3.2. 修改数据类型

具体如图2-4 修改 store\_id 的数据类型所示。

图2-4 修改 store\_id 的数据类型

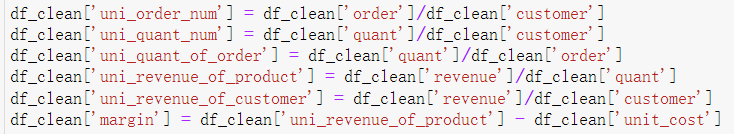
### 2.3.3. 增加计算字段

增加计算字段细则如表2-2 增加计算字段细则所示：

表2-2增加计算字段细则

|  |  |  |
| --- | --- | --- |
| 字段释义 | 字段名称 | 字段公式 |
| 人均订单量 | uni\_order\_num | order/customer |
| 人均购买数量 | uni\_quant\_num | quant / customer |
| 订单平均包含商品量 | uni\_quant\_of\_order | quant / order |
| 商品平均销售额 | uni\_revenue\_of\_product | revenue / quant |
| 人均消费额 | uni\_revenue\_of\_customer | revenue / customer |
| 利润商品平均销售额 | margin | uni\_revenue\_of\_product - unit\_cost |

具体如图2-5 增加计算字段所示。

图2-5 增加计算字段

## 2.4. 数据分类统计

### 2.4.1. 对 city（城市）字段进行分类统计

首先选取city 字段数据，并更改其列名，将index 字段更改为城市名称，将城市对应销量更改为count，最终呈现如图2-6 city字段数据统计列表所示。



图2-6 city字段数据统计列表

最后以city字段数据作为x轴数据，以count字段数据y轴数据，绘制柱状图，结果如图2-7 city字段分类统计所示。

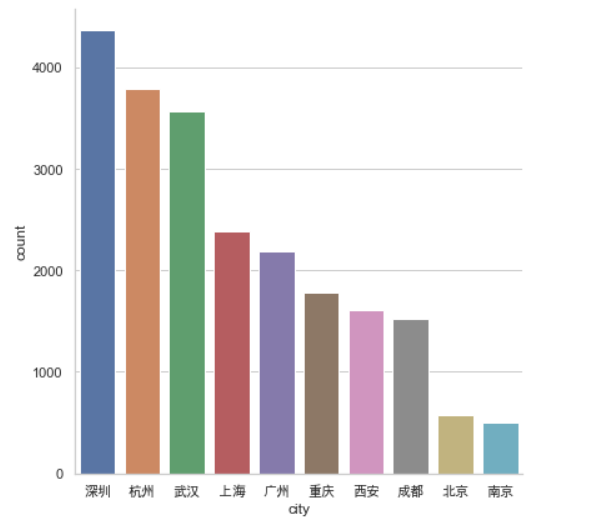


图2-7 city字段分类统计

### 2.4.2. 对 channel（购买方式）字段进行分类统计

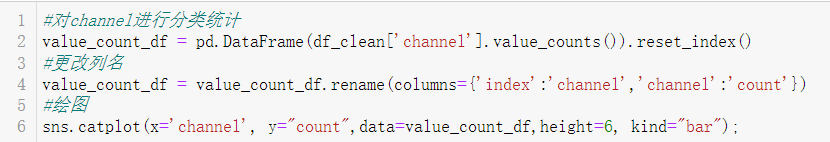
 首先选取 channel 字段数据，并更改其列名，将 index 字段更改为购买方式名称，将购买方式对应销量更改为 count，具体代码见图2-8 channel 分类统计代码所示。

