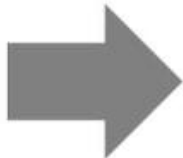


京东消费者分析

Cassie Zhang

Field	Type	Definition
customer_id	Bigint	客户编码
product_id	Bigint	产品编码
action_date	Date	行为时间
action_id	Bigint	行为编码
type	String	行为类别
age_range	Int	年龄分段
gender	String	性别
customer_register_date	Date	客户注册日期
customer_level	Int	会员级别
city_level	Int	会员城市级别
brand	String	产品品牌
shop_id	Bigint	店铺编码
category	String	产品类别
product_market_date	Date	产品上市日期
vender_id	Bigint	商家编码
fans_number	Int	粉丝数
vip_number	Int	会员数
shop_register_date	Date	开店时间
shop_category	String	店铺主营
shop_score	Float	店铺打分



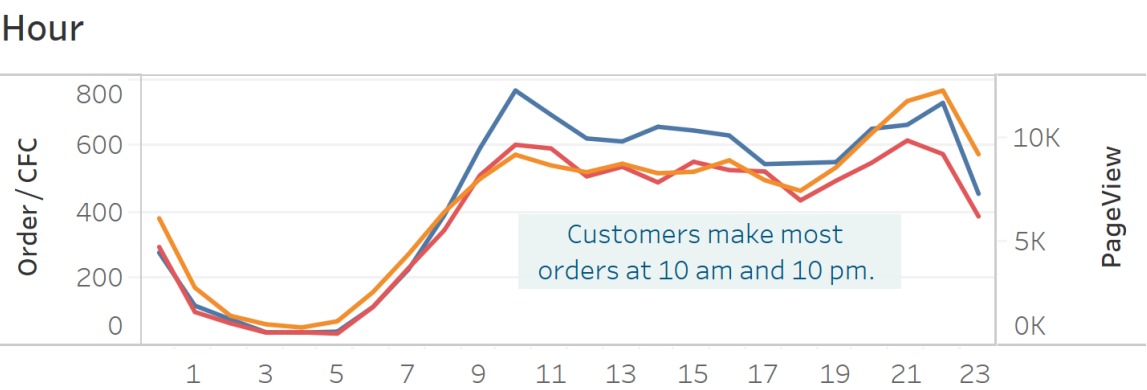
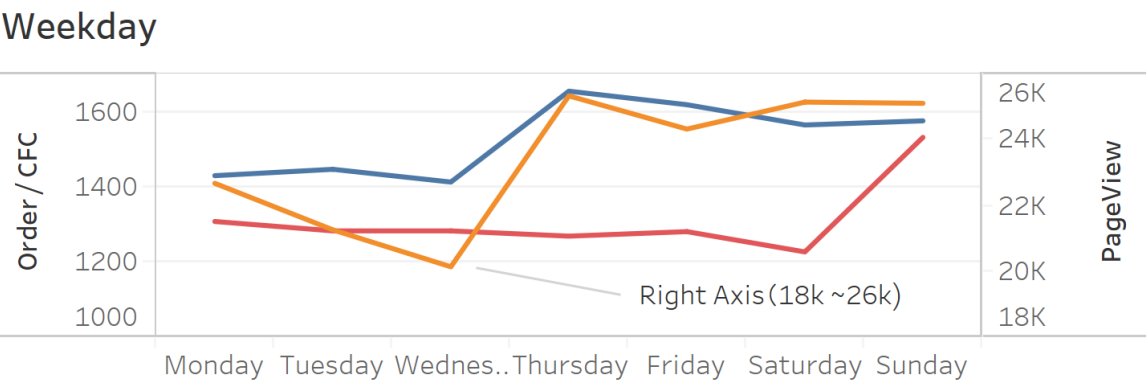
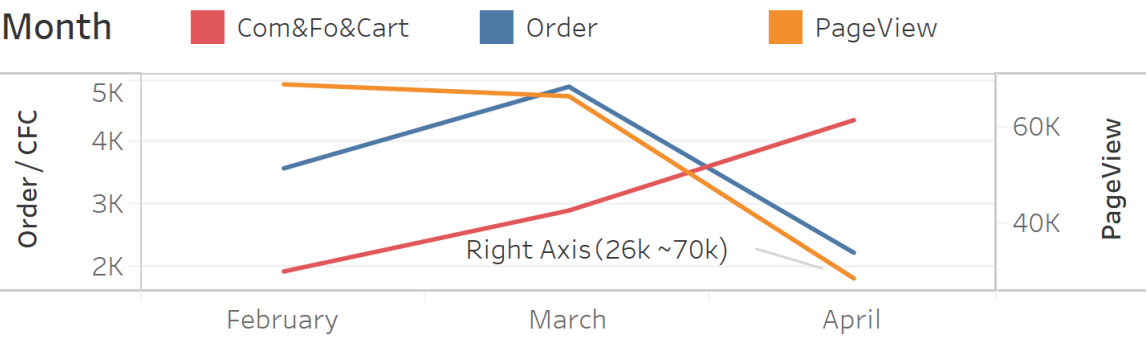
类别	Field	Definition
用户维度	customer_id	客户编码
	age_range	年龄分段
	gender	性别
	customer_register_date	客户注册日期
	customer_level	会员级别
	city_level	会员城市级别
购买行为	action_date	行为时间
	action_id	行为编码
	type	行为类别
产品/店铺	brand	产品品牌
	shop_id	店铺编码
	category	产品类别
	product_market_date	产品上市日期
	vender_id	商家编码
	fans_number	粉丝数
	vip_number	会员数
	shop_register_date	开店时间
	shop_category	店铺主营
	shop_score	店铺打分
	product_id	产品编码



What: 用户的购买趋势是怎样的？

- 选择什么变量？
 - 购买行为: type
 - 行为时间: action_date
- 呈现怎样的数据关系？
 - 趋势关系
- 可以选择怎样的图表？
 - 折线图
 - 热力图

