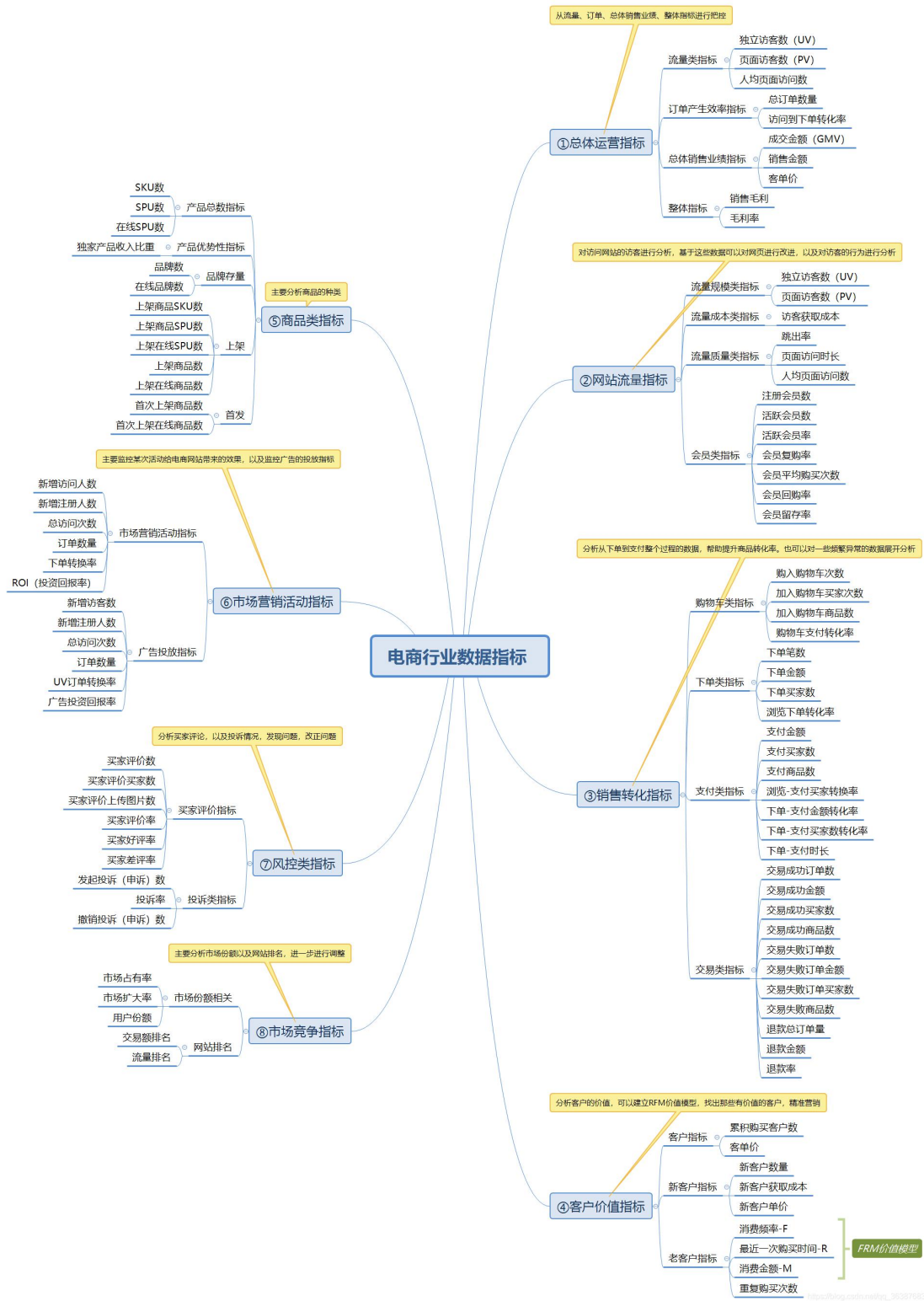




数仓常见指标整理

—— *from my heart* —— 02 大数据启示录

数据分析总是离不开各种指标和术语，最近我花了一周整理了上百个数据分析指标与术语：用户数据指标、行为数据指标、业务数据指标、数据分析术语、统计学常用语、数据报告常用术语。



<https://big.cdnchina.com/50207693>

- (1) UV：独立访问客数
- (2) PV：页面浏览量/阅读量
- (3) VV(Visit View)：访问次数
- (4) 新访问占比
- (5) 实例数
- (6) 访问深度
- (7) 停留时间
- (8) 跳出率
- (9) 退出率
- (10) 产品页转化率
- (11) 加入购物车转化率
- (12) 结算转化率
- (13) 购物车内转化率
- (14) 目标转化率
- (15) IP：独立IP数
- (16) TGI 目标群体指数
- (17) DAU: 日活跃用户
- (18) DNU: 日新增用户

(19) MAU:月活跃用户

(20) 活跃用户留存率

(21) 用户回访率

(1) UV

UV 是指独立访客，表示不同的、通过互联网访问、浏览一个网页的用户。UV 是衡量用户“人数”的重要指标，能够反映来到网站的用户“数量”。比如，一个访客从同一个 IP 地址来访问某个网站 N 次，访问次数记为 N，但 UV（独立访客数）= 1。此外，UV 的定义只与时间有关，与其他任何行为都无关。

(2) PV

PV 是指页面浏览量。用户每一次对网站中的某个页面访问均被记录 1 次。用户对同一个页面的多次访问，访问量累计。PV 的本质是衡量页面被浏览的“绝对数量”。

(3) Visit View

Visit 是指访问次数。该指标反映了有多少“人次”来到网站。Visit 与 UV 相结合可以评估网站来了多少“人”，且黏性如何。

例如，一个网站每天的 UV 是 10，但访问数是 300，反映了该网站的每个 UV 可以带来 30 次访问。Visit 的定义与 UV 类似，只是 Visit 的默认定义时间不是 1 天，而是 30 分钟，即用户在 30 分钟内重复打开网站，Visit 只计为 1；若超过了 30 分钟，则将该次访问记录为一个新的访问。所以，一个 UV 可以产生多个 Visit。

(4) 新访问占比

新访问占比 = 新访问量 / (新访问量 + 老访问量)。新访问占比是用来定义所有访问中新访问的占比情况，反映了站外渠道或网站吸引新用户的能力。新访问占比高意味着市场扩大和新用户不断进入。所以，新访问占比是评估站外引流效果的重要指标，对于以吸引新用户关注为目的的渠道具有重要意义。在新访问占比的公式中，新访问量是指之前没有访问记录，第一次访问该网站的用户数量。如果用户在当天既有第一次访问，又产生了第二次访问，则新访问量和老访问量同时增加 1。

(5) 实例数

实例数是一个特殊的流量指标，它用来衡量站内自定义对象的触发次数(如某个按钮、某个功能区、某个下拉菜单等的触发次数)。

实例数的计数原理是每次监测的代码触发一次，则实例数增加 1。比如，网站设置点击某个功能区为自定义对象，该功能区今日被点击 50 次，则关于该功能区的实例数为 50。

