

京东消费者行为数据分析

1.认识表中的数据

Category	Field	Definition
User	customer_id	Unique customer id
	age_range	Age range
	gender	Gender
	customer_register_date	Register date
	customer_level	Membership level
	city_level	Member city level

Category	Field	Definition
User Purchase Behavior	action_date	Action date
	action_id	Action id
	type	Action type

Category	Field	Definition
Product/Merchant	brand	Brand
	shop_id	Shop id
	category	Product category
	product_market_date	Product market date
	vender_id	Vender id
	fans_number	Number of fans
	vip_number	Number of VIP
	shop_register_date	Shop register date
	shop_category	Shop category
	shop_score	Shop score
	product_id	Product id

2.分析

1) What: 用户的购买趋势是怎样的?

a)选择什么变量?

1.购买行为: type

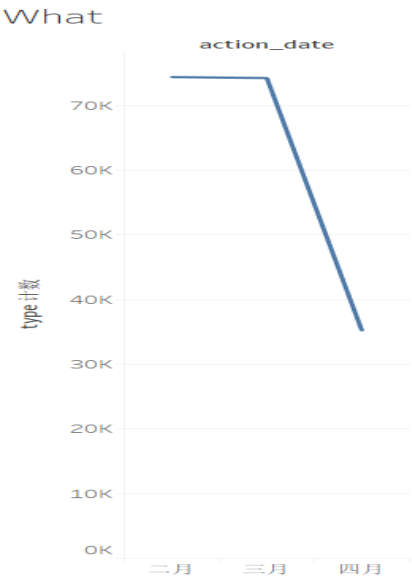
2.行为时间: action_date

b)呈现怎样的数据关系?

趋势关系

c)可以选择怎样的图表?

折线图
答：



分析思路：先看下 2~4 月以来用户购买行为的整体趋势是怎样的？将 type 拖拽至行，对 type 进行总体的计数，同时将 action_date 拖拽至列，然后选择“月”，最后看下总体趋势。

结果：发现这几个月的趋势是下降的，尤其 3~4 月下降非常明显。为了具体弄清出可能潜在的用户购买行为下降原因，又对每个月的每一天进行趋势分析。



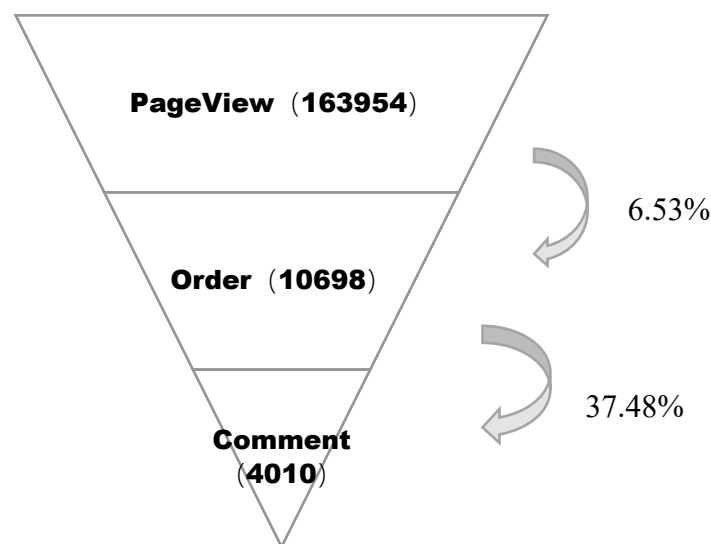
结果：2月呈现先下降后上升趋势，总体是下降的，2.15这天达到最低。
3月整体保持平衡，但在3.27下降非常急剧，可能是系统后台的原因导致用户无法进入页面。
4月依然保持先下降后上升趋势，总体还是下降的。
因为用户购买行为整体下降比较明显，因此我们需要进一步分析影响用户购买行为的因素。

2) Why: 影响用户购买行为的因素都有哪些?

a) 选择什么变量?

1. 购买行为: type
2. 影响因素可思考维度:
 - a. 用户维度: 年龄、性别、城市等维度
 - b. 商家维度: 商家评分、粉丝数、商家类别等维度
- c. 漏斗分析:

因为 follow (关注) 和 savedcart (保存购物车) 不一定和用户 order (下单) 行为挂钩，所以漏斗分析暂不列这两项。可以看出 PageView 由于是第一步，所以在整个可视化分析中，该步骤发生次数也是最多的。从 PageView 到 Order 的转化率只有 6.53%，不是特别高，而从 Order 到 Comment 的转化率有 37.48%，这个转化率还是可以的，考虑到可能商家对用户评论这个行为有奖励措施。这两个转化步骤都是比较重要的，从 PageView 到 Order 是直接关系到商铺的收入，因此这个指标是商家重点关注的，转化率不高需要采取一些措施；而好的 Comment 可以及时帮助商家改进和对其他用户在下单行为前有一定参考意义，故 Order 到 Comment 转化也是很重要的，可以适当地想办法再提升一点。



b) 呈现怎样的数据关系?

