



Distilo

RELATÓRIO DE ANÁLISE

Análise de Conjuntos

A worked example of cross-sell (fictional data).

Gerado em

Cliente

runId

15/08/2026

Sample report · synthetic data

demo-set_mining

Vaccine cross-analysis: what your patients take together

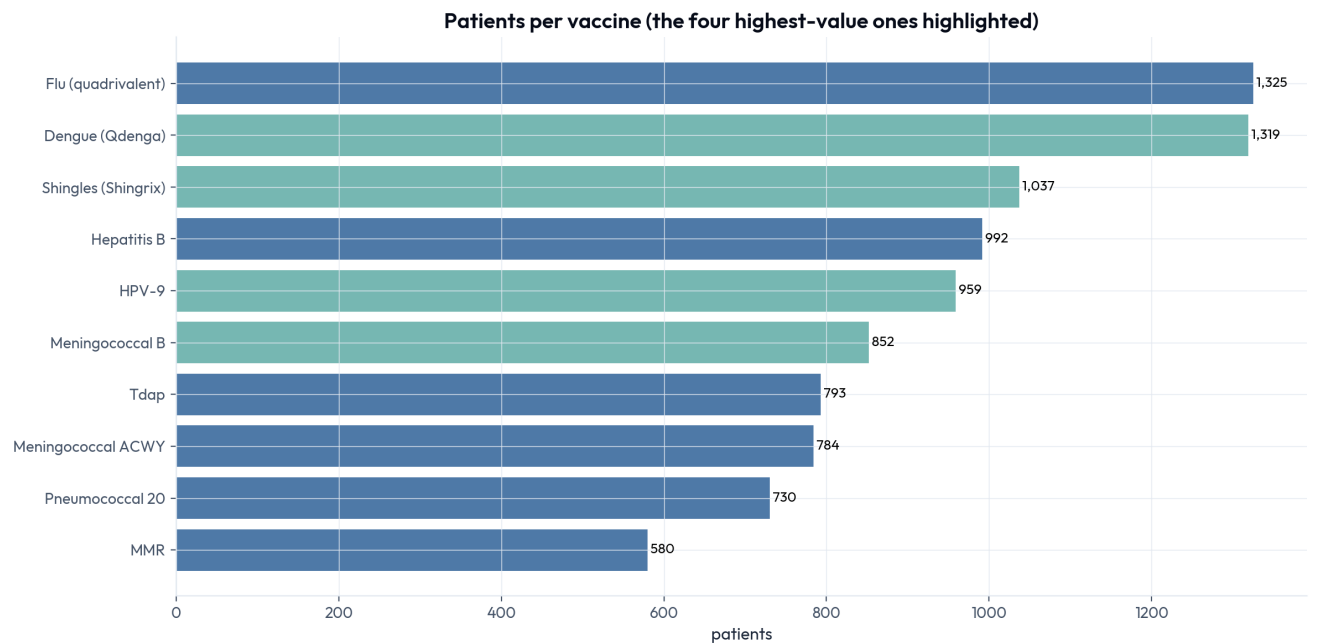
An analysis of 19,247 doses bought by 3,221 patients between 26/12/2023 and 10/04/2026, adding up to R\$ 10,748,740 across the 10 vaccines in the catalogue. The request was to find out which vaccines the same patients tend to take together, so that a WhatsApp cross-sell campaign could be built around the four with the highest margin: Qdenga, HPV-9, quadrivalent flu and shingles.

The answer to what was asked is short: there is no combination. No pair of vaccines travels together in your base, and that kills the idea of “whoever took A wants B”. But the cross-analysis surfaced three things that were not in the question and that change the order of priorities, the first of them urgent: on 09 March sales of your four most expensive vaccines all but stopped, and have not come back.

1. The base is strong and nobody has completed the catalogue

Each patient has taken 2.9 different vaccines on average (median of 3). The record is 6 vaccines, reached by 59 patients. Nobody has all 10, and nobody would be expected to, since part of the catalogue is for a different age group.

