Part 1

This part of the case focuses on working with the patient and practitioner team to optimise our retention. The company has had incredibly strong retention historically due to our effective, medical grade products and the customer experience provided by our skincare concierge (see next slides for high-level patient and prac journey).

We have provided a false dummy dataset of **order level data** (each row is one order) to help assist you in understanding how our existing business is performing. What do you think? What are opportunities to further grow sales from existing patients / practitioners over the next 1-3 years?

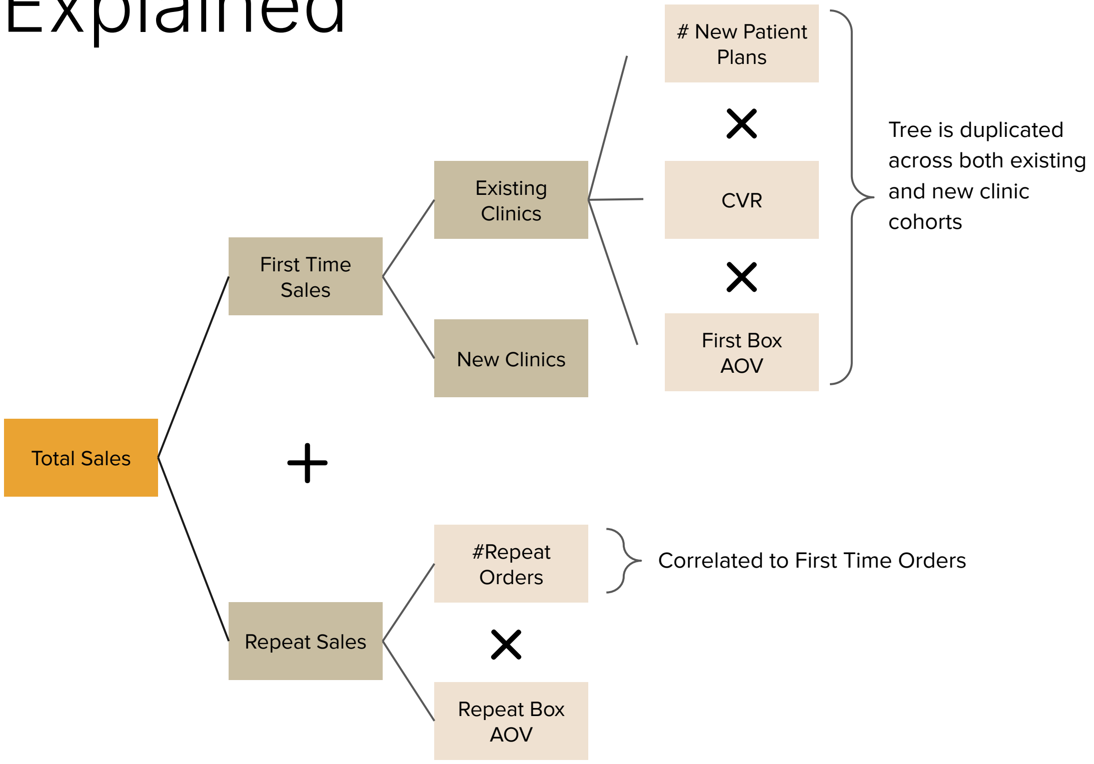
Submission Format
Please use Python to conduct the analysis and pull out your insights - no fancy presentation required (should you be successful, you will work on translating your findings into a presentation as part of Part 2).

We request that your Part 1 submission includes:

- High-level summary of insights and the actionable recommendation
- Any code written to conduct the analysis
- Any visualisations to aide the analysis
- Access to the raw analysis

For example, a submission using a Google Colab notebook containing Python inline code, visualisations, raw analysis and a summary is a suggested format.

Business Drivers Explained



GLOSSARY

New Patient Plans / Referrals / Adds:
Plans Created by a Clinician for a New Patient; also equiv. to # of New Patients Referred to GH or Added to GH

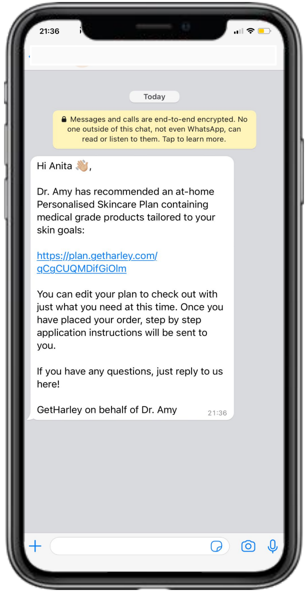
Conversion Rate:
New Patient Orders over New Patient Plans

