

数据分析知识总结

应用

Excel

一些注释:

引用其他表使用 ('表名'! E20) 的格式

单元格锁定使用

H20和H\$20的含义不同H20是锁定的是H, 也就是横向拖动, 结果不变

H\$20锁定的是20行, 也就是纵向拖动结果不变

各种函数:

if(条件, 真返回值, 假返回值)

sum(求和范围)

sumif(判断范围, 判断条件, 求和范围)

sumifs(求和范围, 判断范围, 判断条件)

subtotal(求值类型,求值区间)

日期函数: month(),year(),day(),date(year(),month()-1,day())

date(year(),month()+1,day())-1: 每个月最后一天

if()可以嵌套

名词:

环比: 在这个周期中比上一单位时间增加的比率

同比: 这个周期比上一收起增加的比率

进阶:

1.VLOOKUP(B99,'拌客源数据1-8月'!D:E,2,0),查询B99在表区域对应的所对应的行对应的第二列值 (0的意思是精确查询)

=VLOOKUP(I99&"??",F95:G103,2,FALSE)

通配符'?代表一个字符, *代表任意个字符, 同单元格连接时需要用&连接,例如I99&'%'

2.善用index()和match()进行查找匹配

index(数据区域,行位置, 列位置)

match(元素内容, 搜索区域 (行或列), 0) (0代表精准搜索, 返回对应的行或者列值)

INDEX('拌客源数据1-8月'!\$A:\$I,MATCH(\$B113,'拌客源数据1-8月'!\$I:\$I,0),MATCH(D\$111,'拌客源数据1-8月'!\$A\$1:\$I\$1,0))

下拉, B不变, 113增加, D111不变, 故搜索对应的行结果改变SUMIFS(INDEX('拌客源数据1-8月'!A:

X,0,MATCH('常用函数-完成版'!H\$111,'拌客源数据1-8月'!A1:X1,0)), '拌客源数据1-8月'!

I,'常用函数-完成版'!B112)

这里index行位置为0, 代表不对行进行搜索 (因为要聚合), sumifs三部分分别是, 找GMV对应列, 给出门店列, 给出门店名, 通过门店名对门店列按照GMV求和, index第三个为0就代表横向求和 (对行求和了)

插入-筛选器-切片器 (可独立)

插入-数据透视图 (表),数据透视表和数据透视图的区别

还可以插入迷你图

数据源可以从sql, excel, csv等导入, 然后通过拖拽简历映射关系, 然后可以查看每个表的字段类型, 数据等, 建立好关系后可以方便的进行可视化

1.创建分层结构

2.

tableau

绿色代表连续，蓝色代表离散

绿色代表度量，蓝色代表维度

列一般放维度，行放度量（绿色）

标记有四种：颜色，大小，标签，详细信息

按住ctrl可以复制拖动，有些时候为了视觉效果，需要很多图做支撑

明天复习

python

总结pandas函数库

sql

格式：

select-from-join-where-group by-having-order by-limit

窗口函数：在select后，有几种窗口函数

rank()over(partition by year order by sum asc|desc) 1 1 3 4

dense_rank() 1 1 2 3

row_number() 1 2 3 4

count()

max()

min()

sum()

avg()

first_value()

last_value()

lag()：第一个为待查询的参数列名，第二个为向上偏移的位数，第三个参数为超出最上面边界的默认值,向上查询，
1 2 3， 输出为 0, 1, 2

lead() 1 2 3 输出为 2, 3, 0

子查询：

子查询可以出现在from中，将子查询的结果作为新的表来查询（子查询作为筛选的范围）

或者出现在where中，作为进一步筛选的条件（子查询作为筛选的条件）

表连接：

全连接 outer join （一般没有）

内连接 inner join （inner可省略，自动省略所有null值）

左连接left join （省去左表的null值）

右连接right join （省去右表的null值）

case when then when then end

八股