用户购买趋势分析



Daytime

Hour ..	Monday	Tuesday	Wednesd..	Thursday	Friday	Saturday	Sunday
0	885	836	844	1,008	940	957	973
1	362	356	379	461	439	448	486
2	225	187	197	224	216	233	242
3	131	124	117	156	181	183	180
4	125	124	97	128	146	168	131
5	169	138	173	195	194	182	170
6	342	351	327	471	425	407	385
7	582	577	570	754	704	736	739
8	900	895	839	1,114	985	1,057	1,068
9	1,207	1,095	1,032	1,322	1,320	1,340	1,350
10	1,357	1,411	1,210	1,523	1,482	1,510	1,522
11	1,237	1,319	1,179	1,443	1,434	1,421	1,392
12	1,272	1,252	1,121	1,407	1,325	1,348	1,300
13	1,280	1,243	1,127	1,478	1,373	1,468	1,457
14	1,325	1,177	1,070	1,360	1,334	1,407	1,336
15	1,282	1,102	1,074	1,420	1,451	1,389	1,350
16	1,311	1,194	1,206	1,531	1,431	1,493	1,443
17	1,175	1,049	1,077	1,234	1,263	1,348	1,415
18	1,084	995	957	1,225	1,147	1,319	1,323
19	1,345	1,230	1,120	1,416	1,326	1,324	1,428
20	1,558	1,414	1,367	1,639	1,591	1,674	1,698
21	1,773	1,645	1,611	1,903	1,848	1,841	1,892
22	1,840	1,710	1,628	2,038	1,962	1,931	1,981
23	1,332	1,276	1,220	1,536	1,422	1,522	1,419

Type

☐ Comment

☐ Follow

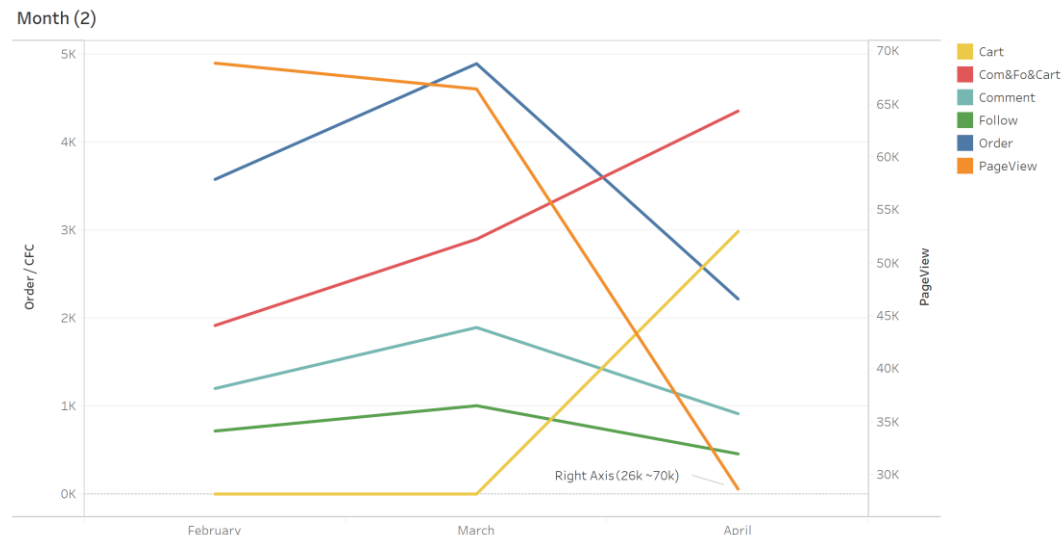
☒ Order

☒ PageView

☐ SavedCart

结论与启示

1. Pageview自2月起逐月下降，需要警示；Order 3月最多，4月较少可能因为数据只统计到4.18；Comment&Follow&Cart 4月最多，进一步分析后发现是SavedCart数量激增，后续可以期待销量的增加（顾客后续可能会为购物车中商品付款）
2. 周四至周日的订单量和浏览量都很高。其中，周四的表现令人意外，商家和平台都应当利用这一点制定营销策略。评论/关注/加购等行为多发生在周末。如何促使消费者为周末加购物车的商品买单，是平台需要考虑的问题。
3. 每日上午十点至晚上十点消费者活动较活跃。制定相应营销策略的同时也要注意这段时期客服质量的把控。
4. 对热力图单独筛选Order和Pageview，发现用户浏览多集中在晚上十点；而下单时间较分散，更多发生在上午十点，下午两点 and 晚上十点也有。值得注意的是周一晚上十点的下单量较为可观，有可能是消费者周末加购物车后思考了一天在周一晚上下单。（如何提高下单率呢？也许可以在用户打开app时推荐“大家都在买的相似好物”）



Order:

Hour..	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	40	42	31	32	40	47	47
1	9	11	20	18	16	26	20
2	12	9	10	16	13	14	5
3	5	6	6	2	7	7	7
4	4	4	7	4	8	6	6
5	6	6	5	8	6	5	6
6	15	14	9	22	18	18	20
7	33	28	33	34	37	35	29
8	57	55	48	61	68	47	55
9	76	73	68	99	79	100	88
10	8	123	108	118	116	101	112
11	10	106	85	96	113	100	91
12	85	91	68	99	105	88	86
13	75	71	90	90	90	104	105
14	99	82	103	110	81	89	94
15	75	87	99	103	119	83	84
16	73	84	79	98	100	94	103
17	75	65	81	75	85	79	85
18	81	72	71	79	84	88	73
19	94	63	60	105	72	68	89
20	86	105	84	93	80	106	98
21	87	89	94	122	91	87	92
22	116	107	84	115	107	97	103
23	45	50	61	77	77	75	77