New Clinics:
New Clinics onboarded in respective year (2023 clinics are considered “New” this year)

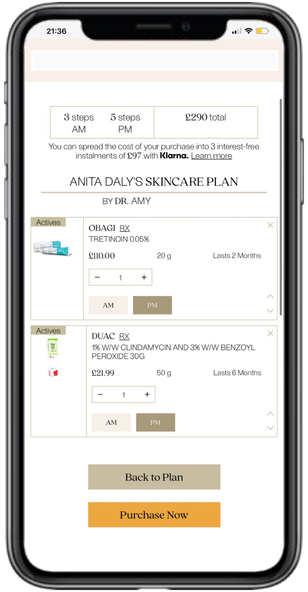
Existing Clinics: Existing Clinics onboarded prior to 2023 or in 2022

We enable patients to access personalised, medical grade skincare recommendations from clinicians digitally, in turn creating a B2B2C acquisition channel for us

1 Clinician Creates Personalised Plan for Their Patient on our platform

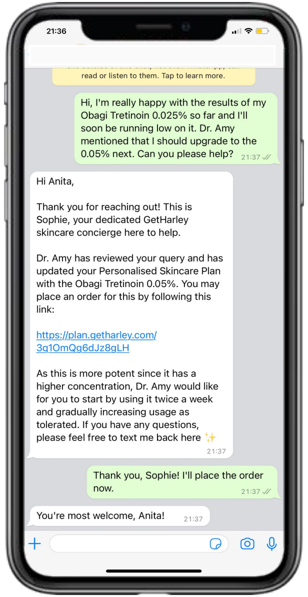


2 Patient Receives Skincare Plan Digitally



3 Patient Checks Out Online

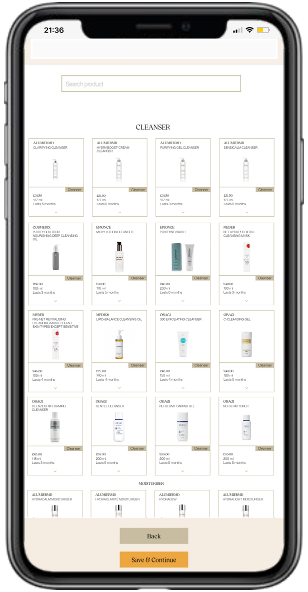
4 Patient Receives Products



5 Patient Enjoys Continuous Care and Replenishes Seamlessly

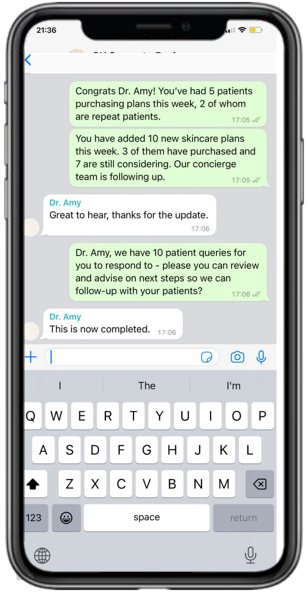
Clinicians provide superior patient care and earn ~30% more in annual net income, enabling win-win-win dynamics for the patient, clinician and us

1 Clinician Integrates In-Clinic Treatments with Skincare Plan

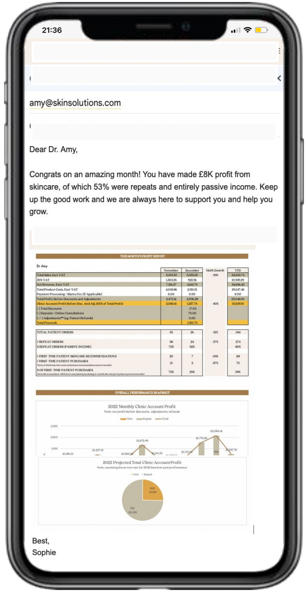


2 Clinician Curates Plan from Limitless Product Shelf

__Current Alternative without us:__ Sparse Inventory Choices in Clinic



3 Clinician Receives Real Time Feedback



4 Clinician Earns Commission Anytime, Anywhere

Part 2

Section 1
Please can you re-summarise the insights that you found in Part 1 in a digestible format for our exec and retention team to take action on.

Section 2
The previous data shared in Part 1 was part of a snapshot of a script run by our analyst at the end of the every month which we use to analyse and interpret patient and practitioner cohorts. This is one of the pain points at Company 2 and the exec team is unhappy with the cadence of reporting and the level of visibility over cohort behavior on the platform.