The most widely taken vaccine is quadrivalent flu, present in 1,325 of the 3,221 patients, closely followed by Qdenga with 1,319. In other words: even the reach champion is in 41% of the base. The two with the highest protocol value, HPV-9 (R\$ 2,234 for the 2-dose protocol) and shingles (R\$ 2,160), are in 30% (959 patients) and 32% (1,037).



Patients with more vaccines on file have spent more, which sounds promising but is close to an arithmetic identity: the average price per dose is practically the same across all groups, between R\$ 551 and R\$ 569, and each distinct vaccine on file brings 2.1 doses on average. The table below describes the base; it does

not prove that offering more vaccines generates that value. On top of that, patients with more vaccines tend to have been customers for longer, and length of relationship is not isolated here.

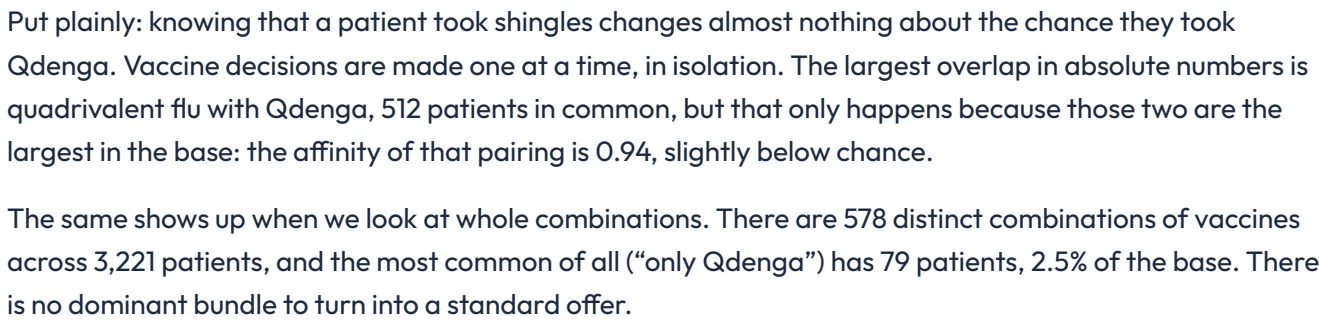
Distinct vaccines	Patients	Average revenue per patient	Revenue per dose
1	430	R\$ 1,334	R\$ 569
2	882	R\$ 2,280	R\$ 558
3	894	R\$ 3,541	R\$ 559
4	639	R\$ 4,601	R\$ 551
5	317	R\$ 5,423	R\$ 569
6	59	R\$ 5,761	R\$ 557

