

# Relatório de análise de dados do Olist Store

Por Luiz Fernando Antonelli Galati  
04/2024

## Introdução

Este relatório descreve os resultados de uma análise de dados de pedidos feitos de forma online em múltiplos marketplaces brasileiros entre 2016 e 2018. Todos os pedidos contaram com a intermediação do Olist Store. Por isso, para simplificar, ao longo do relatório vou falar que os pedidos foram feitos na loja online Olist Store.

Os dados originais foram publicados pelo Olist em: <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>.

## Análise 1

1) Meu primeiro objetivo foi responder à pergunta: **qual a quantidade de itens dos pedidos?** Para isso, contei quantos pedidos tem x itens, com x variando de 1 até o número de itens do(s) pedido(s) com maior quantidade de itens. Isso resultou na seguinte tabela:

Número de itens	Quantidade de pedidos	Porcentagem da quantidade de pedidos
1	88863	90,06%
2	7516	7,62%
3	1322	1,34%
4	505	0,51%
5	204	0,21%
6	198	0,20%
7	22	0,02%
8	8	0,01%
9	3	0,00%
10	8	0,01%
11	4	0,00%
12	5	0,01%
13	1	0,00%
14	2	0,00%
15	2	0,00%
20	2	0,00%
21	1	0,00%
Total Geral	98666	100,00%

Ou seja, 88863 dos 98666 pedidos (90,06%) foram de apenas 1 item, 7516 dos 98666 pedidos (7,62%) foram de 2 itens, e assim por diante. 21 foi a quantidade de itens do pedido com maior número de itens.

Na tabela seguinte, podemos observar a soma acumulada da quantidade de pedidos por número de itens:

Número de itens	Qtd. pedidos (soma acumulada)	Porc. qtd. pedidos (soma acumulada)
1	88863	90,06%
2	96379	97,68%
3	97701	99,02%
4	98206	99,53%
5	98410	99,74%
6	98608	99,94%
7	98630	99,96%
8	98638	99,97%
9	98641	99,97%
10	98649	99,98%
11	98653	99,99%
12	98658	99,99%
13	98659	99,99%
14	98661	99,99%
15	98663	100,00%
20	98665	100,00%
21	98666	100,00%

Com base nesta tabela, concluímos que aproximadamente **99,02% dos pedidos continham 3 ou menos itens** (e que, portanto, **apenas 0,98% dos pedidos continham mais de 3 itens**).

## Análise 2

Meu segundo objetivo foi responder às perguntas:

**(I) quantos consumidores únicos de cada estado do Brasil fizeram pedidos na loja?**

**(II) os três estados mais populosos do Brasil tiveram mais consumidores fazendo pedidos na loja?**

**(III) se um estado é mais populoso do que outro, ele necessariamente teve mais consumidores fazendo pedidos na loja?**

A seguinte tabela responde à primeira pergunta:

Estado	Quantidade de consumidores únicos	Porcentagem do total de consumidores
AC	77	0,08%
AL	401	0,42%
AM	143	0,15%
AP	67	0,07%
BA	3276	3,41%
CE	1311	1,36%
DF	2073	2,16%
ES	1963	2,04%
GO	1951	2,03%
MA	725	0,75%
MG	11255	11,71%
MS	693	0,72%
MT	875	0,91%

PA	949	0,99%
PB	519	0,54%
PE	1604	1,67%
PI	482	0,50%
PR	4882	5,08%
RJ	12377	12,88%
RN	474	0,49%
RO	239	0,25%
RR	45	0,05%
RS	5277	5,49%
SC	3529	3,67%
SE	341	0,35%
SP	40295	41,93%
TO	273	0,28%
<b>Total Geral (Brasil)</b>	<b>96096</b>	<b>100,00%</b>

Para responder às perguntas II e III, extraí, do site do IBGE, as populações de cada estado do Brasil segundo o Censo de 2022. Em seguida, calculei, para cada estado, a porcentagem da sua população em relação ao total da população brasileira, e montei a seguinte tabela, cujos valores de “Porc. do total de consumidores” são iguais aos da tabela anterior:

Estado	População	Porc. do total da população brasileira	Porc. do total de consumidores
AC	830018	0,41%	0,08%
AL	3127683	1,54%	0,42%
AM	3941613	1,94%	0,15%
AP	733759	0,36%	0,07%
BA	14141626	6,96%	3,41%
CE	8794957	4,33%	1,36%
DF	2817381	1,39%	2,16%
ES	3833712	1,89%	2,04%
GO	7056495	3,47%	2,03%
MA	6775805	3,34%	0,75%
MG	20539989	10,11%	11,71%
MS	2757013	1,36%	0,72%
MT	3658649	1,80%	0,91%
PA	8121025	4,00%	0,99%
PB	3974687	1,96%	0,54%
PE	9058931	4,46%	1,67%
PI	3271199	1,61%	0,50%
PR	11444380	5,64%	5,08%
RJ	16055174	7,91%	12,88%
RN	3302729	1,63%	0,49%
RO	1581196	0,78%	0,25%
RR	636707	0,31%	0,05%
RS	10882965	5,36%	5,49%
SC	7610361	3,75%	3,67%
SE	2210004	1,09%	0,35%
SP	44411238	21,87%	41,93%

TO	1511460	0,74%	0,28%
<b>Total geral (Brasil)</b>	<b>203.080.756</b>	<b>100,00%</b>	<b>100,00%</b>

Podemos observar que o estado mais populoso do Brasil, São Paulo, teve a maior porcentagem de consumidores fazendo pedidos no Olist Store. Mas o segundo estado mais populoso do Brasil, Minas Gerais, não teve a segunda maior porcentagem de consumidores fazendo pedidos na loja, e sim a terceira. Já o terceiro estado mais populoso do Brasil, Rio de Janeiro, teve a segunda maior porcentagem de consumidores fazendo pedidos na loja. De qualquer modo, mesmo com essa diferença, pode-se dizer que os três estados mais populosos do Brasil tiveram as três maiores porcentagens de consumidores fazendo pedidos na loja. Portanto, de forma grosseira, **a resposta à segunda pergunta é “sim”**.

Entretanto, essa resposta, apesar de não estar incorreta, é excessivamente simplificadora, pois ignora algumas questões. Por exemplo: em vários estados, há elevada disparidade entre a porcentagem de sua população em relação ao total da população brasileira e a porcentagem de seus consumidores em relação ao total de consumidores que fizeram pedidos na loja.

Analisemos o caso do estado de São Paulo. Embora sua população represente 21,87% do total da população brasileira, 41,93% dos consumidores da loja foram desse estado. Grandes disparidades podem ser vistas também em outros estados.

Além disso, um estado mais populoso do que outro não necessariamente teve mais consumidores fazendo pedidos na loja, nem necessariamente maior percentual relativo de consumidores fazendo pedidos na loja. Confrontemos, por exemplo, os dados da Paraíba (PB) com os de Mato Grosso do Sul (MS). Apesar de a população da Paraíba ser de 3.974.687 pessoas e a de Mato Grosso do Sul ser de 2.757.013 pessoas, este último estado teve 693 consumidores fazendo pedidos, enquanto aquele teve 519. **A resposta à pergunta III, portanto, é não.**

A que se devem essas disparidades? Em outras palavras, podemos formular outras perguntas: **(IV) por que alguns estados tiveram o percentual de consumidores muito maior do que o percentual de sua população em relação ao todo? E, nesse mesmo sentido, por que alguns estados tiveram o percentual de consumidores muito menor do que o percentual de sua população em relação ao todo?** Afinal, se as condições de vida das pessoas fossem perfeitamente iguais em todos os estados, poderíamos supor que, em cada um deles, o percentual de consumidores em relação ao todo seria praticamente igual ao percentual de população em relação ao todo.

**Talvez essas disparidades devam-se, ao menos parcialmente, às diferenças entre as rendas domiciliares per capita dos estados.** Vamos analisar as duas próximas tabelas para ver se isso faz sentido.

A primeira tabela mostra a renda domiciliar per capita de cada estado (esses dados também são do Censo de 2022 e foram extraídos no site do IBGE), juntamente com um indicador que mostra se essa renda é maior ou menor do que a renda domiciliar per capita do Brasil como um todo (que é de R\$ 1893).

