2019年"泰迪杯"数据分析职业技能大赛

A 题

超市销售数据分析

目录

	-		致 据分析职业技能大 费	
A 题	<u> </u>			1
	第一	章 问题	· 描述	1
		1.1 问题	描述	1
		1.2 论文	て的结构安排	2
	第二	章 超市勢	数据预处理	2
		2.1 数据	来源	2
		2.2 超市	数据预处理	2
		2.2.1 缺分	失属性的剔除	3
		2.2.2 异台	常数据的处理	3
			统计	
			个大类商品的销售金额	
		2.3.2 每	个中类商品的促销销售金额和非促销销售金额	4
			鲜类产品和一般产品的每周销售金额	
			位顾客每月的消费额及消费天数	
	第三		数据分析与可视化	
			分析与可视化	
			类商品和一般商品每天销售金额折线图	
			绘制各大类商品销售金额的占比饼图	
			商品和非促销商品销售金额的周环比增长率柱状图	
	第匹		画像与促销策略	
	/\ v		消费情况	
			·类商品的销售情况	
			对商品销售的影响	
	第 五		A	

第一章 问题描述

1.1 问题描述

近年来,随着国家经济的快速发展,天猫,京东,唯品会等各种电商如雨后春笋般涌现,这些电商被称之为新零售业,他们相同的特点:以用户为中心,便捷,海量数据处理能力以及物流和速度快。其按照零售轮转理论,当一种新兴的零售业态出现时,总是采取低价位、低毛利、低市场地位的三低政策,使这一新兴业态逐步走向成功。正是新零售业的快速发展,

消费者购买商品时有了更多的对比和选择,导致超市行业的竞争日益激烈,利润空间不断压缩。

超市凭借地利的优势占据着城市中心点,以覆盖式的方式辐射周边下消费者,占地面积较大,商品种类齐全,旨在于为消费者提供"一站式购物"消费场景,而且部分商品价格较低。超市的优势在互联网时代面前不断在削弱,虽不见客流量减少,但是客单价的下滑是非常严重的,由于电商的特点抢占了市场份额,导致消费者购买力下降。而且,超市面临着节假日促销活动时的排队,也不断影响着消费者购物体验。

超市的经营管理产生了大量数据,可以提升超市的竞争力,为超市的运营及经营策略调整提供重要依据。首先,我们需要从附件所给的数据中,提取客户购买物品的种类和数量,从而对销售数据进行统计分析;之后,通过对顾客什么时间段购买的商品种类分析顾客的消费行为,判断促销对销售的影响;最后分析各大类商品的销售情况,总结其销售规律。分析促销对商品销售的影响,为超市制定销售策略提供建议。

1.2 论文的结构安排

本文共分为五章,各章内容安排如下:

第一章,对论文需要解决的问题进行描述,并简单介绍整篇论文的结构安排。

第二章,针对问题一:从多角度对超市的信息进行分析,对数据预处理后,把处理后的数据按照每个大类商品的销售金额、每个中类商品的促销销售金额和非促销销售金额、每位顾客每月的消费额及消费天数、生鲜类产品和一般产品的每周销售金额进行统计。

第三章,针对问题二:将数据进行分析与可视化,通过折线,柱状以及饼图来进行销售状况分析。

第四章,针对问题三:通过数据得出顾客的消费情况,分析顾客的行为;再根据各大类商品的销售情况,总结其销售规律;并通过分析促销对商品销售的影响,为超市制定销售策略提供建议。

第五章, 总结。

第二章 超市数据预处理

2.1 数据来源

数据来源:附件(某超市的经营数据)数据时间:2015年1月1日至4月30

2.2 超市数据预处理

数据预处理是第一步,同时也是很重要的一步,数据预处理的好坏直接决定着之后数据可视化和分析等步骤能否顺利进行。在对数据进行分析之前和分析的过程中,我们逐渐对题目所给的数的认识逐步加深,并最终得出了一套较为完整的数据预处理流程。

2.2.1 缺失属性的剔除

通过对数据进行观察,我们发现数据集并不是每个属性都是被正常记录的。表 1 就是对数据进行处理后的结果,可得出有两行数据中的销售数量存在缺失值,记为 NAN。这 2 个属性的记录是缺失的,不能为我们提供有价值的信息。因此,我们选择将这 2 个属性从数据中删除。