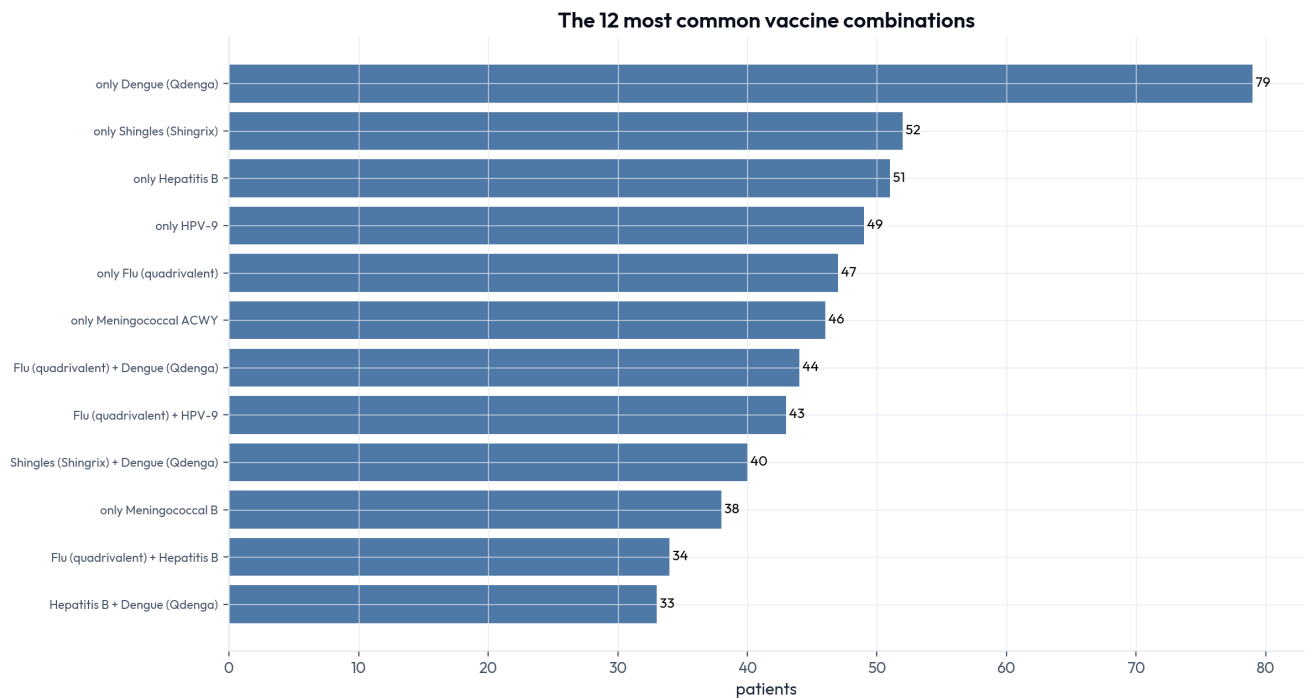
The 430 single-vaccine patients (file `exclusivos_uma_vacina.csv`) bring R\$ 1,334 each, against R\$ 3,337 for the base as a whole. It is the largest unoccupied space in the base, with the caveat above about what the table actually demonstrates.

2. There is no pair of vaccines that travels together

This is where the result contradicts the expectation. All 45 possible pairs were tested, measuring whether someone who took one vaccine is more likely to have taken the other than any patient in the base. The indicator is affinity: 1.00 means pure chance, above 1.50 would mean a genuine pairing.

No pair got past 1.06. The floor was 0.85. Zero pairs above 1.50.





Worth recording a trap that nearly became a recommendation. Twelve of the 45 pairs appear together slightly less often than chance would predict, and it is tempting to read that as “one expensive vaccine postpones the other”. It is not. The effect is mechanical: since each patient has a more or less fixed number of vaccines, choosing one arithmetically reduces the chance of any other appearing, and the effect turns up even in pairs that do not compete for budget (shingles at R\$ 2,160 with hepatitis B at R\$ 540 gives exactly the same pattern as expensive with expensive). There is no signal of trade-off between vaccines here to act on.

The file [overlaps_pares.csv](#) has the 45 pairs with their overlap, affinity in both directions and the full statistical test, in case you want to check any pairing.

3. Sales of the four expensive vaccines stopped on 09 March

This is the most urgent finding in the report, and it was not in the question.

Up to 08 March the clinic was selling 67 doses and R\$ 40,304 a day, averaged over the previous 30 days. From 09 March it started selling 17 doses and R\$ 6,478 a day, and stayed at that level for the following 33 days until the end of the data. Over the observed period that is R\$ 1.12 million less than the earlier pace would have produced.

This is not a general drop in footfall: it is selective, and it hits exactly the vaccines that carry your margin.

