Generic BNPL Pty. Ltd.
Ranking Merchants for Onboarding

Business Model

Buy Now Pay Later

"Buy Now Pay Later services allow users to purchase a product or service at a shop or online and pay it back in **interest free instalments.** Customers who use BNPL services receive their items instantly and they make instalment repayments to the BNPL service provider, not the store that the goods are purchased from.

Payment Flexibility

Allow payments to be split into 5 installments interest-free.

Revenue

Commissions for transactions made using our payment solution

Industry Competitors

afterpay, Latitude Pay, openpay, Klarna., Zip, PayPal

Project Objectives

01 Develop a robust **ranking model** with insights to find top 100 ideal merchants.

02 Create an **automated ingestion pipeline** to extract datasets, apply business rules and output curated datasets for insights.

03 Determine **metrics and heuristics** that greatly separated merchants that should and shouldn't be accepted.

04 Incorporate **external datasets** such as Australian Bureau of Statistics (ABN) datasets to determine merchants' customer base characteristics.

05 Deliver **recommendations and insights** of merchants which drive profit and exposure to the firm.

Team Responsibilities

Xavier Travers
Data Engineer

Oliver Tan
Data Engineer

Tommy Cho
Data Analyst

Ke He
Data Analyst

Glendon Goh
Data Analyst

Project Timeline

Method of Communication

GitHub Kanban Discord

Estimated Hours of Work

~ **20 hours** / week

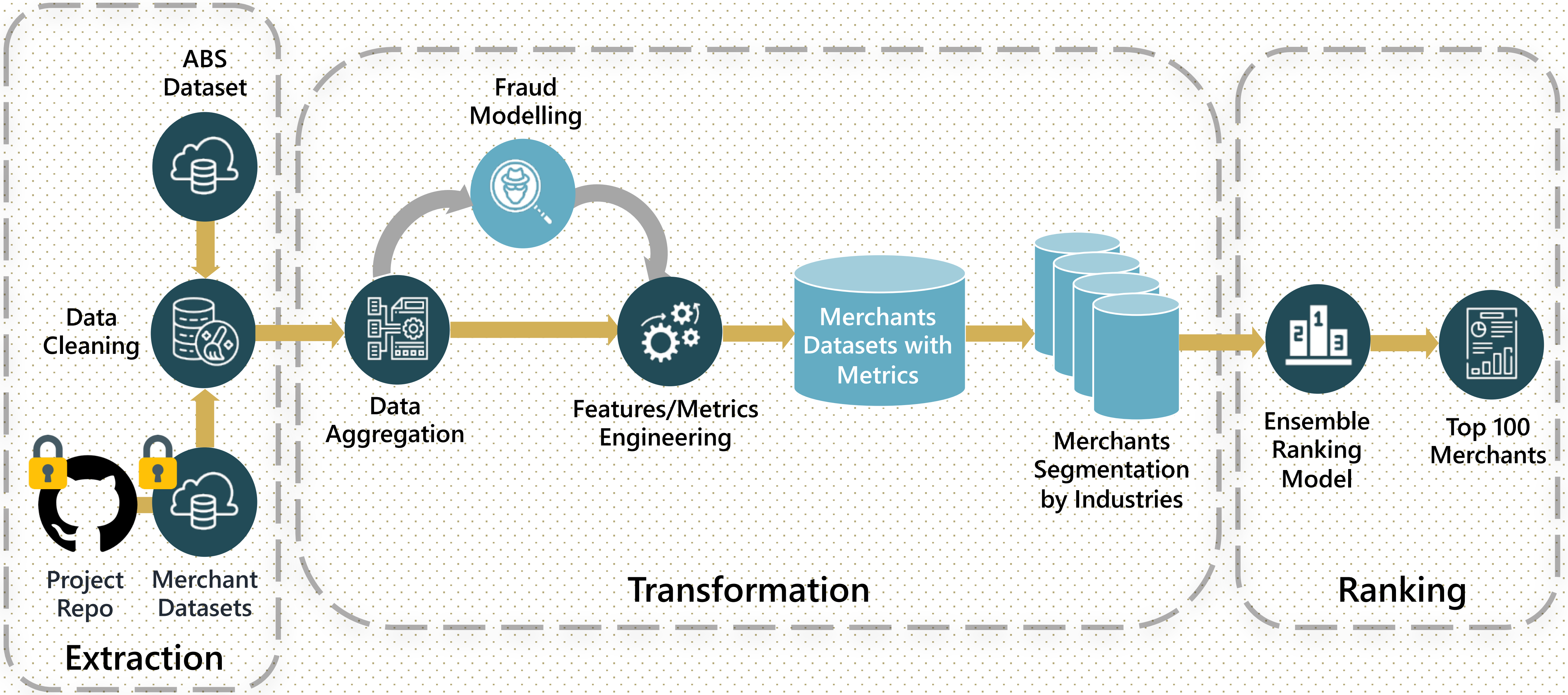
Week 1-2
- ETL Script Development
- In-depth Analysis

