



Transforming Transactional Enrichment with Traditional AI Models and LLM Ensemble for Enhanced Banking Insights

Nov 2024



Agenda

01 **Banco de Crédito e Inversiones (BCI)**
What is BCI?

02 **Diego Pavez**
Who am I and what do I do?

03 **Problem**
The Need for Transactional Enrichment

04 **Solution**
The End-to-End Transactional Enrichment Process

05 **Actionable Insights**
Benefits and Impact of Transactional Enrichment

06 **Questions**



Agenda

01 **Banco de Crédito e Inversiones (BCI)**
What is BCI?

02 **Diego Pavez**
Who am I and what do I do?

03 **Problem**
The Need for Transactional Enrichment

04 **Solution**
The End-to-End Transactional Enrichment Process

05 **Actionable Insights**
Benefits and Impact of Transactional Enrichment

06 **Questions**

Banco de Crédito e Inversiones (BCI)

What is BCI (Bank of Credit and Investments)?



-The Bank was **founded in 1937 (87 years ago)** by Juan Yarur Lolas in order to support small and medium enterprises in Chile.

-Bci Miami is the first international branch of Bci and the second-largest branch of a foreign bank in Florida. It was **founded in 1999 (25 years ago)**

-In **2015 (9 years ago)**, BCI purchased the **City National Bank of Florida**, which has 26 branches in South Florida and Orlando

-In **2018 (6 years ago)**, City National Bank of Florida acquired Miami-based **TotalBank**

-City National Bank of Florida to **purchase Executive National Bank (ENB)** by the end of **2020 (4 years ago)**

-BCI operates mainly in Chile and EEUU, but also has operations in **Peru, Colombia, Brazil, Mexico, and China**

Our service model

BCI



*Personal
Banking*



*Business
Banking*



*Digital Payments
& Services*



*Wealth Management,
including Trust and
Services*

Business Units

*Retail banking
Wholesale banking
Finance division
Insurance Products*

Products

*Multi-currency checking accounts
Money Market account
Working capital
Letters of Credit (Export and Import)
Investments
Time deposits*

Services

*International Trade Solutions
International Payments
Standby Letters of Credit issued in U.S. dollars and euros
Foreign Exchange Services
Import and Export Documentary Collection
Forward Exchange Rate Hedging*

Agenda

01 Banco de Crédito e Inversiones (BCI)
What is BCI?

02 **Diego Pavez**
Who am I and what do I do?

03 **Problem**
The Need for Transactional Enrichment

04 **Solution**
The End-to-End Transactional Enrichment Process

05 **Actionable Insights**
Benefits and Impact of Transactional Enrichment

06 **Questions**

Me

Diego Pavez



I currently work in the Data & Analytics area of the bank, where we are a team of approximately 120 people. The bank has around 12,000 employees in Chile alone. My current responsibilities are focused on customer experience, but I have worked in various areas within the bank. One of those roles involved understanding transactional behavior at the merchant level and helping commercial teams improve the value proposition of our products.



Senior Data Scientist - Present



Machine Learning Engineer



Research Data Scientist



Data Scientist - 2019



<https://www.linkedin.com/in/diegotbot/>



diego.pavez.olave@gmail.com



diego.pavez@bci.cl

Agenda

01 Banco de Crédito e Inversiones (BCI)
What is BCI?

02 Diego Pavez
Who am I and what do I do?

03 Problem
The Need for Transactional Enrichment

04 Solution
The End-to-End Transactional Enrichment Process

05 Actionable Insights
Benefits and Impact of Transactional Enrichment

06 Questions

Problem

How can we understand the customer and improve the user experience?

User Experience

1. Fraud Experience
2. Experience with Bank Representative
3. Security and Privacy Experience
4. Digital Banking Experience
5. Mobile App Usability
6. Complaint Resolution Experience
7. Loan Experience (Personal and Mortgage)
8. **Loyalty Program (Cashback, Credit Card Discounts, Credit Card Rewards)**
9. Investment Experience
10. Transaction Speed and Efficiency
11. Overall Trust and Reliability
12. Account Opening and Onboarding Experience
13. ATM and Branch Experience
14. Fee Transparency
15. Interest Rate Satisfaction
16. Savings Account Experience
17. Billing and Statement Clarity
18. And more..