The analyst will run board_patient_cohorts.py (see py attachment) every month via command line: *python board_patient_cohorts.py ./input/stripe/unified_payments_<current_date>.csv ./output/board_cohorts_<current_date>.csv*

Input data (please find lexicon for the next page):
- unified_payments_<current_date>.csv
- combined_clinic_patient_orders_detailed.csv

Some challenges our analyst runs into while creating this dataset today:
- In order to run the cohort scripts, the analyst must first download up-to-date revenue data from **Stripe (unified_payments_<current_date>)** for the full month (taking 3-4 hours)
- Once this data is downloaded from Stripe, he must match this with the order level data from our **database (combined_clinic_patient_orders_detailed)** and then assign cohorts to each new order (and all historic orders)
- The script that assigns the cohorts in itself has a long running time (6-8 hours), with the majority of the running time spent processing order matching
- Order matching is required because the patient information data within the database does not have unique absolute identifications to match an order to a specific customer (eg name spelling errors/shorten name/nickname/blanks, one person might have multiple cards/billing addresses/emails)
- Also, the Sumup Machine (a new payment method of tappable machines) was introduced from 2022-06 onwards which added additional challenges to reporting:
    - One Sumup Machine will have one unique ‘card fingerprint’, but multiple customers, causing some incorrect order matching
    - For simulated example:
        - When ‘card fingerprint’ = ‘zzzzzzzzzzzzz’, ‘username’ has value of ‘Tim Liu’, ‘James Lee’, ‘Susan sontag’...
        - But modified_username will all return ‘Susan sontag’ as she got the most orders within ‘card fingerprint’ = ‘zzzzzzzzzzzzz’

Please provide a plan to diagnose the issue(s) and increase the data quality and granularity. We’d like to see a high level architecture of the analytic infrastructure, cost / investment and a roadmap on how to build it coupled with the roles and responsibilities of the orgs involved in the project in order to get buy-in and sign-off from the exec on what we need to do / where we need to invest to step-change the analysis.

Is there anything else you can think of that we should build / track to better understand the business problem present by the exec team?

Input Data Dictionary

- **unified_payments_<current_date>.csv:**
  ['id', 'Description', 'Seller Message', 'Created (UTC)', 'Amount', 'Amount Refunded', 'Currency', 'Converted Amount', 'Converted Amount Refunded', 'Fee', 'Taxes On Fee', 'Converted Currency', 'Mode', 'Status', 'Statement Descriptor', 'Customer ID', 'Customer Description', 'Customer Email', 'Captured', 'Card ID', 'Card Last4', 'Card Brand', 'Card Funding', 'Card Exp Month', 'Card Exp Year', 'Card Name', 'Card Address Line1', 'Card Address Line2', 'Card Address City', 'Card Address State', 'Card Address Country', 'Card Address Zip', 'Card Issue Country', 'Card Fingerprint', 'Card CVC Status', 'Card AVS Zip Status', 'Card AVS Line1 Status', 'Card Tokenization Method', 'Shipping Address Line1', 'Shipping Address Line2', 'Shipping Address City', 'Shipping Address State', 'Shipping Address Country', 'Shipping Address Postal Code', 'Disputed Amount', 'Dispute Status', 'Dispute Reason', 'Dispute Date (UTC)', 'Dispute Evidence Due (UTC)', 'Invoice ID', 'Invoice Number', 'Payment Source Type', 'Is Link', 'Destination', 'Transfer', 'Interchange Costs', 'Merchant Service Charge', 'Transfer Group', 'PaymentIntent ID', 'order_id (metadata)', 'plan_id (metadata)', 'patient_id (metadata)', 'order_number (metadata)', 'patient (metadata)', 'source (metadata)', 'user_id (metadata)', 'payload (metadata)', 'appointmentTypeId (metadata)', 'payment_method (metadata)', 'appointment_id (metadata)']

- **combined_clinic_patient_orders_detailed.csv:**
  ['code', 'day', 'month', 'year', 'number', 'patient_id', 'plan_id', 'promo_code', 'sub_total', 'total', 'discount', 'harley_sub_total', 'status', 'order_payment_received_no', 'order_in_clinic_no', 'order_complementary_no', 'new_patient', 'full_name', 'day_of_week', 'public_holiday', 'is_public_holiday', 'existing_patient', 'plan_creation_date', 'plan_creation_time', 'plan_value']