Estado	Renda domiciliar per capita (R\$)	Renda menor ou maior que a média brasileira?
AC	1095	Menor
AL	1110	Menor
AM	1.172	Menor
AP	1520	Menor
BA	1139	Menor
CE	1166	Menor

DF	3357	Maior
ES	1915	Maior
GO	2017	Maior
MA	945	Menor
MG	1918	Maior
MS	2030	Maior
MT	1991	Maior
PA	1282	Menor
PB	1320	Menor
PE	1113	Menor
PI	1342	Menor
PR	2115	Maior
RJ	2367	Maior
RN	1373	Menor
RO	1527	Menor
RR	1425	Menor
RS	2304	Maior
SC	2269	Maior
SE	1218	Menor
SP	2492	Maior
TO	1581	Menor
<b>Brasil</b>	<b>1893</b>	

Podemos observar que unidades da Federação como São Paulo, Rio de Janeiro e Distrito Federal, cujo percentual de consumidores fazendo compras no Olist Store foi bem **maior** que o seu percentual de população em relação ao total da população brasileira, têm renda domiciliar per capita **maior** do que a média brasileira. No outro extremo, estados como Acre, Amazonas e Alagoas, cujo percentual de consumidores fazendo compras no Olist Store foi bem **menor** do que o seu percentual de população em relação ao total da população brasileira, têm renda domiciliar per capita **menor** do que a média brasileira.

Isso me levou a estabelecer a seguinte hipótese: quando a porcentagem de consumidores de um estado é maior do que a porcentagem de população desse estado, a renda domiciliar per capita dele é maior do que a renda domiciliar per capita do Brasil; no outro extremo, quando a porcentagem de consumidores de um estado é menor do que a porcentagem de população desse estado, a renda domiciliar per capita dele é menor do que a renda domiciliar per capita do Brasil. Para testar essa hipótese, construí a seguinte tabela:

Estado	Porc. do total de consumidores	Porc. do total da população	Diferença	É verdadeira a hipótese?
AC	0,08%	0,41%	-80,40%	Sim, caso de menoridade
AL	0,42%	1,54%	-72,91%	Sim, caso de menoridade
AM	0,15%	1,94%	-92,33%	Sim, caso de menoridade
AP	0,07%	0,36%	-80,70%	Sim, caso de menoridade
BA	3,41%	6,96%	-51,04%	Sim, caso de menoridade
CE	1,36%	4,33%	-68,50%	Sim, caso de menoridade
DF	2,16%	1,39%	55,50%	Sim, caso de maioridade
ES	2,04%	1,89%	8,21%	Sim, caso de maioridade
GO	2,03%	3,47%	-41,57%	Não
MA	0,75%	3,34%	-77,39%	Sim, caso de menoridade
MG	11,71%	10,11%	15,80%	Sim, caso de maioridade

MS	0,72%	1,36%	-46,88%	Não
MT	0,91%	1,80%	-49,46%	Não
PA	0,99%	4,00%	-75,30%	Sim, caso de menoridade
PB	0,54%	1,96%	-72,41%	Sim, caso de menoridade
PE	1,67%	4,46%	-62,58%	Sim, caso de menoridade
PI	0,50%	1,61%	-68,86%	Sim, caso de menoridade
PR	5,08%	5,64%	-9,85%	Não
RJ	12,88%	7,91%	62,92%	Sim, caso de maioridade
RN	0,49%	1,63%	-69,67%	Sim, caso de menoridade
RO	0,25%	0,78%	-68,06%	Sim, caso de menoridade
RR	0,05%	0,31%	-85,06%	Sim, caso de menoridade
RS	5,49%	5,36%	2,47%	Sim, caso de maioridade
SC	3,67%	3,75%	-2,00%	Não
SE	0,35%	1,09%	-67,39%	Sim, caso de menoridade
SP	41,93%	21,87%	91,74%	Sim, caso de maioridade
TO	0,28%	0,74%	-61,83%	Sim, caso de menoridade

Para mostrar como a tabela deve ser lida, tomemos três exemplos. O primeiro é o do estado do Acre. Do total de consumidores da base divulgada pela Olist, 0,08% eram desse estado, cuja população representa 0,41% da população total brasileira. Isso significa que, para esse estado, o percentual de consumidores é 80,4% **menor** do que o percentual de população; trata-se do que podemos chamar de um “caso de menoridade”. Resta saber se a hipótese sustentou-se neste caso. Como a renda domiciliar per capita do Acre é **menor** do que a renda domiciliar per capita do Brasil, então sim, a hipótese sustentou-se nesse caso de menoridade.

O segundo exemplo é o do estado de São Paulo. Do total de consumidores da base da Olist, 41,93% eram desse estado, cuja população representa 21,87% da população total brasileira. Isso significa que, para esse estado, o percentual de consumidores é 91,74% **maior** do que o percentual de população; trata-se do que podemos chamar de um “caso de maioridade”. Resta saber se a hipótese sustentou-se neste caso. Como a renda domiciliar per capita de São Paulo é **maior** do que a renda domiciliar per capita do Brasil, então sim, a hipótese sustentou-se nesse caso de maioridade.