(6) 访问深度

访问深度 = PV（或 UV） / 访问次数。访问深度是衡量用户访问质量的重要指标，是指用户在一次浏览网站的过程中，所浏览的网站的页数，可以用来评估用户看了多少个页面。访问深度越大意味着用户对网站的内容越感兴趣。不过，访问深度不一定总是越高越好，访问深度过高可能意味着用户在网站中迷失了方向而找不到目标内容。

(7) 停留时间

停留时间包括两类，第一类是页面停留时间（下一个页面请求的时间戳 - 当前页面时间戳），第二类是网站停留时间（在网站中最后一次请求的时间戳 - 第 1 次请求的时间戳）。通过分析停留时间可以判断网站或某个页面对用户是否具有足够的吸引力。但停留时间并不意味着用户真的“停留”或“浏览”，比如用户可能打开网页后离开计算机，或使用多 Tab 浏览器同时打开多个页面。

(8) 跳出率

跳出率 = 跳出的访问数量 / 落地页访问数量。跳出是指用户在到达落地页之后，没有单击第二个页面就离开网站的行为。跳出率则是指只浏览了一个页面就离开网站的访问次数占总访问次数的百分比。通常来说，跳出率可以用来评估用户进入网站后的第一反应情况。跳出率过高意味着站外流量的质量低，或者是网站的页面设计有问题，导致用户不愿意继续浏览。不过，有些时候跳出率高并不一定是坏事，还需要结合当初设置这个页面的目标进行具体分析。例如，某些电商网站的购物车或者结算页面，单击支付后会跳到第三方支付平台，那么这个网页的跳出率高属于正常现象。而落地页是指访问者当前所处的特定页面。在公式中，“跳出的访问数量”表示的是通过搜索关键词或单击站外链接进入网站仅浏览了第一个页面就离开的访问次数，该页面指的是落地页。而“落地页访问次数”指的是用户访问落地页产生的所有访问次数。简单来理解，跳出率就是指访问了入口页面（例如网站首页）就离开的访问量与所产生总访问量（包含跳出的，也包含进一步深入访问的）的百分比。

(9) 退出率

退出率 = 当前页面退出的访问量 / 当前页面的总访问量。退出是指用户从网站离开，而没有进一步动作的行为。退出率指的是在某个页面退出的访问量占该页面总访问量的比例。通过退出率可以对网站页面进行分析。如果一个页面的退出率比较高，则这个页面的内容或者是页面设计可能存在一些问题，使用户不愿意继续浏览，这种时候就需要及时对页面进行相应的优化。虽然页面的退出率越低越好，但某些特殊情况下出现高退出率也是正常的。比如，网站中的客服页面，当用户的问题得到解答之后退出，也属于正常情况。

退出率和跳出率要做区分，跳出率仅针对是落地页（一般是网站打开后的第一个页面），而网站中的所有页面都有退出率。

(10) 产品页转化率

产品页转化率 = 产品页访问次数 / 总访问次数。通常情况下，用户要完成订单，需要先浏览产品页查看产品信息，确认产品信息之后才能继续购物车流程。所以产品页转化率是用户订单转化过程中的重要指标。在上述公式中，“产品页访问次数”是指产品页被访问的总次数，而“总访问次数”是指网站被访问的总次数。

(11) 加入购物车转化率

加入购物车转化率 = 加入购物车的访问量 / 总访问量。把商品加入购物车是用户进入购物车环节的第一步,用户在这一步确认商品信息、数量等。加入购物车转化率是指将产品加入购物车的访问量占总访问量的比例,比产品页转化率的参考价值可能更高,因为加入购物车说明用户的购物意向更强。所以,该指标通常会用来衡量所有站外和站内运营的业务效果。

(12) 结算转化率

结算转化率 = 结算页 UV / 总 UV。结算是用户在购物车环节的第二步,在这个步骤,用户可以确认订单联系人、送货时间、送货地址、优惠折扣、运费等信息。结算转化率就是到达结算页面的访问占比,结算转化率越高,意味着用户完成订单的概率就越大,所以结算转化率也是相关业务部门的重要指标。

(13) 购物车内转化率

购物车内转化率 = 提交订单的访问量 / 加入购物车的访问量。对于电商的销售环节来说,购物车内转化率是重要的监控指标。一般来说,购物车内转化率应大于给定的阈值(常用 0.6)。如果低于这个阈值,说明可能存在流量作弊问题,或者购物车流程设计有问题(但某些决策周期较长的特殊商品除外,比如空调等大

件商品)。此外,加入购物车转化率和购物车内转化率之间的差值,可以用来衡量产品和业务的提升空间。差值越大,提升空间越大。

(14) 目标转化率

目标转化率就是完成某个目标的访问数占比,目标可以定义为下载、注册、登录、试用、咨询等。目标转化率也可以作为过程衡量指标,比如浏览商品、加入购物车、结算等。

(15) 活跃留存率

指某日新增用户在其后 N 日仍启动该 APP 的用户数,占所选日期新增用户数的比例。

以账号为基础计算:

留存率=活跃账号留存/新增账号留存*100%

以设备为基础计算:

留存率=活跃设备留存/新增设备留存*100%

(16) TGI (Target Group Index): 目标群体指数。

$$\text{TGI指数} = \frac{\text{目标群体中具有某一特征的群体所占比例}}{\text{总体中具有相同特征的群体所占比例}} * 100$$



@朱小五整理

TGI 指数高于 100，代表该类用户该特征比例高于整体水平，即具有更高的相关倾向或偏好；小于 100，则说明该类用户相关倾向较弱；等于 100 表示在平均水平。

TGI 指数常用于用户画像的评判中，它可以清晰地反映不同群体某一特征的关联程度，并进行直观的比较，挖掘更多潜在的用户价值。

- (1) 订单量和商品销售量
- (2) GMV
- (3) 商品销售额和订单金额
- (4) 平均订单金额
- (5) 客单价

- (6) 订单转化率
- (7) 订单有效率
- (8) 支付转化率
- (9) 商品毛利与毛利率
- (10) SPU 和 SKU
- (11) 每订单成本和每有效订单成本
- (12) 活动直接收入与活动间接收入
- (13) 活动收入占比
- (14) 活动拉升比例
- (15) 每优惠券收益与积分兑换收益
- (16) 复购率
- (17) 回购率

(1) 订单量和商品销售量

订单量 = 去重后的订单 ID 数量; 商品销售量 = 销售商品的数量。

订单量和商品销售量的区别在于, 订单量是用来衡量唯一订单的数量, 而商品销售量则是用来衡量商品的总数量。

(2) GMV

GMV 是指 Gross Merchandise Volume，也就是网站成交总金额。一般情况下，电商的 GMV 计算公式为： $GMV = \text{销售额} + \text{取消订单金额} + \text{拒收订单金额} + \text{退货订单金额}$ 。GMV 高的话企业可以获得更多的资金，如果企业具备一定规模的 GMV，就能基于资金周转周期，将这笔资金用于资本运作，比如进行投资获取利润。要注意的是，GMV 包含了未发生实际支付的部分，所以与电商平台的实际经营额可能有不小的差距。

