Drivers of bad reviews and seller problem segmentation

Olist e-commerce · 22/10/2026

This report analyses 1- and 2-star reviews in the Olist store data to explain what drives bad reviews across fulfilment, delivery, products, sellers and regions. After deduplicating to the latest review per order, the analysis shows that the biggest drivers are non-delivery, severe lateness, and multi-item order complexity. Portuguese review comments confirm that the main complaint families are delivery and non-delivery problems, product-quality issues, missing or wrong items, and weak service recovery. Sellers can be grouped into three operational segments: Delivery and fulfilment-risk sellers, Bulky-item and completeness-risk sellers, and Baseline / lower-risk mixed sellers. A blended priority score highlights a relatively small set of sellers that account for a large share of attributable bad reviews and should be targeted first by operations.

Business question and analytical approach

The objective was to explain what drives **bad reviews**, defined as **1- and 2-star reviews**, and to turn those findings into an operational seller-prioritisation view. The work covered structured order, product, seller and regional factors, plus Portuguese review text.
Key analytical choices:

- Reviews were **deduplicated to one row per order**, keeping the **latest review**.
- Delivery metrics used **delivered orders only** when applying the late-order rule.
- The workspace rule was applied consistently: an order is **late when** `order_delivered_customer_date > order_estimated_delivery_date`, using the full timestamps.
- Seller analysis used **single-seller orders** for cleaner attribution.
- Category analysis used **single-category orders** for cleaner attribution.
- Complaint analysis combined an interpretable **rule-based theme framework** with an unsupervised **TF-IDF + NMF topic model**.
- Seller prioritisation blended **impact** (bad-review volume) and **severity** (shrunk bad-review rate).

The final analytical base contains **98,673 reviewed orders** after review deduplication.

Sample	Rows	Bad review rate (%)
all_reviewed_orders	98 673	14.69
delivered_reviewed_orders	95 824	12.81
single_category_delivered_orders	95 050	12.57
bad_comments	10 944	100
seller_segmentation_input	796	13.56

The main drivers of bad reviews

The structured analysis points to a clear hierarchy of drivers.

1. Fulfilment failure and order status

Order status is one of the strongest signals in the data. Among reviewed orders, non-delivered statuses have extremely high bad-review rates:

- **processing**: **92.88%**
- **unavailable**: **84.87%**
- **invoiced**: **82.85%**

- **canceled**: **76.86%**
- **shipped**: **69.67%**
- **delivered**: **12.81%**

This means that unresolved fulfilment failure is a dominant source of 1- and 2-star reviews.

2. Late delivery

Within delivered orders, lateness is the strongest delivery-related driver. Using the workspace rule exactly — **late =** `order_delivered_customer_date > order_estimated_delivery_date` **at full timestamp level** — the bad-review rate rises sharply as lateness worsens:

- **on time or early**: **9.22%**
- **0-3 days late**: **19.12%**
- **3-7 days late**: **61.25%**
- **7+ days late**: **78.44%**

On the cleaner delivered modelling sample restricted to **single-seller** and **single-category** orders, the same pattern holds with nearly identical values:

- **on time or early**: **8.62%**
- **0-3 days late**: **18.92%**
- **3-7 days late**: **61.26%**
- **7+ days late**: **78.31%**

The explanatory logistic models support the same conclusion. Relative to delivered on time or early, severe lateness carries very large odds ratios, and these effects remain stable across 2017 and 2018.