Week 3-4
- Fraud Analysis
- Metrics and Heuristics Research

Week 5-6
- Merchants Segmentation
- Metrics Experimentation
- Ranking System Development

Week 7
- Finalizing Ranking System
- Findings and Project
- Summary Consolidation

Preprocessing & Modelling Pipeline



Preprocessing – Issues and Resolutions

Highly Skewed Distribution of Dollar Value in Transaction Dataset

- Removed transactions outside $1-$2000 as they are **not reflective** of our **future transactions.**
- The range is comparable to our industry competitors (e.g. Afterpay maximum value/transaction is $1.5k)

Post Codes Discrepancies When Joining with ABS Dataset

- Reduce granularity by **grouping postcodes** based on **SA2 Codes.**
- Postcodes in Consumer dataset fully matched the external ABS postcode-SA2 dataset

Missing Information on Median Customer Weekly Income for Certain Postcodes

- Since there are only small amount of missing weekly income information, we decided to **impute** the missing values based on the **median weekly income of all consumers**.

Ideal Merchants

What characteristics separate an ideal and less ideal merchant?

Challenges Faced by BNPL Start-Ups

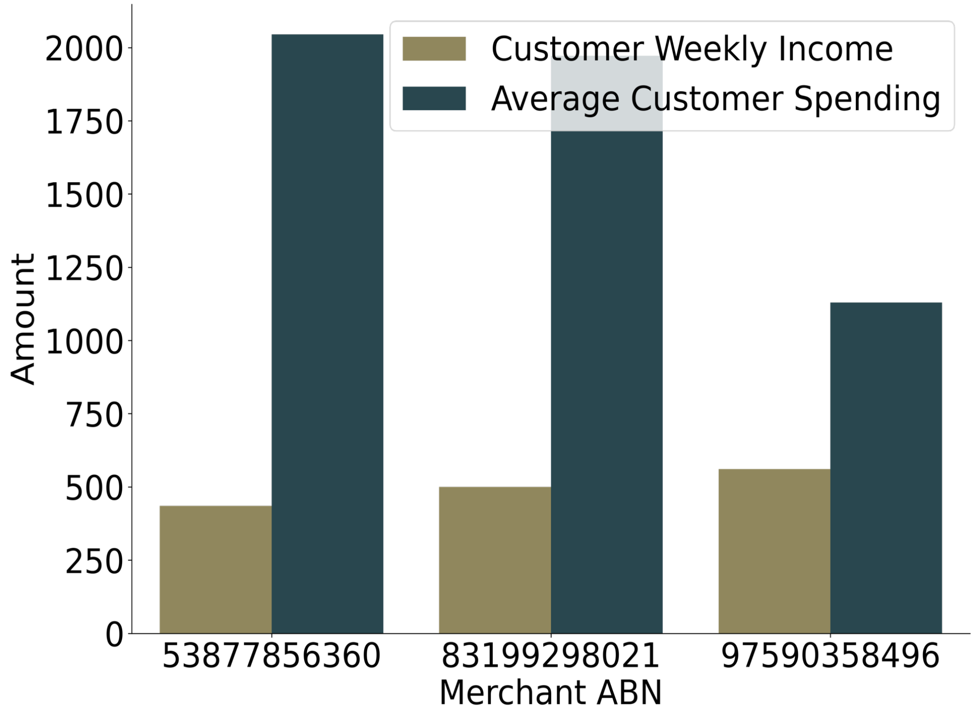
- Exposure & Brand Familiarity
- Customer Demographic
- Merchants’ Sales Performance
- Fraudulent Transactions

Customer Demographic

With flexible financing comes the increasing **risk of default payments** among BNPL consumers. Thus, we need to consider the **repay-ability of our users**. To gauge this information, we will look at the **consumer base** of the merchants.