	顾客编 号	大类编 码	大类名 称	中类编码	中类名称	小类编 码	小类名 称	销售日期	销售月 份	商品编码	规格型 号	商品类型	单位	销售数 量	销售金 额	商品单价	是否促 销
8180	935	21	冲调	2106	蜂蜜/面包酱	210601	蜂蜜	20150121	201501	DW-2106010189	牛魔空 版	12g*8	- 般	NaN	1.0	9.9	9.9
26902	1978	21	冲调	2106	蜂蜜面 包酱	210601	蜂蜜	20150315	201503	DW-2106010189	牛魔空 版	12g*8	般	NaN	1.0	9.9	9.9

表 1 超市数据中销售数量缺失

2.2.2 异常数据的处理

附件中的有些信息虽然被记录,但是不符合实际情况的数据都是异常值。如表 2 所示,其中有编号为 15 的顾客买了散称的凉拌素食 0.001kg 花了 0.02 元,即买了 2 分钱的散称的 1 克熟食,这显然不符实际情况,是个异常值,通过限定销售数量的的范围过滤掉异常的数据。

	顾客编 号	大类 编码	大类 名称	中类编 码	中类名 称	小类编 码	小类名称	销售日期	销售月 份	商品编码	规格 型号	商品 类型	单 位	销售数 量	销售金 额	商品单价	是否 促销
93	48	12	蔬果	1201	蔬菜	120105	鲜调味	20150101	201501	DW-1201050012	散称	生鲜	千克	0.064	0.23	3.56	否
95	31	12	蔬果	1205	其它加工	120501	咸菜干货	20150101	201501	DW-1205010005	散称	生鲜	kg	0.072	0.14	2.00	否
189	15	13	熟食	1301	凉拌熟 食	130101	凉拌素食	20150101	201501	DW-1301010076	散称	联营 商品	kg	0.001	0.02	20.00	否
232	31	12	蔬果	1202	加工豆类	120204	豆干豆皮	20150101	201501	DW-1202048059	散称	生鲜	千克	0.080	0.62	7.76	否
294	44	22	休闲	2202	糕点	220208	散称糕点	20150101	201501	DW-2202080013	散称	一般商品	kg	0.086	1.70	31.80	是

表 2 超市数据中销售数量异常

2.3 数据统计

2.3.1 每个大类商品的销售金额

对超市中每个大类的数据进行合并,最后可得出 15 个大类商品数据,再将他们各自的销售金额进行求和,就得出了表 3 所展示的数据。通过表 3 的数据可以看出,大类商品中日配、蔬果、粮油和肉禽等销售金额较大,烘培、家居、文体等商品销售金额较小。像日配等商品作为日用品和食品,需求量较大,故销售金额大;而家居、烘培等产品可能需求量较少,故

	0	1	. 2	3	4	5	6	7	8	Ç) 1) 1	1 12	2 13	14
name 1	休闲	冲调	家居	家电	文体	日配	水产	洗化	烘焙	負煤	粮油	肉禽	蔬果	酒饮	针织
totalsale	74145.2	13955.6	6311.1	853.9	1970.3	81958.3	2891	38013.8	110.9	5939. 94	60931.9	5 25197.6	7 81375.79	54790.9	5765.9

表 3 每个大类商品的销售金额

2.3.2 每个中类商品的促销销售金额和非促销销售金额

下表 4 是每个中类商品的促销销售金额和非促销销售金额(左侧的数据为促销时候的销售金额,右侧的数据为非促销时的销售数据)。不难看出,促销时的销售金额多于非促销时的销售金额,促销是对商品的销售有一定的帮助。但通过数据可以观察到并非商品促销的时候,其销售金额就一定大,像塑料餐具,巧克力等不同物品,他们的非促销时金额就远远大于促销时的金额,可以知道该类商品在平日里就销售的不错。

国产白酒	8816.5	11699
地垫类	154.8	15
塑料餐具	19.9	224.9
外送面包制	925.4	668.16
女上装	243.3	
女下装	149.6	
女内衣裤	60.6	732.3
女孩玩具	49	62
女袜	62.3	355.3
女鞋	34.8	
奶粉	185.8	1556.9
婴幼儿用品	14.8	704.4
季节性电器	25	76
室内清洁剂	39.6	73.9
巧克力	566.42	4094.76
常温乳品	24837.5	13753.4
常温肉制品	2334	1685.4
床品单/套	302.1	
打扫用品	186	1030.7
拖鞋	186.1	272.9
料理切具	55.3	544.5

