



Distilo

RELATÓRIO DE ANÁLISE

Análise de Conjuntos

Een uitgewerkt voorbeeld van cross-sell (fictieve data).

Gerado em

15/08/2026

Cliente

Voorbeeldrapport · synthetische data

runId

demo-set_mining

Vaccins gekruist: wat je patiënten samen halen

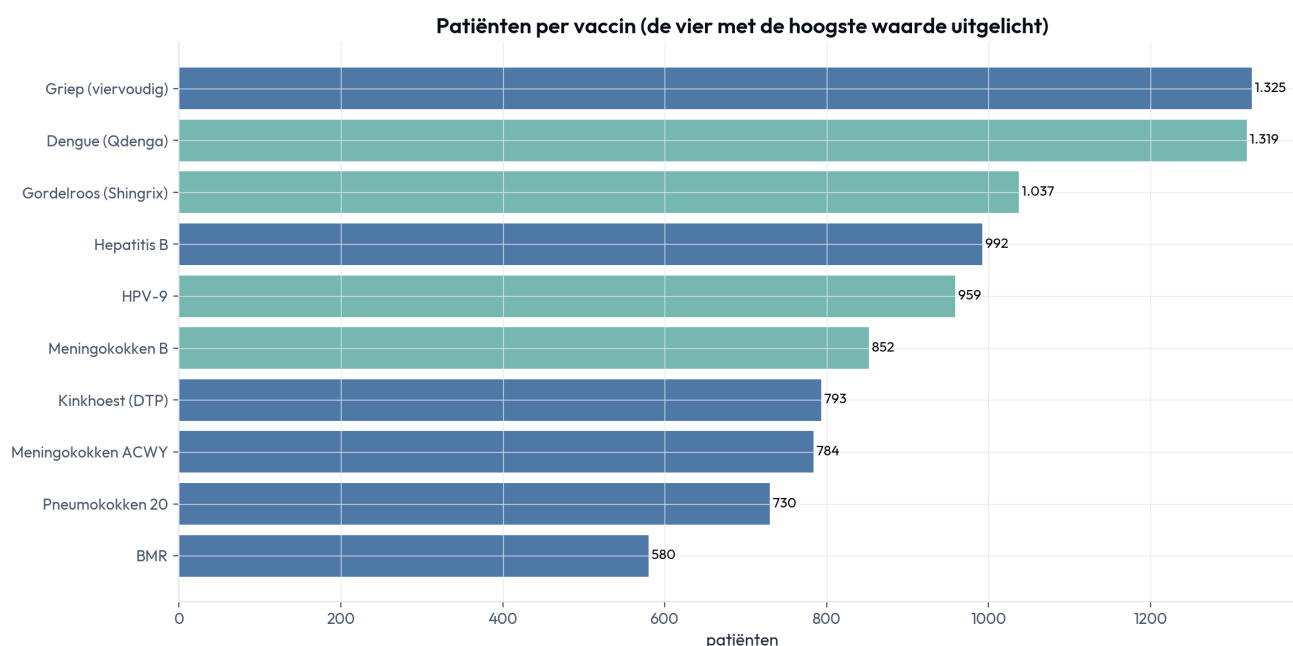
Een analyse van 19.247 doses gekocht door 3.221 patiënten tussen 26-12-2023 en 10-04-2026, samen R\$ 10.748.740 over de 10 vaccins in het assortiment. De vraag was om uit te zoeken welke vaccins dezelfde patiënten doorgaans samen halen, zodat er een cross-sellcampagne via WhatsApp gebouwd kon worden rond de vier met de hoogste marge: Qdenga, HPV-9, griep en gordelroos.

Het antwoord op de gestelde vraag is kort: er is geen combinatie. Geen enkel paar vaccins loopt samen op in je bestand, en dat haalt het idee onderuit dat “wie A haalde, B wil”. Maar de kruising bracht drie dingen boven water die niet in de vraag zaten en die de volgorde van prioriteiten veranderen, waarvan het eerste urgent is: op 9 maart stopte de verkoop van je vier duurste vaccins vrijwel volledig, en die is niet teruggekomen.

1. Het bestand is sterk en niemand heeft het assortiment compleet

Elke patiënt heeft gemiddeld 2,9 verschillende vaccins gehaald (mediaan 3). Het record is 6 vaccins, bereikt door 59 patiënten. Niemand heeft alle 10, en dat zou ook niet verwacht worden, omdat een deel van het assortiment voor een andere leeftijdsgroep is.

Het meest gehaalde vaccin is griep, aanwezig bij 1.325 van de 3.221 patiënten, op de voet gevolgd door Qdenga met 1.319. Met andere woorden: zelfs de koploper zit bij 41% van het bestand. De twee met de hoogste protocolwaarde, HPV-9 (R\$ 2.234 voor het protocol van 2 doses) en gordelroos (R\$ 2.160), zitten bij 30% (959 patiënten) en 32% (1.037).



Patiënten met meer vaccins in het dossier hebben meer uitgegeven, wat veelbelovend klinkt maar bijna een rekenkundige vanzelfsprekendheid is: de gemiddelde prijs per dosis is in alle groepen praktisch gelijk,

tussen R\$ 551 en R\$ 569, en elk apart vaccin in het dossier brengt gemiddeld 2,1 doses mee. De tabel hieronder beschrijft het bestand; hij bewijst niet dat méér vaccins aanbieden die waarde oplevert. Daar komt bij dat wie meer vaccins heeft doorgaans al langer klant is, en de klantduur is hier niet apart gezet.