Loyalty Program

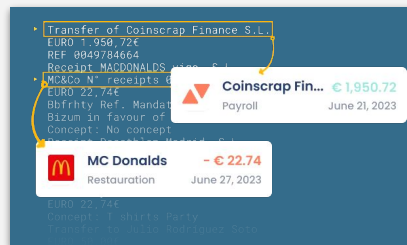
- Negotiate partnership terms with merchants for cashback (How)
- Analyzing Customer Purchase Behavior (How)
- Analyzing our customers' transactions



Problem

We cannot optimally understand customer purchasing behavior due to transactional noise

- Merchant name in different versions (WALMART, WAMART, WM CENTER3)
- Registered company name (REBUILDING TECHNOLOGY PTE LTD is DiDi)
- Company URL (CRUNCHYROLL.COM3)
- Company address (A.21 DE MAYO 2750 CONCEPCI is Copec in Chile)
- Company-associated acronyms (SB, SBX, SBKS is Starbucks)



Agenda

01 Banco de Crédito e Inversiones (BCI)
What is BCI?

02 Diego Pavez
Who am I and what do I do?

03 Problem
The Need for Transactional Enrichment

04 Solution
The End-to-End Transactional Enrichment Process

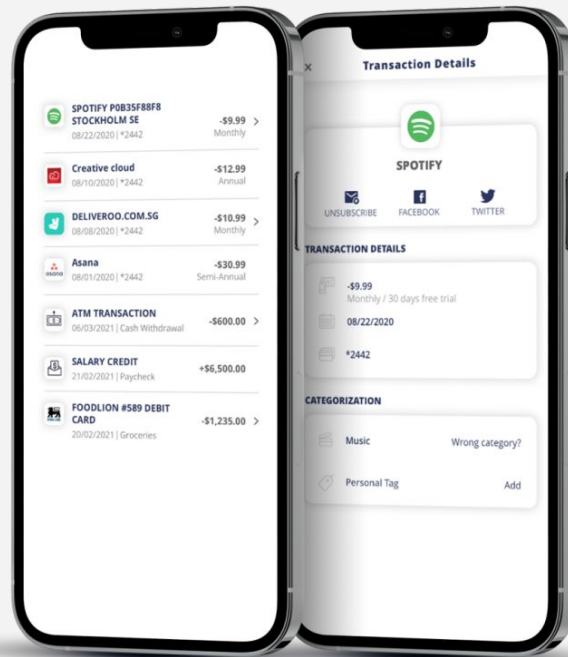
05 Actionable Insights
Benefits and Impact of Transactional Enrichment

06 Questions

Solution

Goal

To **enrich transactions** by transforming raw and ambiguous data into clean, organized, and structured information, adding supplementary details to each transaction to **enhance data understanding and usability**.



Solution

High-Level Merchant Transaction Enrichment

Merchant Names GUI



Interface for loyalty teams to manually map store names to transactional names, building a feedback-enriched training dataset

Data Cleaning and Feature Engineering



Data cleaning and FE involve removing unnecessary elements like stopwords, prepositions, citys, company terms; segmenting businesses by size (small or large); and developing logic to identify merchants within payment gateways

Merchant Identification Logic



We identify merchant names by training a DeBERTaV3 model for inference, applying Jaccard similarity, and using a mode-based heuristic for the final name selection

Ensemble learning



We used LLaMA 3.x with RAG to select the best brand from five options (brand cleaning, classifier, similarity, mode, and not applicable), letting the LLM make the final decision

