

Contagem e análise

Transcrição

Conseguimos ordenar a exibição de um valor por vídeo. O problema é que perdemos informação. Antes, mesmo com dificuldade para contagem, tínhamos a quantidade de vezes que os vídeos foram visualizados. A melhor maneira de recuperá-la é com a elaboração de uma tabela, utilizando as informações importantes das duas exibições, para obtermos a quantidade de visualizações de apenas um vídeo.

Para isso, utilizaremos a função `table()`, de "tabela" em inglês, especificando o banco de dados e a variável de interesse.

```
table(aulas$section_id)
```

No Console, ao executarmos o comando, teremos como retorno uma tabela, em que abaixo do código de cada vídeo encontraremos a respectiva contagem de visualizações:

```

13 aulas[33137,3] <- 3255
14
15 sort(aulas$section_id)
16
17 unique(aulas$section_id)
18
19 length(unique(aulas$section_id))
20
21 table(aulas$section_id)
22
22:1 (Top Level) :

```

6	5	20	15	11	11	10	8	4	1	1	1	1	1	1	2
3960	3961	3962	3963	3964	3965	3966	3967	3968	3969	3972	3973	4007	4008	4009	4011
1	1	1	1	1	1	15	9	9	2	1	1	1	1	16	20
4012	4013	4014	4015	4016	4017	4018	4019	4020	4021	4022	4023	4024	4025	4044	4045
8	10	9	7	8	9	7	6	12	7	4	7	1	5	2	4
4046	4047	4048	4059	4060	4061	4066	4087	4088	4089	4090	4091	4107	4108	4109	4110
2	2	1	1	2	1	2	7	2	2	2	2	11	5	5	5
4111	4112	4113	4114	4115	4116	4117	4118	4119	4120	4122	4123	4129	4130	4131	4132
4	3	3	3	1	1	1	1	1	1	4	1	3	2	1	1
4133	4134	4135	4136	4137	4138	4139	4141	4142	4146	4147	4148	4149	4150	4151	4152
1	12	8	6	4	3	3	1	1	10	7	4	4	4	3	1
4156	4162	4163	4164	4165	4166	4167	4168	4169	4171	4176	4177	4186	4187	4237	
3	3	2	2	3	6	1	1	1	1	11	2	2	1	2	

O último 4237, por exemplo, foi visualizado duas vezes.

O RStudio entregou a tabela que pedimos, porém, sua apresentação é de difícil assimilação. O objetivo é saber quais são os vídeos mais e menos populares, então, buscar o mais assistido deles no meio de todos os que são exibidos no Console é complicado, pois são muitas informações juntas.

Resolveremos esse problema com funções que vimos anteriormente: `sort()` para ordenar, e o *nesting* para alinhar uma função à outra:

```
sort(table(aulas$section_id))
```

No Console, após a execução do comando, veremos os vídeos organizados por número de visualizações, e não por código.

```

15 sort(aulas$section_id)
16
17 unique(aulas$section_id)
18
19 length(unique(aulas$section_id))
20
21 table(aulas$section_id)
22
23 sort(table(aulas$section_id))
24

```

24:1 (Top Level) : R Script :

55	55	55	55	56	56	56	56	56	56	57	57	57	57	57	58
2831	1558	1884	2834	2856	3484	1567	2828	3076	549	1965	176	1534	1883	2822	70
58	59	59	59	59	59	60	60	60	62	62	63	63	63	63	65
548	1436	1594	2821	3077	175	640	3078	547	1529	1366	1539	1566	596	1530	71
65	65	66	66	67	69	69	69	70	71	72	72	72	73	73	74
595	2820	1528	639	1527	1964	638	1526	1593	1750	3079	2836	594	425	2819	3320
74	75	76	77	77	77	78	81	81	81	81	82	83	84	84	84
636	637	1882	593	66	635	2853	634	592	633	632	1877	2835	2818	1825	631
86	86	89	91	93	93	94	97	98	99	101	101	101	103	104	106
1795	1824	2832	1773	1823	2817	2833	630	591	1822	1765	1821	1930	1931	2829	629
108	108	108	109	111	116	116	117	118	118	119	120	121	124	124	126
1820	1764	1819	3683	1771	628	1752	1818	1736	627	1817	626	1816	625	1815	
126	128	131	133	134	138	138	143	148	149	162	165	171	176	198	

Segundo a ordenação, o vídeo mais assistido é o último, 1815, com 198 visualizações. No topo do retorno, estarão os menos assistidos. É importante ressaltar que a **análise quantitativa** é boa para informar onde a empresa deve buscar **insights qualitativos**. Ou seja, direcionaremos a análise qualitativa por meio da análise quantitativa.