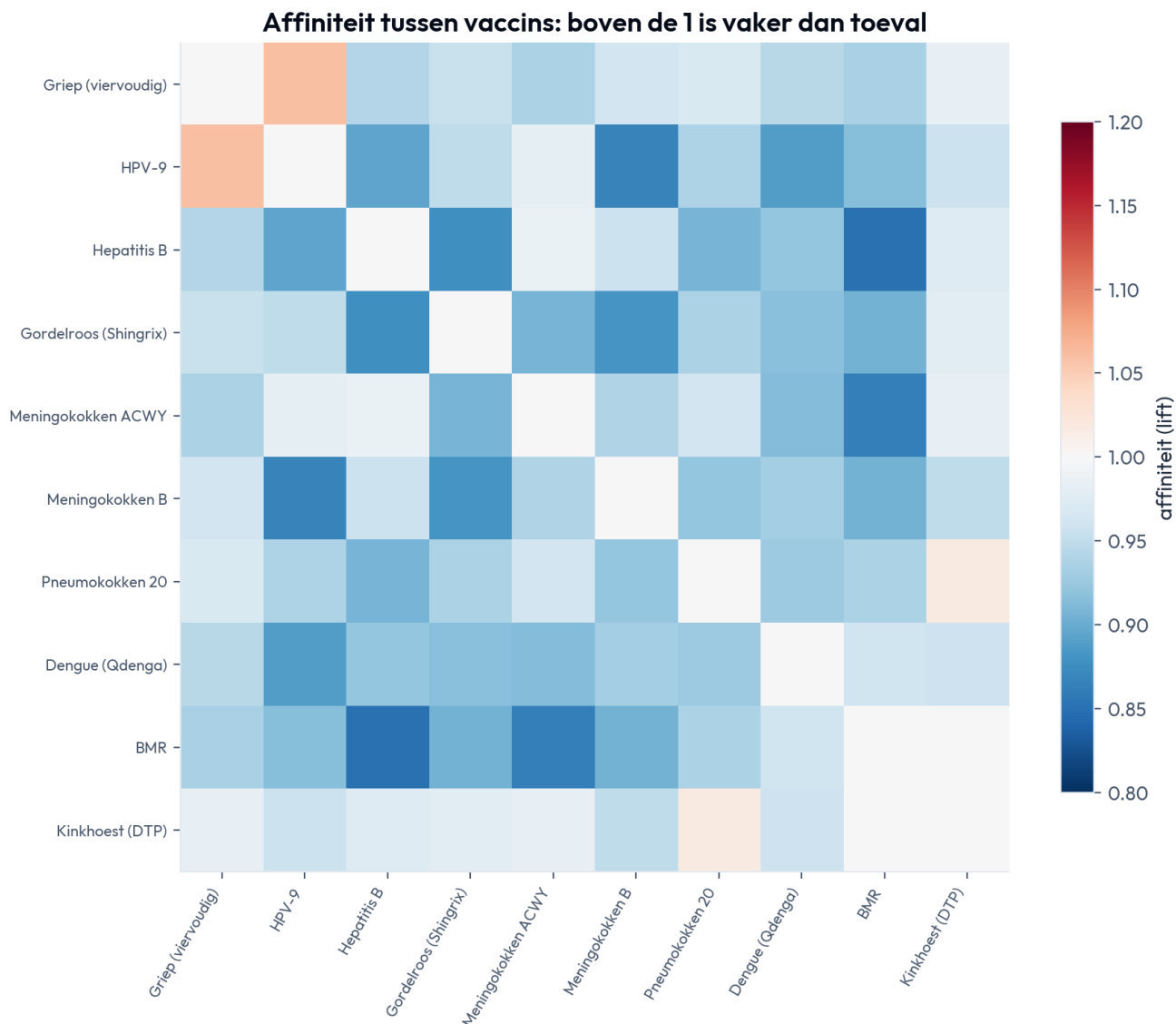
Verschillende vaccins	Patiënten	Gemiddelde omzet per patiënt	Omzet per dosis
1	430	R\$ 1.334	R\$ 569
2	882	R\$ 2.280	R\$ 558
3	894	R\$ 3.541	R\$ 559
4	639	R\$ 4.601	R\$ 551
5	317	R\$ 5.423	R\$ 569
6	59	R\$ 5.761	R\$ 557

De 430 patiënten met één vaccin (bestand `exclusivos_uma_vacina.csv`) brengen elk R\$ 1.334 op, tegen R\$ 3.337 voor het bestand als geheel. Het is de grootste onbenutte ruimte in het bestand, met de kanttekening hierboven over wat de tabel werkelijk aantoont.