表 4 每个中类商品的促销销售金额和非促销销售金额

2.3.3 生鲜类产品和一般产品的每周销售金额

表 5 为生鲜类产品和一般产品的每周销售金额。通过生鲜销售额和一般商品销售额数据的对比,知道一般商品销售额明显大于生鲜销售额,百姓对一般销售额的需求明显多余生鲜销售额。生鲜销售额的数量在 700~10000 范围之内,一般商品销售额范围在 2000~20000 范围。第 4 周为 1 月底,临近春节,百姓提前购置物品,需求量大,都达到了 18 周内销售额最大值,18 周不足一周,只有工作日的前四天,都在忙碌工作,没有过多的时间购买商品,所以两者的销售额都偏低。

1		周数	生鲜销售额	一般商品销售额
2	0	1	6146.66	17369.98
3	1	2	6033.68	18245.82
4	2	3	8393.59	18651.03
5	3	4	9975	20430.03
6	4	5	4598.95	29379.5
7	5	6	7231.42	29001.47
8	6	7	7868.1	46867.61
9	7	8	2897.32	13802.41
10	8	9	4983.32	13853.06
11	9	10	6606.1	14562.27
12	10	11	8596.53	14796.51
13	11	12	6057.88	16163.11
14	12	13	5187.37	13356.1
15	13	14	6582.62	17929.87
16	14	15	6768.69	14196.34
17	15	16	6172.72	17097.76
18	16	17	6972.33	18199.04
19	17	18	736. 42	2627.83

表 5 生鲜类产品和一般产品的每周销售金额

2.3.4 每位顾客每月的消费额及消费天数

下表 6 所展示的是编号为 0-10 的顾客每月的消费额及消费天数,可以知道月数决定了顾客消费的金额和消费天数。顾客都有相同的特点,一月消费天数多,消费金额大。一月是过春节的时候,需求量大,购买的物品多,而其他月份顾客的消费额就很低。可以通过这个表 6 知道哪些月份是销售的旺季,还可以了解到哪些顾客消费额大,购买的商品多,需求大。在表 6 中,多数顾客的二三四月份销费额或者消费天数为 0,考虑到时间因素,可能是顾客这些月份工作忙碌或者外出,故不能消费。

顾客编号 -	一月消费天数	一月消费额	二月消费天数	二月消费额	三月消费天数	三月消费额	四月消费天数	四月消费额
0	4	11.05	0	0	0	0	1	13.6
1	4	12.3	3	30.3	0	0	0	0
2	12	48.7	10	30	0	0	0	0
3	9	67.39	17	360.97	16	68. 87	10	49. 28
4	4	27.9	0	0	0	0	14	154. 94
5	7	57.6	0	0	4	33.5	4	48.34
6	3	23.99	8	65.67	1	11.5	9	49.14
7	6	138.37	0	0	8	61.49	11	52. 07
8	12	190.49	0	0	11	111.85	5	74. 72
9	9	80.4	0	0	0	0	0	0
10	16	220.63	0	0	11	117.03	0	0

表 6 编号 0-10 的顾客每月的消费额及消费天数

第三章 超市数据分析与可视化

3.1 数据分析与可视化简介

数据可视化是指以饼状图等图形的方式展示数据。由于人脑对视觉信息的处理要比书面信息容易得多,使用图表来总结复杂的数据,可以确保对关系的理解要比那些混乱的报告或电子表格更快,这帮可以助用户能够更快地识别模式。交互式数据可视化鼓励用户探索甚至操纵数据,以发现其他因素。这就为使用分析提供了更好的意见,并且交互式可视化能够让决策者深入了解细节层次。这种展示方式的改变使得用户可以查看分析背后的事实。