Pageview:

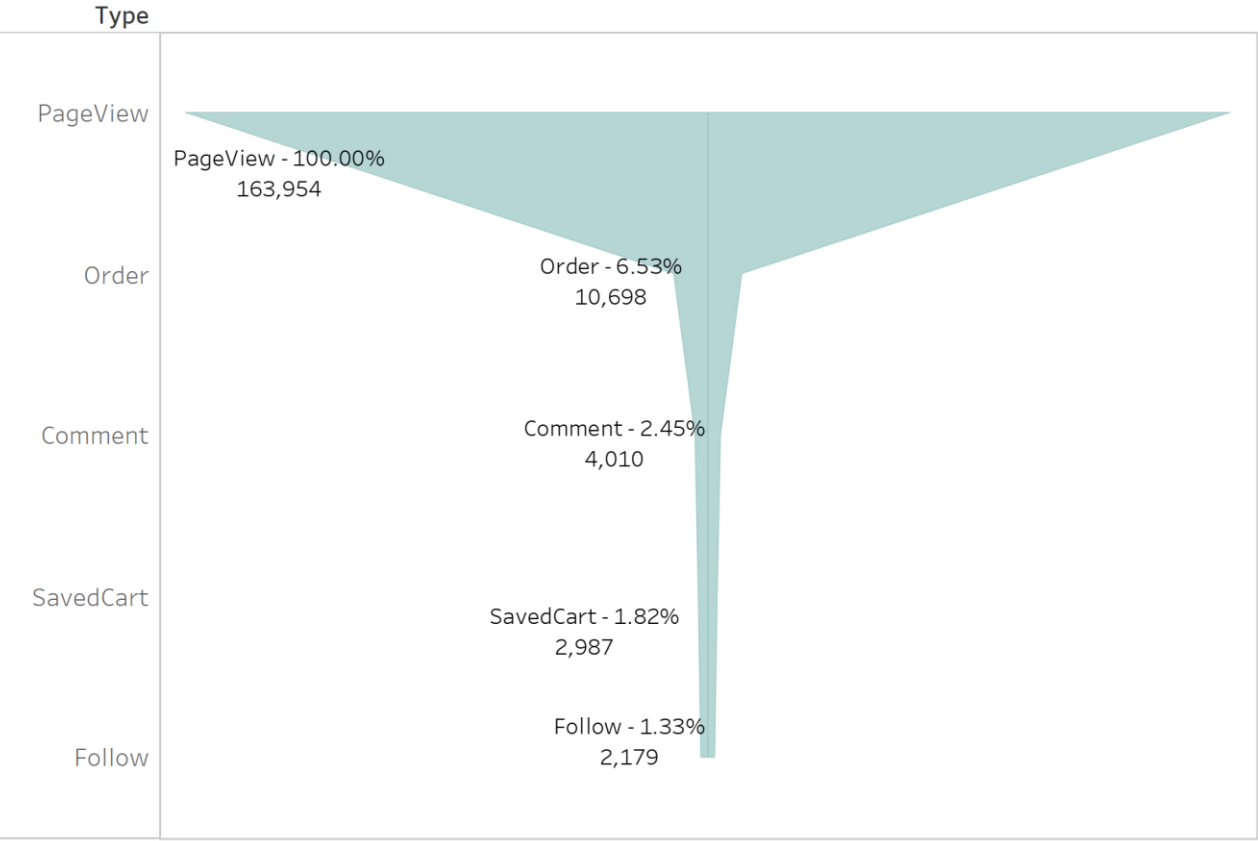
Hour..	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	845	794	813	976	900	910	926
1	353	345	359	443	423	422	466
2	213	178	187	208	203	219	237
3	126	118	111	154	174	176	173
4	121	120	90	124	138	162	125
5	163	132	168	187	188	177	164
6	327	337	318	449	407	389	365
7	549	549	537	720	667	701	710
8	843	840	791	1,053	917	1,010	1,013
9	1,140	1,019	959	1,233	1,221	1,240	1,262
10	1,269	1,288	1,102	1,405	1,366	1,409	1,410
11	1,135	1,213	1,094	1,347	1,321	1,321	1,301
12	1,187	1,161	1,053	1,308	1,220	1,260	1,214
13	1,205	1,172	1,037	1,390	1,293	1,364	1,352
14	1,226	1,095	968	1,250	1,253	1,318	1,242
15	1,207	1,015	979	1,317	1,332	1,306	1,266
16	1,238	1,110	1,127	1,433	1,331	1,399	1,340
17	1,100	984	996	1,159	1,178	1,269	1,330
18	1,003	923	886	1,146	1,063	1,231	1,250
19	1,251	1,167	1,060	1,311	1,254	1,256	1,339
20	1,472	1,309	1,282	1,556	1,503	1,508	1,608
21	1,686	1,556	1,516	1,781	1,757	1,754	1,800
22	1,724	1,603	1,544	1,923	1,855	1,834	1,878
23	1,287	1,226	1,153	1,439	1,357	1,447	1,544

Why: 影响用户购买行为的因素都有哪些?

