

Jupyter Notebook Trainee_BI_OLIST_Dataset

June 17, 2021

1 IMPORT

```
[1]: import pandas as pd
import pandasql
```

2 LOAD DATA

```
[2]: order_items = pd.read_csv('olist_order_items_dataset.csv')
order_reviews = pd.read_csv('olist_order_reviews_dataset.csv')
orders = pd.read_csv('olist_orders_dataset.csv')
products = pd.read_csv('olist_products_dataset.csv')
```

```
[3]: order_items.head()
```

```
[3]:
```

	order_id	order_item_id	\
0	00010242fe8c5a6d1ba2dd792cb16214	1	
1	00018f77f2f0320c557190d7a144bdd3	1	
2	000229ec398224ef6ca0657da4fc703e	1	
3	00024acbcd0a6daa1e931b038114c75	1	
4	00042b26cf59d7ce69dfabb4e55b4fd9	1	

	product_id	seller_id	\
0	4244733e06e7ecb4970a6e2683c13e61	48436dade18ac8b2bce089ec2a041202	
1	e5f2d52b802189ee658865ca93d83a8f	dd7ddc04e1b6c2c614352b383efe2d36	
2	c777355d18b72b67abbef9df44fd0fd	5b51032eddd242adc84c38acab88f23d	
3	7634da152a4610f1595efa32f14722fc	9d7a1d34a5052409006425275ba1c2b4	
4	ac6c3623068f30de03045865e4e10089	df560393f3a51e74553ab94004ba5c87	

	shipping_limit_date	price	freight_value
0	2017-09-19 09:45:35	58.90	13.29
1	2017-05-03 11:05:13	239.90	19.93
2	2018-01-18 14:48:30	199.00	17.87
3	2018-08-15 10:10:18	12.99	12.79
4	2017-02-13 13:57:51	199.90	18.14

```
[4]: order_items = order_items.fillna(0)
```

```
[5]: order_items.shape
```

```
[5]: (112650, 7)
```

```
[6]: order_reviews.head()
```

```
[6]:
```

	review_id	order_id	review_score	review_comment_title	review_comment_message	review_creation_date	review_answer_timestamp
0	7bc2406110b926393aa56f80a40eba40	73fc7af87114b39712e6da79b0a377eb	4	NaN	NaN	2018-01-18 00:00:00	2018-01-18 21:46:59
1	80e641a11e56f04c1ad469d5645fdfdde	a548910a1c6147796b98fdf73dbeba33	5	NaN	NaN	2018-03-10 00:00:00	2018-03-11 03:05:13
2	228ce5500dc1d8e020d8d1322874b6f0	f9e4b658b201a9f2ecdecbb34bed034b	5	NaN	NaN	2018-02-17 00:00:00	2018-02-18 14:36:24
3	e64fb393e7b32834bb789ff8bb30750e	658677c97b385a9be170737859d3511b	5	NaN	Recebi bem antes do prazo estipulado.	2017-04-21 00:00:00	2017-04-21 22:02:06
4	f7c4243c7fe1938f181bec41a392bdeb	8e6bfb81e283fa7e4f11123a3fb894f1	5	NaN	Parabéns lojas lannister adorei comprar pela I...	2018-03-01 00:00:00	2018-03-02 10:26:53

```
[7]: order_reviews = order_reviews.fillna(0)
```

```
[8]: order_reviews.shape
```

```
[8]: (100000, 7)
```

```
[9]: orders.head()
```

```
[9]:
```

	order_id	customer_id
0	e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d
1	53cdb2fc8bc7dce0b6741e2150273451	b0830fb4747a6c6d20dea0b8c802d7ef
2	47770eb9100c2d0c44946d9cf07ec65d	41ce2a54c0b03bf3443c3d931a367089
3	949d5b44dbf5de918fe9c16f97b45f8a	f88197465ea7920adcdbec7375364d82
4	ad21c59c0840e6cb83a9ceb5573f8159	8ab97904e6daea8866dbdbc4fb7aad2c