3. Order complexity and logistics burden

The delivered-order driver model shows that **multi-item orders** remain materially riskier than single-item orders even after controlling for delivery timing, category and region. Direct descriptive rates in the clean delivered sample show:

- **multi-item** delivered orders: **23.61%** bad-review rate
- **single-item** delivered orders: **11.27%**

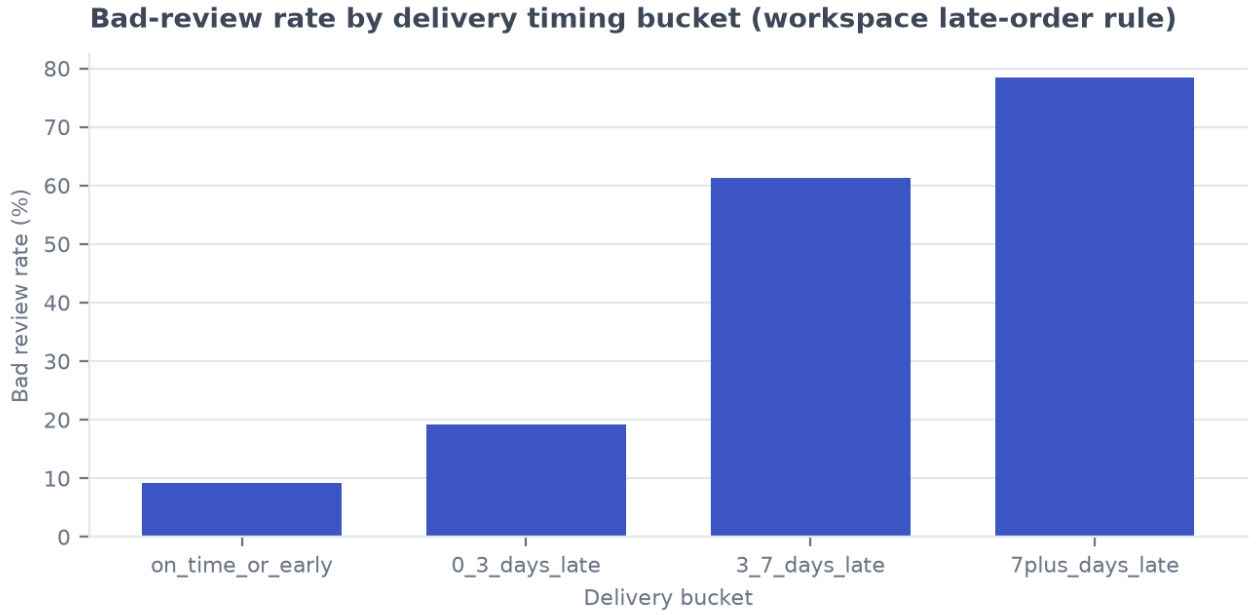
Bulky and heavy items also carry extra risk:

- **bulky 100k+ cm³** delivered orders: **17.26%** bad-review rate
- **under 100k cm³**: **12.25%**
- **heavy 5kg+** delivered orders: **14.01%**
- **under 5kg**: **12.08%**

These patterns suggest that complex fulfilment, larger items, and harder last-mile execution all increase the chance of a bad experience.



Order status	Reviewed orders	Bad reviews	Bad review rate (%)
processing	295	274	92.88
unavailable	595	505	84.87
invoiced	309	256	82.85
canceled	605	465	76.86
shipped	1 032	719	69.67
created	3	2	66.67
approved	2	1	50
delivered	95 832	12 273	12.81



Delivery bucket	Reviewed orders	Bad reviews	Bad review rate (%)
on_time_or_early	88 163	8 131	9.22
0_3_days_late	2 636	504	19.12

Delivery bucket	Reviewed orders	Bad reviews	Bad review rate (%)
3_7_days_late	1 773	1 086	61.25
7plus_days_late	3 252	2 551	78.44

Delivery bucket	Reviewed orders	Bad reviews	Bad review rate (%)
on_time_or_early	85 569	7 373	8.62
0_3_days_late	2 585	489	18.92
3_7_days_late	1 732	1 061	61.26
7plus_days_late	3 200	2 506	78.31

Feature	Coef	Odds ratio
C(item_structure, Treatment(reference="single_item"))[T.multi_item]	1.11	3.04
C(seller_state_group, Treatment(reference="SP"))[T.DF]	0.24	1.28
has_bulky_item_100k_cm3_plus	0.24	1.27
C(customer_state_group, Treatment(reference="SP"))[T.PA]	0.20	1.22
C(customer_state_group, Treatment(reference="SP"))[T.RJ]	0.17	1.19
has_heavy_item_5kg_plus	0.10	1.11
same_state_flag	-0.20	0.82

