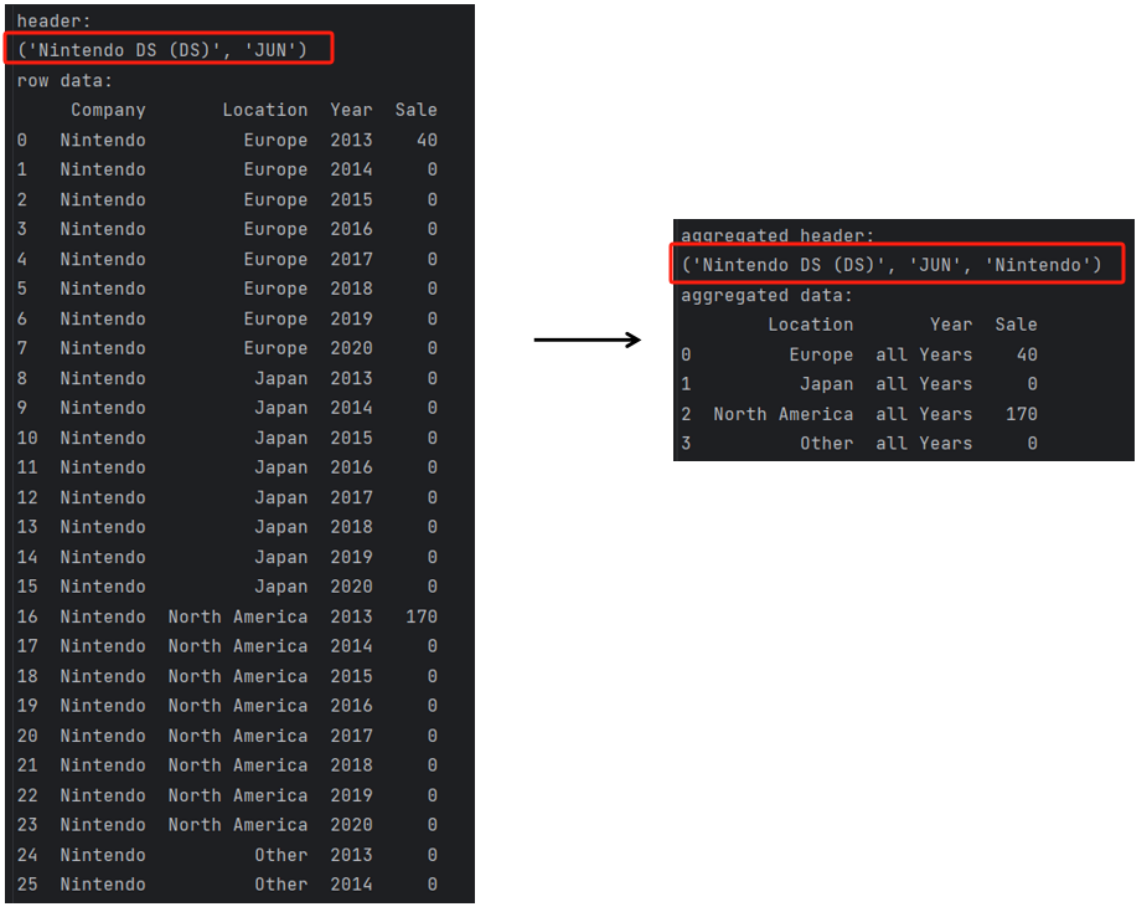
数据处理方式

step1. 枚举header的组合找出全部block data

step2. 对每一个block data，有两种计算insight的方式：

1. no breakdown，直接对整个数据
2. 设置除最后一列（Sale）的每一列为breakdown index，进行分组和聚合

因为最后要生成discription，所以计算前需要先检查主列是否只有一个实体，保证discription里的主语是正确的。如果是的话就提取出来（本质上是将多余的全包含的列提取到header里），这样做会导致insight重复，因为('Nintendo DS(s)', 'JUN', 'Nintendo')这个subspace也会计算出来相同的结果，所以写入的时候要去重



header:
('Nintendo DS (DS)', 'JUN')
row data:

	Company	Location	Year	Sale
0	Nintendo	Europe	2013	40
1	Nintendo	Europe	2014	0
2	Nintendo	Europe	2015	0
3	Nintendo	Europe	2016	0
4	Nintendo	Europe	2017	0
5	Nintendo	Europe	2018	0
6	Nintendo	Europe	2019	0
7	Nintendo	Europe	2020	0
8	Nintendo	Japan	2013	0
9	Nintendo	Japan	2014	0
10	Nintendo	Japan	2015	0
11	Nintendo	Japan	2016	0
12	Nintendo	Japan	2017	0
13	Nintendo	Japan	2018	0
14	Nintendo	Japan	2019	0
15	Nintendo	Japan	2020	0
16	Nintendo	North America	2013	170
17	Nintendo	North America	2014	0
18	Nintendo	North America	2015	0
19	Nintendo	North America	2016	0
20	Nintendo	North America	2017	0
21	Nintendo	North America	2018	0
22	Nintendo	North America	2019	0
23	Nintendo	North America	2020	0
24	Nintendo	Other	2013	0
25	Nintendo	Other	2014	0

aggregated header:
('Nintendo DS (DS)', 'JUN', 'Nintendo')
aggregated data:

	Location	Year	Sale
0	Europe	all Years	40
1	Japan	all Years	0
2	North America	all Years	170
3	Other	all Years	0

一个例子

例如对于header:('Nintendo', 'Nintendo 3DS (3DS)')

block data（筛选后得到的原始数据）：

	Location	Season	Year	Sale
0	Europe	DEC	2013	950
1	Europe	DEC	2014	0
2	Europe	DEC	2015	530
3	Europe	DEC	2016	160
4	Europe	DEC	2017	1400
5	Europe	DEC	2018	200
6	Europe	DEC	2019	40
7	Europe	DEC	2020	10
8	Europe	JUN	2013	550
9	Europe	JUN	2014	440
10	Europe	JUN	2015	650
11	Europe	JUN	2016	560
12	Europe	JUN	2017	170
13	Europe	JUN	2018	170
14	Europe	JUN	2019	0
15	Europe	JUN	2020	30
16	Europe	MAR	2013	1370
17	Europe	MAR	2014	2180
18	Europe	MAR	2015	950
19	Europe	MAR	2016	720
20	Europe	MAR	2017	760
21	Europe	MAR	2018	1190
22	Europe	MAR	2019	320
23	Europe	MAR	2020	70
24	Europe	SEP	2013	660
25	Europe	SEP	2014	0
26	Europe	SEP	2015	210
27	Europe	SEP	2016	0
28	Europe	SEP	2017	200
29	Europe	SEP	2018	0
30	Europe	SEP	2019	0

1. no breakdown：对整个表格计算insight，这组数据里没计算出insight
2. 三组breakdown：

breakdown = 0（以第一列为主列，分组+聚合）：top2

	Location	Sale
0	Europe	14560
1	Japan	15840
2	North Ame...	18270
3	Other	2860

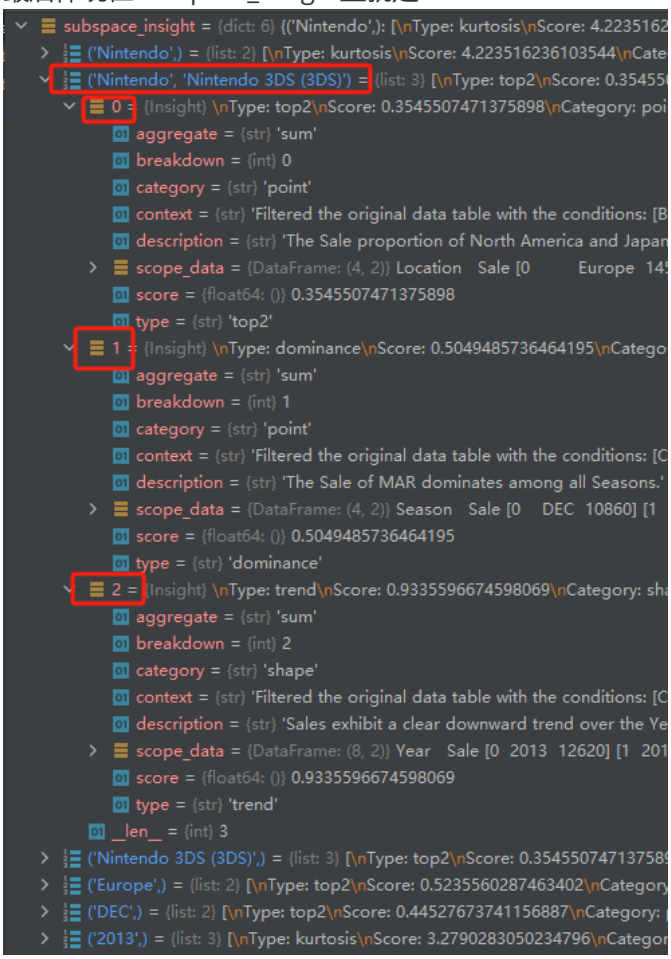
breakdown = 1（以第二列为主列，分组+聚合）：dominance

	Season	Sale
0	DEC	10860
1	JUN	8670
2	MAR	26020
3	SEP	5980

breakdown = 2（以第三列为主列，分组+聚合）： trend

	Year	Sale
0	2013	12620
1	2014	10110
2	2015	8190
3	2016	5620
4	2017	7150
5	2018	5100
6	2019	1790
7	2020	950

最后体现在subspace_insight里就是：



因为对header做枚举的时候是对所有可能的情况做了**组合**，在breakdown的时候又考虑了**顺序**，所以这样做完得到的insight是全的。

比如上面的例子里，在筛选('Nintendo', 'Nintendo 3DS (3DS)')过后得到的数据还有四列，做breakdown就是考虑这个区域里的所有列，分别是location-sale、season-sale、year-sale三组对应关系，得到三组insight。