	order_status	order_purchase_timestamp	order_approved_at	\
0	delivered	2017-10-02 10:56:33	2017-10-02 11:07:15	
1	delivered	2018-07-24 20:41:37	2018-07-26 03:24:27	
2	delivered	2018-08-08 08:38:49	2018-08-08 08:55:23	
3	delivered	2017-11-18 19:28:06	2017-11-18 19:45:59	
4	delivered	2018-02-13 21:18:39	2018-02-13 22:20:29	

	order_delivered_carrier_date	order_delivered_customer_date	\
0	2017-10-04 19:55:00	2017-10-10 21:25:13	
1	2018-07-26 14:31:00	2018-08-07 15:27:45	
2	2018-08-08 13:50:00	2018-08-17 18:06:29	
3	2017-11-22 13:39:59	2017-12-02 00:28:42	
4	2018-02-14 19:46:34	2018-02-16 18:17:02	

	order_estimated_delivery_date
0	2017-10-18 00:00:00
1	2018-08-13 00:00:00
2	2018-09-04 00:00:00
3	2017-12-15 00:00:00
4	2018-02-26 00:00:00

```
[10]: orders = orders.fillna(0)
```

```
[11]: orders.shape
```

```
[11]: (99441, 8)
```

```
[12]: products.head()
```

```
[12]:
```

	product_id	product_category_name	\
0	1e9e8ef04dbcff4541ed26657ea517e5	perfumaria	
1	3aa071139cb16b67ca9e5dea641aaa2f	artes	
2	96bd76ec8810374ed1b65e291975717f	esporte_lazer	
3	cef67bcfe19066a932b7673e239eb23d	bebes	
4	9dc1a7de274444849c219cff195d0b71	utilidades_domesticas	

	product_name_lenght	product_description_lenght	product_photos_qty	\
0	40.0	287.0	1.0	
1	44.0	276.0	1.0	
2	46.0	250.0	1.0	
3	27.0	261.0	1.0	
4	37.0	402.0	4.0	

	product_weight_g	product_length_cm	product_height_cm	product_width_cm
0	225.0	16.0	10.0	14.0
1	1000.0	30.0	18.0	20.0

2	154.0	18.0	9.0	15.0
3	371.0	26.0	4.0	26.0
4	625.0	20.0	17.0	13.0

```
[13]: products = products.fillna(0)
```

```
[14]: products.shape
```

```
[14]: (32951, 9)
```

3 VERIFICAÇÃO DE REPETIÇÕES DOS DADOS

```
[15]: ### HÁ product_id REPETIDOS NA TABELA products?
```

```
product_id_repetidos_df = pandasql.sqldf('''SELECT count(product_id),
                                             product_id
                                             FROM products
                                             GROUP BY product_id
                                             HAVING
↳COUNT(product_id) > 1
                                             ORDER BY
↳count(product_id) DESC;''')
```

```
[16]: product_id_repetidos_df.shape
```

```
[16]: (0, 2)
```

```
[17]: ### HÁ order_id REPETIDOS NA TABELA orders?
```

```
order_id_repetidos_df = pandasql.sqldf('''SELECT count(order_id),
                                             order_id
                                             FROM orders
                                             GROUP BY order_id
                                             HAVING COUNT(order_id)
↳> 1
                                             ORDER BY
↳count(order_id) DESC;''')
```

```
[18]: order_id_repetidos_df.shape
```

```
[18]: (0, 2)
```

```
[19]: ### HÁ review_id REPETIDAS NA TABELA order_reviews?
```

```
review_id_repetidos_df = pandasql.sqldf('''SELECT count(review_id),
                                             review_id
                                             FROM order_reviews
```

```

GROUP BY review_id HAVING
COUNT(review_id) > 1
ORDER BY
↪count(review_id) DESC;'''

```

```
[20]: review_id_repetidos_df.shape
```

```
[20]: (802, 2)
```

```

[21]: ### HÁ order_id REPETIDOS NA TABELA order_reviews?

order_id_repetidos_in_rivew_df = pandasql.sqldf('''SELECT count(order_id),
                                                    order_id
                                                    FROM order_reviews
                                                    GROUP BY order_id
                                                    HAVING
↪COUNT(order_id) > 1
                                                    ORDER BY
↪count(order_id) DESC;''')