图2-8 channel 分类统计代码

最终以 channel 字段数据作为 x 轴数据，以 count 字段数据作为 y 轴数据，绘制柱状图，结构如图2-10 channel 字段分类统计所示。

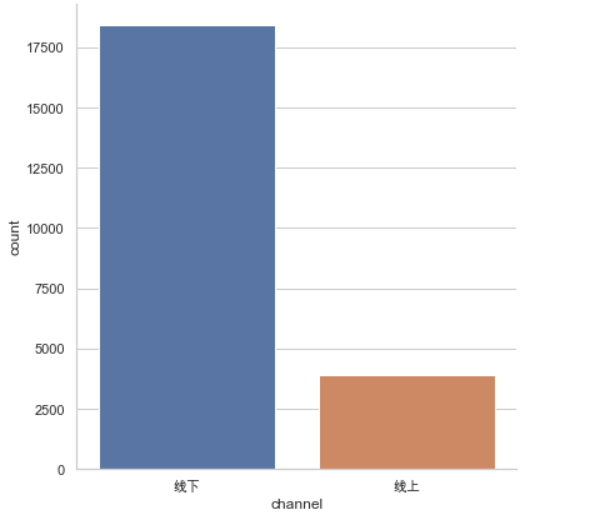


图2-10 channel 字段分类统计

### 2.4.3. 对 gender\_group（性别分类）字段进行分类统计

首先选取 gender\_group字段数据，并更改其列名，将 index 字段更改为性别名称，将性别对应销量更改为 count，具体代码见图2-11 gender\_group分类统计代码所示。

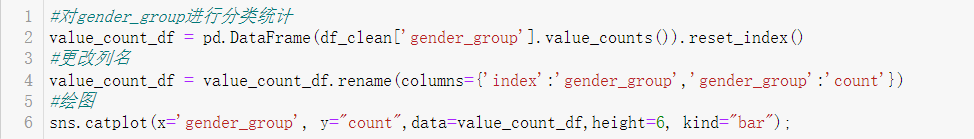


图2-11 gender\_group分类统计代码

最终以 gender\_group字段数据作为 x 轴数据，以 count 字段数据作为 y 轴数据，绘制柱状图，结构如图2-12 gender\_group字段分类统计所示。

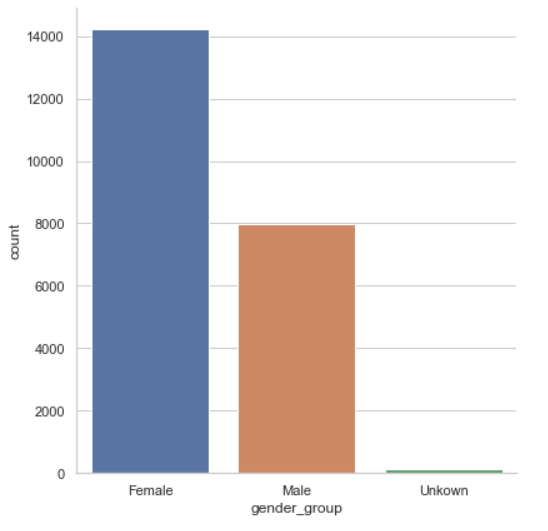


图2-12 gender\_group字段分类统计

### 2.4.4. 对 age\_group（年龄分类）字段进行分类统计

首先设定age-group年龄区间范围，然后选取 age\_group字段数据，并更改其列名，将 index 字段更改为年龄区间，将年龄区间对应销量更改为 count，具体代码见图2-13 age\_group分类统计代码所示。

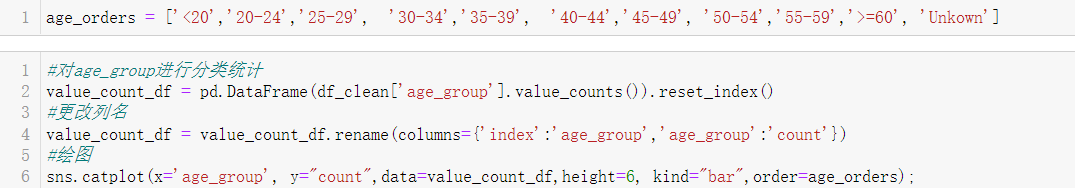


图2-13 age\_group分类统计代码

最终以 age\_group字段数据作为 x 轴数据，以 count 字段数据作为 y 轴数据，绘制柱状图，结构如图2-14 age\_group字段分类统计所示。