Vaccine	Doses/day before	Doses/day after	Change
Meningococcal B	8.1	0.4	-95%
Shingles	9.2	0.7	-93%
HPV-9	8.5	0.7	-92%
Meningococcal ACWY	6.8	0.5	-92%
Qdenga	13.1	1.1	-92%
Hepatitis B	3.5	1.1	-69%
Tdap	4.5	2.6	-43%
MMR	2.9	1.8	-37%
Quadrivalent flu	7.2	5.6	-22%
Pneumococcal 20	3.4	2.9	-13%

Three checks rule out the easy explanations. It is not the file export being cut short, because administrations carried on normally after that date, with April recording 1,751 doses given, practically level with the strongest month in the whole base (January, 1,784). It is not price, because the amounts are identical to before. It is not one channel drying up, because the fall is similar across Google, referrals, Instagram and walk-ins, and also across every payment method.

What changed was the start of treatment: after 09 March only 7 first doses of meningococcal B came in, 9 of meningococcal ACWY, 11 of HPV-9, 11 of shingles and 18 of Qdenga, against 184 of flu, 97 of pneumococcal 20 and 85 of Tdap. In other words, the clinic carried on giving second doses that had already been sold and carried on selling the cheap single-dose vaccines, but stopped starting new patients on the expensive ones. Add to that the fact that new patients coming in had already been falling before this, from 154 in November to 92 in January and 46 in February.

That pattern is what you would expect from a stock shortage of the highest-margin vaccines, but the data does not prove the cause: it could equally be a commercial decision, a member of staff leaving, or a change of supplier (low confidence as to the cause; the pattern itself is high confidence, it comes straight from the data). Finding out which one is the first task of the week, and it is a precondition for everything that follows.

4. R\$ 704,539 already paid for and sitting on the shelf

There are 1,246 doses bought, paid for and never given, spread across 635 patients, adding up to R\$ 704,539 (counted as doses still awaiting administration which, as of the final date in the data, had already passed the normal window for that vaccine and dose, measured by the 95th percentile of the gap between purchase and administration observed in doses that were given). The median overrun beyond that normal window is 126 days.

That is the figure measured with data in hand. If none of those doses has been given in the 127 days the data cannot see, between 10/04 and today, the total rises to 1,452 doses and R\$ 822,944, which is the ceiling and not the estimate.

Vaccine	Value sitting still	Doses	Patients
Shingles	R\$ 177,120	164	121
HPV-9	R\$ 175,369	157	108
Qdenga	R\$ 113,950	215	145
Meningococcal B	R\$ 74,880	117	87
Meningococcal ACWY	R\$ 49,920	104	75
Hepatitis B	R\$ 39,960	222	131
Others (pneumococcal 20, flu, Tdap, MMR)	R\$ 73,340	267	199

The patient column does not add up to 635 because the same person often has a stalled dose of more than one vaccine.

The money is already in the bank, so the loss is not of immediate revenue: it is of incomplete vaccination courses, of the risk of a refund request, and of a patient disappearing without having received what they bought. It is also the best reason for contact the clinic has, because nobody ignores a message saying there is a paid-for dose waiting.

And you can know who is likely to miss before you call. Across all doses bought up to the end of the data, non-attendance is 16.0% among invoice payers against 4.8% among instant-transfer payers, and 13.0% among those who came from Google against 4.3% among those who came by referral; by dose, it rises from 5.9% on the first to 12.7% on the third. Since payment method and channel are fixed per patient in this base, that identifies a profile, it does not explain the reason. Even so it is useful for choosing the contact channel: the file `doses_pagas_nao_aplicadas.csv` already marks each patient as high, medium or low risk, and the 308 at high risk hold R\$ 333,924 of the R\$ 704,539 sitting still.

5. The front desk sells one vaccine at a time

The cross-analysis at patient level says there is no combination. The cross-analysis at visit level says something different and more useful: they never even get offered together.

Of 11,043 purchase visits (same patient, same day), only 45 include more than one vaccine, 0.41%. Across the 17,515 administration visits, only 196 have more than one, 1.12%. The patient comes back to the clinic about 1,474 times a month for a dose, is physically in front of you, and leaves with exactly one vaccine.

