Rebates and Tracing Fee Analysis

Ophthalmic surgical instruments manufacturing leader

Identified excessive claims by designing a robust reconciliation process for distributor rebates/fees and invoices by accounting for variances across quantity, price and customer dimensions

Ophthalmic surgery instruments company needs “visibility” into rebates and fees overpayments

Picture this…

You’re looking for opportunities to enhance profitability by understanding if you are overpaying rebates and fees to the distributors, but you don’t have the mechanism to validate the claimed rebates and fees numbers.

You turn to Accordion.

We partner with your team to accurately understand volume of overpayments and perform a comprehensive investigation including:

1) Establishing a customer mapping methodology based on addresses leveraging geocoding API and LLMs, saving ~10K manhours
2) Extracting data from unstructured PDF documents saving ~50-60 manhours
3) Developing a data model for transaction mapping to detect discrepancies between customer, SKUs referenced and corresponding contracts
4) Creating an inventory model to analyze volumes of secondarily sourced SKUs, along with associated fees and rebates
5) Building a robust reconciliation methodology to analyze discrepancies in contract pricing, acquisition costs, and rebate calculation methodologies.

Your value is enhanced.

- You have setup a streamlined Rebates and Tracing Analysis process, and now you have in-depth visibility into secondary inventory, contract price discrepancies, customer roster discrepancy.
- You reduced data extraction effort from PDF documents resulting in quarterly savings of 50-60 hours
- You were able to create an automated customer mapping process which saved over ~$500K in onetime cost.
- You have identified that you overpaid ~$900K in excess rebates and fees to distributors over a five-month period.
- Furthermore, now you have blueprint and action plan to realize an additional ~$2M in savings, driving significant value creation and operational efficiency improvements.

REBATES AND TRACING FEE RECONCILIATION

KEY RESULT

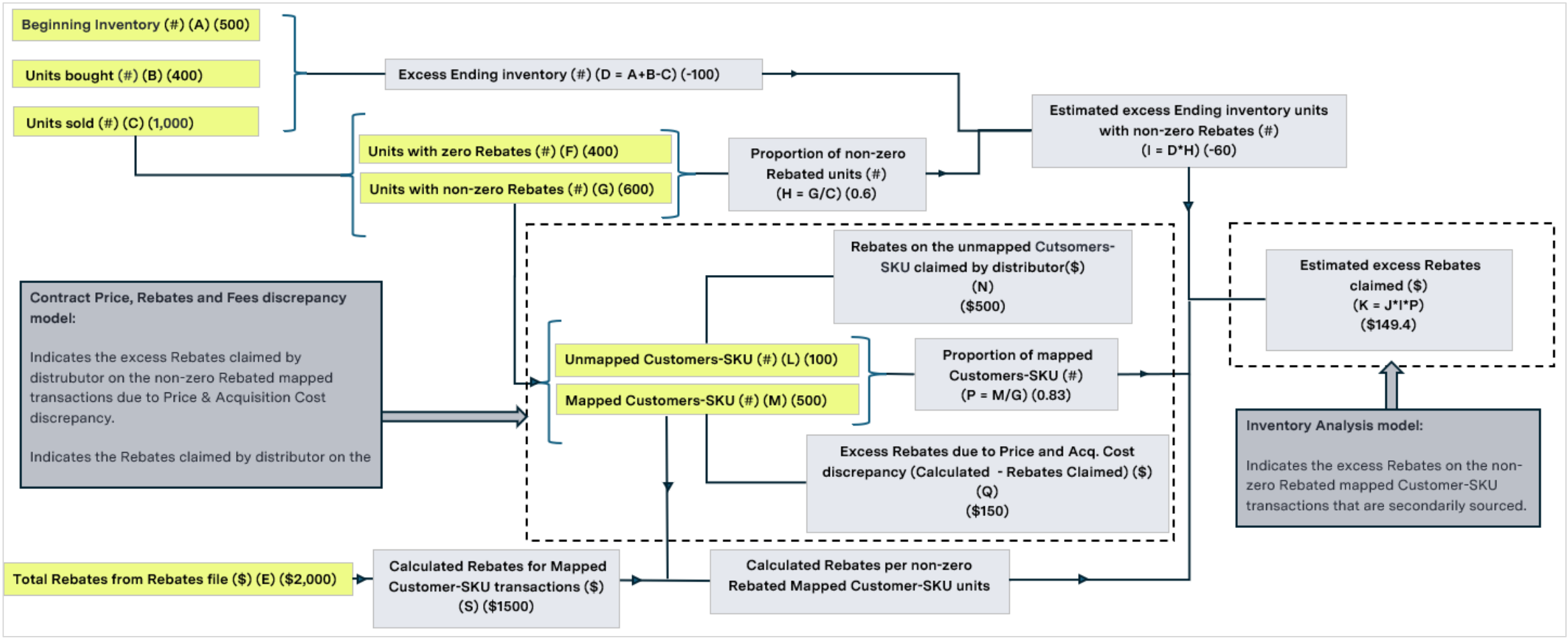
- >$2M potential annual savings
- ~$100K fees overpaid for secondary sourced SKUs identified
- ~10K manhours saved in customers mapping
- ~50-60 manhours saved in extracting data from PDF documents