3.2 生鲜类商品和一般商品每天销售金额折线图

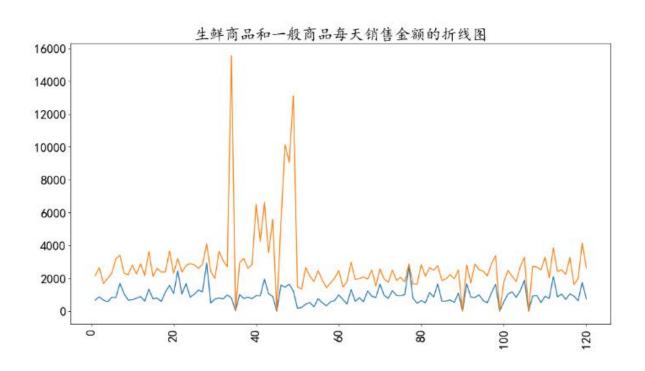


图 1 生鲜类商品和一般商品每天销售金额折线图

分析比较两类产品的销售状况:

图 1 中黄色折线是一般商品每天的销售金额;蓝色折线是生鲜商品每天的销售金额;横轴是时间轴,代表天数;纵轴是指每天销售的金额。观察到黄色折线即一般商品的每天销售金额多于生鲜商品的销售金额,生鲜商品每日的销售金额 1000~2000 范围之间,较为均匀,,没有很大的变化,一般商品的黄色折线在 35 到 45 天数之间变化波动非常大,因为日期从 1 月 1 日算起,35 到 45 的天数恰为春节期间,酒水饮品的购买量大,故一般商品的销售金额大。其余时间较为平均。生鲜商品不受时间或假期的影响,而一般商品受时间的影响较大。

3.3 按月绘制各大类商品销售金额的占比饼图

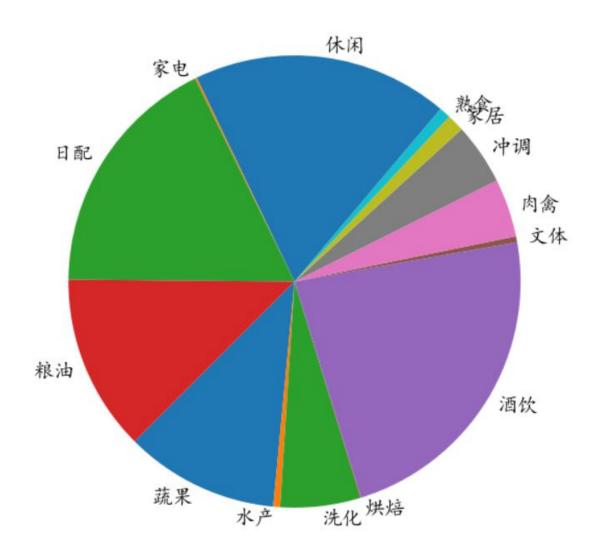


图 2 1月各大类商品销售金额的占比饼图

3.3.1 1月销售情况分析:

上图 2 形象的展示了 1 月各大类商品销售金额占比,在该饼图中以休闲、粮油、蔬果、酒饮和日配为主要部分,接着就是洗化、肉禽和冲调,而水产、文体、家电、熟食、烘培,家居商品占比最少。像文体,家居等已经购买过的商品或者需求量较小的商品销售金额少。 2 月份是春节期间,根据百姓的习惯,会提前购置春节所需的商品,例如酒饮品等。

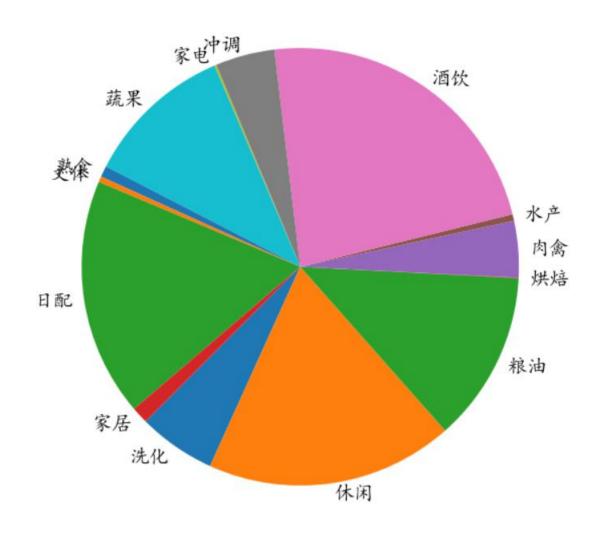


图 3 2月各大类商品销售金额的占比饼图

3.3.2 2月销售情况分析:

上图 3 是 2 月各大类商品销售金额的占比饼图。2 月的占比情况大致和 1 月情况相差不大,总的还是休闲,粮油等日用品食品占比重大,文体等不常用商品占比重小。2 月为春节期间,有些商品销售金额比重减小,例酒饮和粮油稍许减少,休闲和肉禽增大了销售金额比重,其余的像日配,水产等商品基本不变。

