

# 超市数据分析报告

2020 年 12 月 19 日

## 目录

1	项目背景	1
2	数据获取	2
3	数据模型选定	2
3.1	产品类别	3
3.2	销售额	3
3.3	利润	3
4	数据整理	3
5	描述分析	4
5.1	各产品类别与销售额的关系	4
5.2	销售额按照地区与细分消费划分	5
5.3	销售额与利润的关系	7
5.4	利润与地区的关系	7
5.5	各类别产品利润分布情况	8
5.6	订单日期及销售额	8
6	总结	9

## 1 项目背景

自 2015 年 Q1 至 2018 年 Q4 零售行业主要呈现三大趋势:

- 线上线下融合加速落地。就表现形式而言，主要由基于消费体验重构的融合、供应链效率提升与渠道下沉以及消费场景延伸是线上线下融合的三类形式。

- 社交电商异军突起。模式上主要分为现有流量入口的商业价值挖掘和平台化运营两大类。
- 泛零售品类不断扩展。横向扩充与纵向延伸同步推进。此外，从消费端趋势来看，以代际变迁与消费升级为核心的基本趋势将长期持续存在。本报告是基于该大环境为载体的，分析办公相关某商超领域的数据表现。

## 2 数据获取

现有 9959 条超市数据，其中  $\text{利润率} = \text{销售额} / \text{利润}$  (·) 表示负值。

## 3 数据模型选定

维度：地区、城市、订单日期、类别和子类别等；度量：产品类别、销售额、利润和订单量等；度量和相应的维度关系如图 1 所示。



图 1: 度量和相应的维度关系

### 3.1 产品类别

根据产品的不同类别分析其平均销售量。

### 3.2 销售额

1. 分析各个地区的销售额分布图，根据销售额最多的地区看城市分布，观测是否与认知中城市 GDP 及产能相对应与匹配；
2. 根据各个类别的商品进行销售额分析，再细分至公司与个人，观测采购的类别商品与该区域的公司分布与人群有怎样的线性关系；
3. 按照时间维度去看销售额分布情况，对照客观存在活动内容及运营数据反观数据表现；