2. Er is geen vaccinpaar dat samen optrekt

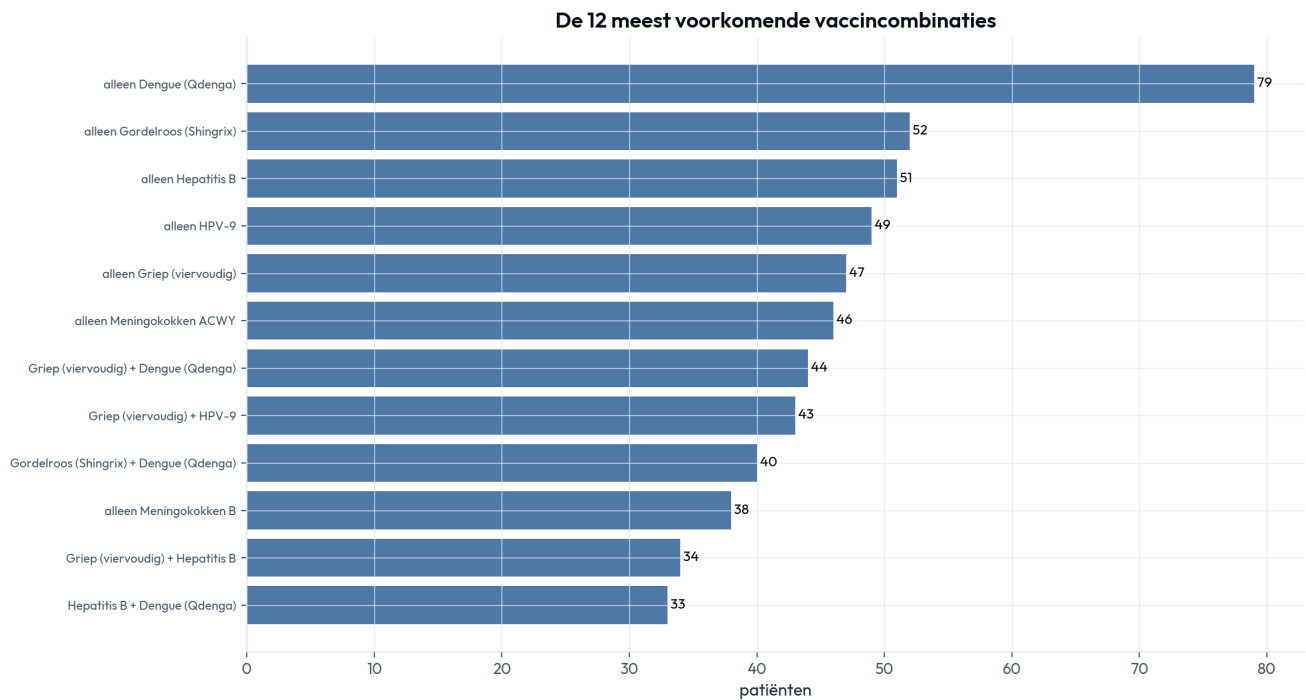
Hier spreekt het resultaat de verwachting tegen. Alle 45 mogelijke paren zijn getoetst, waarbij is gemeten of wie het ene vaccin haalde een grotere kans heeft het andere gehaald te hebben dan een willekeurige patiënt uit het bestand. De maat is affiniteit: 1,00 is puur toeval, boven de 1,50 zou een echt duo betekenen.

Geen enkel paar kwam boven de 1,06. De bodem lag op 0,85. Nul paren boven de 1,50.



In gewone taal: weten dat een patiënt gordelroos heeft gehad, verandert vrijwel niets aan de kans dat hij Qdenga heeft gehad. Vaccinbeslissingen worden een voor een genomen, los van elkaar. De grootste overlap in absolute aantallen is griep met Qdenga, 512 gemeenschappelijke patiënten, maar dat komt alleen doordat die twee de grootste van het bestand zijn: de affiniteit van dat duo is 0,94, iets onder toeval.

Hetzelfde blijkt als we naar hele combinaties kijken. Er zijn 578 verschillende vaccincombinaties onder 3.221 patiënten, en de meest voorkomende van allemaal (“alleen Qdenga”) heeft 79 patiënten, 2,5% van het bestand. Er is geen dominant pakket om tot standaardaanbod te maken.



Het is goed een valkuil vast te leggen die bijna een aanbeveling werd. Twaalf van de 45 paren komen iets minder vaak samen voor dan toeval zou voorspellen, en het is verleidelijk dat te lezen als “een duur vaccin stelt het andere uit”. Dat is het niet. Het effect is mechanisch: omdat elke patiënt een min of meer vast aantal vaccins heeft, verkleint het kiezen van het ene rekenkundig de kans dat een ander verschijnt, en het effect duikt zelfs op bij paren die niet om hetzelfde budget concurreren (gordelroos van R\$ 2.160 met hepatitis B van R\$ 540 geeft precies hetzelfde patroon als duur met duur). Er is hier geen enkel signaal van uitruil tussen vaccins om op te handelen.

Het bestand `overlaps_pares.csv` bevat de 45 paren met hun overlap, affiniteit in beide richtingen en de volledige statistische toets, mocht je een duo willen nakijken.

3. De verkoop van de vier dure vaccins stopte op 9 maart

Dit is de meest urgente bevinding van het rapport, en hij stond niet in de vraag.

Tot 8 maart verkocht de kliniek 67 doses en R\$ 40.304 per dag, gemiddeld over de voorgaande 30 dagen. Vanaf 9 maart werd dat 17 doses en R\$ 6.478 per dag, en dat bleef zo de volgende 33 dagen tot het einde van het bestand. Over de waargenomen periode is dat R\$ 1,12 miljoen minder dan het eerdere tempo zou hebben opgeleverd.

Het is geen algemene daling van de drukte: het is selectief, en het raakt precies de vaccins die je marge dragen.