比较关系

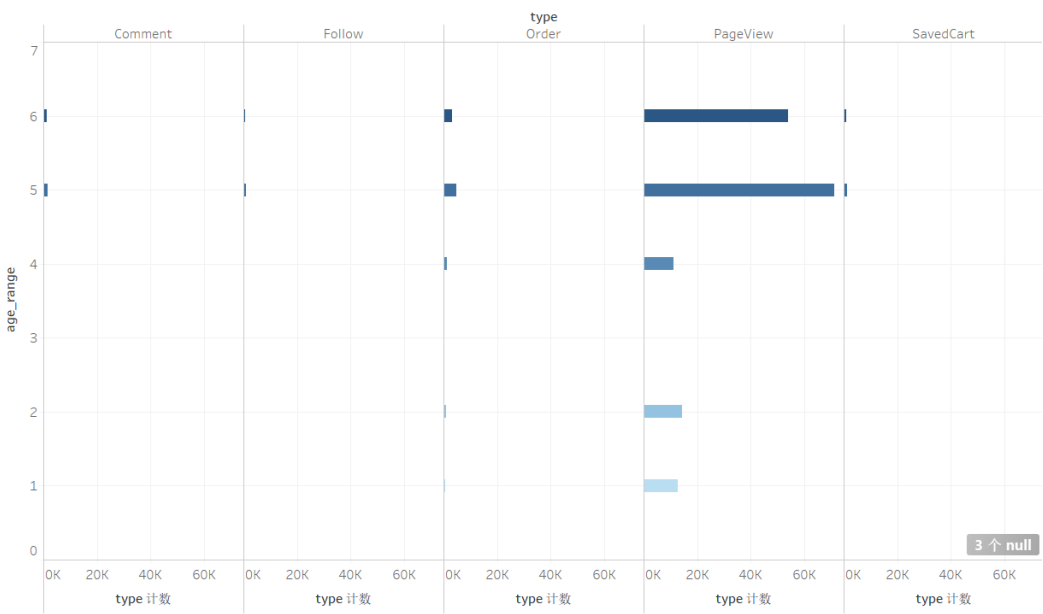
c) 可以选择怎样的图表?

条形图、柱状图

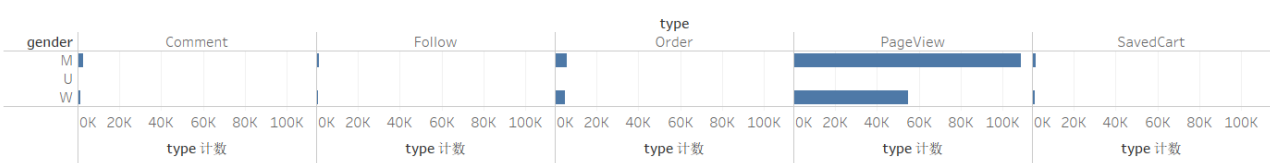
答:

用户角度:

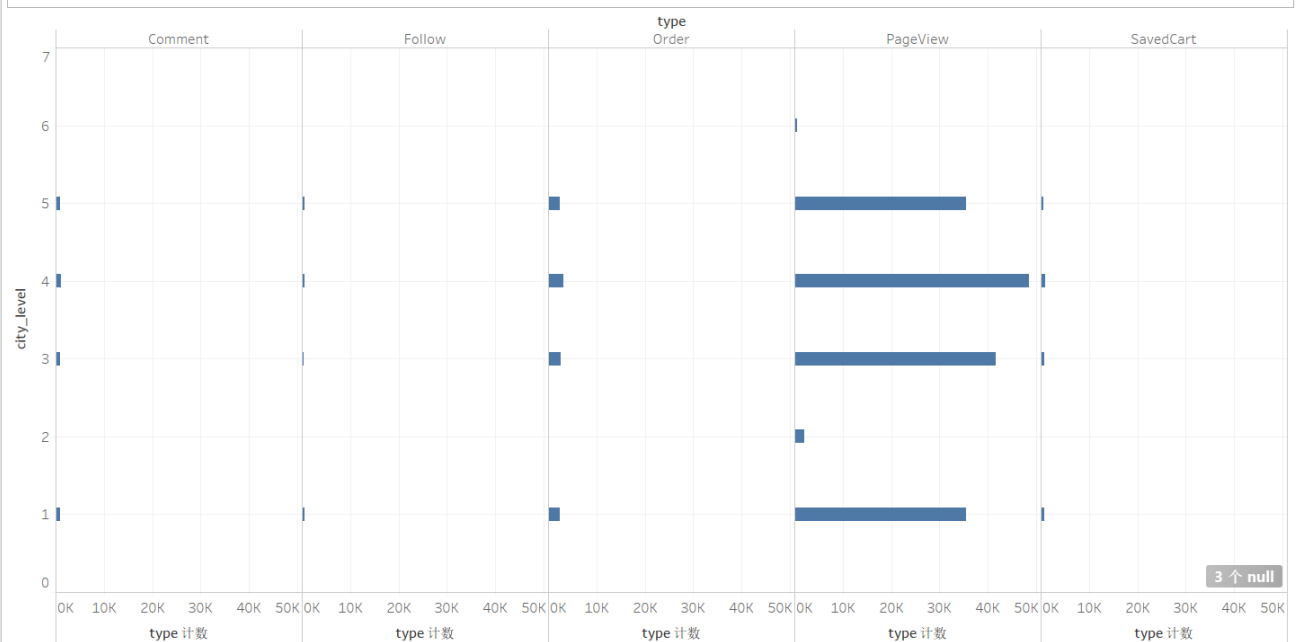
How-用户年龄分段



How-用户性别



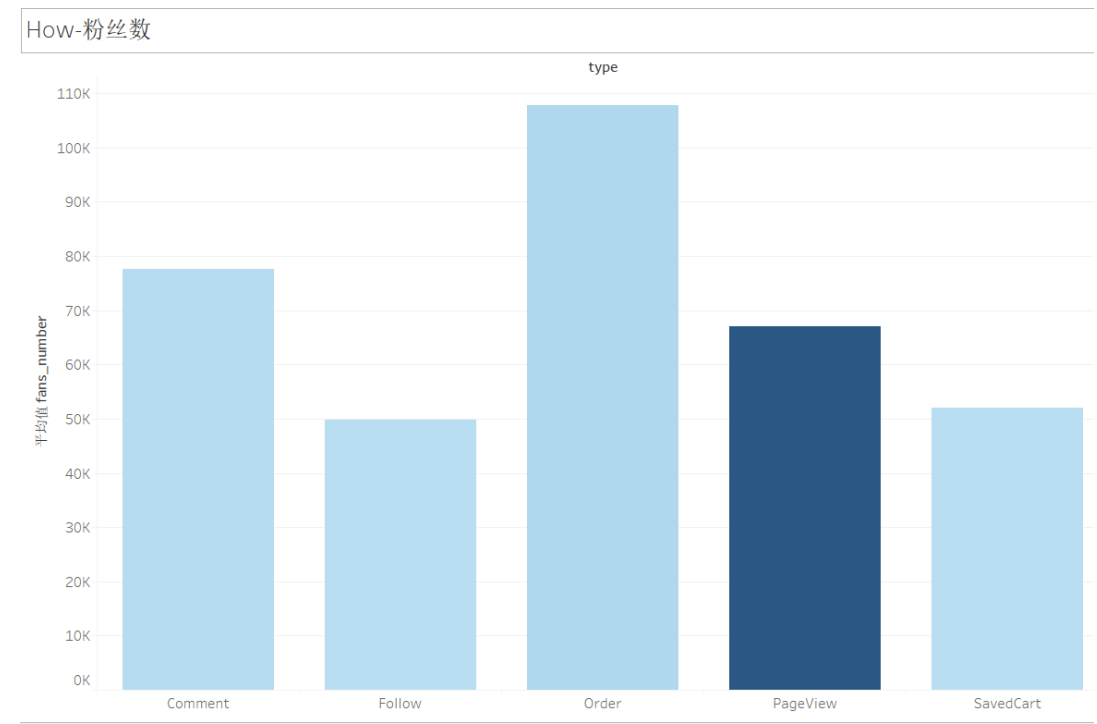
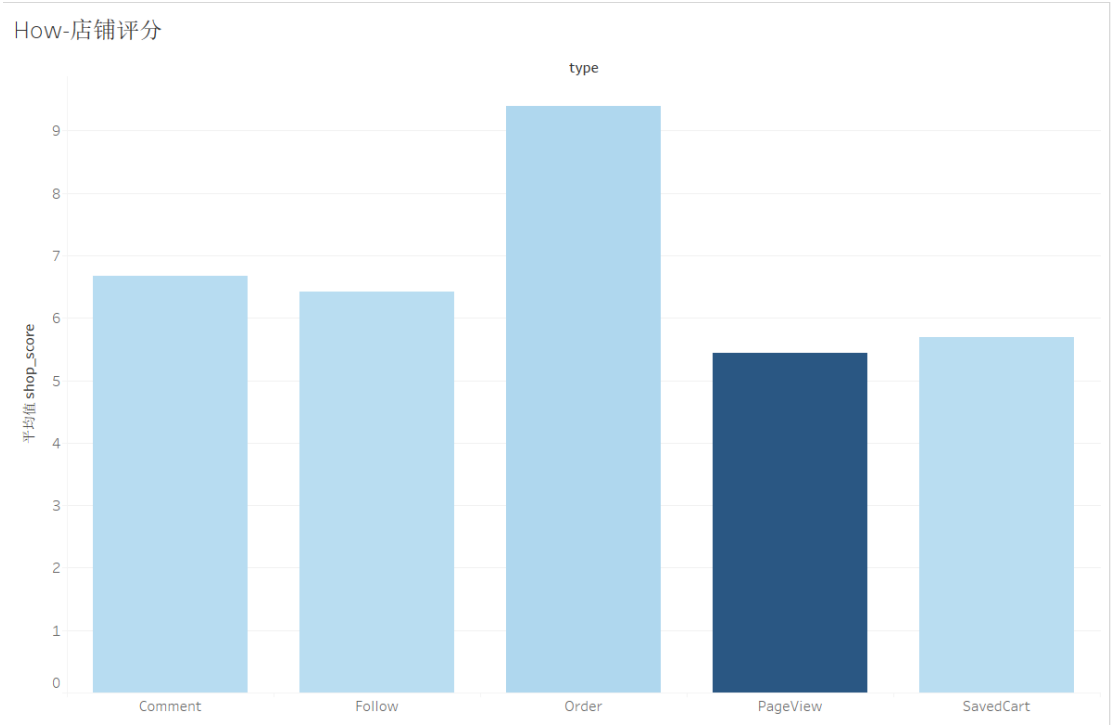
How-用户城市