- 选择什么变量?
 - 购买行为: type
- 影响因素可思考维度:
 - 用户维度: 年龄, 性别, 城市等维度
 - 商家维度: 商家评分/粉丝数/商家类别等维度
 - 产品维度:
 - 漏斗分析: 从页面访问->存购物车->下单->关注, 可以结合多者之间的转化关系得出订单量和页面访问之间的转化关系 (加分项)
- 呈现怎样的数据关系?
 - 比较关系
- 可以选择怎样的图表?
 - 柱状图, 环形图

漏斗分析

Funnel Analysis



对用户行为进行漏斗分析，发现并不是严格按照页面访问->存购物车->下单->评论->关注的流程。可能是有不少用户是没有加购物车就直接下单的。可以看到从访问到下单的转化率仅有6.53%，应考虑提升这一指标。

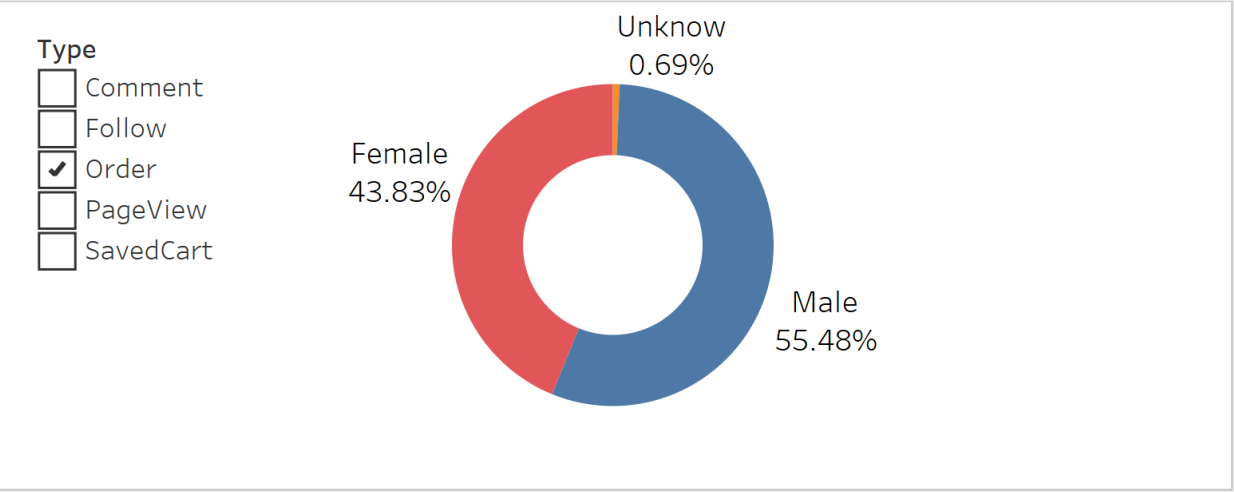
转化率

Results are computed along Table (across).

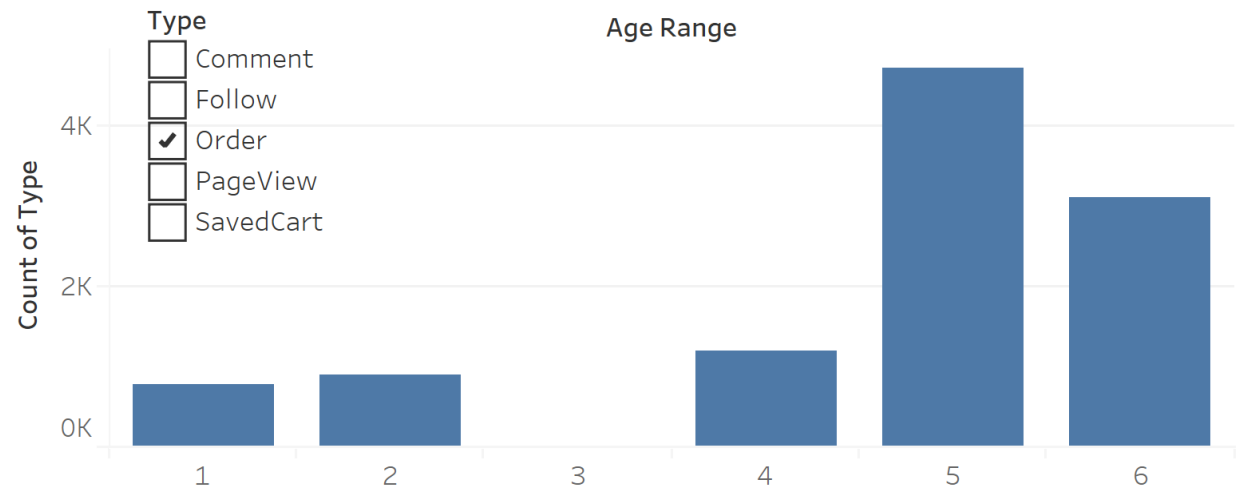
$$\text{COUNT}([Type]) / \text{WINDOW_MAX}(\text{COUNT}([Type])) * 100$$