VALUE LEVERS PULLED

- Inventory analysis
- Rebates analysis
- Tracing fees analysis
- Gen AI based PDF Data Extraction and Geocoding based address matching and validation

Methodology/Approach

The calculated rebates from the Contract Price, Rebates and Fees discrepancy model are pulled into the Inventory Analysis model to estimate the excess rebates claimed for the negative ending inventory SKUs



Methodology/Approach - Generative AI for PDF Data Extraction

Generative AI for PDF Data Extraction

- Data Extraction: Extracted data from approximately 58 PDF documents containing a mix of scanned and digital pages using OpenAI's GPT-4 model.

Impact:

- Compared to traditional OCR techniques and open-source Python libraries, the use of a single generative AI pipeline for extracting unstructured data from PDFs significantly reduced manual effort, saving an estimated 40–50 hours.

Read PDFs	PDF to Image – Image Preprocessing	Image Encoding	Input to GPT – 4o Model	Final Output
• Load and read PDF documents using open-source python libraries such as: PDF Plumber	• Convert the PDF documents to Images. The PDF documents were a mix of scanned and digital documents. • Apply OCR techniques using the python library: pytesseract to improve image readability. • Some of the OCR techniques include enhancing contrast, converting images to gray scale & binary scale etc.	• Convert the processed images to base64 encoded images to feed to the model.	• Feed the Base64 images into the GPT- 4o model along with engineered prompts.	• The final output from the GPT- 4o model is a JSON Object. • Convert the JSON object to python data frame and export the data frame to a local spreadsheet.

Note: A scanned PDF is essentially an image of the document. It is created by scanning a physical document, turning each page into an image. A digital PDF is created directly from digital sources (e.g., Word documents, spreadsheets). The content, such as text, images, and formatting, is stored in a structured format.

Methodology/Approach - Generative AI for PDF Data Extraction

Generative AI for Geocoded Address Validation:

- Address Validation: Validated around 20,000 unique addresses utilizing OpenAI's GPT-4 model.

Impact:

- Customer address validation for 20,000 addresses using the GPT-4 model resulted in a time savings of approximately 500 hours of manual work.