Item structure	Reviewed orders	Bad reviews	Bad review rate (%)
multi_item	8 087	1 909	23.61
single_item	86 296	9 724	11.27

Bulky flag	Reviewed orders	Bad reviews	Bad review rate (%)
bulky_100k_cm3_plus	1 367	236	17.26
under_100k_cm3	93 016	11 397	12.25

Heavy flag	Reviewed orders	Bad reviews	Bad review rate (%)
heavy_5kg_plus	11 881	1 665	14.01
under_5kg	82 502	9 968	12.08

Product and regional risk patterns

Even after accounting for delivery timing and complexity, some categories and regions remain riskier than others.

Higher-risk categories by contribution and rate

Among large single-category reviewed orders, the biggest contributors to bad reviews include:

- **bed_bath_table**: **1,489** bad reviews, **16.33%** bad-review rate
- **health_beauty**: **1,086**, **12.48%**
- **computers_accessories**: **1,053**, **15.96%**
- **furniture_decor**: **984**, **15.88%**
- **sports_leisure**: **970**, **12.76%**
- **watches_gifts**: **829**, **14.97%**

The delivered-order model also showed elevated relative risk for categories such as **office_furniture**, **computers_accessories**, **telephony**, **bed_bath_table**, and **furniture_decor**.

Regional patterns

Bad-review rates vary meaningfully across customer states with sufficient volume. Examples:

- **RJ**: **20.69%**
- **CE**: **19.68%**
- **BA**: **18.89%**
- **SP**: **12.64%**
- **PR**: **12.45%**

In the clean delivered model, **same-state fulfilment** is protective: the odds ratio for the same-state flag is **0.82**. Direct delivered-order rates also support this:

- **different-state** delivered orders: **13.64%** bad-review rate
- **same-state** delivered orders: **9.99%**

So some of the regional effect likely reflects network distance, route difficulty, and service-level variation.

Category en	Reviewed orders	Bad reviews	Bad review rate (%)
bed_bath_table	9 116	1 489	16.33
health_beauty	8 701	1 086	12.48
computers_accessories	6 598	1 053	15.96
furniture_decor	6 197	984	15.88
sports_leisure	7 602	970	12.76
watches_gifts	5 537	829	14.97
housewares	5 741	732	12.75
telephony	4 141	632	15.26
auto	3 842	544	14.16
toys	3 803	472	12.41

Customer state	Reviewed orders	Bad reviews	Bad review rate (%)
RJ	12 687	2 625	20.69
CE	1 326	261	19.68
BA	3 340	631	18.89
PE	1 635	274	16.76
DF	2 128	321	15.08
ES	2 006	300	14.96
GO	2 007	297	14.80
SC	3 609	528	14.63
RS	5 443	728	13.37
MG	11 554	1 538	13.31
SP	41 472	5 242	12.64
PR	5 019	625	12.45

State match	Reviewed orders	Bad reviews	Bad review rate (%)
different_state	60 428	8 240	13.64
same_state	33 955	3 393	9.99

Feature	Coef	Odds ratio
C(item_structure, Treatment(reference="single_item"))[T.multi_item]	1.11	3.04
C(seller_state_group, Treatment(reference="SP"))[T.DF]	0.24	1.28
has_bulky_item_100k_cm3_plus	0.24	1.27
C(customer_state_group, Treatment(reference="SP"))[T.PA]	0.20	1.22

Feature	Coef	Odds ratio
C(customer_state_group, Treatment(reference="SP"))[T.RJ]	0.17	1.19
has_heavy_item_5kg_plus	0.10	1.11
same_state_flag	-0.20	0.82