### 3.3 利润

1. 通过不同的地区与时间，结合折扣情况，看利润的数据表现；
2. 通过不同的类别，看利润分布情况；

## 4 数据整理

将数据进行格式调整，本次处理主要针对带有括号或货币符号的数值进行处理；

## 5 描述分析

### 5.1 各产品类别与销售额的关系

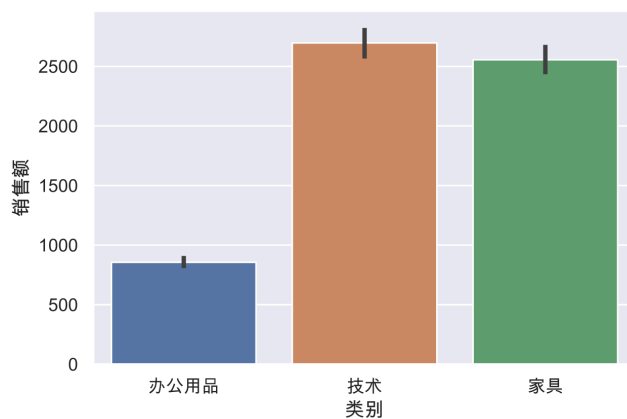


图 2: 各产品类别的销售额

从图 2中可以看出，不同类别的产品销售额的平均值是不一样的。其中，技术类产品最高，而办公用品类最低。

## 5.2 销售额按照地区与细分消费划分

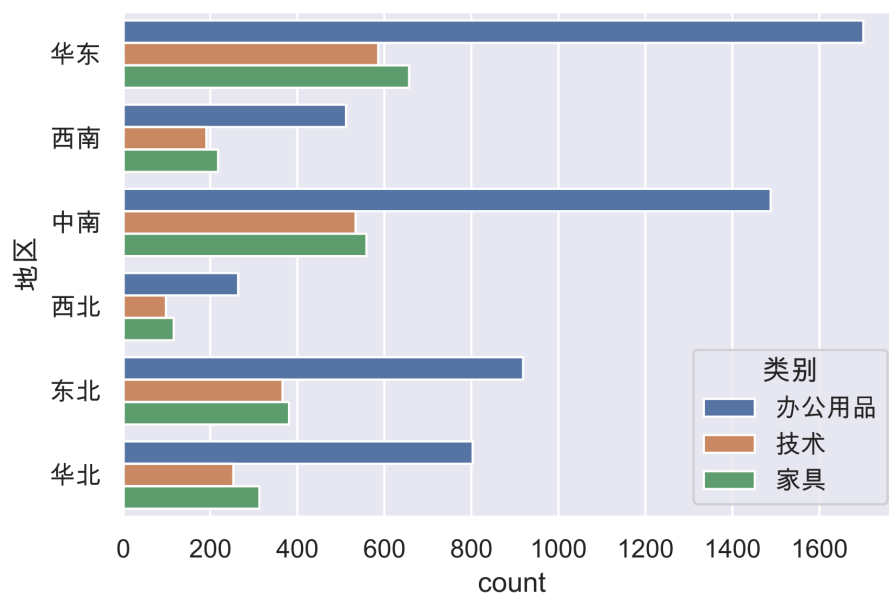


图 3: 各地区不同产品类别的订单量

从该超市的消费区域来看，华东地区的订单量最多，在各个地区中，办公用品占比最大。

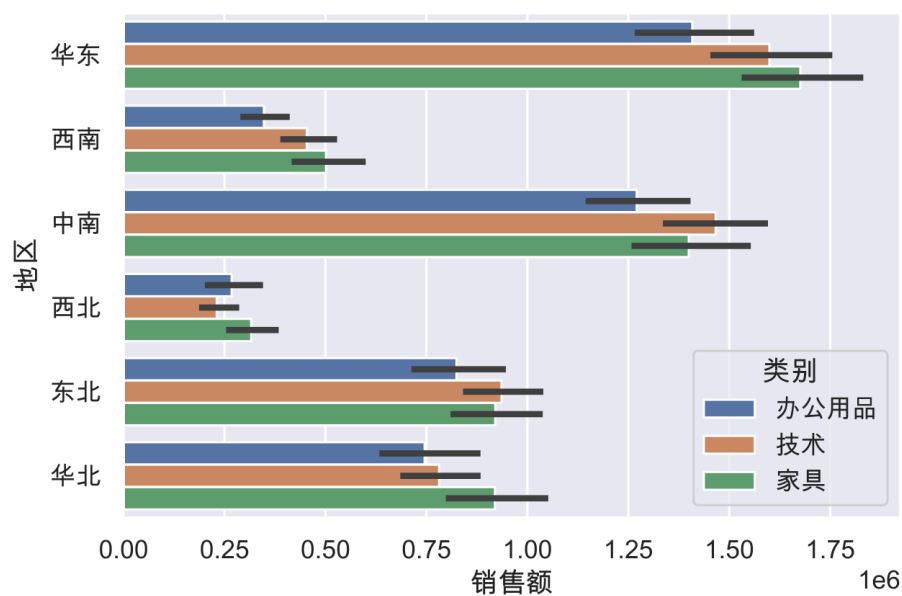


图 4: 各地区不同产品类别的销售额

类似地，华东地区的销售额最多，在各个地区中，办公用品占比最大。

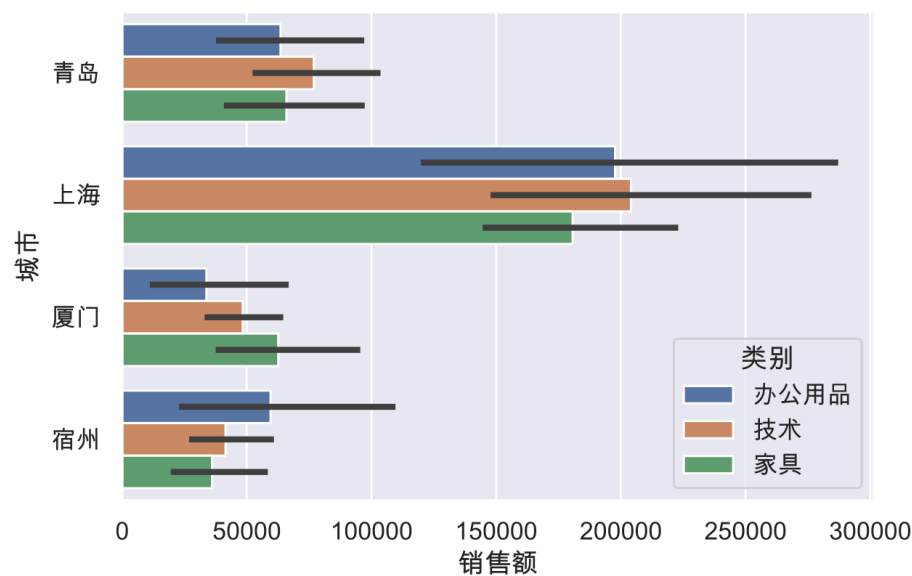


图 5: 华东地区销售额前 4 城市不同产品类别的销售额

继续细分华东区销售额分布，其中四个城市的销售额数据表现较为突出

5.3 销售额与利润的关系