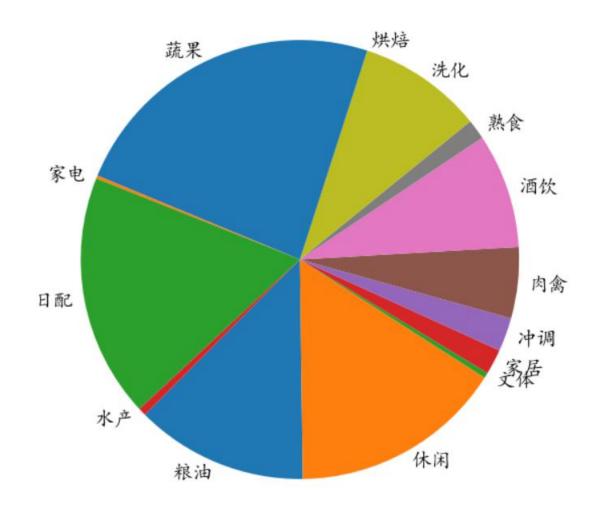


图 4 3 月各大类商品销售金额的占比饼图

3.3.3 3月销售情况分析:

图 4 是 3 月超市个大类销售金额情况,3 月的比重和 1,2 月比重显然不同,以休闲,粮油,日配和蔬果为大比重,酒饮在过了节假日后明显的减小,以洗化,家居为首的商品增大,其他的不常用品则和之前一样,销售金额比重小。

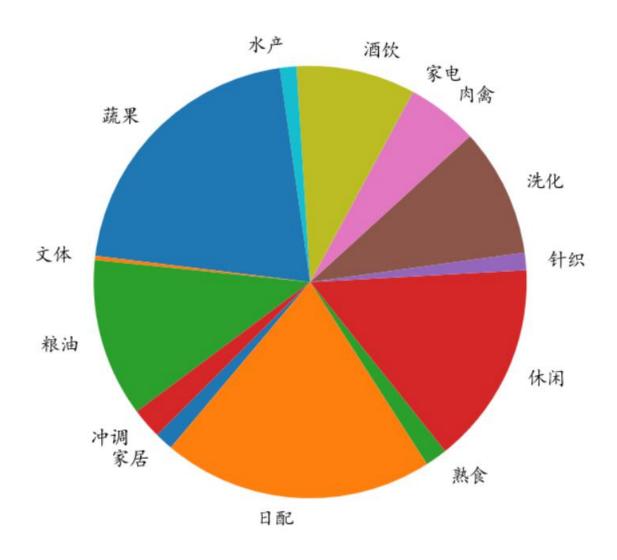


图 5 4月各大类商品销售金额的占比饼图

3.3.4 4月销售情况分析:

4月各大类商品销售金额的占比饼图在上图 5 所表现。3月和 4月比重部分偏差不大,仍然 是粮油,日配等常用品占大比重,水产,针织商品逐渐增大,酒饮商品愈来愈小。

3.4 促销商品和非促销商品销售金额的周环比增长率柱状图

图 7 是促销商品销售金额的周环比增长率柱状图,图 8 是非促销商品销售金额的周环比增长率柱状图。横轴是日期周数,纵轴是增长率。利用了图 6 的公式,分别计算得出促销和非促销时销售金额的环比增长率。第 7,9,11 周的增长率为负,意味着其在变小,相较于非促销时期,促销时期的增长率明显增大,但仍有几周是减小的,并且第六周无明显变化。(18 周只有几天,异常值较大,故没有考虑在范围之内。)

