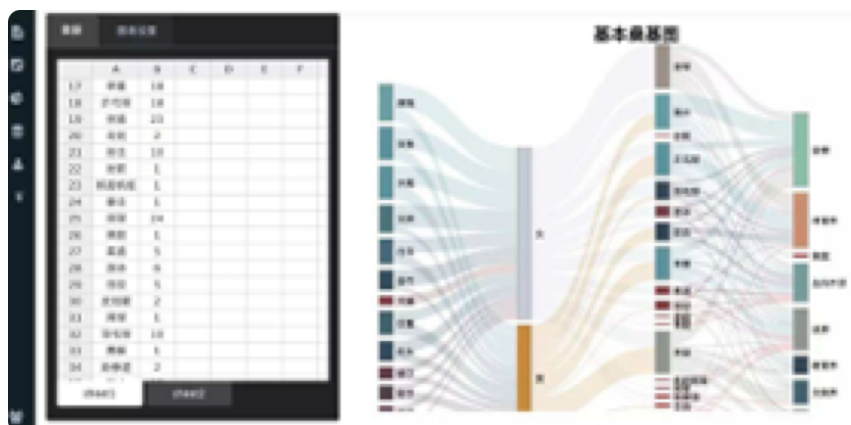
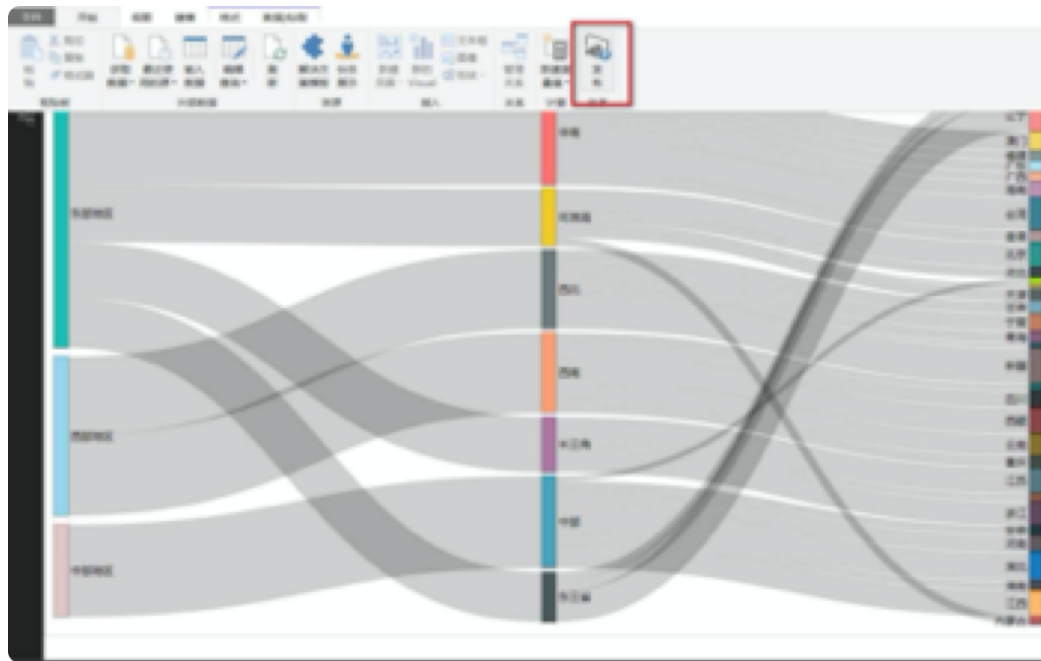


● 桑基图和决策树的比较

桑基图（Sankey diagram），即桑基能量分流图，也叫桑基能量平衡图。它是一种特定类型的流程图，图中延伸的分支的宽度对应数据流量的大小，通常应用于能源、材料成分、金融等数据的可视化分析。因1898年Matthew Henry Phineas Riall Sankey绘制的“蒸汽机的能源效率图”而闻名，此后便以其名字命名为“桑基图”。

桑基图最明显的特征就是，始末端的分支宽度总和相等，即所有主支宽度的总和应与所有分出去的分支宽度的总和相等，保持能量的平衡。



桑基图因为有平衡的问题，首末段平衡，所以很容易联想到“环图”，环的首末相联且相等，所以桑基很容易发现回路，就是图内流量汇聚的汇聚点（通常都是业务的聚集点，比如电商中的支付环节），也就是用于发现业务中的“重点”

决策树(Decision Tree) 是在已知各种情况发生概率的基础上，通过构成决策树来求取净现

值的期望值大于等于零的概率，评价项目风险，判断其可行性的决策分析方法，是直观运用概率分析的一种图解法。由于这种决策分支画成图形很像一棵树的枝干，故称决策树。在机器学习中，决策树是一个预测模型，他代表的是对象属性与对象值之间的一种映射关系。Entropy = 系统的凌乱程度，使用算法ID3, C4.5和C5.0生成树算法使用熵。这一度量是基于信息学理论中熵的概念。

决策树是一种树形结构，其中每个内部节点表示一个属性上的测试，每个分支代表一个测试输出，每个叶节点代表一种类别。



决策树在用户路径分析上直观明了，就是从用户起点（需求端）到用户末端（消费端）的最短路径

做用户增长，主要是就增加这个过程，对于单个用户来说，就是消除长路径（努力做推荐，让用户直接到达消费端，大部分时间推荐会比搜索的效率高一一些，毕竟不是所有人都去用搜索，搜索还是有成本的），对于总用户数来说，就是为用户对接多个消费品（对消费品排序，对不同的用户排序，嗯，做推荐，呵呵），然后对实际消费反打标签给用户，这样反复循环。当然，实际的过程和算法比这个复杂的多。

那是不是桑基图就没用了，桑基图可以发现流程中的“环”，也就是流程中的关键点（流量汇聚点，嗯，你又猜对了，流量消峰，然后控制流量进入它该去的地方，比如搜索结果，支付行为，都是这样），如果你的桑基图都是平行线，没有汇聚，那恭喜你，你没有平台化业务，都是管道流程业务，想做大业务就是要扩大入口，然后改善每个流程环节的效率，因为用户只有走完和离开两个结果（这类业务，做好漏斗效率模型就行了）

神策和友盟都在路径分析上采用桑基，所以可以更好的服务平台型业务，对复杂流程型业务帮助较多，对单观察擦较少（电商类适合）

百度和腾讯采用的决策树，强调C端用户个体行为，观察单体消费更多，对平台型业务考虑的并不多（垂直类app适合）

- 1、决策树在多维度决策的时候可以转化成桑基，就是每个维度会对最终觉得占比，最终占比最高的维度会影响最终觉得
- 2、短路径上，决策树算法更多，比如欧拉回路问题（该算法可以查询最有效的路径，但是没有最优解的），汉明距离可以用于群体决策参考，反打用户标签就是用户的相似程度更高，消费偏好差不多或者完全是从众型消费
- 3、决策树也可以使用自然流量对最终消费需求排序，然后使用协同过滤来做推荐是比较快的方式，比如你浏览过某类产品，系统各种推荐和打折促销秒杀，结合2中的汉明距离，可以筛出一批类似用户都是消费该类产品的价值转化，高级的一些方式还可以把推荐算法集成到客户端sdk中，在本地的sqlite中对服务器的粗排序进行细排序推荐