影响用户行为的因素分析 - 用户维度

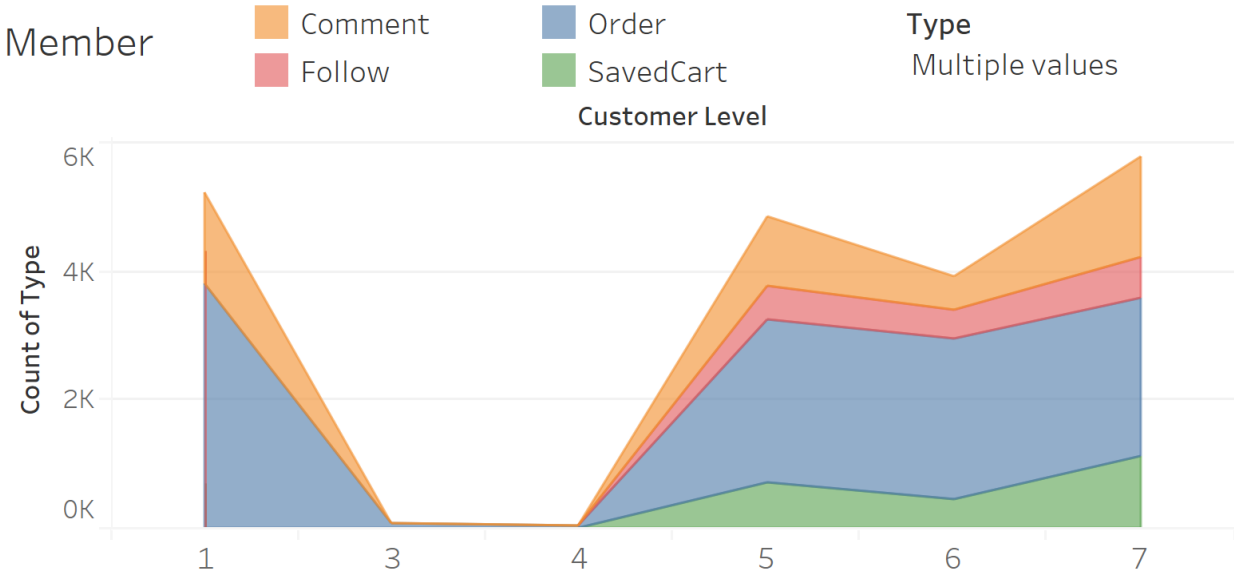
Gender



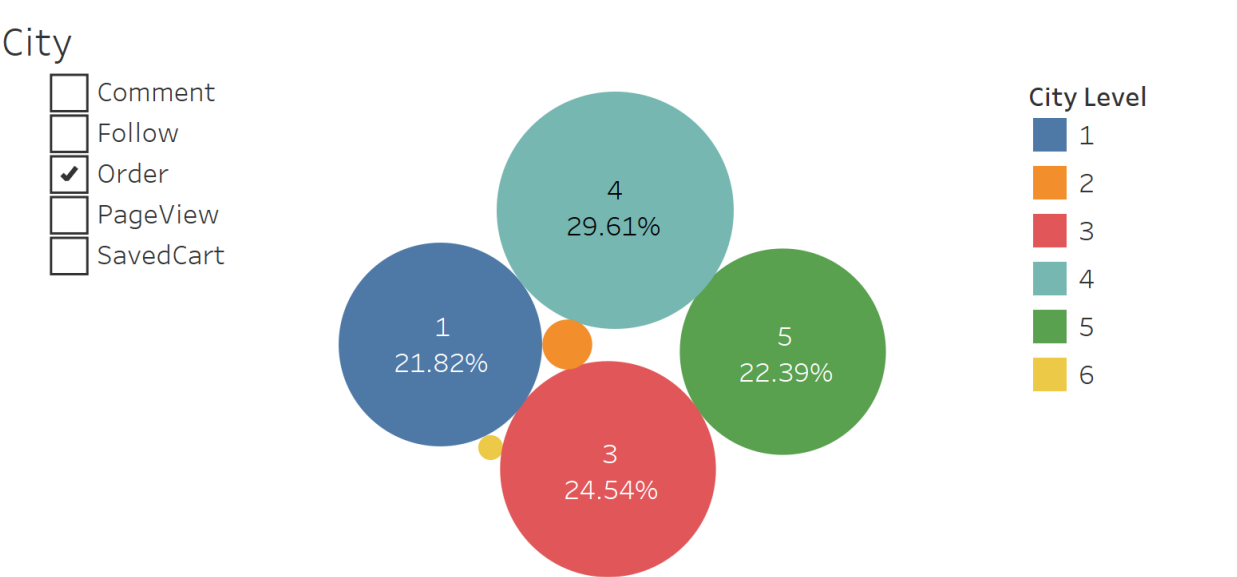
Age Group



Member



City



影响用户购买行为的因素 – 用户维度

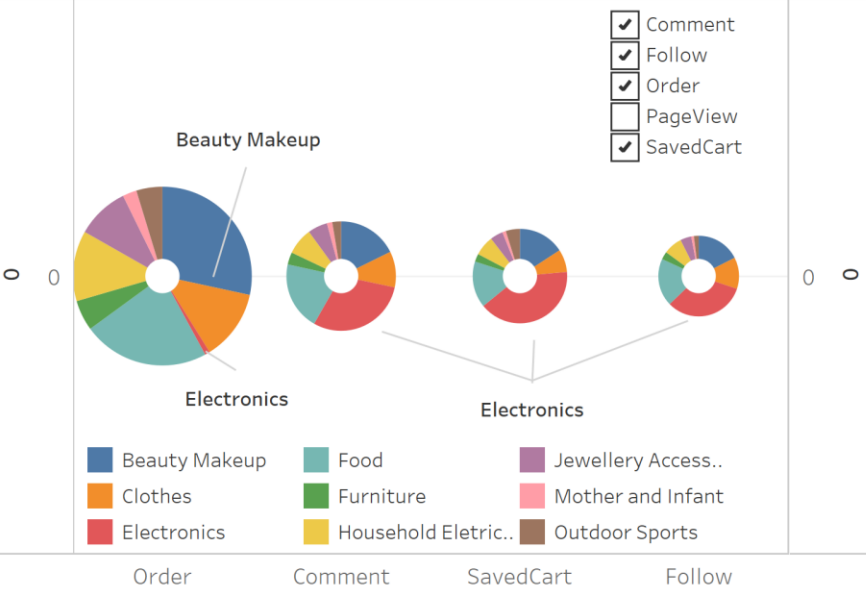
从用户角度来看，影响用户购买行为的因素有：**性别，年龄，会员等级，城市等级**