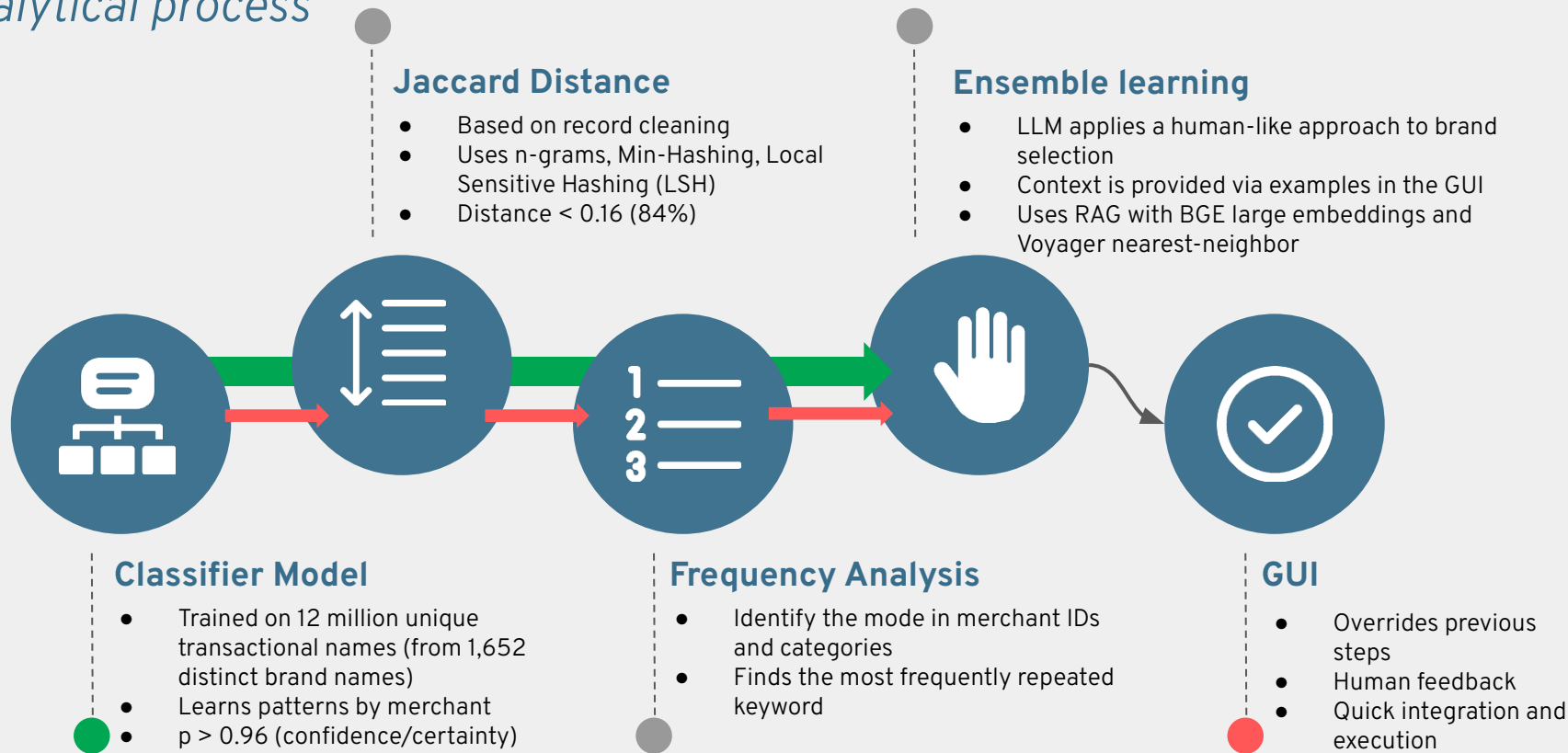
Final Business Name Cleanup



With the final brand identified, we perform cleanups using brand knowledge from the GUI and heuristics, grouping by acquirer ID and merchant categories