(3) 商品销售额和订单金额

商品销售总额 = $\sum (\text{商品销售单价} \times \text{销售量})$ ；订单总金额 = 商品销售总额 - 总运费 - 优惠凭证金额 - 其他折扣（如满减）。商品销售总额和订单总金额都是评估商品销售收入的指标。商品销售额侧重于对总收入的评估，而订单总金额则侧重于对用户的实际付款进行评价。在上面的公式中，运费是指未满足免邮所需要支付的配送费用，而优惠凭证金额则是指通过优惠券、积分兑换、会员卡等可作为金额使用的抵充金额。

(4) 平均订单金额

平均订单金额 = $\text{有效订单总金额（已成交）} / \text{消费总人数}$ 。平均订单金额是指每一个有效订单的金额。通过分析平均订单金额可

以对网站的订单价值进行评估。平均订单金额越高，意味着每个订单能够为企业带来的收益越高。网站的订单价值也就越大。在上面的公式中，“有效订单总金额”是指去除取消、作废、未支付等无效状态订单的总金额。“消费总人数”则是指在网站产生了购买行为的总人数。

(5) 客单价

客单价 = 一定时期内的平均订单金额。注意客单价的本质是在“一定时期内，每位顾客消费的平均订单金额”，离开了“一定时期”这个范围，客单价是没有任何意义的。客单价能衡量一个网站在某一段时间内的销售情况，在流量转化都不变的情况下，高客单价也就意味着高的销售额。但是客单价并不是越高越好，需要结合客单价的变化趋势来确定。客单价低的商品能够吸引流量，提升转化，但对增加销售额没有什么帮助。低客单价不利于一个网站的长久发展，因为客单价越低意味着商品的毛利越少，这是不利于企业盈利的。

(6) 订单转化率

订单转化率 = 产品订单的访问量 (UV) / 总访问量 (总 UV) 。

订单转化率高, 说明网站的运营水平高, 有更多的用户愿意购买网站中的商品。

(7) 订单有效率

订单有效率 = 有效订单量 / 订单总量。在电商企业, 通常去除取消、作废、未支付、审核未通过等无效状态的订单为有效订单。

订单有效率用来衡量订单有效比例的重要指标。一般来说, 订单有效率从下单后开始随时间下降, 直到所有订单完成妥投才处于稳定状态。妥投是指快递物件已经妥善投递, 和签收的差别在于, 妥投是放到收件指定地点, 还未签名。

(8) 支付转化率

支付转化率 = 完成支付的用户数 / 需要支付的用户数。支付是用户完成购物的重要步骤, 更是企业产生真实销售价值的关键。支付转化率高意味着企业会有更高的销售收益。需要注意的是, 支付转化率无法对货到付款的订单及时评估, 所以这个指标通常只针对线上。

(9) 商品毛利与毛利率

毛利 = 商品妥投销售额 - 商品批次进货成本。毛利率 = 毛利 / 商品妥投销售额。毛利是指商品销售收入减去商品原进价后的余额，是反映商品利润情况最重要的指标，也是电商中自营商品最重要的效果指标之一。而毛利率则是指毛利占商品销售收入或营业收入的百分比，毛利率能够反映企业经营的全部商品、大类商品甚至是某个商品的差价水平，是核算企业经营成果和价格制定是否合理的依据。毛利与毛利率综合反映了商品的盈利空间和变化趋势。需要注意的是，在电商中，所有毛利的计算基本是以妥投状态为计算准则的。

(10) SPU 和 SKU

SPU 是指标准化产品单元。是商品信息聚合的最小单位，是一组可复用、易检索的标准化信息的集合，该集合描述了一个产品的特性。通俗点讲，属性值、特性相同的商品就可以称为一个 SPU。SKU 则是库存量单位，SKU 是物理上不可分割的最小存货单元。举例说明一下 SPU 和 SKU 的关系，某个品牌的秋季新品大衣，是一个 SPU。而该秋季新品大衣，有红色、黄色三种颜色，对应的尺寸有 M、L 和 XL，则该秋季新品大衣有 6 个 SKU，分别是红色 M 秋季新品大衣、红色 L 秋季新品大衣、红色 XL 秋季

新品大衣、黄色 M 秋季新品大衣、黄色 L 秋季新品大衣、黄色 XL 秋季新品大衣。可以看到，1 个 SPU，对应了 6 个 SKU。

(11) 每订单成本和每有效订单成本

每订单成本 = 费用 / 订单量；每有效订单成本 = 费用 / 有效订单量。每订单成本是指用户每完成一个订单，企业需要付出的费用。通过分析每订单成本，企业可以发现在营销费用控制和营销过程中存在的问题。如果每订单成本过高，则说明企业投入了相应的营销费用，却没有达到预期的结果。例如，订单量比预期低很多，而导致每订单成本大幅增加。而每有效订单成本与每订单成本的计算逻辑相似，不同点在于每有效订单成本仅包含有效状态的订单成本，是针对企业级别的真实评估指标。

(12) 活动直接收入与活动间接收入

活动直接收入是指单纯通过促销活动所带来的收入，即用户促销活动期间购买了促销产品而为企业带来的收入。活动间接收入则是指用户通过促销活动引入且订单属于非活动商品给企业带来的收入。

(13) 活动收入占比

活动收入占比 = (活动直接收入 + 活动间接收入) / 全站订单成交金额。分析活动收入占比可以评估营销活动是否达到预期效果。如果活动收入占比的值较低,说明活动流程或者优惠等方面可能存在问题,不能对用户产生足够的吸引力,下一次活动需要对此进行优化。

(14) 活动拉升比例

活动拉升比例 = (活动期间收入 / 非活动期间收入) - 1。活动拉升比例是指活动对全站销售的拉升情况,收入也可以换成订单、销量等指标。通过活动拉升比例可以评估活动对网站其他相关指标的提升、促进效果。通常来说,活动拉升比例不能使用活动收入占比来评估,因为活动促销期间本来应该通过正常流程和渠道进行购买的用户反而会通过促销渠道进行下单,这会导致活动拉升比例的结果与实际不符,对活动效果作出错误评估。另外,通常情况下,在计算活动拉升比例时会发现,使用收入来衡量不如订单量和销量明显,是因为活动一般是促销活动,客单价较低,会影响收入的提升效果。

(15) 每优惠券收益与积分兑换收益

每优惠券收益 = 优惠券带来的订单成交金额 / 优惠券的数量；积分兑换收益 = 使用积分兑换的订单成交金额 / 积分兑换量。发放优惠券和积分兑换是电商网站常用的促销手段。这两个指标能反映优惠促销对销售的提升效果如何。但值得注意的是，在实际业务中，用户可能会在一个订单中同时使用优惠券和积分，所以可能会出现订单金额重复计算的问题。