That does not contradict section 2, it completes it: the absence of affinity in the records over 28 months measures what the patient accumulated; the front desk measures what was offered. If nobody offers the

second vaccine at the moment of administration, there will never be a basket for the cross-analysis to find. It is the cheapest opportunity in this report because the cost of contact is zero: the patient is already there.

6. Before sending any message, two checks

These two limitations condition everything in section 7. They come first on purpose.

Confirm what happened on 09 March before selling HPV-9, shingles, Qdenga or meningococcal B. The entire cross-sell campaign points at vaccines whose treatment starts have all but ceased for five months. If the reason is stock or supplier, sending offers now creates interested patients the clinic cannot serve, which is worse than not sending. The check takes a morning: verify physical stock, the situation with the supplier, and the administration schedule for those four vaccines. If they are available, follow the plan as written. If they are not, start with fronts 1, 2 and 4, which do not depend on them.

The data has no age, and that is the most serious limitation of this work. The brief asked for a focus on adults, but the file carries no date of birth and no age, so no list here is filtered by age group. HPV-9 and shingles have well-defined age indications, and offering HPV-9 to a 70-year-old or shingles to a 25-year-old burns both the message and the trust. Before each send, reception needs to check the age on the record. When the patient is outside the range, the `alternativas_se_fora_da_faixa` column in `campanha_onda1.csv` already gives the next high-margin vaccine they are missing. Of the first 600 in the queue, 379 have an alternative and 221 have none; of those 221, 93 received precisely an age-sensitive offer, and for them the rule is simple: if they are outside the range, do not push, offer the flu vaccine from front 4 and move on.

7. The plan, in four fronts

With no affinity between vaccines, the criterion for choosing the offer stops being history and becomes value. For cross-sell the rule is direct: each patient gets one offer only, the most expensive of the four highest-margin vaccines they have not yet taken, and the queue is ordered by how much they have already spent at the clinic, which is the best available signal of willingness to pay. That covers 3,187 of the 3,221 patients, since only 34 have all four. The file `campanha_onda1.csv` has that queue ready, with the vaccine to offer, the protocol value and the age-range alternative.

A patient can appear in more than one front. The three contact lists add up to 3,264 names for 2,388 distinct patients (486 patients are in both recovery and flu, 390 in cross-sell and flu, 14 in recovery and cross-sell). The rule for living together is one message per patient per week, in this order: recovery first, cross-sell second, flu last.

Front 1, starts tomorrow morning: offer the second vaccine at the desk

No list and no system needed. At each administration, reception looks at the record to see which vaccine the patient is missing and asks the question there and then, with the vaccination card in hand. That is about 1,474 administrations a month. If 5% of them turn into a second dose, that is 74 additional doses a month; at 10%, 147. Taking a conservative average of R\$ 300 a dose across the vaccines that remain available, that is somewhere between R\$ 22,200 and R\$ 44,100 a month, recurring, against zero cost of contact (low

confidence on the rate: the clinic has never offered at the desk, so there is no track record of its own; the basis of the calculation, 1,474 monthly administrations, is high confidence). Management measures every Friday how many of the week's visits left with more than one vaccine, starting from a baseline of 1.12%, and the target for the first week is to reach 3%. If after two weeks the number has not moved, the problem is the script and not the patient: management takes two shifts at the desk to listen to the conversation and rewrite the question. Not doing this means carrying on paying for WhatsApp to reach people who were already inside the clinic.