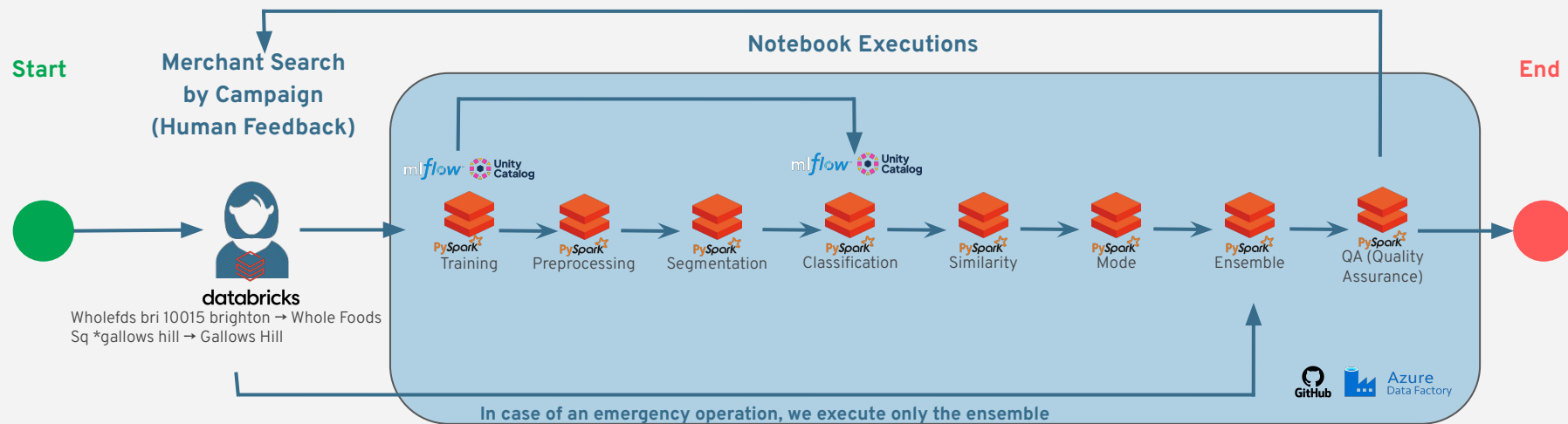
Solution

Analytical process



Solution

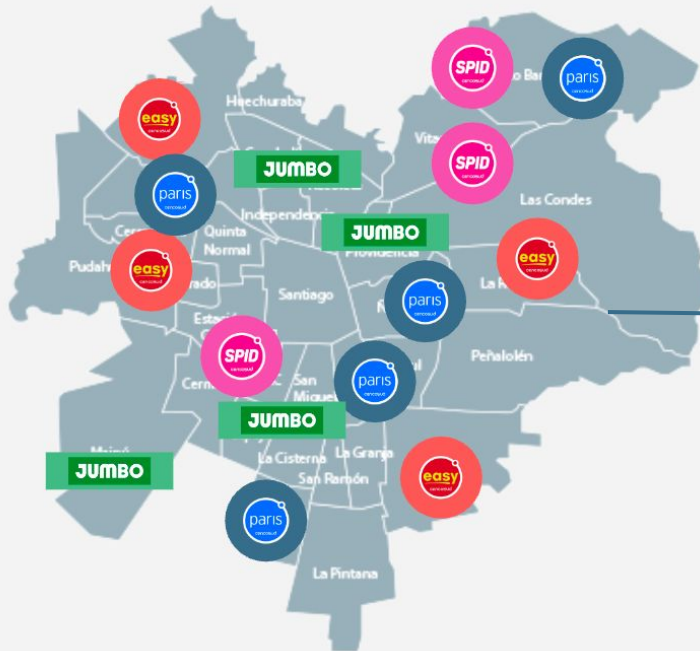
Process flow



Solution

Zoom Classification

“When training a classification model with 1,600 labels, feature engineering is essential. Techniques like n-grams and data augmentation help standardize transaction examples, and including merchants not to classify gives the model full context.”



Paris Alto las Con =
[Alto], [Con]
embedding
0.2]

[Paris],
[0.9, 0.5,

Easy Quil =
embedding

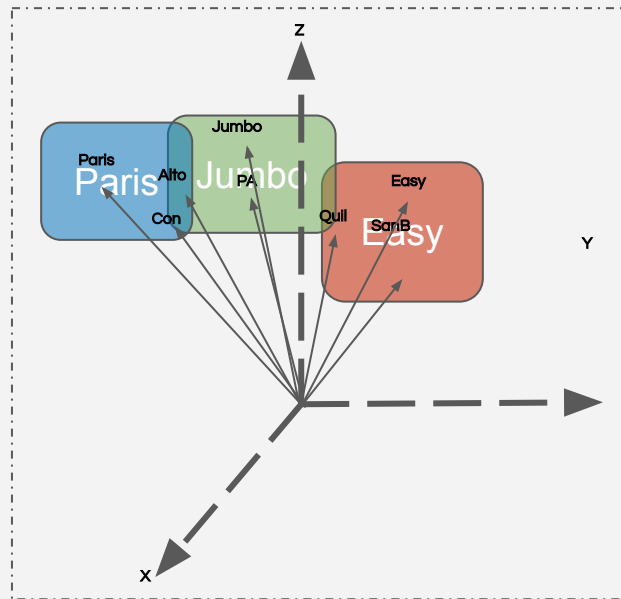
[Easy], [Quil]
[0.8, -0.3]