What customers say in Portuguese review comments

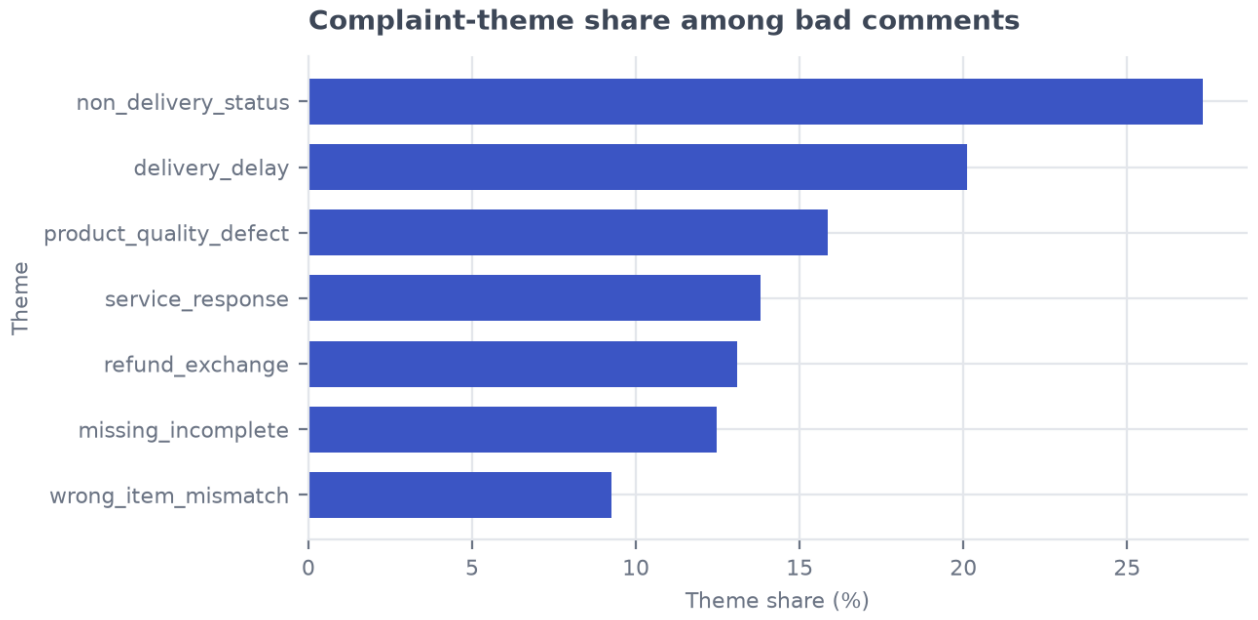
The text analysis supports the structured findings and adds customer-language detail.
On **10,944** deduplicated bad reviews with comments, the rule-based theme system covers **75.85%** of comments. The largest complaint families are:

- **non_delivery_status**: **27.30%**
- **delivery_delay**: **20.12%**
- **product_quality_defect**: **15.87%**
- **service_response**: **13.80%**
- **refund_exchange**: **13.08%**
- **missing_incomplete**: **12.46%**
- **wrong_item_mismatch**: **9.25%**

The unsupervised NMF topic model tells the same story in a different way. The biggest discovered topic is **Delivery problem + contact / refund friction** with **4,761 comments (43.50%)**. Other large topics include **Wrong item, quantity mismatch, or partial delivery** (**21.43%**) and **Poor quality / not as advertised** (**9.94%**).
The comments show that bad reviews are not only about delivery delays themselves. Customers also describe:

- not receiving the order at all
- receiving the wrong item or missing units
- poor quality or defects
- struggling to get an answer, refund, or resolution

That means **service recovery** is part of the bad-review problem, not just logistics execution.