Fetch Latitudes & Longitudes	Identify Direct Matches	Distance Calculation	Mapped & Unmapped Segregation	Further Validation using LLM
• Fetch the latitudes and longitudes of all the BVI addresses and the corresponding Distributors/GPOs/IDNs using Google's Geocoding API.	• Identify the direct matches between the BVI addresses and the corresponding Distributor, GPO, or IDN addresses by comparing them directly.	• Calculate the distance for the rest of the BVI addresses against Distributors or GPOs or IDNs using the extracted latitudes and longitudes using python's open-source library: geodesic from geopy.distance. • Considering the huge volume of addresses and high run time for the distance calculation, the code was optimized to calculate the distances only for records where the zip codes were matching.	• Considering a threshold distance of 200m, we segregated the addresses as "Mapped" and "Unmapped" addresses.	• All the "Mapped" addresses were further validated using manual eyeballing and using the GPT-4 Model. • The LLM prompts were engineered to check and return the following details: a. Similarity match % b. Similarity match flag c. If the similarity match flag is "No" for any 2 addresses, reason for no match. • These were further segregated to "Direct Matches" and "Possible Matches".

ILLUSTRATIVE

Rebates and fees differences

Rebate variance due to Price & Customer discrepancy

Differences arising because of mismatch in prices for mapped customers

Differences due to unmapped customers

Rebate Analysis Bucket (Jan'24 – May'24)	Variance
Rebate Variance – Overall	**$201,287**
1.Driven by Contract Price or Acq. Cost or Rebates calculation mismatches	$87,164
1.1. Only Contract Price mismatch	$16,631
1.2. Only Acq. Cost mismatch	$370
1.3. Both Contract Price and Acq. Cost mismatch	$69,535
1.4. Rebates calculation by distributor mismatch	$628
2. Driven by unmapped Customer-SKU transactions	$114,123
2.1. Customer not available in Roster files	$75,607
2.2. SKUs not available in Price List	$38,515

Deep dive into reasons behind rebates variance

Fee variance due to Price & Customer discrepancy

Investigation for fees variance

Fee Analysis Bucket (Jan'24 – May'24)	Variance
Fee Variance - Overall	**$19,172 - $22,922**
1. Driven by Calculated Fee mismatch (Primarily sourced SKUs)	$518
2. Driven by Fee claimed on secondarily sourced SKUs	$18,651
3. Driven by Tracing transactions	
3.1. Considering no SKUs as secondarily sourced	$3
3.2. Considering all SKUs as secondarily sourced	$3,750

Drivers of tracing fee variance

ILLUSTRATIVE

Inventory variance

Inventory Variance over Months

Distributor	SKU	Parameter	Jan'24	Feb'24	Mar'24	Apr'24	May'24
Distributor 1	SKU 1	Beginning Inventory (#) (A)	4,500	0	5,000	0	0
		SKUs bought (#) (B)	300	8000	10,500	5,000	1,500
		SKUs Sold (#) (C)	5,200	3,000	34,586	15,084	900
		Ending Inventory (#) (A+B-C)	(400)	5,000	(19,086)	(10,084)	600

Thorough Inventory analysis for each distributor and SKU

Negative inventory reset to zero at the start of the new month.

Negative ending inventory indicates potential secondarily sourced SKUs

Major SKUs with negative ending inventory for Distributor 1 for the month of Jan'24

		Jan'24					
SKU No.	SKU Description	Excess Ending Inventory (#)	SKU Per unit Cost ($)	Excess Ending inventory ($)	Average rebate per unit ($)	Estimated excess Rebates ($)	Estimated Fees ($)
SKU 1	Description 1	(400)	1.50	(600)	0.85	175.00	229.00
SKU 2	Description 2	(3,472)	8.90	(30,901)	0.13	451.36	28.00
SKU 3	Description 3	(572)	4.60	(2,631)	0.71	406.12	418.00
SKU 4	Description 4	(763)	50.00	(38,150)	0.03	22.89	332.00
SKU 5	Description 5	(800)	2.40	(1,920)	-	-	8.00

SKU characteristics

Estimated excess rebates and fees associated with secondarily sourced SKUs

Learnings

1) Implementing Geocoding on scale
2) Prompt Engineering required for the proper validation of mapped addresses
3) Extracting structured data from unstructured sources like PDFs leveraging Gen AI
4) Developing an inventory model to track daily position of inventory and identify days where the inventory turned negative indicating the possibility of secondary sourcing