Jumbo PA =
embedding

[Jumbo], [PA]
[0.9, 0.1]

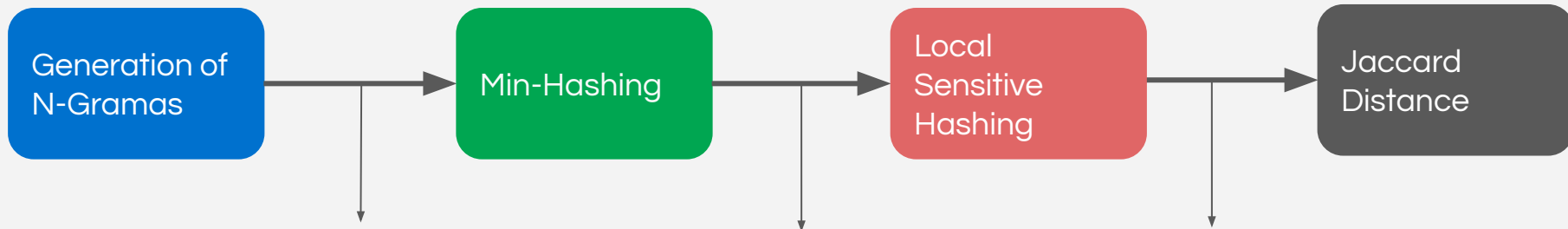
Easy SanB =
embedding

[Easy], [SanB]
[0.6 -0.7]



Solution

Zoom Jaccard Distance



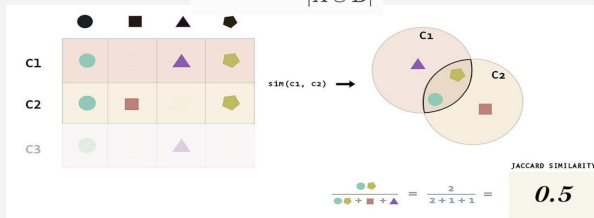
Document Matrix with tri-grams

Signatures: Integer vector representing text similarity

Pairs of signatures to be tested for similarity

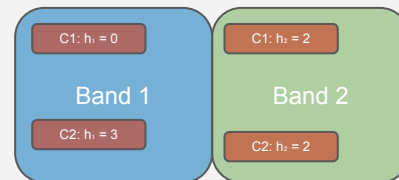
[Paris], [Alto], [Con] =
[PAR][ARI][RIS][ISA][SAL][ALT][LTO]..
[Paris][Isal] = [PAR][ARI][RIS][ISA][SAL]

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$



Shingle	C1	C1
PAR	1	1
ARI	1	1
RIS	1	1
ALT	1	0
LTO	1	0
CON	1	0
ISA	0	1
SAL	0	1

Signature of C1:
 $h_1(C1) = \text{minimum of } \{3, 5, 7, 0, 2, 4\} = 0$
 $h_2(C1) = \text{minimum of } \{5, 8, 2, 5, 8, 2\} = 2$
Signature of C2:
 $h_1(C2) = \text{minimum of } \{3, 5, 7, 6, 8\} = 3$
 $h_2(C2) = \text{minimum of } \{5, 8, 2, 5, 8\} = 2$



Number of matching hashes: 1 (h_2 is the same in both)
(Band 2)
Total hashes: 2
Estimated similarity: $1/2 = 0.5$
Actual Jaccard Similarity:
Common shingles: {PAR, ARI, RIS}
Total unique shingles: 8
Actual similarity: $3/8 = 0.375$

Solution

Zoom Frequency Analysis

Segmentation

Cleaning

Frequency

Identifier



Sta Isabel Vina D Vina del Mar
Santa Isabel C DEC. De Tango
Santa Isabel 14 de Antofagasta

[SANTA, ISABEL, VINA, DEL, MAR]
[SANTA, ISABEL, TANGO]
[SANTA, ISABEL, ANTOFAGASTA]



Company

Unique ID 93.834.000



Spid San Damian Santiago
Spid 35 El Llano Santiago
Spid Apoquindo Santiago

Words

[SANTA]
[ISABEL]
[ANTOFAGASTA]
[VINA]
[TANGO]