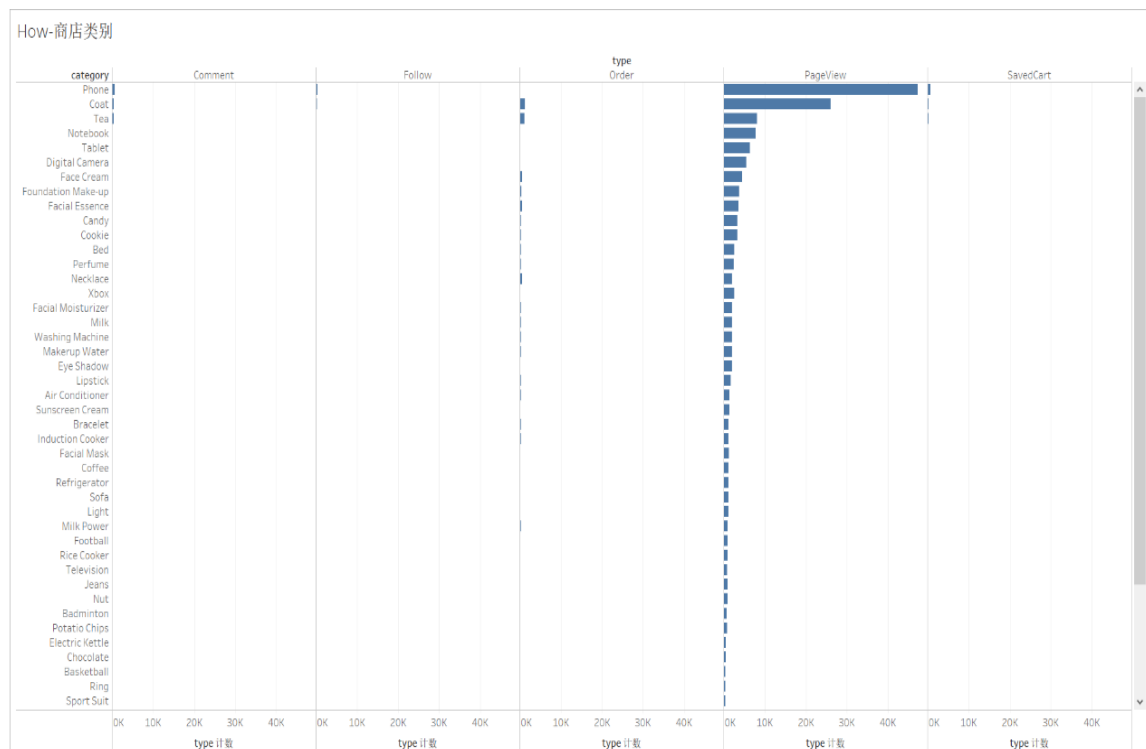


分析思路：从用户角度，我主要选择了用户年龄分段、用户性别和用户城市这三个主要维度进行分析，看各行为类别的发生次数。

结果：PageView 的次数都是最多的。用户年龄分段 5 的购买行为发生次数最多，用户年龄分段 3 最少；男性用户购买行为发生次数比女性用户多；城市级别 4 的