Theme	Comments	Theme share (%)
non_delivery_status	2 988	27.30
delivery_delay	2 202	20.12
product_quality_defect	1 737	15.87

Theme	Comments	Theme share (%)
service_response	1 510	13.80
refund_exchange	1 431	13.08
missing_incomplete	1 364	12.46
wrong_item_mismatch	1 012	9.25

Topic ID	Topic label	Comment count	Comment share (%)	Top terms
1	Delivery problem + contact / refund friction	4 761	43.50	errado, contato, aguardo, mercadoria, dinheiro, correios, pedi, porem, outro, saber, resposta, gostaria
0	Wrong item, quantity mismatch, or partial delivery	2 345	21.43	comprei, unidades, duas, outro, comprei duas, entregaram, somente, comprei dois, kit, tres, mes, modelo
3	Poor quality / not as advertised	1 088	9.94	qualidade, pessima, pessima qualidade, ruim, gostei, material, baixa, foto, baixa qualidade, diferente, original, inferior
5	Long wait / still not arrived	789	7.21	nada, hoje, chegar, mes, passou, esperando, previsao, ver, faz, deram, data, nenhuma
6	Missing units or incomplete order	608	5.56	dois, comprei dois, paguei, somente, itens, faltando, entregaram, kits, porem, consta, item, caixa
4	General dissatisfaction / do not recommend	514	4.70	recomendo, ninguem, gostei, pessimo, super, empresa, atendimento, funciona, dentro, comprado, enviaram, porque
2	Waiting for response / resolution	428	3.91	aguardando, retorno, hoje, resposta, solucao, sobre, posicao, empresa, recebimento, chegar, posso, envio
7	Defect and exchange / return request	411	3.76	defeito, troca, funciona, relogio, trocar, devolucao, antes, solicitei, bem, quebrado, porem, devolver

Theme	Late delivered order	On time or early delivered order	Non delivered order
delivery_delay	33.54	13.12	23.90
missing_incomplete	4.73	17.77	5.25
non_delivery_status	49.69	13.34	42.68
product_quality_defect	7.52	21.40	8.83
refund_exchange	6.94	13.38	22.85
service_response	14.08	13.32	15.19
wrong_item_mismatch	2.38	14.01	2.59

Theme pair	Comments	Share of bad comments (%)
delivery_delay + non_delivery_status	657	6
non_delivery_status + service_response	457	4.18
delivery_delay + service_response	345	3.15
refund_exchange + service_response	342	3.12
product_quality_defect + refund_exchange	328	3
product_quality_defect + service_response	274	2.50
non_delivery_status + refund_exchange	272	2.49
delivery_delay + product_quality_defect	246	2.25

Theme pair	Comments	Share of bad comments (%)
missing_incomplete + service_response	244	2.23
delivery_delay + refund_exchange	239	2.18
wrong_item_mismatch + refund_exchange	230	2.10
delivery_delay + missing_incomplete	176	1.61
wrong_item_mismatch + product_quality_defect	169	1.54
non_delivery_status + missing_incomplete	165	1.51
wrong_item_mismatch + service_response	163	1.49

Seller segmentation and prioritisation

The seller analysis focuses on **796 sellers** with at least **20 reviewed single-seller orders**, covering **86.07%** of attributable bad reviews in the seller-analysis population.
A 3-segment solution gave the best balance of interpretability and stability.

Segment 1 — Delivery and fulfilment-risk sellers

- **178 sellers**
- **16,432** reviewed orders
- **3,350** bad reviews
- average bad-review rate: **21.39%**
- share of segmented bad reviews: **29.27%**

This is the highest-risk segment. These sellers over-index on late delivery, non-delivery and delivery-related complaint themes.

Segment 2 — Bulky-item and completeness-risk sellers

- **91 sellers**
- **8,499** reviewed orders
- **1,274** bad reviews
- average bad-review rate: **14.71%**
- average heavy-item share: **75.10%**
- average bulky-item share: **11.34%**

This segment is exposed to large-item logistics and incomplete-order issues.

Segment 3 — Baseline / lower-risk mixed sellers

- **527 sellers**
- **59,487** reviewed orders
- **6,820** bad reviews
- average bad-review rate: **9.98%**
- share of segmented bad reviews: **59.59%**

This segment is lower risk on average, but still contributes many bad reviews because of its scale.