环比增长率 = 本次数据 - 上次数据 × 100% 上次数据

图 6 环比增长率公式

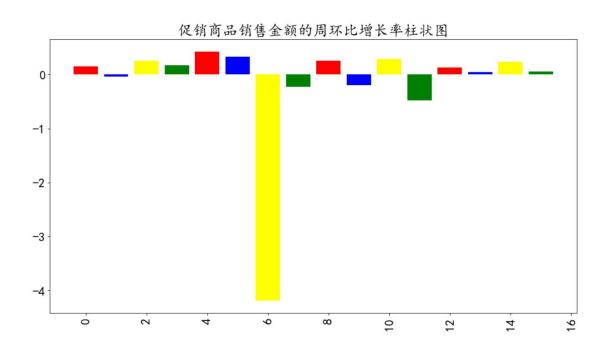


图 7 促销品销售金额的周环比增长率柱状图

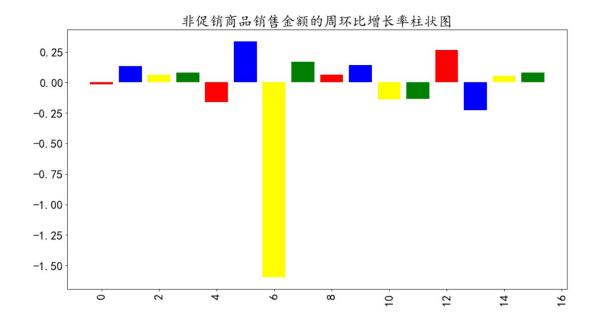


图 8 非促销品销售金额的周环比增长率柱状图

第四章 用户画像与促销策略

4.1 顾客消费情况

4.1.1 用户画像介绍

用户画像又称用户角色(Persona),作为一种勾画目标用户、联系用户诉求与设计方向的有效工具,用户画像在各领域得到了广泛的应用。我们在实际操作的过程中往往会以最为浅显和贴近生活的话语将用户的属性、行为与期待联结起来。作为实际用户的虚拟代表,用户画像所形成的用户角色并不是脱离产品和市场之外所构建出来的,形成的用户角色需要有代表性能代表产品的主要受众和目标群体。

一般的,用户画像在产品没有上线、市场前景较为模糊、产品需求还需探索的阶段,定性化的用户画像能有效地节省时间、资源,在较短的时间通过桌面研究、访谈等定性化的方法来获得用户画像是一种比较可行和最优的方式。而事实上,用户画像是一种能将定性与定量方法很好结合在一起的载体,通过定量化的前期调研能获得一个对于用户群较为精准的认识,在后期的用户角色的建立中能很好地对用户优先顺序进行排序,将核心的、规模较大的用户着重突出出来。定性化的方法虽然无法对不同单位的特征作数量上的比较和统计分析,但能对观察资料进行归纳、分类、比较,进而对某个或某类现象的性质和特征作出概括,在角色建构的过程中定性化的方式能获得大量用户的生活情境、使用场景、用户心智等资料,进而形成活生生的用户类型。基于后台数据的支持和挖掘,可以用户画像选择将定量化和定性化方法相结合来创建用户画像。

用户画像是在创造一系列的"典型"或者"象征性"的用户,但用户画像的一个更高层次的功用在于使用用户画像融合边缘情况的行为或需求。

4.1.2 前十顾客消费情况数据分析和用户画像

表7是消费额前十的消费金额数据,图9是10位顾客每个人的消费情况柱状图。表7可以清楚的知道顾客消费金额,图9则是可以很直观的将每位顾客消费额画在轴上表达出来。

1	顾客编号	销费金额
2	12	1838.19
3	52	3589.15
4	74	1660.37
5	108	2340.32
6	210	1989.73
7	372	1505.88
8	395	1820.20
9	986	2611.24
10	1177	13597.06
11	1385	2451.50

表 7 顾客消费情况

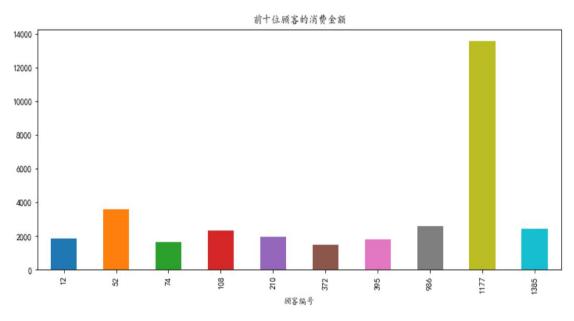
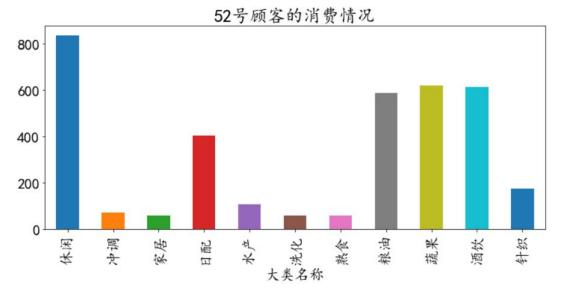
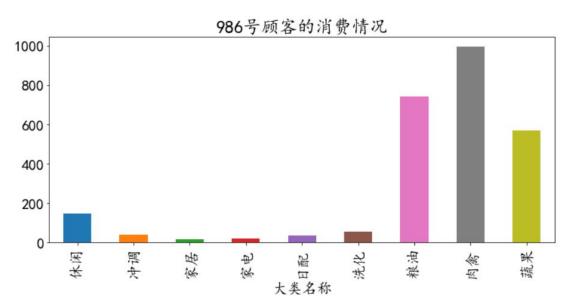