(16) 复购率

按照用户计算： 在某时间窗口内重复消费用户（消费两次及以上的用户）在总消费用户中占比

复购率 = 单位时间内购买次数大于 1 的人 / 所有购买的人

•

重复消费用户的定义，又分为两种：

（1）按天非去重，即一个用户一天产生多笔付款交易，则算重复消费用户。

（2）按天去重，即一个用户一天产生多笔交易付款，只算一次消费，除非在统计周期内另外一天也有消费，才算重复消费用户。

•

按天非去重，是目前 b2c 网站统计数据常用计算方法，相对计算出来的重复购买率要高于第二种。

按照交易计算：在某时间窗口内重复消费交易次数在总交易次数中占比

复购率 = 单位时间内复购次数/所有购买的人

•

举例：比如在某个季度中。一共产生了 100 笔交易，其中有 20 个人有了二次购买，这 20 人中的 10 个人又有了三次购买，则重复购买次数为 30 次，重复购买率为 30%。

•

(17) 回购率

是某一个时间窗口内消费的用户，在下一个时间窗口仍旧消费的占比。

•

举例：按月作为统计窗口，在 1 月份总共有 1000 名消费用户，其中有 300 名用户在 2 月份依然有过消费，则回购率为 30%。

•

(1) 曝光量

(2) 点击率 (CTR)

- (3) 到达率
- (4) 转化率 (CVR)
- (5) CPA
- (6) CPC
- (7) CPD
- (8) CPM
- (9) 每 UV 成本和每 PV 成本
- (10) ROI
- (11) ARPU (Average Revenue Per User) : 每用户平均收入。
- (12) ARRPV (Average Revenue Per Paying User) : 每付费用户平均收益。
- (13) CAC(Customer Acquisition Cost): 获客成本【获取一个客户所花费的成本】
- (14) CPR(Cost Per Response): 每回应成本【以浏览者的每一个回应计费】
- (15) ADPV(Advertisement Page View): 载有广告的 pa-geview 流量
- (16) ADimp(ADimpression): 单个广告的展示次数

(1) 广告曝光量

广告曝光量是指站外广告对用户展示的次数。该指标是衡量广告效果的初级指标,通常是用来衡量展示类广告。广告曝光量越大,意味着广告被用户点击的概率更大,也有更多机会对用户进行转化。从技术角度来说,广告曝光量等于跟踪代码被加载的次数。但广告曝光不一定代表广告就被用户看到,而只是说明广告被加载并展示了出来,至于用户究竟会不会注意到,与广告形式、广告素材、广告位置等因素都有关。

(2) 广告点击率

广告点击率 = 广告点击量 / 广告曝光量。通常用 CTR 来表示广告点击率,该指标能够反映用户对当前广告的喜好程度、所投放的媒介用户质量与所投放的广告的匹配度,也可以用来说明企业投入的广告费用创造了多大的价值。但要注意的是,如果广告曝光量较低,该公式算出来的点击率将没有太大的意义。

(3) 到达率

到达率 = 到达量 / 落地页访问量。到达率是用来衡量站外流量到达网站的比例,反映的是用户从点击站外广告后到达网站的情况,所以到达数据仅针对站外标记广告的落地页产生。到达率越高,说明在广告点击与网站到达之间流失的用户越少,可以进行

转化的用户更多。而到达率过低，则可能意味着用户质量较差，或网站落地页加载较慢，导致用户在页面完全加载出来之前直接退出或无法正确统计数据。公式中，“到达量”是指通过站外广告到达落地页的用户数量；“落地页访问量”是指访问落地页的总用户数量。

(4) 广告转化率

广告转化率 = 转化量 / 到达量。转化是指让通过站外广告进入网站的用户产生注册、购买等行为，转化标志一般是指某些特定页面，例如注册成功页、购买成功页等，这些页面的浏览量称为转化量。广告转化率是用于评估企业广告投放效果的最重要的指标，转化率高意味着所投放的广告为企业带来了更多的收益。

(5) CPA

CPA 不是指考试的那个证书，而是指每次行动付费。这个行动通常是网站特定的转化目标，例如注册、咨询、放入购物车等，然后按照转化目标的数量付费。

(6) CPC

CPC 是指每次点击付费广告，是最常见的一种广告形式，也是部分展示类广告、SEM 广告的主流投放形式。CPC 是衡量广告单位成本支出的重要质保，企业只需要按照广告被点击的次数付费即可，不用再为广告的显示次数付费。

(7) CPD

CPD 是指按天收费。CPD 是很多传统广告媒介普遍使用的费用计算方式，根据展示的天数进行收费，不对展示期间任何广告效果给承诺。

(8) CPM

CPM 表示的是每千人成本。CPM 的含义是指，在广告投放过程中，平均每一千人看到或听到某广告一次，一共需要多少广告成本。

(9) 每 UV 成本和每 PV 成本

每 UV 成本 = 广告费用 / UV 数；每 PV 成本 = 广告费用 / PV 数。
每 UV 成本可以比较真实地反映站外广告每导入一个访问需要

花多少钱，而每 PV 成本则主要反映通过站外广告每产生一次访问需要花多少钱。

(10) ROI

$ROI = \text{利润} / \text{费用}$ ，或 $ROI = \text{成交金额} / \text{费用}$ 。ROI 是指投资回报率，即投入费用能够带来的收益比例，是评估收入产出效果的核心指标。ROI 高表示企业的投入得到了更高的产出，ROI 低则意味着企业的投放策略或网站设计存在问题，用户可能不愿意在网站进行注册、购买等活动。当 ROI 低的时候，需要进行相应的调整，避免上述问题给网站带来用户的流失风险。上述两个公式，在电商行业往往会利用第二个，因为一些电商企业的利润往往是负数，使用利润来计算会得出负的 ROI。而基于第二个公式计算 ROI 时，ROI 这个指标更多评估的是每单位费用所带来的销售额。

之前我曾经总结过[用户行为分析的 5 类指标](#)，这里的行为数据指标便直接引用前文^[1]。

用户访问类的指标有 13 个：PV、UV、DV、日新增用户数、获客成本、用户访问时长、人均页面访问量、人均浏览页数、平均访问页面、访问来源、平均停留时间、跳出率、搜索访问次数占比。（其中 PV、UV、日新增用户数前文已介绍，其余指标具体含义如下图所示。）

指标	指标含义
PV（Page View）	即页面浏览量，是衡量网络新闻的主要指标
UV（Unique Visitor）	即独立用户，是指通过互联网访问、浏览某网页的自然
DV（Depth of Visit）	即访问深度，是指用户在一次浏览网站的过程中浏览了
日新增用户数	在某个时间段内使用网站或应用的新用户总数（统计周期
获客成本	获取一个客户所花费的平均成本
用户访问时长	一次会话持续的时间 @公众号：凹凸数据
人均页面访问量	人均页面访问量=通过浏览器访问的 Session 来记录的访
人均浏览页数	每个唯一身份访问者（UV）页面浏览量（PV）的平均数
平均访问页面	平均每次访问浏览的页面数量，平均访问页数=浏览量/访
访问来源	也称为“推荐来源”， 通常互联网用户会通过各种渠道，就是“访问来源”
平均停留时间	访客在一次访问中，平均打开网站的时长，即每次访问一个页面的平均值
跳出率	只访问了入口页面（如网站首页）就离开的访问量与所 跳出率=访问一个页面后离开网站的次数/总访问次数
搜索访问次数占比	搜索访问次数就是访客点击搜索结果到达网站的次数。指 /总访问次数