Assim, o banco de dados é filtrado, poupando a empresa de assistir aos 1983 vídeos da amostra e permitindo que ela atente somente aos dez mais e menos assistidos, verificando características em comum entre eles para identificar e aplicar o que funciona.

Por exemplo, pode ser que os vídeos mais assistidos tenham em comum o assunto, boa didática do professor, boa execução técnica, e assim por diante. Não tem como fazermos inferências qualitativas com base no que analisamos quantitativamente. Porém, poderemos filtrar o que a empresa deve assistir, economizando tempo, dinheiro e trabalho de funcionários que analisam qualitativamente os vídeos.

Neste sentido, em vez de analisarem quase dois mil vídeos, elas analisarão somente os dez mais populares. Geralmente, os vídeos com poucas visualizações também resultam de uma combinação de fatores, que não conseguimos explicar por meio da análise que estamos fazendo. Isto é, não podemos dizer, especificamente, que estão relacionadas à didática do professor, problemas técnicos na execução, informações irrelevantes ao aluno ou a problemas na plataforma da empresa que impedem o acesso dos alunos ao curso.

Não poderemos afirmar categoricamente sobre isso, mas forneceremos atalhos importantes à empresa para que ela descubra a causa das poucas visualizações de um vídeo específico. Aplicaremos a ordenação aos **cursos**, veremos quais têm o maior e o menor número de alunos inscritos. Repetiremos o modelo anterior, especificando diretamente, de maneira ordenada, quais são os cursos com mais e menos alunos inscritos. Substituiremos `section_id` por `course_id`:

```
sort(table(aulas$course_id))
```

Após executarmos o comando no Console, veremos os cursos organizados em ordem crescente, em relação à inscrição de alunos, de modo que veremos que o curso 88 obteve 1593 matrículas, por exemplo.

The screenshot shows an R Studio window with a script editor and a console. The script editor contains the following R code:

```

16
17 unique(aulas$section_id)
18
19 length(unique(aulas$section_id))
20
21 table(aulas$section_id)
22
23 sort(table(aulas$course_id))
24
25

```

The console output shows a sorted table of course IDs and their corresponding number of enrollments. The output is as follows:

93	94	95	96	99	100	101	102	102	105	109	110	111	111	111	112
185	234	110	157	117	142	71	56	196	213	149	187	113	198	206	100
112	112	113	113	114	115	116	118	118	118	124	126	130	136	137	140
114	94	218	55	225	168	12	215	165	23	68	8	73	124	95	81
141	145	149	154	154	155	159	159	165	167	168	170	174	185	188	189
118	137	85	77	167	179	57	79	62	178	152	162	122	5	112	75
196	197	199	207	207	220	225	237	242	246	247	250	253	262	264	267
72	116	125	46	28	96	226	54	37	30	76	208	170	52	91	115
268	270	279	282	285	285	285	286	288	292	293	293	306	307	324	330
47	6	64	60	65	27	104	24	174	50	18	59	7	154	67	63
336	338	359	378	383	387	390	393	405	462	476	489	493	496	604	605
34	58	106	135	134	78	31	88								
609	653	696	728	809	884	999	1593								

É um curso de destaque, com uma diferença relevante de número de matrículas em relação ao segundo mais acessado. Se formos ao topo do retorno, no Console, teremos os cursos menos acessados. Não sabemos por quê um curso é popular, da mesma forma que não sabíamos o motivo da popularidade dos vídeos. Analisamos somente números, e concluímos que o código 88 é o que possui maior número de matrículas.

No *feedback* que faremos à empresa que nos contratou, diremos que os cursos 88, 31 e 78 foram os três mais cursados, e entregar a eles os números dos dez cursos mais acessados para que eles entendam o porquê deles serem populares. Pode ser por uma combinação de fatores, como o mercado estar exigindo determinada qualificação, adquirida por meio desses cursos, ser muito bem feito, ou ainda o professor possuir ótima didática, que conquista os alunos e faz com que eles indiquem para seus amigos, e por aí vai.

Vale ressaltar que os cursos com poucas matrículas não indicam que eles sejam ruins, problemáticos ou mal executados. Pode ser que o curso seja novo na amostra, então ainda não deu tempo de ser cursado. Somente a empresa terá essas informações, e poderá diferenciar isso na análise qualitativa que fará desses cursos.