Priority concentration

Using the blended priority score:

- top **20** priority sellers account for **1,218** bad reviews (**10.64%** of segmented bad reviews)
- top **50** account for **2,981** (**26.05%**)
- top **100** account for **4,682** (**40.91%**)

Among the **top 50 priority sellers**, **36** belong to the **Delivery and fulfilment-risk sellers** segment. This is the clearest first wave for operations.

Seller segment ID	Seller segment label	Sellers	Reviewed orders	Bad reviews	Avg bad review rate	Avg bad review rate shrunk	Avg non delivered share	+13 more columns
2	Delivery and fulfilment-risk sellers	178	16 432	3 350	0.21	0.19	0.04	…
1	Bulky-item and completeness-risk sellers	91	8 499	1 274	0.15	0.14	0.02	…
0	Baseline / lower-risk mixed sellers	527	59 487	6 820	0.10	0.11	0.01	…

Seller segment label	Sellers in top50	Bad reviews	Avg priority score
Delivery and fulfilment-risk sellers	36	1 753	92.68
Baseline / lower-risk mixed sellers	10	872	90.30
Bulky-item and completeness-risk sellers	4	356	94.03

Top n sellers	Bad reviews	Share of segmented bad reviews (%)
20	1 218	10.64
50	2 981	26.05
100	4 682	40.91

Seller ID	Seller segment label	Reviewed orders	Bad reviews	Bad review rate (%)	Bad review rate shrunk (%)	Priority score
2eb70248d66e0e3ef83659f71b244378	Delivery and fulfilment-risk sellers	196	99	50.51	47.11	99.06
1ca7077d890b907f89be8c954a02686a	Delivery and fulfilment-risk sellers	112	69	61.61	54.36	98.49
7c67e1448b00f6e969d365cea6b010ab	Bulky-item and completeness-risk sellers	966	245	25.36	25.13	98.12
88460e8ebdecbfecb5f96018339819 30	Delivery and fulfilment-risk sellers	228	66	28.95	27.72	97.05
897060da8b9a21f655304d50fd935913	Delivery and fulfilment-risk sellers	308	78	25.32	24.62	96.86
7040e82f899a04d1b434b795a43b4617	Delivery and fulfilment-risk sellers	205	56	27.32	26.11	96.11
a49928bcdf77c55c6d6e05e09a9b4ca5	Delivery and fulfilment-risk sellers	95	37	38.95	34.57	95.85
710e3548e02bc1d2831dfc4f1b5b14d4	Delivery and fulfilment-risk sellers	133	43	32.33	29.90	95.79
54965bbe3e4f07ae045b90b0b8541f52	Delivery and fulfilment-risk sellers	73	32	43.84	37.37	95.38
855668e0971d4dfd7bef1b6a4133b41b	Delivery and fulfilment-risk sellers	303	67	22.11	21.59	95.32
1835b56ce799e6a4dc4eddc053f04066	Delivery and fulfilment-risk sellers	378	81	21.43	21.04	95.29
8444e55c1f13cd5c179851e5ca5ebd00	Delivery and fulfilment-risk sellers	97	34	35.05	31.41	95.07

Seller ID	Seller segment label	Reviewed orders	Bad reviews	Bad review rate (%)	Bad review rate shrunk (%)	Priority score
18a349e75d307f4b4cc646a691ed4216	Delivery and fulfilment-risk sellers	118	36	30.51	28.08	94.79
972d0f9cf61b499a4812cf0bfa3ad3c4	Delivery and fulfilment-risk sellers	73	30	41.10	35.22	94.79
3c7c4a49ec3c6550809089c6a2ca9370	Delivery and fulfilment-risk sellers	137	38	27.74	25.96	94.76

Evaluation and confidence in the findings

The main findings were tested for robustness before deployment.

Driver-model quality

- overall fulfilment model ROC AUC: **0.7523**
- delivered clean model ROC AUC: **0.7337**

These are strong enough for explanatory business use.