```

```
[22]: order_id_repetidos_in_rivew_df.shape
```

```
[22]: (555, 2)
```

```

[23]: ### HÁ order_id REPETIDOS NA TABELA order_items?

order_id_repetidos_in_order_items_df = pandasql.sqldf('''SELECT
↪count(order_id),
                                                    order_id
                                                    FROM order_items
                                                    GROUP BY
↪order_id
                                                    HAVING
↪COUNT(order_id) > 1
                                                    ORDER
↪BY count(order_id) DESC;''')

```

```
[24]: order_id_repetidos_in_order_items_df.shape
```

```
[24]: (9803, 2)
```

```

[25]: ### HÁ product_id REPETIDOS NA TABELA order_items?

product_id_repetidos_in_order_items_df = pandasql.sqldf('''SELECT
↪count(product_id),
                                                    product_id FROM
↪order_items

```

```

GROUP BY_
↪product_id

HAVING_
↪COUNT(product_id) > 1

ORDER_
↪BY count(product_id) DESC;''')

```

```
[26]: product_id_repetidos_in_order_items_df.shape
```

```
[26]: (14834, 2)
```

VERIFICANDO SE HÁ LINHAS REPETIDAS NA TABELA `order_reviews`

```
[27]: linhas_repetidas_order_reviews = order_reviews[order_reviews.duplicated()]
```

```
[28]: linhas_repetidas_order_reviews.head()
```

```
[28]: Empty DataFrame
Columns: [review_id, order_id, review_score, review_comment_title,
review_comment_message, review_creation_date, review_answer_timestamp]
Index: []
```

4 PRODUTOS QUE TIVERAM AVALIAÇÃO MÁXIMA

```
[29]: five_stars_items_df = pandasql.sqldf('SELECT * FROM order_reviews WHERE_
↪review_score = 5;')
```

```
[30]: five_stars_items_df.shape
```

```
[30]: (57420, 7)
```

4.0.1 ### filtrando avaliações que tiveram comentários

```
[31]: five_stars_order_df = five_stars_items_df.fillna(0)
```

```
[32]: five_stars_order_df.head()
```

```
[32]:
      review_id      order_id \
0  80e641a11e56f04c1ad469d5645fdfde  a548910a1c6147796b98fdf73dbeba33
1  228ce5500dc1d8e020d8d1322874b6f0  f9e4b658b201a9f2ecdecbb34bed034b
2  e64fb393e7b32834bb789ff8bb30750e  658677c97b385a9be170737859d3511b
3  f7c4243c7fe1938f181bec41a392bdeb  8e6bfb81e283fa7e4f11123a3fb894f1
4  07f9bee5d1b850860defd761afa7ff16  e48aa0d2dcec3a2e87348811bcfdf22b

      review_score review_comment_title \
0                5                    0
```

```

1          5          0
2          5          0
3          5          0
4          5          0

```

```

                                review_comment_message review_creation_date \
0                                0 2018-03-10 00:00:00
1                                0 2018-02-17 00:00:00
2          Recebi bem antes do prazo estipulado. 2017-04-21 00:00:00
3  Parabéns lojas lannister adorei comprar pela I... 2018-03-01 00:00:00
4                                0 2017-07-16 00:00:00

```

```

review_answer_timestamp
0 2018-03-11 03:05:13
1 2018-02-18 14:36:24
2 2017-04-21 22:02:06
3 2018-03-02 10:26:53
4 2017-07-18 19:30:34

```

4.0.2 ### verificando repetições

```
[33]: five_stars_order_df.head()
```

```

[33]:          review_id          order_id \
0  80e641a11e56f04c1ad469d5645fdfde  a548910a1c6147796b98fdf73dbeba33
1  228ce5500dc1d8e020d8d1322874b6f0  f9e4b658b201a9f2ecdecbb34bed034b
2  e64fb393e7b32834bb789ff8bb30750e  658677c97b385a9be170737859d3511b
3  f7c4243c7fe1938f181bec41a392bdeb  8e6bfb81e283fa7e4f11123a3fb894f1
4  07f9bee5d1b850860defd761afa7ff16  e48aa0d2dcec3a2e87348811bcfdf22b

```

```

review_score review_comment_title \
0          5          0
1          5          0
2          5          0
3          5          0
4          5          0