O terceiro exemplo é o do estado de Goiás. Do total de consumidores da base da Olist, 2,03% eram desse estado, cuja população representa 3,47% da população total brasileira. Isso significa que, para esse estado, o percentual de consumidores é 41,57% **menor** do que o percentual de população. Como a renda domiciliar per capita de Goiás é **maior** do que a renda domiciliar per capita do Brasil, então não, a hipótese não se sustentou nesse caso.

A leitura completa da tabela indica que em 5 dos 27 casos analisados a hipótese não se sustentou. Portanto, a conclusão geral é que a hipótese não se sustenta. Entretanto, 16 dos 21 estados cuja porcentagem de consumidores é menor do que a respectiva porcentagem de população têm renda domiciliar per capita menor do que a renda domiciliar per capita do Brasil inteiro; e 6 dos 6 estados cuja porcentagem de consumidores é maior do que a respectiva porcentagem de população têm renda domiciliar per capita maior do que a renda domiciliar per capita do Brasil inteiro. Portanto, a hipótese, apesar de não ser verdadeira em todos os casos, é verdadeira em muitos deles, e parece indicar que **um estado que tem renda domiciliar per capita maior do que a média brasileira tende a ter também relativamente mais pessoas fazendo compras na loja** (ou seja, tende a ter o percentual de pessoas fazendo compras na loja maior do que seria esperado caso todas as pessoas em todos os estados vivessem sob as

mesmas condições); e um estado que tem renda domiciliar per capita menor do que a média brasileira tende a ter também relativamente menos pessoas fazendo compras na loja (ou seja, tende a ter o percentual de pessoas fazendo compras na loja menor do que seria esperado caso todas as pessoas em todos os estados vivessem sob as mesmas condições).

É de suma importância deixar claro que isso não significa que uma renda domiciliar per capita maior do que a média seja a causa do maior acesso relativo das pessoas à loja em alguns estados, nem que uma renda domiciliar per capita menor do que a média seja a causa do menor acesso relativo das pessoas à loja em outros estados. Eu apenas encontrei uma relação entre essas duas variáveis (renda domiciliar per capita e percentual de consumidores); não significa que haja relação causal entre elas.

Entretanto, este estudo aponta para a **possibilidade de uma alta renda domiciliar per capita de um estado implicar maior presença relativa de consumidores desse estado nas lojas online** e para a **possibilidade de uma baixa renda domiciliar per capita de um estado implicar menor presença relativa de consumidores desse estado nas lojas online**. Para saber se essas implicações são válidas, são necessários mais dados e estudos.

### Análise 3

Meu terceiro objetivo foi analisar a média dos valores pagos pelos pedidos em cada estado. Fui guiado pela seguinte pergunta: **(VI) os consumidores dos estados com maior renda domiciliar per capita pagaram mais pelos pedidos que fizeram?**

Primeiramente, construí a seguinte tabela:

Estado	Soma dos valores pagos pelos pedidos (R\$)	Média dos valores pagos pelos pedidos (R\$)
AC	19.680,62	234,29
AL	96.962,06	227,08
AM	27.966,93	181,60
AP	16.262,80	232,33
BA	616.645,82	170,82
CE	279.464,03	199,90
DF	355.141,08	161,13
ES	325.967,55	154,71
GO	350.092,31	165,76
MA	152.523,02	198,86
MG	1.872.257,26	154,71
MS	137.534,84	186,87
MT	187.029,29	195,23
PA	218.295,85	215,92
PB	141.545,72	248,33
PE	324.850,44	187,99
PI	108.523,97	207,11
PR	811.156,38	154,15
RJ	2.144.379,69	158,53
RN	102.718,13	196,78
RO	60.866,20	233,20

RR	10.064,62	218,80
RS	890.898,54	157,18
SC	623.086,43	165,98
SE	75.246,25	208,44
SP	5.998.226,96	137,50
TO	61.485,33	204,27
<b>Total (Brasil)</b>	<b>16.008.872,12</b>	<b>154,10</b>

Isso significa que os consumidores do estado da Bahia, por exemplo, pagaram em média R\$ 170,82 pelos pedidos que fizeram.