△点击查看大图

用户转化类指标共有 15 个：最近购买间隔、购买频率、购买商品种类、平均每次消费额、单次最高消费额、日应用下载量、一次会话用户数、用户会话次数、漏斗转化—第一步进入次数、漏斗转化—中间步进入次数（漏斗中）、进漏斗转化—进入率（漏斗中）、漏斗转化—进入次数（漏斗中）、漏斗转化—进入率（漏斗中）、漏斗转化—退出次数、漏斗转化—退出率，这些指标具体含义如下图所示。

指标	指标含义
最近购买间隔	用户最近一次购买距离当前的天数
购买频率	用户在一段时间内购买的次数
购买商品种类	用户在一段时间内购买的商品种类
平均每次消费额	也称为客单价，计算公式：平均每次消费额=消费总额/消费次数
单次最高消费额	用户在一段时间内购买的单次最高支付金额
日应用下载量	单个用户账户下载的应用数量
一次会话用户数	即下载完 App，仅打开过产品一次，且该次使用数量 @公众号：凹凸数据
用户会话次数	用户会话（Session），是用户在时间窗口内的所有行为，包括搜索商品、浏览商品、下单并且支付、最后退出的次数
漏斗转化—第一步进入次数	第一步进入次数=含有此页面的访问次数
漏斗转化—中间步进入次数（漏斗中）	在漏斗上一步进入次数中以上一步为前链进入此步骤的次数
进漏斗转化—进入率（漏斗中）	进入率=在漏斗上一步进入次数中以上一步为前链进入此步骤的次数/上一步进入次数
漏斗转化—进入次数（漏斗中）	在当前步骤的进入次数中，没有进入漏斗下一步的次数
漏斗转化—进入率（漏斗中）	进入率=在当前步骤的进入次数中没有进入漏斗下一步的次数/当前步骤的进入次数
漏斗转化—退出次数	在当前步骤的进入次数中，没有进入漏斗下一步的次数
漏斗转化—退出率	退出率=当前步骤退出次数/当前步骤进入次数

△点击查看大图

用户留存类指标共有 15 个：**用户留存率、渠道留存率、次日留存率、退出率、活跃度、活动参与率、活跃交易用户数、DAU、MAU、用户回访率、用户流失率、功能使**

用率、GMV、复购率、退货率，这些指标具体含义如下图所示。

指标	指标含义
用户留存率	留存用户数占当时新增用户数的比例（这里的“当时”指的是用户首次使用的时间点，一般统计周期为天）
渠道留存率	不同用户渠道的用户留存率
次日留存率	计算公式：次日留存率=当时新增用户中在第 2 天依然使用网站或 App 的用户数/新增用户数， 次日留存分析需要结合产品的新手引导设计和新用户转化路径等一起进行
退出率	计算公式：退出率=从该页退出的页面访问数/进入该页的页面访问数。 如，某商品页进入 PV 为 1000，该页直接关闭的访问数为 300，则退出率为 30%
活跃度	线上产品的用户在线时长以及登录频次
活动参与率	参加活动人数占总人数的比重
活跃交易用户数	和活跃用户一样，活跃交易用户也可以区分成首单用户（第一次消费）、复购用户、流失消费用户等
DAU	日活跃用户数量（Daily Active User），指的是一个统计日内登录或使用某产品且立不重复的用户数 @公众号：凹凸数据
MAU	即月活跃用户人数（Monthly Active Users）
用户回访率	用户回访率=回访用户数/统计周期内流失的用户数×100%
用户流失率	如果在某段时间内消费的人群中，一部分人在一个流失周期后变为流失人群，流失人群人数与消费人群的总人数之比被定义为流失率。首先要确定如何定义流失，不同运营人员对流失行为的定义不同
功能使用率	产品重要功能的使用指标，如收藏率、点赞率、评论率等
GMV	一段时间内的成交总额
复购率	计算复购率有两种方法。 1. 复购率=重复购买客户数量/客户样本数量。例如，客户样本 100 人，其中有 60 人重复购买（不用考虑重复购买了几次），复购率=60/100，即 60% 2. 复购率=客户购买行为次数（或交易次数）/客户样本数量。例如，客户样本 100 人，其中 60 人重复购买，这 60 人中有 40 人重复购买 1 次（即购买 2 次），有 20 人重复购买 2 次（即购买 3 次），复购率=（40*1+20*2）/100，即 80%
退货率	计算公式：退货率=产品售出后由于各种原因被退回的数量/同期售出的产品数量

△点击查看大图

用户社交类指标共有 8 个：**好友数量、帖子数量、看帖数量、回复数量、分享数量、点赞数量、转发数量、评论数量**，这些指标具体含义如下图所示。

指标	指标含义
好友数量	用户分享渠道可以细分为微信好友/群、微信朋友
帖子数量	发帖数量级按自定义时间窗口统计
看帖数量	浏览帖数量级按自定义时间窗口统计
回复数量	回复数量级按自定义时间窗口统计
分享数量	分享数量级按自定义时间窗口统计
点赞数量	点赞数量级按自定义时间窗口统计
转发数量	转化数量级按自定义时间窗口统计
评论数量	评论数量级按自定义时间窗口统计

- (1) 注册会员数
- (2) 激活会员数
- (3) 购买会员数
- (4) 可营销会员数
- (5) 会员营销费用和费率
- (6) 会员营销收入
- (7) 用券类指标

- (8) 单位成本指标
- (9) 会员活跃度
- (10) 新会员数和老会员数
- (11) ARPU 和 ARPPU
- (12) 会员复购率
- (13) 会员平均购买次数
- (14) 会员的 RFM 指标
- (15) 会员生命周期相关指标
- (16) 会员生命周期转化率
- (17) 会员生命周期剩余价值
- (18) 会员留存率
- (19) 会员流失率
- (20) 会员异动比

(1) 注册会员数

注册会员数就是指在网站上注册过的会员数量。注册会员数可以反映网站所覆盖的整体会员规模。但注册会员数只能反映网站目前会员的注册数量，无法对质量进行评估。

(2) 激活会员数

与注册会员数相比，激活会员数有一个特定的激活动作，这个动作往往决定了用户是否真的成为了企业会员。常见的代表性动作包括：手机验证、身份验证、单击确认链接等。此外，会员激活率 = 激活会员数 / 注册会员数。会员激活率是评估会员注册质量的一个重要指标，也是网站运营能力的体现。会员激活率高说明网站对已注册会员的运营策略得当，会员激活率低则意味着网站对于已注册会员没有投入足够的精力来运营，或者是激活流程繁琐导致用户放弃激活。

(3) 购买会员数

是指在网站购买过商品的会员。某种情况下，购买会员也会被称为活跃会员。购买会员数越多意味着会员为企业带来的利润更多，同样这也是网站运营良好的表现。此外，如果企业内部，转化周期和步骤比较长，还会细分更多的转化状态指标，比如充值会员、妥投会员等。