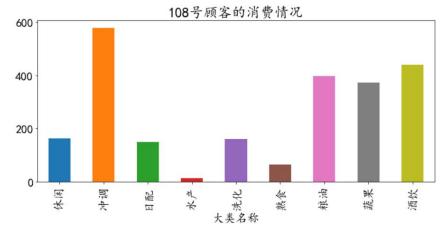
图 9 顾客消费情况柱状图



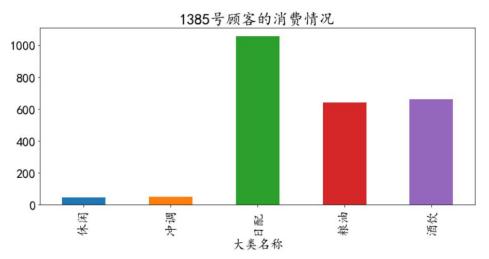
52 号顾客看其购买的物品多为休闲,粮油,蔬果等日常用品,且其休闲品多于粮油等商品,可以推测其为青年,为自己购买每日所需的用品和自己所需要的休闲品。



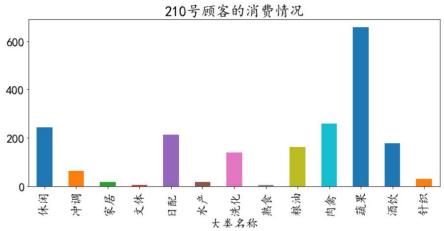
986 号顾客以肉禽,粮油,蔬果为主,其余商品很少,可猜测其实家庭妇女,照顾家庭购买商品。为这类顾客可以推荐日常所需物。



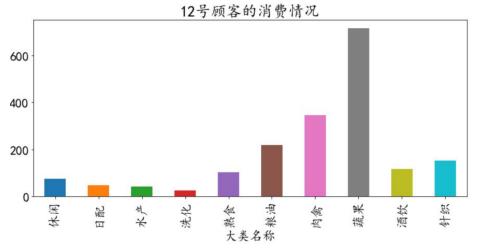
108 号顾客也是以粮油,蔬果为主,但是其冲调消费和洗化消费情况大,可以猜测其家庭有孩子,需要奶粉,纸品。为该类顾客推荐婴儿等商品。



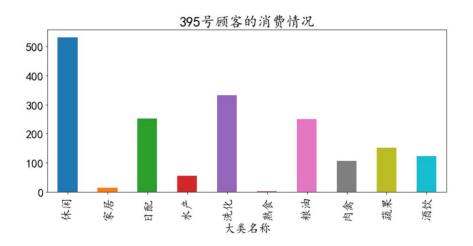
1385 顾客除了日常品之外,推测其爱喝酒,其酒饮商品的消费基本与粮油持平,可以 为该类顾客推荐酒类饮品。



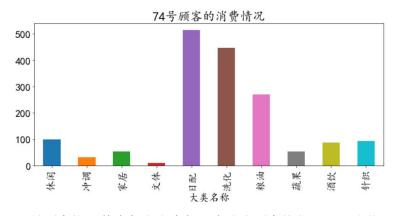
210 顾客除了日用品驻外,蔬果购买多,很喜欢酒饮和蔬果。可以为此类顾客推荐水果一类商品。