Uma maneira de tentar responder à pergunta (VI) é: em uma tabela, ordenar os estados segundo a média dos valores pagos pelos pedidos; depois, em outra tabela, ordenar os estados segundo a renda domiciliar per capita; por fim, comparar as duas tabelas.

A seguinte tabela mostra os estados ordenados segundo a média dos valores pagos pelos pedidos (da menor média para a maior média):

Estado	Soma dos valores pagos nos pedidos	Média dos valores pagos nos pedidos
SP	5.998.226,96	137,50
PR	811.156,38	154,15
MG	1.872.257,26	154,71
ES	325.967,55	154,71
RS	890.898,54	157,18
RJ	2.144.379,69	158,53
DF	355.141,08	161,13
GO	350.092,31	165,76
SC	623.086,43	165,98
BA	616.645,82	170,82
AM	27.966,93	181,60
MS	137.534,84	186,87
PE	324.850,44	187,99
MT	187.029,29	195,23
RN	102.718,13	196,78
MA	152.523,02	198,86
CE	279.464,03	199,90
TO	61.485,33	204,27
PI	108.523,97	207,11
SE	75.246,25	208,44
PA	218.295,85	215,92
RR	10.064,62	218,80
AL	96.962,06	227,08
AP	16.262,80	232,33
RO	60.866,20	233,20
AC	19.680,62	234,29
PB	141.545,72	248,33
<b>Total Geral</b>	<b>16.008.872,12</b>	<b>154,10</b>



Os dez estados com menor média de valores pagos pelos pedidos estão assinalados em amarelo. Os dez estados com maior média de valores pagos pelos pedidos estão assinalados em verde.

A seguinte tabela mostra os estados ordenados segundo a renda domiciliar per capita (da menor renda para a maior renda):

Estado	Renda domiciliar per capita (R\$)	Menor ou maior que a média brasileira?
MA	945	Menor
AC	1095	Menor
AL	1110	Menor
PE	1113	Menor
BA	1139	Menor
CE	1166	Menor
AM	1.172	Menor
SE	1218	Menor
PA	1282	Menor
PB	1320	Menor
PI	1342	Menor
RN	1373	Menor
RR	1425	Menor
AP	1520	Menor
RO	1527	Menor
TO	1581	Menor
ES	1915	Maior
MG	1918	Maior
MT	1991	Maior
GO	2017	Maior
MS	2030	Maior
PR	2115	Maior
SC	2269	Maior
RS	2304	Maior
RJ	2367	Maior
SP	2492	Maior
DF	3357	Maior
<b>Brasil</b>	<b>1893</b>	

Os 16 estados com renda domiciliar per capita **menor** do que a renda domiciliar per capita do Brasil como um todo estão assinalados em **verde**; os 11 estados com renda domiciliar per capita **maior** do que a renda domiciliar per capita do Brasil como um todo estão assinalados em **amarelo**.

Comparando as duas tabelas, podemos observar que, dos 10 estados com menor média de valores pagos pelos pedidos, 9 têm renda domiciliar per capita maior do que a do Brasil como um todo. No outro extremo, dos 10 estados com maior média de valores pagos pelos pedidos, 10 têm renda domiciliar per capita menor do que a do Brasil como um todo. Portanto, **a tendência parece ser a seguinte: os consumidores dos estados com menor renda domiciliar per capita pagaram mais pelos pedidos que fizeram, enquanto os consumidores dos estados com maior renda domiciliar per capita pagaram menos pelos pedidos que fizeram.**

Por que isso ocorreu, se, como vimos na análise anterior, uma maior renda domiciliar per capita parece estar associada a uma maior quantidade relativa de consumidores fazendo pedidos na loja? Uma hipótese para explicar isso é que é maior a desigualdade dos estados com renda domiciliar per capita menor do que a renda brasileira. Se isso for verdadeiro, talvez, nesses estados, a quantidade de pessoas que acessam as lojas online seja relativamente menor, porque é grande a quantidade de pessoas que não têm poder aquisitivo suficiente para acessar a internet e essas lojas; entretanto, as relativamente poucas pessoas que as acessam têm poder aquisitivo alto, já que a desigualdade de renda do estado é grande, e por isso gastam mais nos pedidos que fazem.

Mais uma vez, cabe assinalar que tanto a tendência explicitada acima, realçada em **negrito**, quanto a hipótese postulada para explicá-la devem ser objeto de mais estudos para terem sua validade testada.