Temporal robustness

The biggest signals remain stable across **2017** and **2018**:

- non-delivery remains roughly **36x to 41x** riskier than delivered on time or early
- delivered **7+ days late** remains roughly **34x to 45x** riskier
- multi-item risk remains materially elevated in both years
- same-state fulfilment remains protective in both years

Segmentation stability

The final 3-segment seller solution is stable enough for action:

- silhouette score: **0.1296**
- median adjusted Rand index across 30 reruns: **0.9393**
- minimum adjusted Rand index across reruns: **0.8975**

The silhouette is modest, which is expected in overlapping operational data, but the segment structure is repeatable and business-meaningful.

Main caveats

- These are explanatory associations, not causal estimates.
- The text-theme framework is strongest as a **multi-label complaint-family system**.
- Some complaint comments remain in `other_or_unclear` or mix several complaint types in one message.

Model	Rows used	Bad review rate (%)	Roc auc	Brier score	Pseudo r2 mcfadden	Aic
overall_fulfilment_model	97 917	14.19	0.75	0.09	0.22	62 664.30
delivered_clean_driver_model	94 383	12.33	0.73	0.09	0.18	57 909.40

Model	Comparison	2017	2018
delivered_clean_driver_model	0_3_days_late_vs_on_time_or_early	2.88	2.29
delivered_clean_driver_model	3_7_days_late_vs_on_time_or_early	16.39	18.66
delivered_clean_driver_model	7plus_days_late_vs_on_time_or_early	33.92	44.55
delivered_clean_driver_model	bulky_100k_plus	1.10	1.53
delivered_clean_driver_model	multi_item_vs_single_item	2.70	3.38

Model	Comparison	2017	2018
delivered_clean_driver_model	office_furniture_vs_health_beauty	1.56	2.10
delivered_clean_driver_model	same_state_flag	0.87	0.78
overall_fulfilment_model	delivered_0_3_days_late_vs_on_time_or_early	2.91	2.20
overall_fulfilment_model	delivered_3_7_days_late_vs_on_time_or_early	16.14	18.19
overall_fulfilment_model	delivered_7plus_days_late_vs_on_time_or_early	33.56	43.49
overall_fulfilment_model	multi_item_vs_single_item	3.12	3.86
overall_fulfilment_model	non_delivered_vs_on_time_or_early	36.46	41.08

N sellers	N clusters	Silhouette score	Median assigned center distance	Median distance margin	Pct with positive margin
796	3	0.13	2.34	0.80	100

Runs compared	Min ari vs final	Median ari vs final	Max ari vs final	Pct runs ari 1
30	0.90	0.94	1	3.33

Recommendations

1. Tackle delivery and fulfilment-risk sellers first

This is the highest-severity seller segment and dominates the top priority list. Immediate actions:

- weekly monitoring of late share and non-delivered share
- seller-specific escalation for high-priority sellers
- SLA review for dispatch, carrier handoff and exception handling
- rapid resolution workflow for shipped-but-stalled and unavailable orders

2. Create a separate playbook for bulky-item and completeness-risk sellers

These sellers face different operational failure modes. Actions:

- packaging and pick/pack accuracy audits
- completeness checks for multi-unit and large-item orders
- carrier suitability review for heavy or bulky items
- proactive delivery messaging and issue triage for high-friction item classes

3. Prioritise multi-item order quality control

Because multi-item orders show much higher bad-review rates, operations should review:

- picking and consolidation accuracy
- partial shipment communication
- missing-item prevention and proof-of-completeness controls

4. Focus category interventions where both risk and volume are meaningful

High-value categories for investigation and action include:

- **bed_bath_table**
- **computers_accessories**
- **furniture_decor**
- **telephony**
- **watches_gifts**
- **office_furniture**

These categories combine substantial bad-review contribution with elevated rates.