用户购买行为最多，城市级别 6 最少。
商家角度：





分析思路：从商家角度，我主要选择了商铺评分、粉丝数和商铺类别这三个主要指标进行分析，看各行为类别的发生次数。

结果：PageView 的次数都是最多的。用户各购买行为均普遍发生在商铺评分 5 分以上，其中，评分越高的用户选择下单的几率越大；同样，商铺粉丝数越多，用户选择下单的几率也越大；而“电话”、“大衣”和“茶”是用户购买行为发生次数最多的前三个商铺类别，推测这段时间用户对这三种类别的产品可能需求更高。

漏斗分析：以上。

3) How: 京东的业务部门可以怎么做，提升订单量？

a)问题拆解？

京东的用户都由哪些群体组成？他们具有怎样的特征？

b)模型选择？

1.基于你的问题，你会选择哪个数据分析模型？

聚类分析模型

c)变量选择？

1.你会将哪些变量放进模型中？

用户的性别、年龄分段、城市级别

d)模型建立与解读？

1.你会如何建立你的模型？关注什么输出的结果？

用 tableau 建立聚类模型，关注群集分类（描述）。

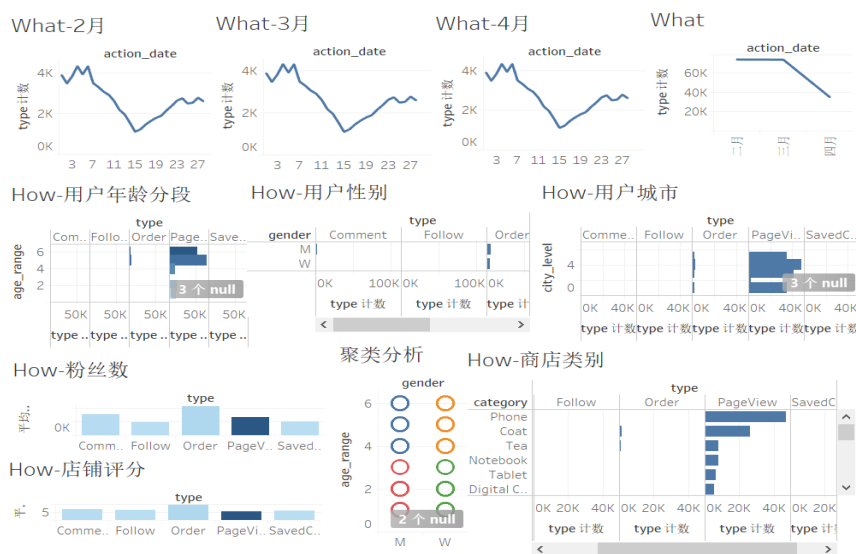
2.使用 Tableau 进行模型建立，并尝试进行业务解读。

答：

分析思路: 将 `gender` 拖拽至列, `age_range` 拖拽至行, 然后取消“聚合度量”, 在“分析”中选择“群集”拖拽至画布, 在弹出的窗口将需要分析的变量放进去, 这里我又将 `city level` 放进去, 一共创建 4 个群集。

群集	项数	平均值	age_range	city_level	总计	最常见 gender
群集 1	98018	5.3637			3.2799	M
群集 2	54999	5.2003			3.3082	W
群集 3	22036	1.5432			3.3011	M
群集 4	8246	1.4932			3.3096	W
未建立群集	56					

结果：可以看到一共有 4 个主要的群集，分别是年龄分段=5、城市级别=3 的男性群体，年龄分段=5、城市级别=3 的女性群体，年龄分段在 1-2 之间、城市级别=3 的男性及女性群体。



业务解读：通过对京东用户群体进行聚类分析，发现用户主要集中在 2 个群集，分别是年龄分段=5、城市级别=3 的男性集体和年龄分段=5、城市级别=3 的女性集体，因此可以结合这个年龄段的用户特征和这个城市级别的用户特征对用户进行相应的推广和营销策略（因为在这里不清楚年龄分段=5 和城市级别=3 的具体含义，故不展开详述）。