(4) 可营销会员数

是指在所有会员中，可以通过一定手段进行会员营销来满足企业特定需求的会员数。比如，会员具有邮箱、微信、QQ、手机号

等可识别并可接触的信息点,具备这些信息中的任何一种便可以形成可营销会员。

(5) 会员营销费用和费率

会员营销费用特指企业进行会员营销时的投入。而会员营销费率是指会员营销费用占会员营销收入的比例。分析费率的目的是对营销费用的支出情况进行监督,确保其不超出计划指标。

(6) 会员营销收入

会员营销收入是指通过会员营销渠道(包括电子邮件、短信、会员通知、特定会员优惠码、线下二维码等)和相关的会员运营活动产生的费用。会员营销收入越高,说明企业针对会员的营销活动越好。在评估会员营销收入时,要记得让用户有特定的标志,这样才能区分营销的效果,和其他渠道区分开。

(7) 用券类指标

优惠券是很多电商进行会员营销的主要形式。用券类指标主要包括:用券会员比例(使用优惠券下单的会员数/总会员数)、用

券金额比例（使用优惠券下单的金额占总下单金额的比例）、优惠券订单比例（使用优惠券下单量占总下单量的比例）。

(8) 单位成本指标

单位成本是指企业为完成单个特定目标所要付出的费用。这个目标可以是注册、订单等。对电商企业来说，单位成本对于分析企业管理水平具有重要的参考意义。常用的单位成本指标包括：每注册成本（注册一个用户需要多少成本）、每订单成本（每获得一个订单需要多少成本）、每会员成本（每获得一个会员需要多少成本）。

(9) 会员活跃度

会员活跃度是指会员的活跃程度。会员活跃度是企业会员营销水平的体现。会员活跃度越高，说明企业的会员营销活动效果越好。通常，活跃度是通过企业定义的业务关键因素来判定的，例如注册、登录、查看商品、停留时间等。

(10) 新会员数和老会员数

通常新会员是指第一次购买的用户,而老会员则是指购买 2 次或 2 次以上的会员。对于新会员,有些时候会加如时间限制,比如一周内新产生的购物会员都算作新会员,无论购买几次。而老会员则是企业产生销售的最重要保证,老会员的数量越多,企业的销售规模越大。

(11) ARPU 和 ARPPU

ARPU 即 Average Revenue Per User, 指的是一个时期内(通常为一个月或一年)电信运营企业平均每个用户贡献的通信业务收入,其单位为元/户。从计算的角度看,ARPU 值的大小取决于两个因素,业务收入和用户数量,相对用户数量,业务收入越高,ARPU 值越大。同时 ARPU 值也反映企业用户结构状况,当用户构成中高端客户占的比重越高,ARPU 值就越高。而 ARPPU 则是指每付费用户平均收益。即统计周期内,付费用户对产品产生的平均收入。

(12) 会员复购率

会员复购率是指在一定时期内,购买 2 次或 2 次以上的会员比例。通过复购率可以评估网站对用户是否具有足够的黏性。复购率高

说明用户对网站的忠诚度越高，会员对网站的价值也就越大，反之则越低。

(13) 会员平均购买次数

会员平均购买次数 = 会员的订单总数 / 会员总数。会员平均购买次数也能够反映会员对网站的忠诚度，平均购买次数越高，说明会员对网站的忠诚度越高，会员的价值也就越大。

(14) 会员的 RFM 指标

[1] 会员消费频次：是指用户在一定时间内的消费频率按照次数做统计的结果；

[2] 最近一次购买的时间：是指用户最后一次产生购买行为的时间；

[3] 最近一次购买的金额：是指用户最后一次产生购买行为时支付的金额。

会员消费频次和最近一次购买的时间，可以拿来有效分析会员对于企业的消费黏性。消费频次高、最近一次购买的时间离当前时间近，说明会员的消费黏性大，对于企业来说，这些消费黏性大的用户具有更高的价值，需要投入更大的精力去维护。而最近一

次购买的金额越大，则说明用户最近一次的消费能力越高。根据“二八法则”，20%的老会员贡献80%的消费额。

(15) 会员生命周期相关指标

- [1] 总价值：用户在整个生命周期内下单金额总和；
 - [2] 订单量：用户在整个生命周期内下单量之和；
 - [3] 平均订单价值：用户在整个生命周期内的下单金额 / 下单量。
- 由于会员的生命周期相关指标突破了时间的限制，能从整体上获得会员的宏观状态，所以是衡量会员宏观价值的重要指标。

(16) 会员生命周期转化率

会员生命周期转化率 = 会员在完整的生命周期内完成的订单 / 到达网站的次数。该指标可以衡量用户是否具有较高的转化率。会员生命周期转化率越高，表明会员在生命周期内为企业贡献的价值越大，反之则越小。

(17) 会员生命周期剩余价值

该指标是预测性的指标，能预测用户在生命周期内还能产生多少价值，可以采用特定的算法和模型进行预测。如果某个会员的生

命周期剩余价值较高,则意味着企业需要投入精力继续对该会员进行维护。

(18) 会员留存率

是指新会员首次在平台或 APP 上进行登录或消费后,在某一周期内再次登录或消费的会员比例。留存率一般有次日留存率、三日留存率、七日留存率、三十日留存率等。在电商领域,常用消费数据作为会员留存率的统计依据。

(19) 会员流失率

会员流失率 = 流失会员数 / 购买会员数。会员流失是指会员不再购买或消费企业的相关产品、业务和服务。会员流失是正常现象,但需要关注流失率的走向。在正常情况下,会员流失率应该是一个比较小的比例,如果突然开始上升,企业需要多加关注。

(20) 会员异动比

会员异动比 = 新增会员数 / 流失会员数。如果会员异动比 > 1,说明新增会员数大于流失会员数,企业处于上升发展期;如果 <

1, 则说明会员增长跟不上会员流失的速度, 企业将面临会员枯竭危机。

- (1) 库存量
- (2) 库存金额和平均库存金额
- (3) 库存可用天数
- (4) 库存周转率
- (5) 库存周转天数
- (6) 库龄
- (7) 滞销金额
- (8) 残次商品占比
- (9) 缺货率

(1) 库存量

库存量是指一定时间内全部库存商品的数量。由于在库存量的定义中, 包括多种状态的商品, 比如正常可售卖商品、已被订购但未发货商品、调拨未出库商品、调拨未入库商品、残次商品等,

所以有些时候会出现商品有库存但无法销售的情况。通常情况下，为了保证商品在一定程度上可以满足用户的购买需求，同时不出现商品积压，企业都会定义安全库存量、最低库存量和最高库存量，目的是要保证商品在一定程度上可满足用户购买需求，同时不至于造成商品积压。如果库存量低于最低库存量，说明商品需要及时补货；如果库存量高于最高库存量，则可能会导致库存积压，需要及时调整营销策略。