购买行为较多的用户特征为：**男性，5/6年龄组，1/5/6/7会员等级，4/3/5/1城市等级**

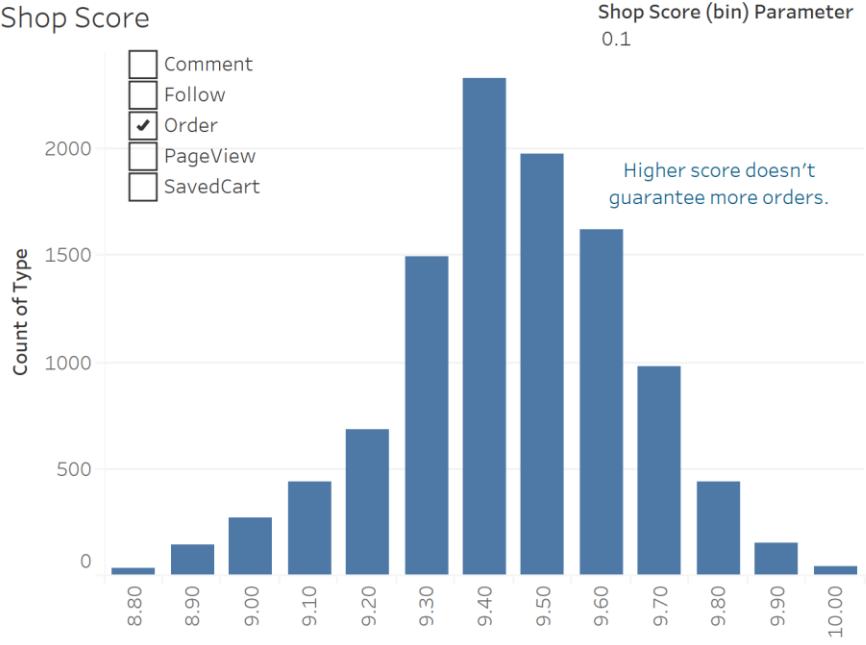
另外，平台应深入挖掘一下年龄组3和会员等级2/3/4的用户为何购买行为非常少。

影响用户行为的因素分析 - 商家维度

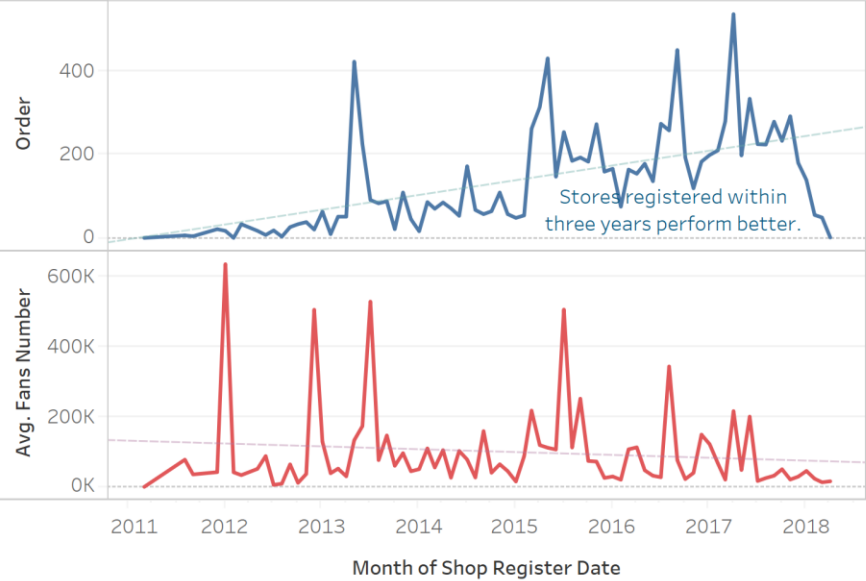
Shop Category



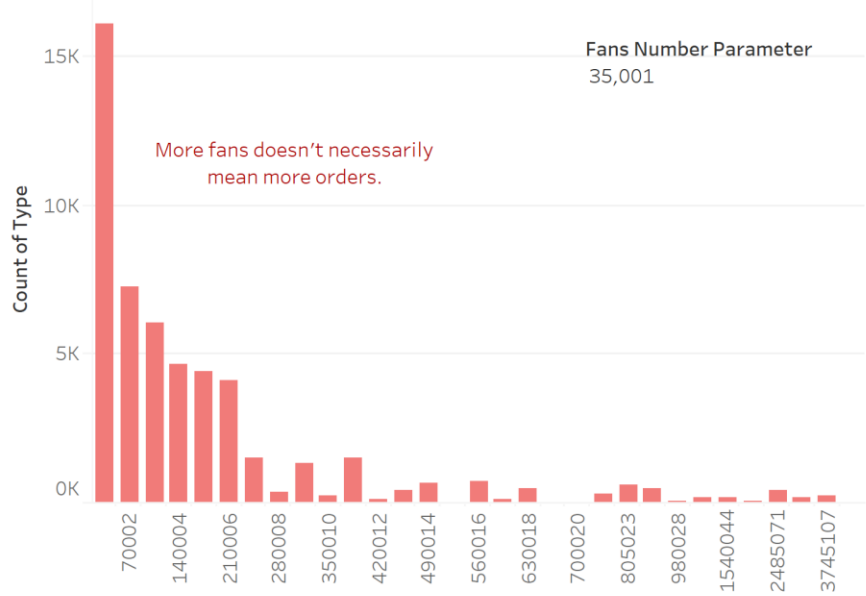
Shop Score



Shop Register Date



Fans number



影响用户购买行为的因素 – 商家维度

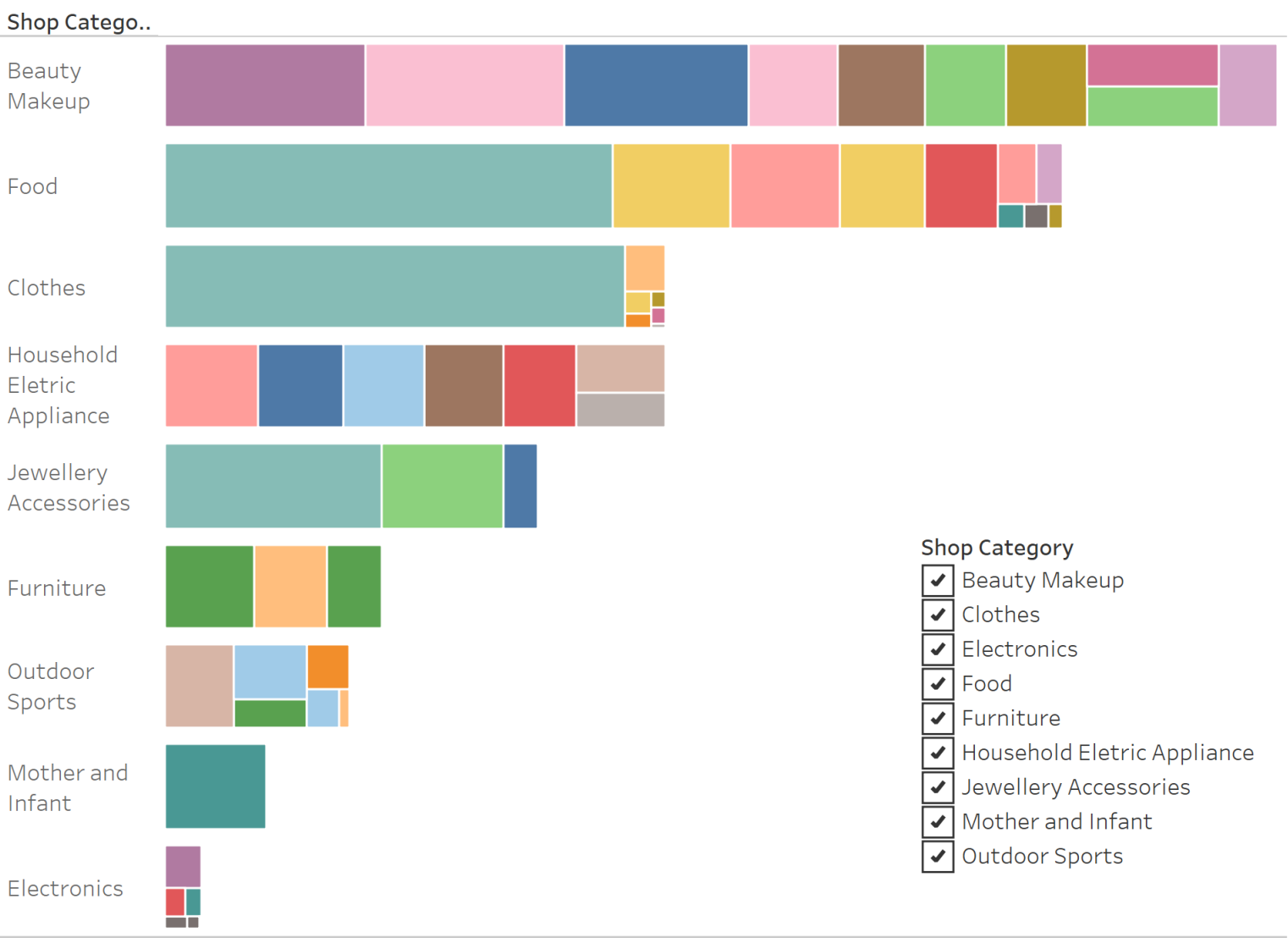