Vaccin	Doses/dag ervoor	Doses/dag erna	Verandering
Meningokokken B	8,1	0,4	-95%
Gordelroos	9,2	0,7	-93%
HPV-9	8,5	0,7	-92%
Meningokokken ACWY	6,8	0,5	-92%
Qdenga	13,1	1,1	-92%
Hepatitis B	3,5	1,1	-69%
Kinkhoest (DTP)	4,5	2,6	-43%
BMR	2,9	1,8	-37%
Griep	7,2	5,6	-22%
Pneumokokken 20	3,4	2,9	-13%

Drie controles sluiten de makkelijke verklaringen uit. Het is geen afgekapte export, want de toedieningen liepen na die datum gewoon door, met 1.751 gezette doses in april, vrijwel gelijk aan de sterkste maand van het hele bestand (januari, 1.784). Het is geen prijs, want de bedragen zijn identiek aan daarvoor. Het is geen kanaal dat opdroogde, want de daling is vergelijkbaar bij Google, doorverwijzing, Instagram en inloop, en ook bij alle betaalwijzen.

Wat veranderde was het starten van een behandeling: na 9 maart kwamen er nog maar 7 eerste doses meningokokken B binnen, 9 van meningokokken ACWY, 11 van HPV-9, 11 van gordelroos en 18 van Qdenga, tegen 184 van griep, 97 van pneumokokken 20 en 85 van kinkhoest. De kliniek bleef dus tweede doses zetten die al verkocht waren en bleef de goedkope eenmalige vaccins verkopen, maar stopte met het starten van nieuwe patiënten op de dure. Daar komt bij dat de instroom van nieuwe patiënten daarvoor al daalde, van 154 in november naar 92 in januari en 46 in februari.

Dat patroon is wat je zou verwachten bij een voorraadtekort van de vaccins met de hoogste marge, maar de data bewijst de oorzaak niet: het kan net zo goed een commerciële beslissing zijn, het vertrek van een medewerker of een wisseling van leverancier (laag vertrouwen over de oorzaak; het patroon zelf is hoog vertrouwen, dat komt rechtstreeks uit de data). Uitzoeken welke van de drie het is, is de eerste taak van de week, en het is een voorwaarde voor alles wat erna komt.

4. R\$ 704.539 al betaald en blijven liggen

Er zijn 1.246 doses gekocht, betaald en nooit gezet, verdeeld over 635 patiënten, samen R\$ 704.539 (geteld als doses die nog op toediening wachten en die op de einddatum van het bestand het normale venster voor dat vaccin en die dosis al waren gepasseerd, gemeten aan het 95e percentiel van de tijd tussen aankoop en toediening bij doses die wél gezet zijn). De mediane overschrijding van dat normale venster is 126 dagen.

Dat is het cijfer gemeten met data in de hand. Als geen van die doses is gezet in de 127 dagen die het bestand niet ziet, tussen 10-04 en vandaag, loopt het totaal op naar 1.452 doses en R\$ 822.944, en dat is het plafond, niet de schatting.

Vaccin	Blijven liggen	Doses	Patiënten
Gordelroos	R\$ 177.120	164	121
HPV-9	R\$ 175.369	157	108
Qdenga	R\$ 113.950	215	145
Meningokokken B	R\$ 74.880	117	87
Meningokokken ACWY	R\$ 49.920	104	75
Hepatitis B	R\$ 39.960	222	131
Overige (pneumokokken 20, griep, kinkhoest, BMR)	R\$ 73.340	267	199

De patiëntenkolom telt niet op tot 635, omdat dezelfde persoon vaak een blijven liggen dosis van meer dan één vaccin heeft.

Het geld is al binnen, dus het verlies is niet dat van directe omzet: het is een onvolledige vaccinatiereeks, het risico op een terugvordering, en een patiënt die verdwijnt zonder gekregen te hebben waarvoor hij betaalde. Het is ook de beste aanleiding voor contact die de kliniek heeft, want niemand negeert een bericht dat er een betaalde dosis klaarstaat.

En je kunt vooraf weten wie waarschijnlijk niet komt opdagen. Over alle gekochte doses tot het einde van het bestand is het niet-verschijnen 16,0% bij wie per factuur betaalt tegen 4,8% bij directe overboeking, en 13,0% bij wie via Google kwam tegen 4,3% bij wie via doorverwijzing kwam; per dosis loopt het op van 5,9% bij de eerste naar 12,7% bij de derde. Omdat betaalwijze en kanaal in dit bestand vastliggen per patiënt, identificeert dat een profiel en verklaart het geen reden. Toch is het bruikbaar om het contactkanaal te kiezen: het bestand [doses_pagas_nao_aplicadas.csv](#) merkt elke patiënt al aan als hoog, gemiddeld of laag risico, en de 308 met hoog risico houden R\$ 333.924 van de R\$ 704.539 vast.

5. De balie verkoopt één vaccin tegelijk

De kruising op patiëntniveau zegt dat er geen combinatie is. De kruising op bezoekeniveau zegt iets anders en nuttigers: ze worden nooit eens samen aangeboden.

Van 11.043 aankoopbezoeken (zelfde patiënt, zelfde dag) bevatten er maar 45 meer dan één vaccin, 0,41%. Van de 17.515 toedieningsbezoeken hebben er maar 196 meer dan één, 1,12%. De patiënt komt ongeveer 1.474 keer per maand naar de kliniek voor een prik, staat fysiek voor je, en vertrekt met precies één vaccin.