Frequency

[3]
[3]
[1]
[1]
[1]

SANTA ISABEL

Words

[SPID]
[SAN]
[DAMIAN]
[LLANO]
[APOQUINDO]

Frequency

[3]
[1]
[1]
[1]
[1]

SPID

Solution

Zoom Ensemble Model



+



Input String: "TARGET BOSTON CENTRAL"

Possible Options: ['TARGET', 'WALMART', 'COSTCO']

Store Name: **TARGET**

Input String: "STARBUCKS CAMBRIDGE"

Possible Options: ['DUNKIN', 'STARBUCKS', 'PEET'S COFFEE']

Store Name: **STARBUCKS**

Input String: "WHOLE FOODS MARKET CHARLESTOWN"

Possible Options: ['WHOLE FOODS', 'TRADER JOE'S', 'STOP & SHOP']

Store Name: **WHOLE FOODS**

Input String: "BEST BUY FENWAY"

Possible Options: ['BEST BUY', 'APPLE STORE', 'MICRO CENTER']

Store Name: **BEST BUY**

Input String: "CVS PHARMACY BACK BAY"

Possible Options: ['CVS PHARMACY', 'WALGREENS', 'RITE AID']

Store Name: **CVS PHARMACY**

Solution

Result

“Using **LLMs to decide the best option** improves model ensemble performance, **but it is less effective** if asked to **create a brand name from scratch.**”

	Analytical Asset 1.0	Analytical Asset 2.0	Analytical Asset 3.0
%Accuracy in Loyalty Program Universe (N = 7.5k)*	77%	92%	96%
%Accuracy in Ground Truth Dataset (N = 32k)	76%	91%	95%
%Accuracy in Random Sample (N = 100)**	66%	76%	89%

* Brands with BCI Transactions: 276 out of 499 historically in the program

** Large-Sized Merchants

Agenda

01 Banco de Crédito e Inversiones (BCI)
What is BCI?

02 Diego Pavez
Who am I and what do I do?

03 Problem
The Need for Transactional Enrichment

04 Solution
The End-to-End Transactional Enrichment Process

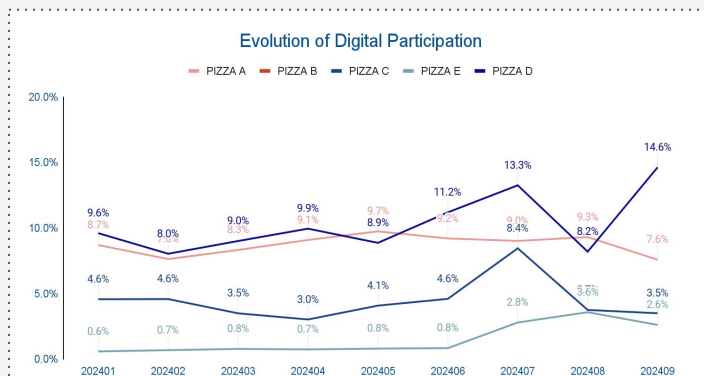
05 Actionable Insights
Benefits and Impact of Transactional Enrichment

06 Questions

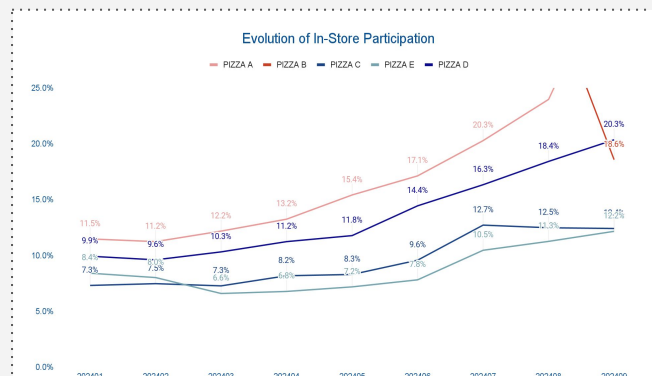
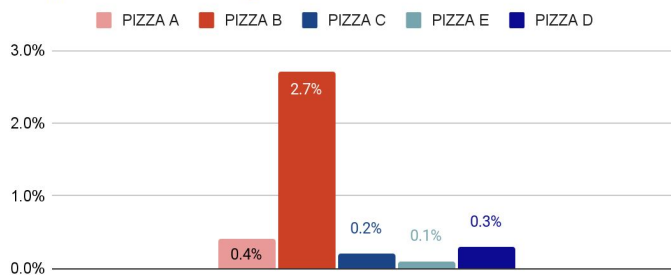
Actionable Insights