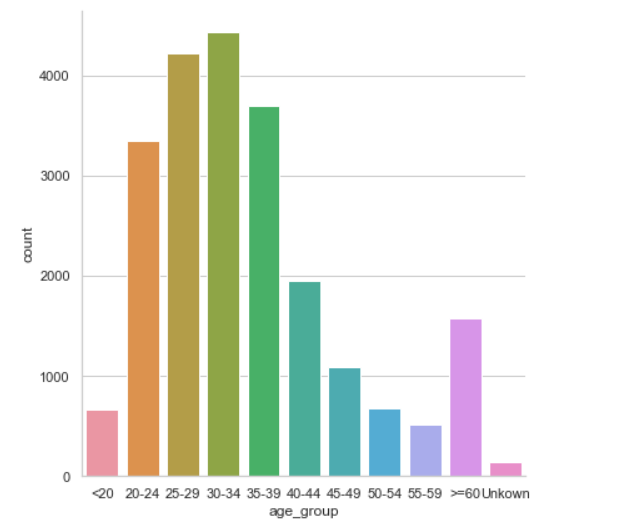


图2-14 age\_group字段分类统计

### 2.4.5. 对 revenue（营收）字段进行分类统计

首先，如图2-15 revenue字段数据描述所示，确定绘制直方图时，x轴的数据选取区间。

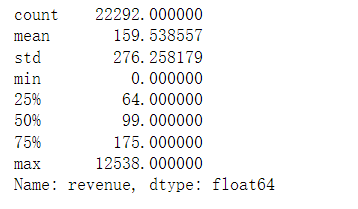


图2-15 revenue字段数据描述

最后以限定范围的 revenue字段数据绘制直方图，结果如图2-16 revenue 字段分类统计所示。

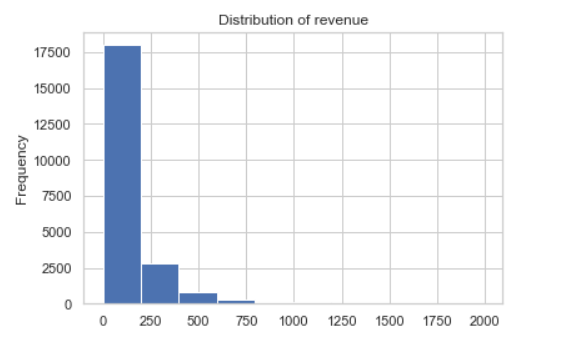


图2-16 revenue 字段分类统计

### 2.4.6. 对 wkd\_ind（购买时间）字段进行分类统计

#### （1）销量表现

将 wkd\_ind字段数据作为x轴数据，绘制直方图，结果如图2-17销量表现所示。

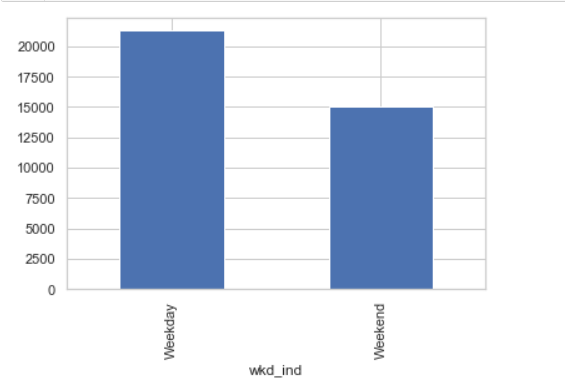


图2-17销量表现

#### （2）渠道表现

将 channel字段与wkd\_ind字段进行聚合，通过柱状图展示不同销售渠道在不同时间的购买情况对比，具体如图2-18 渠道表现所示。

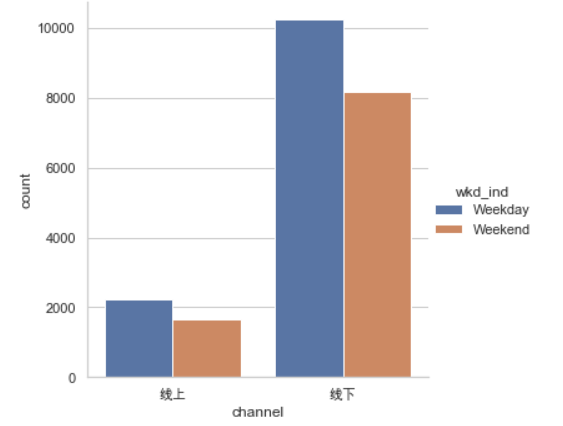


图2-18 渠道表现

#### （2）性别表现

将 gender\_group字段与wkd\_ind字段进行聚合，通过柱状图展示不同性别在不同时间的购买情况对比，具体如图2-19 性别表现所示。

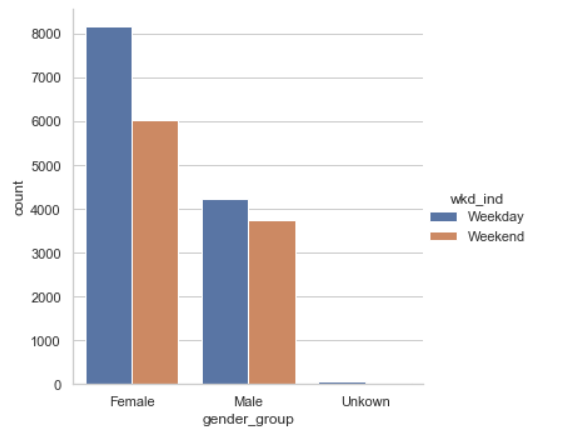


图2-19 性别表现