```

```

                                review_comment_message review_creation_date \
0                                0 2018-03-10 00:00:00
1                                0 2018-02-17 00:00:00
2          Recebi bem antes do prazo estipulado. 2017-04-21 00:00:00
3  Parabéns lojas lannister adorei comprar pela I... 2018-03-01 00:00:00
4                                0 2017-07-16 00:00:00

```

```

review_answer_timestamp
0 2018-03-11 03:05:13
1 2018-02-18 14:36:24

```

```

2      2017-04-21 22:02:06
3      2018-03-02 10:26:53
4      2017-07-18 19:30:34

```

```

[34]: repetição_de_review_id_df = pandasql.sqldf('''SELECT count(review_id),
                                                    review_id,
                                                    order_id,
                                                    review_score
                                                    FROM five_stars_order_df
                                                    GROUP BY review_id
                                                    HAVING COUNT(review_id) > 1
                                                    ORDER BY count(review_id) DESC;''')

```

```

[35]: repetição_de_review_id_df.shape

```

```

[35]: (407, 4)

```

```

[36]: repetição_de_review_id_df.head()

```

```

[36]:      count(review_id)      review_id \
0          3  c444278834184f72b1484dfe47de7f97
1          3  7b606b0d57b078384f0b58eac1d41d78
2          3  70509c441d994fa03d6c1457930c9024
3          3  69a1068c3128a14994e3e422e4539e04
4          3  44e9f871226d8a130de3fc39dfbdf0c5

```

```

      order_id  review_score
0  df56136b8031ecd28e200bb18e6ddb2e      5
1  f3028a8f41ea1ee2b461420913663f97      5
2  748f5318aa38169f5901337bfec3afea      5
3  175e7244409a1c34674cdd3407e25808      5
4  1a8d422cb8cdae5221a5a8eb371c49cd      5

```

```

[37]: repetição_de_order_id_df = pandasql.sqldf('''SELECT count(order_id),
                                                    order_id,
                                                    review_id,
                                                    review_score
                                                    FROM five_stars_order_df
                                                    GROUP BY order_id
                                                    HAVING COUNT(order_id) > 1
                                                    ORDER BY count(order_id) DESC;''')

```



```
↪count(order_id) DESC;''')
```

ORDER BY

```
[38]: repetição_de_order_id_df.shape
```

```
[38]: (251, 4)
```

```
[39]: repetição_de_order_id_df.head()
```

```
[39]:      count(order_id)      order_id \
0          3  df56136b8031ecd28e200bb18e6ddb2e
1          2  ff850ba359507b996e8b2fbb26df8d03
2          2  fe041ba1c9f54016432fa6ee91709dbc
3          2  fd95ae805c63c534f1a64589e102225e
4          2  f7c5c5ff5045e13c98901bbbf8e871d4

      review_id  review_score
0  c444278834184f72b1484dfe47de7f97      5
1  80f25f32c00540d49d57796fb6658535      5
2  117b4a19af2b247561f0075f73d6eea1      5
3  eba9ae334beb044f1045c073555df70e      5
4  c92bb5ae7ac1487097e9d330d4e972c7      5
```

```
[40]: five_stars_items_df['order_id'].value_counts()
```

```
[40]: df56136b8031ecd28e200bb18e6ddb2e      3
2f8f31eb2f7b6572836d662a6625c8e4      2
03515a836bb855b03f7df9dee520a8fc      2
8e0ace07f58e07d34ada126b1de0d6a1      2
d07ea9203ae0a6ea15a2eacc87b4c0f7      2
..
58aa4a2e2d2b63972e1c4508ed5f7cd7      1
c3303e89dcbc08a6ba79213095eca5f0      1
b9fce2067f47956396b6782d9bf146ca      1
c31010ef2672e6bae06cf6e265b98138      1
67f2267a555a1794c011883d8fd86430      1
Name: order_id, Length: 57168, dtype: int64
```

```
[41]: five_stars_items_df['review_id'].value_counts()
```

```
[41]: 39b4603793c1c7f5f36d809b4a218664      3
44e9f871226d8a130de3fc39dfbdf0c5      3
0c76e7a547a531e7bf9f0b99cba071c1      3
69a1068c3128a14994e3e422e4539e04      3
1fb4ddc969e6bea80e38deec00393a6f      3
..
c2fd3562e52aa9e4742503b4582b891d      1
a7aba82959025672895ede66985b14a8      1
```

```
0163abd2c1f21f5a8a339e64aaee5abe    1
d88cc94b002f1df7b3b3c9d304e88f9f    1
49a4192653b51e31de164464afbc6e42    1
Name: review_id, Length: 57000, dtype: int64
```