Dat spreekt paragraaf 2 niet tegen, het vult hem aan: het ontbreken van affiniteit in het dossier over 28 maanden meet wat de patiënt heeft opgebouwd; de balie meet wat er is aangeboden. Als niemand het

tweede vaccin aanbiedt op het moment van toedienen, ontstaat er nooit een mandje dat de kruising kan vinden. Het is de goedkoopste kans in dit rapport, omdat de contactkosten nul zijn: de patiënt is er al.

6. Voordat je enig bericht verstuurt: twee controles

Deze twee beperkingen bepalen alles in paragraaf 7. Ze staan er met opzet vóór.

Bevestig wat er op 9 maart gebeurde voordat je HPV-9, gordelroos, Qdenga of meningokokken B verkoopt. De hele cross-sellcampagne wijst naar vaccins waarvan het starten van behandelingen al vijf maanden vrijwel stilligt. Is de reden voorraad of leverancier, dan levert nu een aanbod versturen geïnteresseerde patiënten op die de kliniek niet kan helpen, en dat is erger dan niet versturen. De controle kost een ochtend: fysieke voorraad nakijken, de situatie bij de leverancier, en de toedieningsagenda voor die vier vaccins. Zijn ze beschikbaar, volg dan het plan zoals het er staat. Zijn ze dat niet, begin dan met sporen 1, 2 en 4, die er niet van afhangen.

Het bestand bevat geen leeftijd, en dat is de ernstigste beperking van dit werk. De opdracht vroeg om focus op volwassenen, maar het bestand bevat geen geboortedatum en geen leeftijd, dus geen enkele lijst hier is op leeftijdsgroep gefilterd. HPV-9 en gordelroos hebben een duidelijk afgebakende leeftijdsindicatie, en HPV-9 aanbieden aan een 70-jarige of gordelroos aan een 25-jarige verbrandt zowel het bericht als het vertrouwen. Vóór elke verzending moet de balie de leeftijd in het dossier controleren. Valt de patiënt buiten de leeftijdsgroep, dan geeft de kolom `alternativas_se_fora_da_faixa` in `campanha_onda1.csv` al het volgende vaccin met hoge marge dat hij mist. Van de eerste 600 in de rij hebben er 379 een alternatief en 221 niet; van die 221 kregen er 93 juist een leeftijdsgevoelig aanbod, en voor hen is de regel simpel: valt hij buiten de groep, forceer dan niets, bied het griepvaccin uit spoor 4 aan en ga door.

7. Het plan in vier sporen

Zonder affiniteit tussen vaccins is het criterium voor het aanbod niet langer de historie maar de waarde. Voor cross-sell is de regel rechttoe rechtaan: elke patiënt krijgt één aanbod, het duurste van de vier vaccins met de hoogste marge dat hij nog niet heeft gehad, en de rij is geordend op wat hij al bij de kliniek heeft uitgegeven, wat het beste beschikbare signaal is van betalingsbereidheid. Dat dekt 3.187 van de 3.221 patiënten, aangezien maar 34 alle vier hebben. Het bestand `campanha_onda1.csv` bevat die rij kant-en-klaar, met het aan te bieden vaccin, de protocolwaarde en het alternatief per leeftijdsgroep.

Een patiënt kan in meer dan één spoor voorkomen. De drie contactlijsten tellen samen 3.264 namen voor 2.388 verschillende patiënten (486 patiënten staan in zowel terughalen als griep, 390 in cross-sell en griep, 14 in terughalen en cross-sell). De regel om ze naast elkaar te laten bestaan is één bericht per patiënt per week, in deze volgorde: terughalen eerst, cross-sell daarna, griep als laatste.

Spoor 1, begint morgenochtend: het tweede vaccin aan de balie aanbieden

Geen lijst en geen systeem nodig. Bij elke toediening kijkt de balie in het dossier welk vaccin de patiënt mist en stelt de vraag ter plekke, met het vaccinatieboekje in de hand. Dat zijn ongeveer 1.474 toedieningen per maand. Wordt 5% daarvan een tweede dosis, dan zijn dat 74 extra doses per maand; bij 10% zijn het er 147.

Uitgaande van een voorzichtige gemiddelde prijs van R\$ 300 per dosis onder de vaccins die beschikbaar blijven, is dat tussen R\$ 22.200 en R\$ 44.100 per maand, terugkerend, tegen nul contactkosten (laag vertrouwen wat het percentage betreft: de kliniek heeft nooit aan de balie aangeboden, dus er is geen eigen historie; de basis van de berekening, 1.474 toedieningen per maand, is hoog vertrouwen). Het management meet elke vrijdag hoeveel bezoeken van die week met meer dan één vaccin vertrokken, vanaf een uitgangspunt van 1,12%, en het doel voor de eerste week is 3% halen. Beweegt het getal na twee weken niet, dan ligt het aan het script en niet aan de patiënt: het management draait twee diensten aan de balie om het gesprek te horen en de vraag te herschrijven. Dit niet doen betekent doorgaan met betalen voor WhatsApp om mensen te bereiken die al in de kliniek stonden.