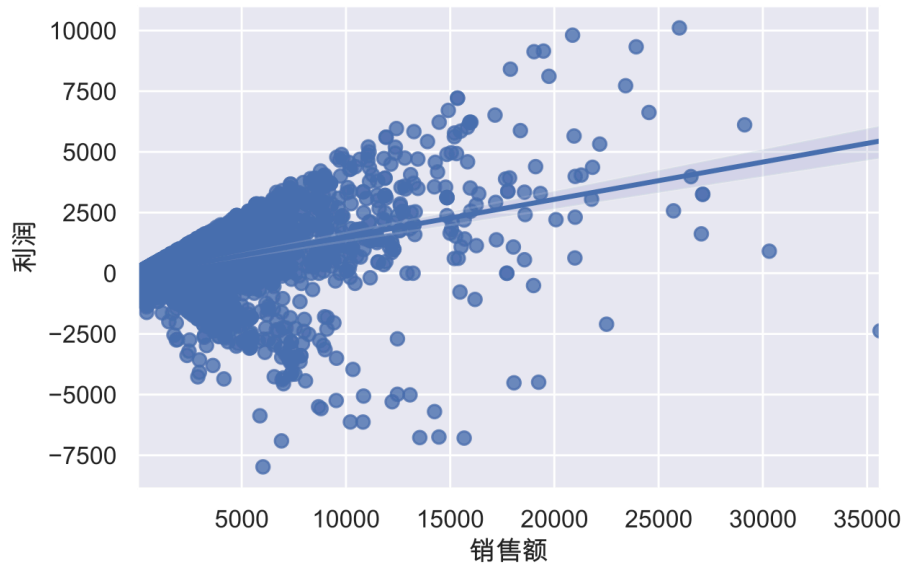


图 6: 销售额与利润的关系

销售额和利润在被产品名称细分的情况下，呈现非常明显的线性关系 ( $p$  值  $< 0.01$ )。

5.4 利润与地区的关系

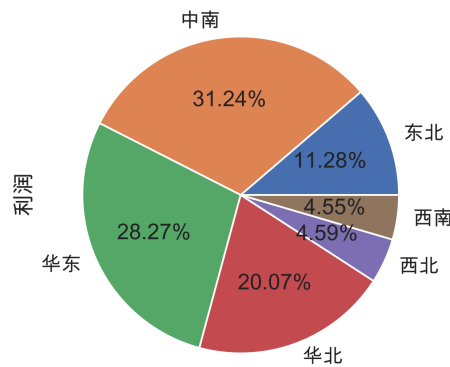


图 7: 各地区的利润占比

地区为中南的利润最多，中南、华东、华北所占的比重超过了 66.66%，达到了 77.31%。可见，即使销售额与利润成明显的线性关系，但利润在不

同地区还是呈现了稍有不同的数据表现。

### 5.5 各类别产品利润分布情况

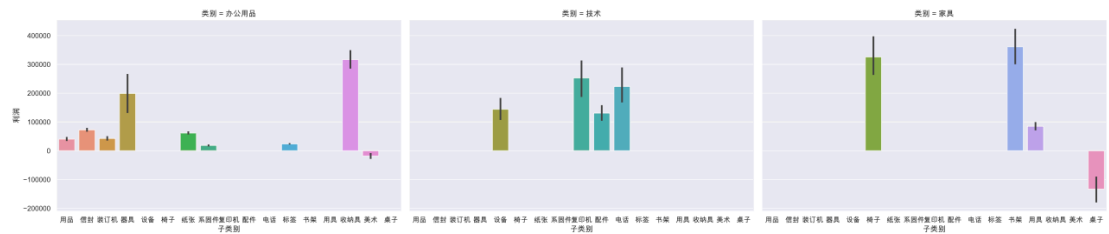


图 8: 各类别产品的利润分布

由图可见，三大类产品，办公用品利润最高，其中收纳具的利润最高，远高于其他品类，但其中美术的利润较低，家具与技术类的产品数据表现基本相似，但家具类桌子的利润相对较低。

### 5.6 订单日期及销售额

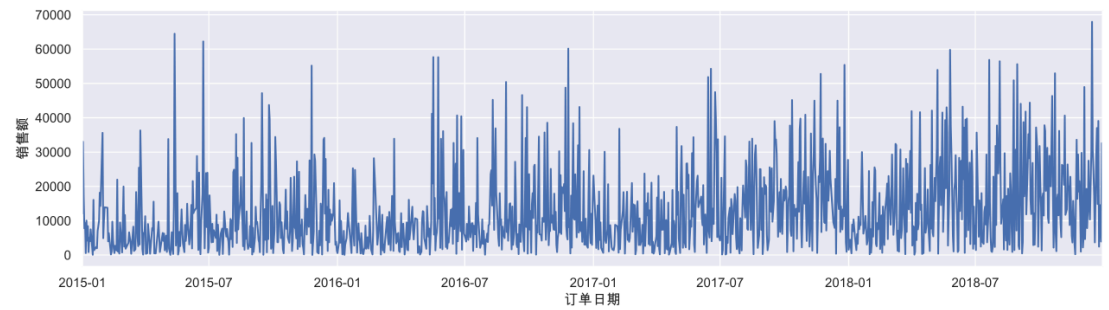


图 9: 销售额与日期的关系

由订单分布图可见，自 2015 年至 2018 年，每年 6 月与 11 月销售额将处于一个次波峰与波峰，其他时间，动态波动；



## 6 总结

结合本次样本（办公相关产品在 2015 年 2018 年数据）数据表现整体呈稳步上升趋势，结合国家的政策来看，在 2014 年下半年发布的鼓励“大众创业，万众创新”，以长江三角洲为发射点的创业之光直接映射到各领域、地区。

在未来，我认为地区地区间消费日趋均衡，高质量发展是方向，一方面，由于网络零售和零售赋能带来的普惠效应，消费市场的地域鸿沟得以逐渐弥合（在消费数量和品质方面均有体现）；另一方面，在行业端与消费端的共同驱动下，消费品质与服务体验得以持续提升，朝着高质量发展的方向前进。

## 参考文献

- [1] JiaweiHan, MichelineKamber, JianPei, 裴健, 坎伯, 韩家炜, 孟小峰, and 范明. 数据挖掘概念与技术. 机械工业出版社, 2012.
- [2] Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.
- [3] Michael Waskom and the seaborn development team. mwaskom/seaborn, September 2020.
- [4] 邱南森 and 张伸. 《数据之美: 一本书学会可视化设计》. 交通与港航, 03(v.4):65–65, 2017.
- [5] 陈为, 沈则潜, and 陶煜波. 数据可视化. 2013.