4.0.3 ### VERIFICANDO AVALIAÇÕES REPETIDAS NO MESMO PEDIDO QUE TIVERAM 5 ESTRELAS, POIS ALGUÉM PODE AVALIAR O PRODUTO E COMENTAR DEPOIS

```
[42]: reviews_orders = pandasql.sqldf('''SELECT order_id,
                                     review_id
                                     FROM five_stars_items_df;''')
```

```
[43]: reviews_order_repetidos = reviews_orders[reviews_orders.duplicated()]
```

```
[44]: reviews_order_repetidos.shape
```

```
[44]: (0, 2)
```

5 UNINDO AS TABELA five_stars_items COM order_items PELA CHAVE order_id para saber qual produto teve mais avaliações

```
[45]: five_stars_order_items = pandasql.sqldf('''SELECT five_stars_order_df.
    ↳ review_score,
                                     order_items.order_id,
                                     order_items.order_item_id,
                                     order_items.product_id
    FROM order_items
    JOIN five_stars_order_df
    ON
    ↳ five_stars_order_df.order_id = order_items.order_id;''')
```

```
[46]: five_stars_order_items.head()
```

```
[46]:
```

	review_score	order_id	order_item_id	\
0	5	00010242fe8c5a6d1ba2dd792cb16214	1	
1	5	000229ec398224ef6ca0657da4fc703e	1	
2	5	00042b26cf59d7ce69dfabb4e55b4fd9	1	
3	5	000576fe39319847cbb9d288c5617fa6	1	
4	5	00061f2a7bc09da83e415a52dc8a4af1	1	

	product_id
0	4244733e06e7ecb4970a6e2683c13e61
1	c777355d18b72b67abbeef9df44fd0fd
2	ac6c3623068f30de03045865e4e10089

```
3 557d850972a7d6f792fd18ae1400d9b6
4 d63c1011f49d98b976c352955b1c4bea
```

```
[47]: five_stars_order_items.shape
```

```
[47]: (63626, 4)
```

```
[48]: five_stars_order_items.describe()
```

```
[48]:
```

	review_score	order_item_id
count	63626.0	63626.000000
mean	5.0	1.145821
std	0.0	0.550436
min	5.0	1.000000
25%	5.0	1.000000
50%	5.0	1.000000
75%	5.0	1.000000
max	5.0	15.000000

6 QUAL FOI O PRODUTO QUE RECEBEU MAIOR NÚMERO DE 5 ESTRELAS?

```
[49]: five_stars_mais_avalidados = pandasql.sqldf('''SELECT COUNT(product_id) AS
↳Número_de_avaliações_máximas,
                                product_id
                                FROM five_stars_order_items
                                GROUP BY product_id
                                ORDER BY
↳COUNT(product_id) DESC;''')
```

```
[50]: five_stars_mais_avalidados.head()
```

```
[50]:
```

	Número_de_avaliações_máximas	product_id
0	272	422879e10f46682990de24d770e7f83d
1	271	aca2eb7d00ea1a7b8ebd4e68314663af
2	241	389d119b48cf3043d311335e499d9c6b
3	238	99a4788cb24856965c36a24e339b6058
4	213	368c6c730842d78016ad823897a372db

```
[51]: five_stars_mais_avalidados.shape
```

```
[51]: (23242, 2)
```

7 PRODUTO MAIS AVALIADO

```
[52]: produto_mais_avaliado = pandasql.sqldf('''SELECT * FROM_
      ↪five_stars_mais_avaliados LIMIT 1;''')
```

```
[53]: produto_mais_avaliado.head()
```

```
[53]:
```