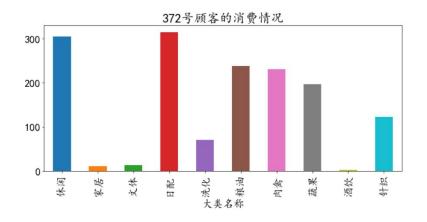
12 号顾客同样得也是蔬果消费大,也为此类顾客推荐蔬果商品。



395 号顾客推测其为女性。除了家居,水产和熟食外,其他几类商品购买量大,尤以休闲和 洗化最多。



74号顾客推测其为都市女青年。为该类顾客推荐日配,洗化,粮油等日用品。



372 号顾客为家庭购买粮油,肉禽,休闲,蔬果等商品,可以为该类主推家庭所需用品。

4.2 各大类商品的销售情况

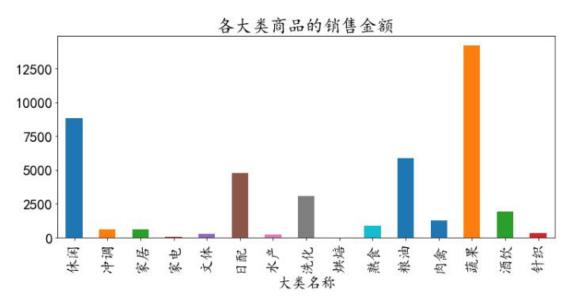


图 10 各大类商品的销售金额柱状图

销售情况分析及销售规律:

情况分析:各大类商品的销售金额柱状图如上所示,以蔬果,休闲,粮油,日配,洗化,酒水,肉禽等商品依次减少,前者多为日常所需用品,后者少的销售金额为不常购买品,多数为家中已有商品,需求量小。

销售规律:综合图 2,图 3,图 4,图 5以及图 10,可以知道在春节等节假日酒饮商品和休闲销售大,在 3 月和 4 月时候,休闲和酒饮销售减少,日配,粮油,蔬果,洗化都为日常用品,无论是在节假日还是平常,变化不大。

4.3 促销对商品销售的影响和建议

图 **11** 是促销对商品销售的影响,通过柱状图的明显对比,可看出促销对销售额的影响巨大,当促销时达到了 **30000+**,而非促销时仅为 **10000**,两者差距明显,促销可以帮助超市销售额有很大的帮助。

超市可能有端架、堆头陈列不规范,单品多而杂,无较明显且与之对应的 POP,标示太乱、不明显;卖场杂物乱堆、散货随处乱放,商品卫生普遍较差,商品排面维护不到位;仓库商品摆放无序、随意性较大,散货、坏货乱堆乱放,未划区标明摆放;排队时间久;可能人员专业知识差等等一系列原因。

建议:通过上述数据分析,可在节假目前的时间多添置酒饮和休闲商品;确保仓库中日配等日用品数量充足,为不同的顾客推荐他们常购买的商品;还可以开通网页让顾客在网上选择商品,超市送货到家;增加商品多样性;向农贸市场学习,增加蔬果,生鲜等商品,时刻保持商品的新鲜;给厂家联系,让厂家大力支持,这样也会宣传知名度也可提高销售,你要经常与厂家代表不断沟通,不断给厂家要政策,容易提高销售。

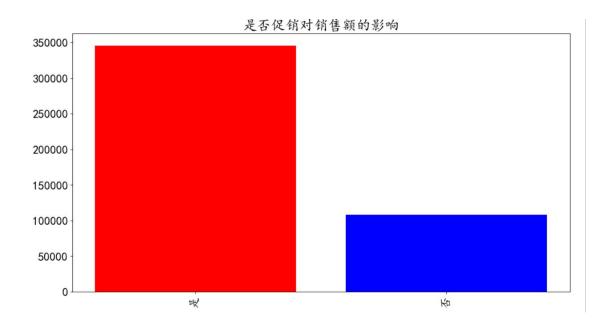


图 11 促销对商品销售的影响

第五章 总结

超市分为几大块,首选搞明白,销售量大的主要是菜和生活用品,你可以把往年销售量统计出来,今年可以写出怎么搞促销活动,首选要给厂家联系,让厂家大力支持,这样也会宣传知名度也可提高销售,你要经常与厂家代表不断沟通,不断给厂家要政策,这样很容易提高销售。

按照不同的时间,按月份周数进行划分,并根据不同类型的顾客介绍不同的商品,并且利用超市凭借地利的优势占据着城市中心点,以覆盖式的方式辐射周边下消费者,适当的促销让周围的顾客和消费者时常光顾超市,提高知名度,保证商品高质量,让顾客感到满意,从而增加市场份额,让消费者购买力上升。