### 2.4.7. 对 product（产品类型）字段进行分类统计

#### （1）销售额表现

通过选取product字段中revenue类型数据经过排序后绘制柱状图以展示不同产品的销售额，具体见图2-20销售额表现所示。

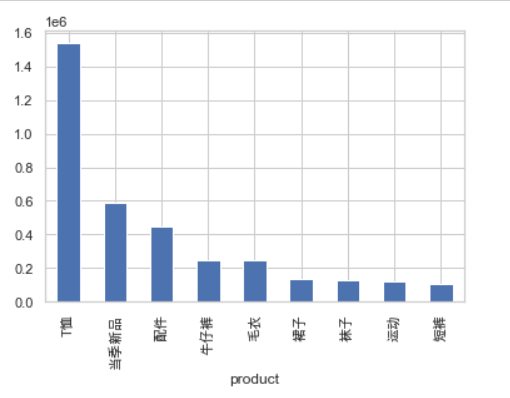


图2-20销售额表现

#### （2）销售量表现

通过选取product字段中quant类型数据经过排序后绘制柱状图以展示不同产品的销售量，具体见图2-21销售量表现所示。

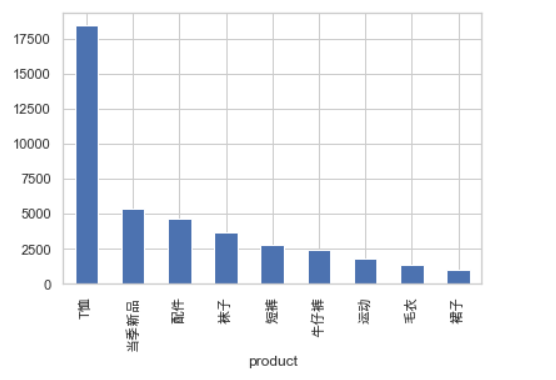


图2-21销售量表现

#### （3）利润表现

通过选取product字段中margin类型数据经过排序后绘制柱状图以展示不同产品的利润，具体见图2-22利润表现所示。

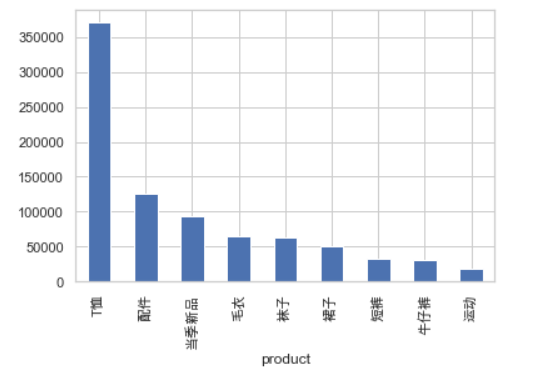


图2-20利润表现

# 3. 总结

通过对10座城市的22292条数据进行清洗处理数据，最终生成柱状图可视化分析整体销售情况（包括产品销售数量、销售金额、顾客人数等）与时间（工作日和周末）变化的关系，分析不同类别产品的销售情况，得出以下结论：

1. 数据主要集中在深圳、杭州、武汉三所城市；
2. 选择线下购买的用户远多于线上购买；
3. 女性客户相较男性客户稍多；
4. 客户的年龄段主要集中的40岁以下；
5. 客户选择工作日购买大于周末购买总次数；
6. T恤时最畅销的产品，占绝对领先地位，在销售额、销售数量和利润指标上都遥遥领先；
7. 当季新品和配件占据销售额、销售数量和利润指标上占据亚季军，其中当季新品的销售额和数量较配件高，但利润较配件低；
8. 毛衣、裙子和配件的平均利润最高，都在80元左右；
9. 商品的销售额与成本之间是成正向关的。

提出建议：适当增加 T恤、当季新品和配件库存量。