从商家角度来看，影响用户购买行为的因素可能有：店铺主营类别，商家评分，粉丝数，开店时间

通过分析发现：

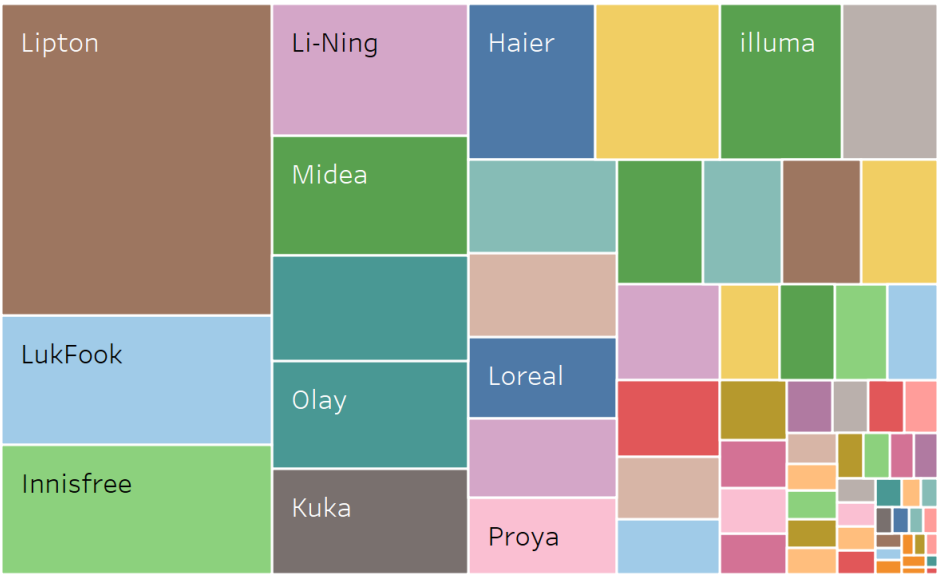
1. 虽然电子产品的浏览/关注/加购/点评数量都是最高的，但订单数量却很少。订单数量最多的是美妆，食品，服装和家电。看起来似乎更像是女性用户会购买的产品。然而事实上在京东购物的男性用户更多。（可以深挖一下原因，比如买礼物？）
2. 并不是店铺评分越高，订单越多。订单量多的店铺集中在9.4分上下。可能是因为高分店铺数量其实并不多。
3. 数据集是2018年2月至4月18日的，发现于2015-2017年开店的店铺订单量较高
4. 并不是粉丝越多，订单越多。