Spoor 2, week 1: de blijven liggen doses terughalen

De balie neemt de 635 patiënten uit `doses_pagas_nao_aplicadas.csv`, dat al geordend is van het hoogste blijven liggen bedrag naar het laagste, en stuurt persoonlijke WhatsApp-berichten met wat er betaald is en openstaat, met twee concrete tijdstippen om uit te kiezen. Wie in de kolom `risco_de_falta` als hoog risico staat, krijgt een telefoontje in plaats van een bericht, omdat dat het profiel is dat historisch niet komt opdagen. Met een quotum van 60 contacten per dag is de lijst in 11 werkdagen rond. Wie binnen 3 dagen niet reageert, krijgt een telefoontje; wie twee telefoontjes niet opneemt, krijgt een bericht met een afspraaklink en verlaat de actieve rij. Het doel is 190 tot 254 patiënten opnieuw in te plannen, een terughaalpercentage van 30% tot 40% van de lijst (laag vertrouwen: dat is een gangbaar richtcijfer voor een al betaalde dosis, de kliniek heeft geen eigen historie). Op vrijdag telt één getal: hoeveel blijven liggen doses van de lijst af zijn, met een doel van 90 in de eerste week. De totale kosten komen op ongeveer R\$ 2.900 aan berichten en telefoontjes, tegen R\$ 704.539 aan betaalde doses die openstaan. Draait dit spoor niet, dan is het risico geen gemiste omzet: het is een patiënt die een jaar later zijn geld terugvraagt.

Spoor 3, week 2, afhankelijk van de voorraadcontrole: cross-sell naar de eerste 600

De eigenaar of de commercieel manager neemt de eerste 600 uit `campanha_onda1.csv` (afkapgrens van R\$ 5.050 historische omzet) en stuurt een adviserende, niet promotionele WhatsApp met het vaccin uit de kolom `vacina_ofertada`. Die partij valt uiteen in 215 aanbiedingen van HPV-9, 180 van gordelroos, 141 van Qdenga en 64 van griep, met een gemiddelde protocolwaarde van R\$ 1.718 en een brutopotentieel van R\$ 1.030.730 als iedereen zou kopen, wat niet gaat gebeuren en niet als doel gebruikt moet worden. Quotum van 60 contacten per dag, 10 werkdagen. Bij een conversie van 3% tot 6% (laag vertrouwen, geen eerdere campagne in het bestand om op te ijken) ligt het verwachte resultaat tussen 18 en 36 protocollen, oftewel R\$ 30.900 tot R\$ 61.800, tegen ongeveer R\$ 2.550 kosten: een rendement van 12 tot 24 keer de inleg. Wie binnen 5 dagen niet reageert, krijgt een tweede bericht met sociale bewijskracht, en zonder reactie na nog eens 5 dagen een telefoontje; zonder reactie op het telefoontje gaat hij achteraan in de rij en wordt hij pas het kwartaal erop opnieuw benaderd. Het vrijdaggetal is hoeveel afspraken er uit de partij zijn gekomen, met een doel van 6 in de eerste week. Dit spoor niet draaien laat R\$ 30.900 tot R\$ 61.800 in de rij staan, en de kosten van uitstel lopen op omdat het bestand veroudert: de meest recente patiënt zit al 127 dagen zonder aankoop, en hoe kouder het contact, hoe lager de conversie.

Het is de moeite waard de 600 in twee helften van 300 te splitsen om het argument te toetsen, want niemand in de kliniek weet welk argument werkt: de ene helft krijgt het bericht dat draait om het risico dat het vaccin voorkomt, de andere het bericht dat draait om de uitkomst (het gezin beschermd, een zorgeloze reis, geen uitval). Met 18 tot 36 conversies over de hele partij sluit de toets nu statistisch niet, maar hij geeft

wel richting voor de volgende golven, die groter zullen zijn: na deze 600 blijven er nog 2.587 patiënten in de rij.

Spoor 4, week 3: griep op orde

Griep is jaarlijks volgens de klinische kalender, en het is het vaccin met de breedste indicatie in het assortiment, dus het heeft een eigen spoor nodig: volgens de regel van het duurste aanbod zouden van de 1.896 patiënten die nooit griep hebben gehad er maar 64 het in spoor 3 aangeboden krijgen. Het bestand `gripe_em_atraso.csv` voegt de twee groepen samen, 2.029 patiënten: 133 die het gehad hebben en er meer dan 12 maanden niet op zijn teruggekomen, en 1.896 die het nooit hebben gehad. De balie stuurt WhatsApp-herinneringen, te beginnen bij de 133 die over tijd zijn, quotum van 80 per dag, waarmee beide groepen in 26 werkdagen rond zijn, met een telefoontje voor wie binnen 7 dagen niet reageert. Bij de groep die over tijd is, is de verwachting 25,2% respons (laag vertrouwen: dat is het percentage uit de enige cohort in het bestand met een volledige 365 dagen opvolging, en dat zijn maar 147 patiënten); bij wie het nooit heeft gehad geldt dezelfde marge van 3% tot 6% als bij cross-sell. Samen is dat tussen R\$ 17.100 en R\$ 27.930 bij R\$ 190 per dosis, tegen R\$ 1.015 kosten. Laat 10% van de lijst erbuiten, zonder iets te ontvangen, als controlegroep: het verschil in conversie tussen wie wel en wie niets kreeg is de enige manier om te weten of de campagne werkte of dat deze mensen sowieso waren teruggekomen. Het vrijdaggetal is hoeveel griepdoses er die week zijn ingepland, doel van 20 in de eerste. Dit spoor niet draaien laat het enige jaarlijks terugkerende vaccin zonder enig contact, en dat is de goedkoopste ingang om wie gestopt is met kopen weer te activeren.