5. Treat service recovery as part of the root-cause programme

Customers frequently mention lack of response, refund friction and unresolved incidents. Improvements should include:

- faster customer-contact SLAs

- clearer refund/exchange progress visibility
- escalation rules when delivery or quality incidents remain unresolved

6. Use regional routing and same-state fulfilment where possible

Same-state fulfilment is associated with better outcomes. Where operationally feasible:

- reduce long-distance routing for sensitive categories
- review regional carrier performance in higher-risk states such as **RJ**, **CE**, and **BA**
- use route- or region-specific intervention tracking

7. Work the priority list in waves

A practical sequence is:

- **Wave 1:** top 20 priority sellers
- **Wave 2:** top 21-50 sellers
- **Wave 3:** segment-level process fixes for bulky-item and multi-item issues

This balances speed of impact with broader operational improvement.

Seller ID	Seller segment label	Reviewed orders	Bad reviews	Bad review rate (%)	Bad review rate shrunk (%)	Priority score
2eb70248d66e0e3ef83659f71b244378	Delivery and fulfilment-risk sellers	196	99	50.51	47.11	99.06
1ca7077d890b907f89be8c954a02686a	Delivery and fulfilment-risk sellers	112	69	61.61	54.36	98.49
7c67e1448b00f6e969d365cea6b010ab	Bulky-item and completeness-risk sellers	966	245	25.36	25.13	98.12
88460e8ebdecbfecb5f96018339819 30	Delivery and fulfilment-risk sellers	228	66	28.95	27.72	97.05
897060da8b9a21f655304d50fd935913	Delivery and fulfilment-risk sellers	308	78	25.32	24.62	96.86
7040e82f899a04d1b434b795a43b4617	Delivery and fulfilment-risk sellers	205	56	27.32	26.11	96.11
a49928bcdf77c55c6d6e05e09a9b4ca5	Delivery and fulfilment-risk sellers	95	37	38.95	34.57	95.85
710e3548e02bc1d2831dfc4f1b5b14d4	Delivery and fulfilment-risk sellers	133	43	32.33	29.90	95.79
54965bbe3e4f07ae045b90b0b8541f52	Delivery and fulfilment-risk sellers	73	32	43.84	37.37	95.38
855668e0971d4dfd7bef1b6a4133b41b	Delivery and fulfilment-risk sellers	303	67	22.11	21.59	95.32
1835b56ce799e6a4dc4eddc053f04066	Delivery and fulfilment-risk sellers	378	81	21.43	21.04	95.29
8444e55c1f13cd5c179851e5ca5ebd00	Delivery and fulfilment-risk sellers	97	34	35.05	31.41	95.07
18a349e75d307f4b4cc646a691ed4216	Delivery and fulfilment-risk sellers	118	36	30.51	28.08	94.79
972d0f9cf61b499a4812cf0bfa3ad3c4	Delivery and fulfilment-risk sellers	73	30	41.10	35.22	94.79
3c7c4a49ec3c6550809089c6a2ca9370	Delivery and fulfilment-risk sellers	137	38	27.74	25.96	94.76

Top n sellers	Bad reviews	Share of segmented bad reviews (%)
20	1 218	10.64
50	2 981	26.05

Top n sellers	Bad reviews	Share of segmented bad reviews (%)
100	4 682	40.91

Seller segment ID	Seller segment label	Sellers	Reviewed orders	Bad reviews	Avg bad review rate	Avg bad review rate shrunk	Avg non delivered share	+13 more columns
2	Delivery and fulfilment-risk sellers	178	16 432	3 350	0.21	0.19	0.04	...
1	Bulky-item and completeness-risk sellers	91	8 499	1 274	0.15	0.14	0.02	...
0	Baseline / lower-risk mixed sellers	527	59 487	6 820	0.10	0.11	0.01	...

Conclusions

Bad reviews in this dataset are driven first by **fulfilment failure and severe lateness**, then by **order complexity**, **large-item logistics**, **specific product categories**, and **regional network effects**. Customer comments confirm that many bad experiences are multi-step failures: delivery issues, wrong or incomplete orders, weak service recovery, and refund friction often appear together. Operationally, the strongest first move is to focus on the relatively small group of **Delivery and fulfilment-risk sellers**, then build a separate intervention track for **Bulky-item and completeness-risk sellers**, while also tightening multi-item controls, category-specific action plans, and customer recovery processes.