Enhancing the loyalty program offer

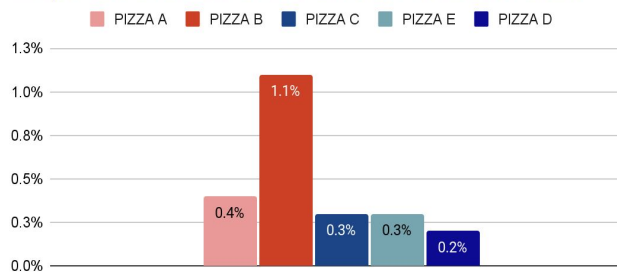
“With improved data, we can see the full picture and determine which merchants are best suited for our loyalty program”



Trx per Merchant - Digital Purchases vs Market Share



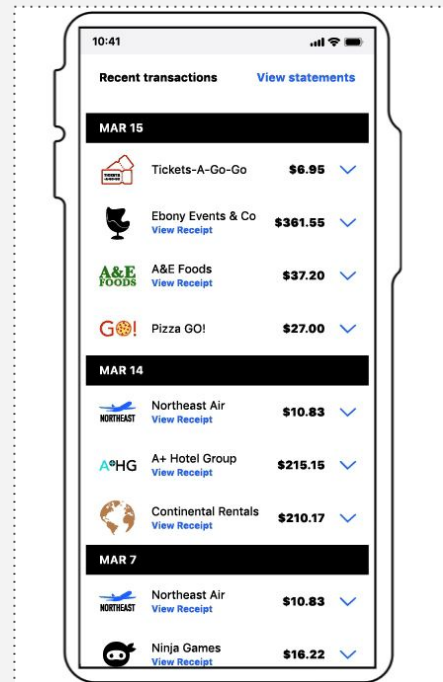
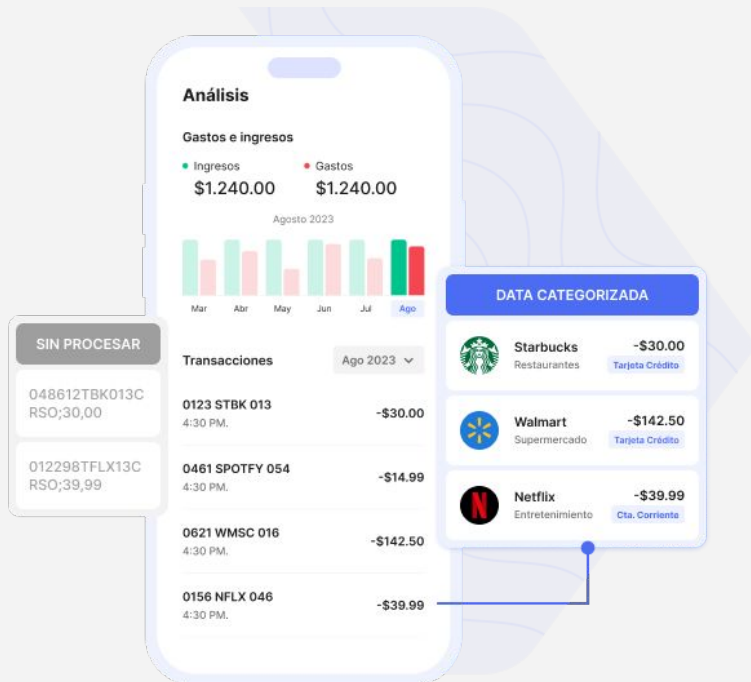
Trx per Merchant - In-Store Purchases vs Market Share