影响用户行为的因素分析 - 产品维度

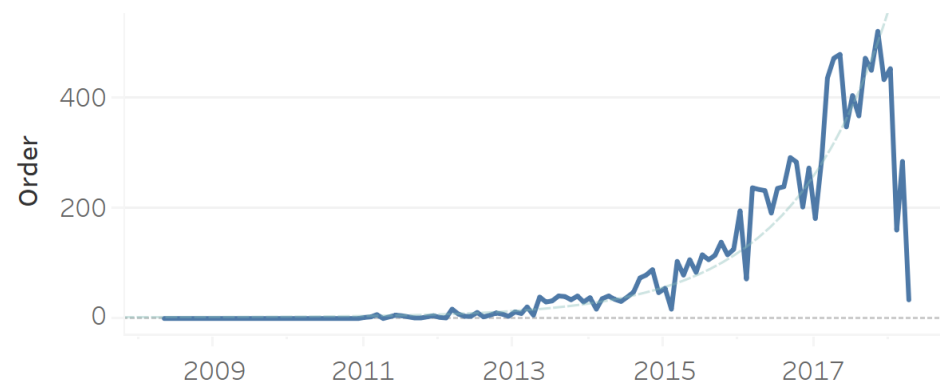
Category



Brand



Product Market Date

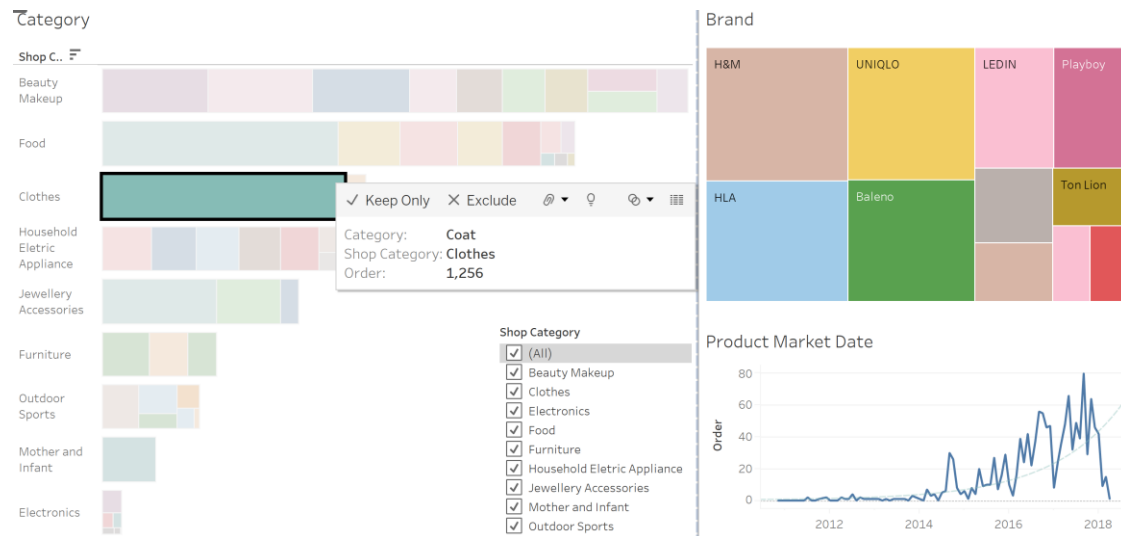


影响用户购买行为的因素 – 产品维度

从用户角度来看，影响用户购买行为的因素有：产品品牌，产品类别，产品上市日期。

可以在仪表盘中查看，不同店铺类别最热销的产品类别，以及不同产品类别最热销的品牌，也可以查看不同品牌最热销的产品类别。比如下图可以看到，大衣卖得最好的品牌是H&M, HLA, Uniqlo, Baleno。

还可以发现，总地来说，新上市地产品销量更高。



How: 京东的业务部门可以怎么做，提升订单量？

聚类分析

- 目的：进行客群分析，找到目标客群
- 模型选择： K-means Clustering Model
(经过试验，选择k=3)
- 变量选择：会员级别，城市级别，性别，年龄组
如果有三方数据，可以再加入消费者收入，教育，爱好，使用京东的RFM数据等等。

由于数据限制，本次客群聚类的结果并不是很好，但仍可以看出Cluster 1 的客群是主力客群，更容易产生购买行为。其性别为男，年龄组5/6，会员等级为7，城市等级为4。京东可以target在这类人群，以提高订单量。

结果：

Variables: City Level
Customer Level
Gender
Age Range
Level of Detail: City Level, Customer Level
Scaling: Normalized

Summary Diagnostics

Number of Clusters: 3
Number of Points: 183691
Between-group Sum of Squares: 14754.0
Within-group Sum of Squares: 11675.0
Total Sum of Squares: 26429.0

Clusters	Number of Items	Centers Age Range	Most Common		
			City Level	Customer Level	Gender
Cluster 1	153308	5.3045	4	7	M
Cluster 2	13265	1.5047	5	1	M
Cluster 3	17118	1.5486	3	7	M
Not Clustered	137				

逻辑回归分析

- 目的：预测用户是否会产生购买行为，以判断是否要推送营销信息
- 模型选择：Logistic Regression
- 变量选择：会员级别，城市级别，性别，年龄组，历史购买/浏览/加购/关注/评论数据
如果有三方数据，可以再加入消费者收入，教育，爱好，使用京东的RFM数据等等。

实际研究过程可以选择某个品牌来进行，用逻辑回归找出最有可能购买该品牌产品的用户，向其发送相应的营销广告。

线性回归分析

- 目的：预测目标用户的购买金额
- 模型选择：Linear Regression
- 变量选择：
 - Y: 订单金额
 - X: 会员级别，城市级别，性别，年龄组，历史购买金额/浏览/加购/关注/评论数据如果有三方数据，可以再加入消费者收入，教育，爱好，使用京东的RFM数据等等。

实际研究过程可以选择某个产品类别来进行，用线性回归预测消费者可能下单的金额。

本数据集中缺少销售额数据，故无法进行线性回归分析。