8. Wat deze lezing nog kan veranderen

Het bestand eindigt op 10-04-2026, 127 dagen geleden. Elke telling van “dagen zonder aankoop” draagt die ondergrens, en de meest recente patiënt verschijnt met 127 dagen terwijl hij op de laatste dag in het bestand nog kocht. Het is ook waarom het aantal blijven liggen doses is gemeten op de einddatum van het bestand en niet vandaag: meten tegen vandaag zou het totaal opblazen met R\$ 118.405 die misschien al gezet zijn zonder dat het bestand het laat zien.

Griep gedraagt zich in dit bestand niet als jaarlijkse herhaalaankoop, ook al is het dat in de klinische praktijk wel. Van de 393 intervallen tussen herhaalde griepaankopen haalt er geen enkele de 300 dagen, en het mediane interval is 83 dagen, in lijn met alle andere vaccins (tussen 76 en 117 dagen). Het bestand legt de jaarcyclus dus niet vast, waarschijnlijk omdat het waargenomen venster te kort is. Spoor 4 is gebouwd op de klinische regel van 12 maanden en niet op een herhaalpatroon dat de data had laten zien.

Elke regel is één gekochte dosis, en het hele protocol wordt in één keer gekocht. De tweede en derde dosis dragen dezelfde aankoopdatum als de eerste en worden later gezet. Daarom is er niemand die een halve reeks heeft gekocht: wat er wel is, is een gekochte en niet gezette dosis, en dat is het onderwerp van spoor 2. “Heeft vaccin X gehad” betekent in dit rapport “heeft het protocol van X gekocht”.

De conversiepercentages zijn marktreferenties, niet die van jouw bedrijf. Het bestand legt geen eerdere campagne vast, dus de marges van 3% tot 6% en 30% tot 40% zijn beredeneerde schattingen. De projecties gaan er ook van uit dat de conversie vergelijkbaar is bij vaccins die twaalf keer in prijs verschillen, van R\$ 190 tot R\$ 2.234; converteren patiënten in de praktijk minder op de dure, dan valt de geprojecteerde

omzet richting de helft terug en komt het rendement tussen 6 en 12 keer de kosten, nog altijd hoog genoeg om het spoor de moeite waard te maken. Herbereken de rij na de eerste golf met de werkelijke conversie.

Sorteren op historische omzet bedient meer dan één wijk minder goed. De eerste 600 hebben een gemiddelde omzet van R\$ 6.350, tegen R\$ 3.337 voor het bestand, en drie dimensies raken scheef. Naar betaalwijze komt 21,4% van de patiënten met directe overboeking erin tegen 13,6% van wie per factuur betaalt. Naar kanaal 20,6% van wie via doorverwijzing kwam tegen 15,0% van wie via Google kwam. Naar wijk 23,7% van de bewoners van Parque Industrial tegen 14,2% van die in Jardim das Acácias. Factuur en verzekering zijn de betaalwijzen die het sterkst samenhangen met een lager inkomensprofiel, dus deze rij praat eerst met wie al winstgevender is. Een minimumquotum van 50 per wijk zou symbolisch zijn: dat verschuift 3 patiënten op 600. Een quotum van 60 per wijk verschuift 38 patiënten en brengt het gebied werkelijk in balans. Is het doel om inkomen te balanceren en niet alleen geografie, dan moet het quotum per betaalwijze, met een minimumaantal contacten onder patiënten met factuur en verzekering.

Geen affiniteit betekent niet dat cross-sell niet werkt. Het betekent dat de vaccinhistorie niet helpt kiezen wat je aanbiedt. De keuze wordt protocolwaarde en klinische geschiktheid, en dat is wat de lijsten doen. Gaat de registratie leeftijd en de reden van weigering bij elk contact vastleggen, dan kan de volgende ronde echt segmenteren.

Bijlage

Geproduceerde bestanden

Bestand	Wat erin staat	Volgorde en vertekening
<code>campanha_onda1.csv</code>	3.187 patiënten met het aanbod met de hoogste waarde dat elk mist, plus alternatieven per leeftijdsgroep	op historische omzet aflopend, bedient wie minder uitgeeft en wie per factuur of verzekering betaalt minder goed
<code>doses_pagas_nao_aplicadas.csv</code>	635 patiënten met een betaalde en verlopen dosis, met bedrag, dagen over het venster en een markering voor no-showrisico	op blijven liggen bedrag aflopend, bedient wie een goedkope dosis heeft liggen minder goed
<code>doses_paradas_por_vacina.csv</code>	de 10 vaccins met doses, blijven liggen bedrag en patiënten, om de tabel in paragraaf 4 na te kijken	op blijven liggen bedrag aflopend
<code>gripe_em_atraso.csv</code>	2.029 patiënten zonder griep op orde: 133 meer dan 12 maanden over tijd en 1.896 die het nooit hadden	gegroepeerd op situatie, daarna historische omzet aflopend
<code>alvos_qdenga.csv</code>	1.902 patiënten zonder Qdenga (protocol R\$ 1.060)	op historische omzet aflopend
<code>alvos_hpv9.csv</code>	2.262 patiënten zonder HPV-9 (protocol R\$ 2.234)	op historische omzet aflopend
<code>alvos_zoster.csv</code>	2.184 patiënten zonder gordelroos (protocol R\$ 2.160)	op historische omzet aflopend
<code>alvos_gripe.csv</code>	1.896 patiënten zonder griep (dosis R\$ 190)	op historische omzet aflopend
<code>exclusivos_uma_vacina.csv</code>	430 patiënten die maar één vaccin haalden, met welk vaccin dat was	op historische omzet aflopend
<code>pacientes_multivacina.csv</code>	59 patiënten met 6 vaccins, het maximum in het bestand, kandidaten voor doorverwijzing en onderzoek	op historische omzet aflopend
<code>overlaps_pares.csv</code>	de 45 vaccinparen met overlap, Jaccard, affiniteit in beide richtingen en de statistische toets	op overlap aflopend
<code>summary_by_id.csv</code>		op totale omzet aflopend