Actionable Insights

Improving the digital experience

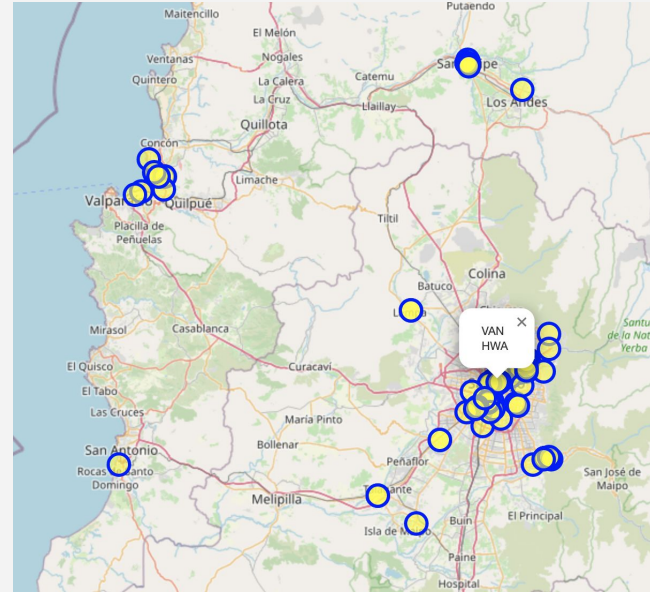
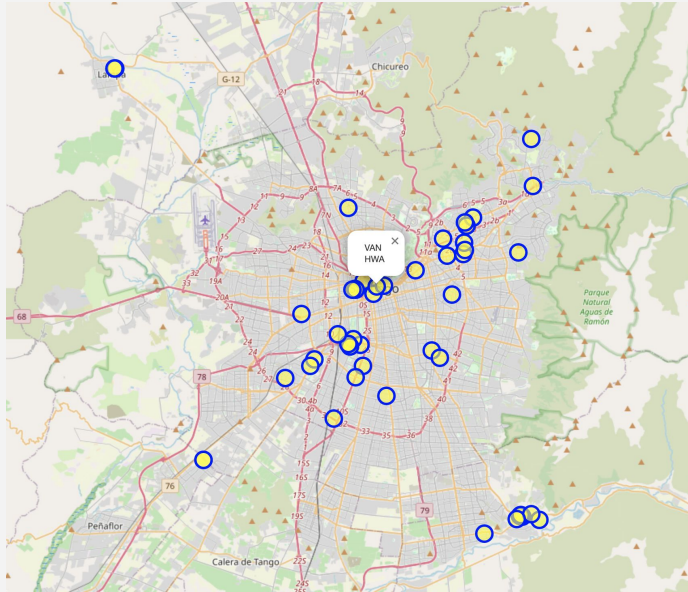
“Robust merchant name identification improving user experience by clarifying purchases, reducing confusion, and improving transaction understanding”



Actionable Insights

Gaining deep customer insights through highly granular geolocation data

“Understanding customer behavior helps us personalize their experience with the bank and create a more tailored ecosystem of benefits. It also enables us to offer merchants specialized banking products and proprietary payment solutions”



Next steps

Final thoughts

- *While large language models (LLMs) are extremely powerful, they are not trained to understand the specific context of most companies (aside from the largest and most well-known ones). For an LLM to significantly improve its accuracy and relevance within a particular business context, it is essential to fine-tune* the model with data and context tailored to the objectives at hand. Additionally, it is recommended to enhance the model with Retrieval-Augmented Generation (RAG) techniques to provide real-time, relevant context.*
- *Validations for these types of projects are a highly manual process. While we can semi-automate this by calculating similarities using various techniques (such as Dice, Levenshtein, among others), additional validation can be achieved through agents and specialized LLMs. This involves fine-tuning, prompt engineering, and context inclusion to validate brand choices or identify potential issues with improperly generated business names, ideally leveraging different base LLM models.*
- *Beyond traditional techniques for finding a transactional business name, LLMs open the door to a new approach to ensemble learning, with capabilities and orchestrations limited only by the user's imagination. This allows for combinations that can drastically enhance outcomes.*

Agenda

01 Banco de Crédito e Inversiones (BCI)
What is BCI?

02 Diego Pavez
Who am I and what do I do?

03 Problem
The Need for Transactional Enrichment

04 Solution
The End-to-End Transactional Enrichment Process

05 Actionable Insights
Benefits and Impact of Transactional Enrichment

06 Questions

Thanks

Transforming Transactional Enrichment with Traditional AI Models and LLM Ensemble for Enhanced Banking Insights

Nov 2024