(2) 库存金额和平均库存金额

库存金额就是全部库存产品按入库成本价格计算的总金额。平均库存金额 = (期初库存金额 + 期末库存金额) / 2，注意平均库存金额是一个时点指标。

(3) 库存可用天数

库存可用天数 = 库存商品数量 / 期内每日商品销售数量。库存可用天数主要反映当前库存可以满足供应的天数，是仓库备货能力的体现。可用天数越多，说明仓库备货能力越强。但过长的可用天数可能意味着商品滞销，而过短则会导致缺货，所以需要保持在一定的范围内。此外，不同商品的可用天数需要根据库存周转天数来定义，即按时间划分。例如，商品库存周转天数是 30 天，

则可以将库存天数划分为 7 天以内、8-14 天、15-30 天、30 天以上等,

(4) 库存周转率

库存周转率 = 年度销售产品成本 / 当年平均库存金额。库存周转率是指某一时间段内库存货物周转的次数,是反映一定期间(一年或半年)库存周转快慢程度的指标,周转率越大说明销售情况越好。

(5) 库存周转天数

库存周转天数 = $360 / \text{库存周转率}$ 。库存周转天数是指企业从取得存货 / 产品入库开始,直到消耗 / 销售为止所经历的天数。周转天数越少,说明企业零库存或者是存货变现的速度越高。通常,用 360 作为一年的计算周期。

(6) 库龄

库龄 = 出库时间 - 入库时间。一般意义上的库龄是指商品库存时间,仓储中的商品从进入仓库开始就产生库龄。仓储系统是按照先进先出、先进先销的原则出库,所以同一个商品的库龄要

按照其相应进货批次的时间计算。当库龄的时间过长，意味着商品进入了滞销阶段，企业此时需要及时对商品的销售策略做调整。

(7) 滞销金额

滞销金额是指商品周转天数超过其应售卖的周期，导致商品无法销售出去的情况。从企业的商品周转快慢程度来看：周转快的为畅销，周转慢的为平销，无周转则为滞销。购买量为 0 是滞销的首要特征。滞销一方面会导致资金积压，另一方面则导致产品过季、过保质期而无法继续销售。

(8) 残次商品占比

残次商品占比 = 残次商品数量 / 总库存商品数量。残次是指由于商品的搬运、装卸、库存、物流、销售等主客观因素所造成的商品外包装损坏、商品损坏、附件丢失等影响商品二次销售的情况。

(9) 缺货率

缺货率 = 缺货数量 / 顾客订货数量。缺货是相对于滞销而言的另一个极端，缺货说明库存商品数量低于用户的购买数量，这会导致用户无法购买的窘境，给企业带来损失。

(1) 配送业务量

(2) 配送满足率

(3) 配送准确率

(4) 满载率

(1) 配送业务量

是指企业根据用户的订单需求，对商品进行挑选、包装、出库、配送等服务，并按时送达到指定地点的商品数量，包括同城配送和区域配送。配送业务量的多少也是企业盈利能力的一个体现，配送业务量越大，意味着企业能获得更多的利润，反之则说明企业在商品销售等环节可能存在问题导致业务量减少。

(2) 配送满足率

配送满足率 = 实际可配送的订单量 / 需求配送订单量。配送满足率是指实际可用的配送资源与配送需求的比例, 通常情况下该值应 > 1, 只有在特殊情况 (如双十一) 才会出现 < 1。

(3) 配送准确率

配送准确率 = 准确配送的订单量 / 总配送订单量。如果配送准确率较低, 企业可能会面临大量的客诉或退货。

(4) 满载率

满载率 = 车辆实际载重量 / 车辆额定载重量。该指标是合理安排配送资源的重要依据之一。车辆利用率越高, 则有利于企业降低物流成本, 提高物流配送效率。



大数据启示录

数仓常见术语

from my heart

大数据启示录

绝对数和相对数

绝对数：是反应客观现象总体在一定时间、一定地点下的总规模、总水平的综合性指标，也是数据分析中常用的指标。比如年 GDP，总人口等等^[4]。

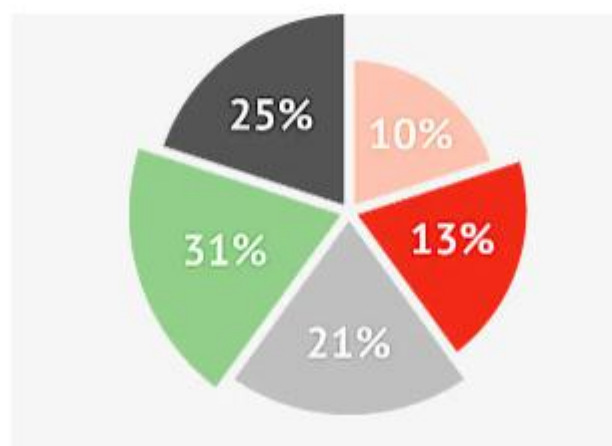
相对数：是指两个有联系的指标计算而得出的数值，它是反应客观现象之间的数量联系紧密程度的综合指标。相对数一般以倍数、百分数等表示。相对数的计算公式：

相对数=比较值（比数）/基础值（基数）

百分比和百分点

百分比：是相对数中的一种，它表示一个数是另一个数的百分之几，也称为百分率或百分数。百分比的分母是 100，也就是用 1%作为度量单位，因此便于比较。

百分点：是指不同时期以百分数的形式表示的相对指标的变动幅度，1%等于 1 个百分点。



频数和频率

频数：一个数据在整体中出现的次数。

频率：某一事件发生的次数与总的事件数之比。频率通常用比例或百分数表示。

比例与比率

比例：是指在总体中各数据占总体的比重，通常反映总体的构成和比例，即部分与整体之间的关系。

比率：是样本(或总体)中各不同类别数据之间的比值，由于比率不是部分与整体之间的对比关系，因而比值可能大于 1。

变量

变量来源于数学，是计算机语言中能储存计算结果或能表示值抽象概念。变量可以通过变量名访问。

连续变量

在统计学中，变量按变量值是否连续可分为连续变量与离散变量两种。在一定区间内可以任意取值的变量叫连续变量，其数值是连续不断的，相邻两个数值可作无限分割，即可取无限个数值。如：年龄、体重等变量。