	Número_de_avaliações_máximas	product_id
0	272	422879e10f46682990de24d770e7f83d

8 UNINDO A TABELA order_items COM A COLUNA products.product_category_name PELA CHAVE product_id PARA SABER AS CATEGORIAS MAIS VENDIDAS

```
[54]: order_category_df = pandasql.sqldf('''SELECT order_items.order_id,
      order_items.order_item_id,
      order_items.product_id,
      products.product_category_name
      FROM order_items
      JOIN products
      ON order_items.
      ↪product_id = products.product_id''')
```

```
[55]: order_category_df.head()
```

```
[55]:
```

	order_id	order_item_id	\
0	00010242fe8c5a6d1ba2dd792cb16214	1	
1	00018f77f2f0320c557190d7a144bdd3	1	
2	000229ec398224ef6ca0657da4fc703e	1	
3	00024acbcdcf0a6daa1e931b038114c75	1	
4	00042b26cf59d7ce69dfabb4e55b4fd9	1	

	product_id	product_category_name
0	4244733e06e7ecb4970a6e2683c13e61	cool_stuff
1	e5f2d52b802189ee658865ca93d83a8f	pet_shop
2	c777355d18b72b67abbef9df44fd0fd	moveis_decoracao
3	7634da152a4610f1595efa32f14722fc	perfumaria
4	ac6c3623068f30de03045865e4e10089	ferramentas_jardim

```
[56]: order_category_df.shape
```

```
[56]: (112650, 4)
```

9 QUAIS AS TOP 5 CATEGORIAS QUE MAIS VENDERAM PRODUTOS?

```
[57]: five_stars_category_mais_vendido = pandasql.sqldf('''SELECT SUM(order_item_id)
↳AS Vendas,
                                FROM
↳product_category_name
                                GROUP BY
↳product_category_name
                                ORDER
↳BY SUM(order_item_id) DESC;''')
```

```
[58]: five_stars_category_mais_vendido.head(10)
```

```
[58]:
```

	Vendas	product_category_name
0	13665	cama_mesa_banho
1	11540	moveis_decoracao
2	11081	beleza_saude
3	9932	esporte_lazer
4	9874	informatica_acessorios
5	9051	utilidades_domesticas
6	6594	relogios_presentes
7	5874	ferramentas_jardim
8	5161	telefonica
9	4881	automotivo

```
[59]: top_five_category = pandasql.sqldf('''SELECT * FROM
↳five_stars_category_mais_vendido
                                LIMIT 5;''')
```

10 TOP 5 CATEGORIAS QUE MAIS VENDERAM PRODUTOS

```
[60]: top_five_category.head()
```

```
[60]:
```

	Vendas	product_category_name
0	13665	cama_mesa_banho
1	11540	moveis_decoracao
2	11081	beleza_saude
3	9932	esporte_lazer
4	9874	informatica_acessorios

11 Quais pedidos (order_id) com review acima de 3 venderam o maior número de produtos distintos?

```
[61]: above_three_stars_reviews_df = pandasql.sqldf('SELECT * FROM order_reviews_┐
      ↪WHERE review_score > 3;')
```

```
[62]: above_three_stars_reviews_df.head()
```

```
[62]:
```

	review_id	order_id \
0	7bc2406110b926393aa56f80a40eba40	73fc7af87114b39712e6da79b0a377eb
1	80e641a11e56f04c1ad469d5645fdfdde	a548910a1c6147796b98fdf73dbeba33
2	228ce5500dc1d8e020d8d1322874b6f0	f9e4b658b201a9f2ecdecbb34bed034b
3	e64fb393e7b32834bb789ff8bb30750e	658677c97b385a9be170737859d3511b
4	f7c4243c7fe1938f181bec41a392bdeb	8e6bfb81e283fa7e4f11123a3fb894f1

	review_score	review_comment_title \
0	4	0
1	5	0
2	5	0
3	5	0
4	5	0

	review_comment_message	review_creation_date \
0		0 2018-01-18 00:00:00
1		0 2018-03-10 00:00:00
2		0 2018-02-17 00:00:00
3	Recebi bem antes do prazo estipulado.	2017-04-21 00:00:00
4	Parabéns lojas lannister adorei comprar pela I...	2018-03-01 00:00:00

	review_answer_timestamp
0	2018-01-18 21:46:59
1	2018-03-11 03:05:13
2	2018-02-18 14:36:24
3	2017-04-21 22:02:06
4	2018-03-02 10:26:53

```
[63]: above_three_stars_reviews_df.shape
```

```
[63]: (76620, 7)
```

```
[64]: above_three_stars_reviews_df = above_three_stars_reviews_df.fillna(0)
```