Front 2, week 1: recovering the stalled doses

Reception takes the 635 patients in `doses_pagas_nao_aplicadas.csv`, which already comes ordered from the largest stalled amount to the smallest, and sends individual WhatsApp messages saying what has been paid for and is outstanding, with two concrete time slots to choose from. Those marked high risk in the `risco_de_falta` column get a phone call instead of a message, because that is the profile that historically does not turn up. A quota of 60 contacts a day closes the list in 11 working days. Anyone who does not reply within 3 days gets a call; anyone who does not answer two calls gets a message with a booking link and leaves the active queue. The target is to rebook between 190 and 254 patients, a recovery of 30% to 40% of the list (low confidence: that is a general benchmark for an already-paid dose, the clinic has no recovery track record of its own). On Friday the number that matters is how many stalled doses have left the list, with a target of 90 in the first week. The total cost comes to around R\$ 2,900 across messages and calls, against R\$ 704,539 of paid-for doses outstanding. If this front does not run, the risk is not lost revenue: it is a patient asking for their money back a year later.

Front 3, week 2, conditional on the stock check: cross-sell to the first 600

The owner or the sales manager takes the first 600 in `campanha_onda1.csv` (a cut-off of R\$ 5,050 in historical revenue) and sends a consultative, not promotional, WhatsApp offering the vaccine in the `vacina_ofertada` column. That batch breaks down into 215 offers of HPV-9, 180 of shingles, 141 of Qdenga and 64 of quadrivalent flu, with an average protocol value of R\$ 1,718 and a gross potential of R\$ 1,030,730 if everyone bought, which will not happen and should not be used as a target. A quota of 60 contacts a day, 10 working days. At a conversion of 3% to 6% (low confidence, no previous campaign in the data to calibrate against), the expected result is between 18 and 36 protocols, or R\$ 30,900 to R\$ 61,800, against an approximate cost of R\$ 2,550, a return of 12 to 24 times what was spent. Anyone who does not reply within 5 days gets a second message with social proof, and with no reply after another 5 days, a phone call; with no answer to the call, they go to the back of the queue and are only contacted again the following quarter. The Friday number is how many bookings came out of the batch, with a target of 6 in the first week. Not running this front leaves R\$ 30,900 to R\$ 61,800 sitting in the queue, and the cost of delaying grows because the base is ageing: the most recent patient is already 127 days without a purchase, and the colder the contact, the lower the conversion.

It is worth splitting the 600 into two halves of 300 to test the argument, since nobody at the clinic knows which one works: half get the message centred on the risk the vaccine avoids, half the message centred on the outcome (protecting the family, a trip without worry, time without being off sick). With 18 to 36 conversions across the whole batch the test will not settle statistically now, but it gives the direction for the following waves, which will be larger: 2,587 patients remain in the queue after these 600.

Front 4, week 3: flu up to date

Flu is annual on the clinical calendar, and it is the vaccine with the broadest indication in the catalogue, so it needs a front of its own: under the most-expensive-offer rule, of the 1,896 patients who have never had a flu vaccine only 64 would be offered it in front 3. The file `gripe_em_atraso.csv` brings the two groups together, 2,029 patients: 133 who have had it and have gone more than 12 months without repeating, and 1,896 who have never had it. Reception sends WhatsApp reminders starting with the 133 who are overdue, a quota of 80 a day, closing both groups in 26 working days, with a call to anyone who does not reply within 7 days. Among those overdue, the expectation is a 25.2% return (low confidence: that is the rate observed in the only cohort in the base with a full 365 days of follow-up, and it is just 147 patients); among those who have never had it, the same 3% to 6% range as cross-sell applies. Together, that is between R\$ 17,100 and R\$ 27,930 at R\$ 190 a dose, against a cost of R\$ 1,015. Leave 10% of the list out, receiving nothing, as a control group: the difference in conversion between those who received something and those who did not is the only way to know whether the campaign worked or whether these people would have come back anyway. The Friday number is how many flu doses were booked in the week, with a target of 20 in the first. Not running this front leaves the only annually repeating vaccine with no contact at all, and it is the cheapest way back in for reactivating people who stopped buying.

8. What else could change this reading

The data ends on 10/04/2026, 127 days ago. Every “days without buying” count carries that floor, and the most recent patient shows up with 127 days even though they bought on the last day in the file. It is also why the number of stalled doses was measured at the final date of the data and not today: measuring against today would inflate the total by R\$ 118,405 that may already have been given without the file showing it.