Bestand	Wat erin staat	Volgorde en vertekening
	1 regel per patiënt, één ja/nee-kolom per vaccin, plus omzet, recentheid en blijven liggen doses	

De vier `alvos_*.csv` -lijsten overlappen elkaar sterk: bij elkaar opgeteld geven ze 8.244 regels voor maar 3.187 verschillende patiënten. Gebruik `campanha_onda1.csv` om te versturen, want die lost de overlap al op met één aanbod per patiënt, en de `alvos_*` -lijsten alleen om per vaccin op te zoeken.

De 12 paren met de grootste overlap

Affiniteit 1,00 is toeval. De Jaccard-kolom meet hoeveel de twee groepen overlappen ten opzichte van beide samen.

Vaccin A	Vaccin B	Beide gehad	Jaccard	Affiniteit
Griep	Qdenga	512	0,24	0,94
Griep	HPV-9	418	0,22	1,06
Griep	Gordelroos	408	0,21	0,96
Gordelroos	Qdenga	389	0,20	0,92
Griep	Hepatitis B	384	0,20	0,94
Hepatitis B	Qdenga	375	0,19	0,92
HPV-9	Qdenga	349	0,18	0,89
Griep	Meningokokken B	337	0,18	0,96
Meningokokken B	Qdenga	325	0,18	0,93
Griep	Kinkhoest (DTP)	321	0,18	0,98
Qdenga	Kinkhoest (DTP)	311	0,17	0,96
Griep	Meningokokken ACWY	302	0,17	0,94

Het paar met de hoogste affiniteit van het hele bestand, griep met HPV-9 (1,06), doorstaat de statistische toets niet: met 45 paren die tegelijk getoetst worden, lijken sommige puur door geluk af te wijken van toeval, en na correctie voor dat aantal toetsen overleeft dit paar niet.

Hoe elk getal tot stand kwam

- **Een patiënt “heeft” een vaccin** wanneer er minstens één aankoopregel voor op zijn naam staat, gezet of niet. De 10 verzamelingen komen uit de kolom `vacina`, en de koppelsleutel is `paciente_id`. Er zijn geen dossiers samengevoegd: de 3.221 identificatoren corresponderen 1 op 1 met de 3.221 namen.

- **Affiniteit** tussen A en B is de kans op B onder wie A heeft, gedeeld door de kans op B in het hele bestand. Significantie via de exacte toets van Fisher, met het criterium gecorrigeerd voor de 45 getoetste paren.
- **Een blijven liggen dosis** is een regel die nog op toediening wacht en waarvan de tijd tussen aankoop en de einddatum van het bestand het 95e percentiel overschreed van de tijd tussen aankoop en toediening bij datzelfde vaccin en dezelfde dosis. Het zijn er 1.246; gemeten tegen vandaag in plaats van het einde van het bestand zouden het er 1.452 zijn, een verschil dat het bestand niet kan bevestigen.
- **De breuk van 9 maart** vergelijkt het daggemiddelde van de voorgaande 30 dagen met dat van de volgende 33 dagen, per vaccin en per dosis, steeds op aankoopdatum.
- **Een bezoek** is de combinatie van patiënt en datum: van aankoop voor de 11.043 verkoopbezoeken, van toediening voor de 17.515 behandelbezoeken.
- **Griep over tijd** zijn de patiënten van wie de laatste griepaankoop ouder is dan 12 maanden vóór het einde van het bestand, plus wie het nooit kocht. Het percentage van 25,2% komt uit de cohort van wie de eerste griepaankoop vóór 10-04-2025 ligt, de enige met een volledig jaar opvolging binnen het bestand.
- **Protocolwaarde** is de prijs per dosis maal het aantal doses in de reeks. De prijzen liggen vast in het hele bestand, zonder variatie per patiënt of per periode.
- **Omzet** telt de kolom **valor** over alle regels op, inclusief doses die nog niet gezet zijn, aangezien de aankoop is geregistreerd.

* **Over dit voorbeeld.** Dit is een echt rapport uit de Distilo-pipeline, gedraaid op synthetische data voor een fictief bedrijf. Dat bedrijf is Braziliaans: een vaccinatiekliniek. Daarom staan de bedragen in Braziliaanse real, zijn de wijk- en patiëntnamen Braziliaans, en weerspiegelt een deel van het assortiment (Qdenga tegen dengue, bijvoorbeeld) een Braziliaans vaccinatieschema. Het is gepubliceerd zoals het geproduceerd is en niet aangepast, omdat de analyse is wat we je laten zien en de analyse niet verandert met het land. Je eigen rapport is geschreven in jouw taal, over jouw bedrijf, in jouw valuta.