12 UNINDO COLUNA order_id.above_three_stars_reviews_df COM TABELA order_items PARA SELECIONAR OS PRODUTOS QUE TIVERAM AVALIAÇÃO ACIMA DE 3

```
[65]: above_three_stars_order_df = pandasql.sqldf('''SELECT order_items.order_id,
                                                    order_items.order_item_id,
                                                    order_items.product_id,
                                                    ↵
↪above_three_stars_reviews_df.review_score
                                                    FROM order_items
                                                    JOIN ↵
↪above_three_stars_reviews_df
                                                    ON order_items.
↪order_id = above_three_stars_reviews_df.order_id;''')
```

```
[66]: above_three_stars_order_df.head()
```

```
[66]:
```

	order_id	order_item_id	\
0	00010242fe8c5a6d1ba2dd792cb16214	1	
1	00018f77f2f0320c557190d7a144bdd3	1	
2	000229ec398224ef6ca0657da4fc703e	1	
3	00024acbcd0a6daa1e931b038114c75	1	
4	00042b26cf59d7ce69dfabb4e55b4fd9	1	

	product_id	review_score
0	4244733e06e7ecb4970a6e2683c13e61	5
1	e5f2d52b802189ee658865ca93d83a8f	4
2	c777355d18b72b67abbeef9df44fd0fd	5
3	7634da152a4610f1595efa32f14722fc	4
4	ac6c3623068f30de03045865e4e10089	5

```
[67]: above_three_stars_order_df.shape
```

```
[67]: (85007, 4)
```

13 UNINDO A COLUNA products.product_category_name À TABELA above_three_stars_order_df PARA SABER AS CATEGORIAS

```
[68]: above_three_stars_order_df = pandasql.sqldf('''SELECT ↵
↪above_three_stars_order_df.order_id,
                                                    ↵
↪above_three_stars_order_df.order_item_id,
                                                    ↵
↪above_three_stars_order_df.product_id,
```

```

        ↪above_three_stars_order_df.review_score,
        products.
        ↪product_category_name
        FROM_
        ↪above_three_stars_order_df
        JOIN products
        ON_
        ↪above_three_stars_order_df.product_id = products.product_id;'''

```

```
[69]: above_three_stars_order_df.shape
```

```
[69]: (85007, 5)
```

```
[70]: above_three_stars_order_df.head()
```

```
[70]:
```

	order_id	order_item_id	\
0	00010242fe8c5a6d1ba2dd792cb16214	1	
1	00018f77f2f0320c557190d7a144bdd3	1	
2	000229ec398224ef6ca0657da4fc703e	1	
3	00024acbcd0a6daa1e931b038114c75	1	
4	00042b26cf59d7ce69dfabb4e55b4fd9	1	

	product_id	review_score	product_category_name
0	4244733e06e7ecb4970a6e2683c13e61	5	cool_stuff
1	e5f2d52b802189ee658865ca93d83a8f	4	pet_shop
2	c777355d18b72b67abbeef9df44fd0fd	5	moveis_decoracao
3	7634da152a4610f1595efa32f14722fc	4	perfumaria
4	ac6c3623068f30de03045865e4e10089	5	ferramentas_jardim

13.0.1 Pedidos (order_id) com review acima de 3 com o maior número de produtos distintos

```

[71]: more_products_by_order = pandasql.sqldf('''SELECT COUNT(product_id) AS_
        ↪quantidade_de_produtos_distintos,
        order_id
        FROM_
        ↪above_three_stars_order_df
        GROUP BY order_id
        ORDER BY_
        ↪COUNT(product_id) DESC;''')

```

```
[72]: more_products_by_order.head()
```

```
[72]:
```

	quantidade_de_produtos_distintos	order_id
0	15	9ef13efd6949e4573a18964dd1bbe7f5
1	15	428a2f660dc84138d969ccd69a0ab6d5

2	12	c05d6a79e55da72ca780ce90364abed9
3	12	aef2bc3845eb581cecce74b769d46c6a
4	12	637617b3ffe9e2f7a2411243829226d0