Flu does not behave like an annual repurchase in this base, despite being annual in clinical practice. Of the 393 intervals between repeat flu purchases, none reaches 300 days, and the median interval is 83 days, in line with every other vaccine (between 76 and 117 days). In other words, the file does not record the annual cycle, probably because the observed window is too short. Front 4 was built on the clinical rule of 12 months, and not on a renewal pattern the data had shown.

Each row is one dose bought, and the whole protocol is bought at once. The second and third doses carry the same purchase date as the first and are given later. That is why there is nobody who has bought half a course: what exists is a dose bought and not given, which is the subject of front 2. “Took vaccine X” in this report means “bought the protocol for X”.

The conversion rates are market benchmarks, not your business’s. The data records no previous campaign, so the ranges of 3% to 6% and 30% to 40% are educated guesses. The projections also assume similar conversion across vaccines that vary twelvefold in price, from R\$ 190 to R\$ 2,234; if in practice patients convert less on the expensive ones, the projected revenue falls to near half and the return lands between 6 and 12 times the cost, still high enough for the front to be worth running. After the first wave, recalculate the queue with the real conversion.

Sorting by historical revenue under-serves more than one neighbourhood. The first 600 have average revenue of R\$ 6,350, against R\$ 3,337 for the base, and three dimensions come out uneven. By payment method, 21.4% of instant-transfer patients get in against 13.6% of invoice payers. By channel, 20.6% of

those who came by referral against 15.0% of those who came from Google. By neighbourhood, 23.7% of residents of Parque Industrial against 14.2% of those in Jardim das Acácias. Invoice and insurance are the payment methods most associated with a lower-income profile, so this queue tends to speak first to those who are already more profitable. A minimum quota of 50 per neighbourhood would be symbolic: it would move 3 patients in 600. A quota of 60 per neighbourhood moves 38 patients and genuinely balances the territory. If the goal is to balance income and not just geography, the quota has to be by payment method, guaranteeing a minimum number of contacts among invoice and insurance patients.

Absence of affinity does not mean cross-sell does not work. It means the vaccine history does not help choose which one to offer. The choice becomes protocol value and clinical eligibility, which is what the lists do. If the records start capturing age and the reason for refusal at each contact, the next round can segment for real.

Appendix

Files produced

File	What it holds	Order and bias
<code>campanha_onda1.csv</code>	3,187 patients with the highest-value offer each one is missing, plus age-range alternatives	ordered by historical revenue descending, under-serves lower spenders, invoice and insurance payers
<code>doses_pagas_nao_aplicadas.csv</code>	635 patients with a paid and overdue dose, with value, days beyond the window and a no-show risk marker	ordered by stalled value descending, under-serves those with a cheap dose stalled
<code>doses_paradas_por_vacina.csv</code>	the 10 vaccines with doses, stalled value and patients, to check the table in section 4	ordered by stalled value descending
<code>gripe_em_atraso.csv</code>	2,029 patients without flu up to date: 133 more than 12 months overdue and 1,896 who never had it	grouped by situation, then historical revenue descending
<code>alvos_qdenga.csv</code>	1,902 patients without Qdenga (protocol R\$ 1,060)	ordered by historical revenue descending
<code>alvos_hpv9.csv</code>	2,262 patients without HPV-9 (protocol R\$ 2,234)	ordered by historical revenue descending
<code>alvos_zoster.csv</code>	2,184 patients without shingles (protocol R\$ 2,160)	ordered by historical revenue descending
<code>alvos_gripe.csv</code>	1,896 patients without quadrivalent flu (dose R\$ 190)	ordered by historical revenue descending
<code>exclusivos_uma_vacina.csv</code>	430 patients who took only one vaccine, with which one it was	ordered by historical revenue descending
<code>pacientes_multivacina.csv</code>	59 patients with 6 vaccines, the ceiling of the base, candidates for referrals and research	ordered by historical revenue descending
<code>overlaps_pares.csv</code>	the 45 vaccine pairs with overlap, Jaccard, affinity in both directions and the statistical test	ordered by overlap descending
<code>summary_by_id.csv</code>	1 row per patient, one yes/no column per vaccine, plus revenue, recency and stalled doses	ordered by total revenue descending