离散变量

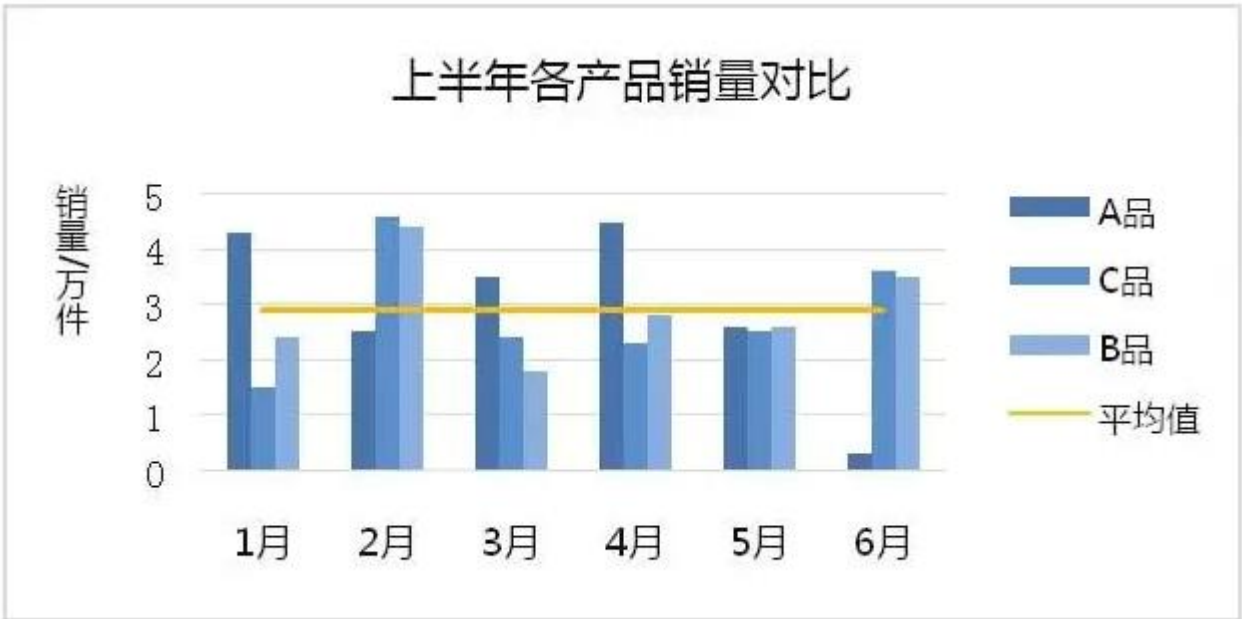
离散变量的各变量值之间都是以整数断开的，如人数、工厂数、机器台数等，都只能按整数计算。离散变量的数值只能用计数的方法取得。

定性变量

又名分类变量：观测的个体只能归属于几种互不相容类别中的一种时，一般是用非数字来表达其类别，这样的观测数据称为定性变量。可以理解成可以分类别的变量，如学历、性别、婚否等。

均值

即平均值，平均数是表示一组数据集中趋势的量数，是指在一组数据中所有数据之和再除以这组数据的个数。



中位数

对于有限的数集，可以通过把所有观察值高低排序后找出正中间的一个作为中位数。如果观察值有偶数个，通常取最中间的两个数值的平均数作为中位数。

缺失值

它指的是现有数据集中某个或某些属性的值是不完全的。

异常值

指一组测定值中与平均值的偏差超过两倍标准差的测定值，与平均值的偏差超过三倍标准差的测定值，称为高度异常的异常值。

方差

是衡量随机变量或一组数据时离散程度的度量。概率论中方差用来度量随机变量和其数学期望（即均值）之间的偏离程度。统计中的方差（样本方差）是每个样本值与全体样本值的平均数之差的平方值的平均数。在许多实际问题中，研究方差即偏离程度有着重要意义。方差是衡量源数据和期望值相差的度量值。

标准差

又常称均方差，是离均差平方的算术平均数的平方根，用 σ 表示。标准差是方差的算术平方根。标准差能反映一个

数据集的离散程度。平均数相同的两组数据，标准差未必相同。

皮尔森相关系数

皮尔森相关系数是用来反映两个变量线性相关程度的统计量。相关系数用 r 表示，其中 n 为样本量，分别为两个变量的观测值和均值。 r 描述的是两个变量间线性相关强弱的程度。 r 的绝对值越大表明相关性越强。

相关系数

相关系数是最早由统计学家卡尔·皮尔逊设计的统计指标，是研究变量之间线性相关程度的量，一般用字母 r 表示。由于研究对象的不同，相关系数有多种定义方式，较为常用的是皮尔森相关系数。

特征值

特征值是线性代数中的一个重要概念。在数学、物理学、化学、计算机等领域有着广泛的应用。设 A 是向量空间的一个线性变换，如果空间中某一非零向量通过 A 变换后所得到的向量和 X 仅差一个常数因子，即 $AX=kX$ ，则称 k 为 A 的特征值， X 称为 A 的属于特征值 k 的特征向量或特征矢量。

倍数和番数

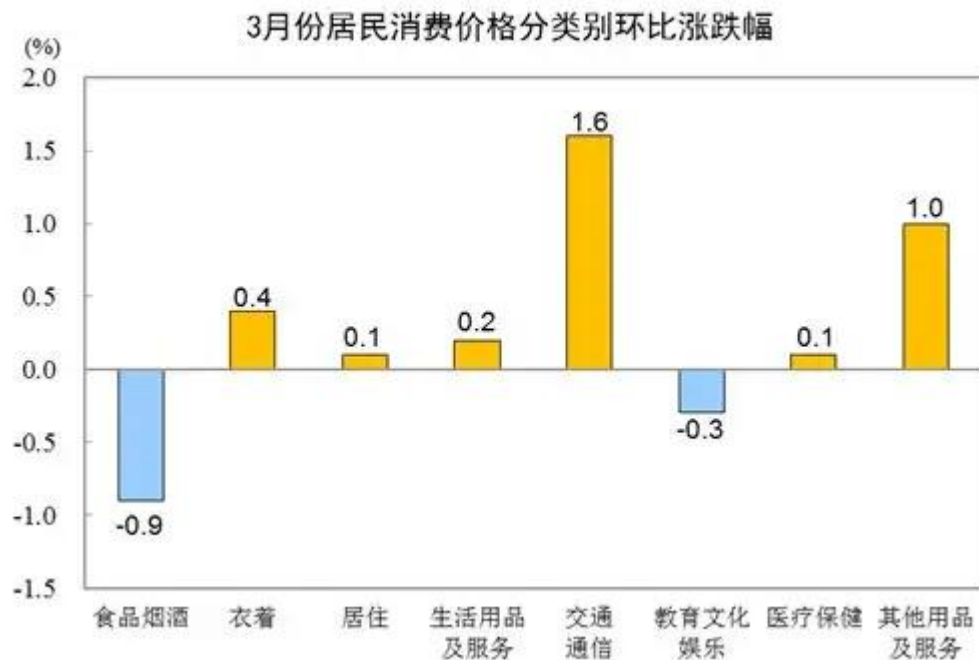
倍数：用一个数据除以另一个数据获得，倍数一般用来表示上升、增长幅度，一般不表示减少幅度。

翻 n 番：指原来数量的 2 的 n 次方。

同比和环比

同比：指的是与历史同时期的数据相比较而获得的比值，反应事物发展的相对性。

环比：指与上一个统计时期的值进行对比获得的值，主要反映事物的逐期发展的情况。



增量：增长的绝对量=现期量-基期量

增速：增长速度=（现期量-基期量）÷基期量

增长率：增量与基期量之比。

增幅：即增长的幅度，也可理解为增量。

基期和现期

基期：被用作参照物的时期称为基期，描述基期的量即为基期量。

现期：相对于基期的称为现期，描述现期的量即为现期量。

YTD:截止到今天为止今年的

LY:last year 去年

YoY——跟上年相比

MAT (moving annual total)：年度动态变化总值

Q4/Q1：4 季度/1 季度

GDP:国内生产总值

GNH (gross national happiness)：国民幸福指数

GNP:国民生产总值