The four `alvos_*.csv` lists overlap heavily with each other: added together they give 8,244 rows for only 3,187 distinct patients. Use `campanha_onda1.csv` to send, since it already resolves the overlap with one offer per patient, and the `alvos_*` lists only for looking things up by vaccine.

The 12 pairs with the largest overlap

Affinity of 1.00 is chance. The Jaccard column measures how much the two groups overlap relative to the two of them combined.

Vaccine A	Vaccine B	Took both	Jaccard	Affinity
Quadrivalent flu	Qdenga	512	0.24	0.94
Quadrivalent flu	HPV-9	418	0.22	1.06
Quadrivalent flu	Shingles	408	0.21	0.96
Shingles	Qdenga	389	0.20	0.92
Quadrivalent flu	Hepatitis B	384	0.20	0.94
Hepatitis B	Qdenga	375	0.19	0.92
HPV-9	Qdenga	349	0.18	0.89
Quadrivalent flu	Meningococcal B	337	0.18	0.96
Meningococcal B	Qdenga	325	0.18	0.93
Quadrivalent flu	Tdap	321	0.18	0.98
Qdenga	Tdap	311	0.17	0.96
Quadrivalent flu	Meningococcal ACWY	302	0.17	0.94

The pair with the highest affinity in the whole base, quadrivalent flu with HPV-9 (1.06), does not pass the statistical test: with 45 pairs tested at once, some look different from chance purely by luck, and once the criterion is adjusted for that number of tests, this pair does not survive.

How each number was obtained

- **A patient “has” a vaccine** when there is at least one purchase row for it in their name, given or not. The 10 sets come from the `vacina` column, and the matching key is `paciente_id`. No records were merged: the 3,221 identifiers map 1 to 1 with the 3,221 names.
- **Affinity** between A and B is the chance of having B among those who have A, divided by the chance of having B across the whole base. Significance by Fisher’s exact test, with the criterion adjusted for the 45 pairs tested.
- **A stalled dose** is a row still awaiting administration whose time between purchase and the final date of the data exceeded the 95th percentile of the purchase-to-administration gap observed for that same

vaccine and dose. There are 1,246; measuring against today rather than the end of the data would give 1,452, a difference the file cannot confirm.

- **The 09 March break** compares the daily average of the previous 30 days with the daily average of the following 33 days, by vaccine and by dose, always by purchase date.
 - **A visit** is the combination of patient and date: of purchase for the 11,043 sales visits, of administration for the 17,515 service visits.
 - **Flu overdue** means patients whose last flu purchase predates 12 months before the end of the data, plus those who never bought it. The 25.2% rate comes from the cohort whose first flu purchase predates 10/04/2025, the only one with a full year of follow-up inside the data.
 - **Protocol value** is the dose price multiplied by the number of doses in the course. Prices are fixed across the whole base, with no variation by patient or by period.
 - **Revenue** sums the `valor` column across all rows, including doses not yet given, since the purchase was recorded.
-

* **About this sample.** This is a real report from the Distilo pipeline, run over synthetic data for a fictional company. That company is Brazilian: a vaccination clinic, so the amounts are in Brazilian reais, the neighbourhood and patient names are Brazilian, and part of the catalogue (Qdenga against dengue, for instance) reflects a Brazilian vaccination schedule. It is published as it was produced rather than adapted, because the analysis is what we are showing you and the analysis does not change with the country. Your own report is written in your language, about your business, in your currency.