Median Customer Weekly Income

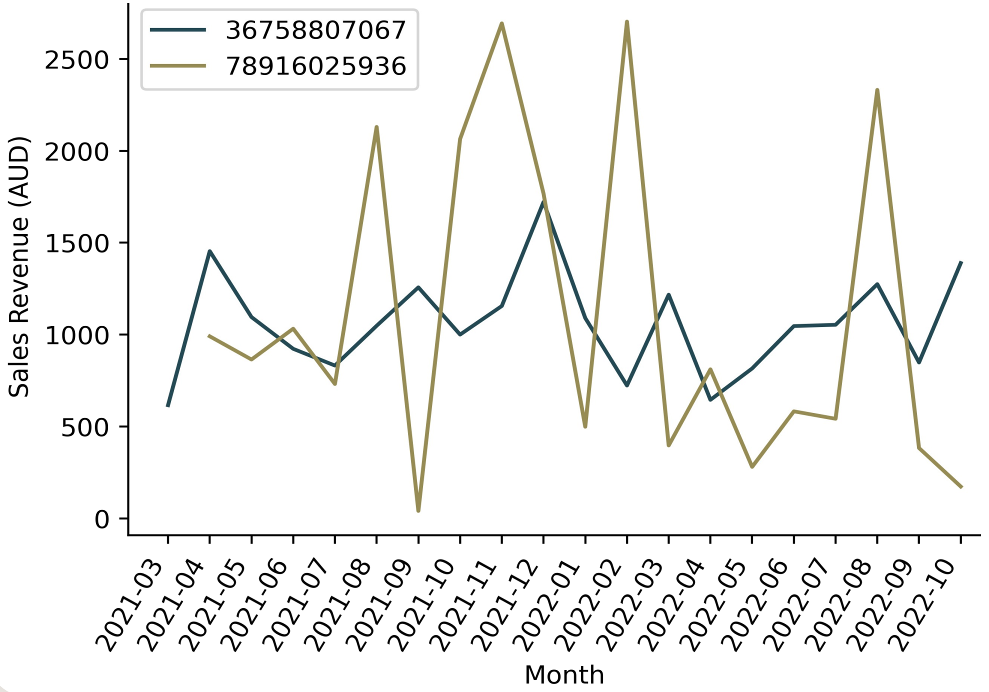
- Assuming customer's income approximates to the median weekly income by their postcode in the ABS dataset.
- From the plot, we see that the first two merchants have very **high average customer spending (~$2000)**, but their customer base median weekly income is around $400
- Taking their living cost into account, these customers **can barely afford** their weekly installment. (~$400/week)
- Higher income ➢ Lower risk of default payment



Merchants’ Sales Performance

With limited resources, we need to select merchants such that three forms of positive financial performance are maximized overall: **stability**, and **profit potential**. Ideally, we’re looking for merchants who maximize these metrics.

Monthly Revenue Between Two Merchants



Average Monthly Commission

- Measures our firm’s **profit potential** and prioritizes a consistent large monthly commission from merchants.
- Strongly correlated to other revenue metrics.

Average Monthly Orders

- Order frequency is an indicator for demand.

Standard Deviation of Monthly Revenue

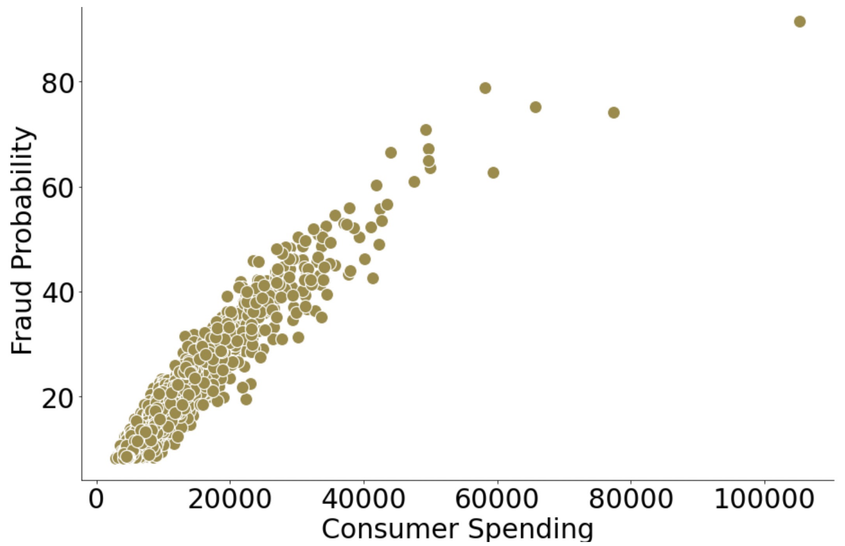
- Strong measure of **stability**.
- Between merchants with similar **profit potential**, the revenue **stability** is not necessarily the same.

Fraudulent Transactions

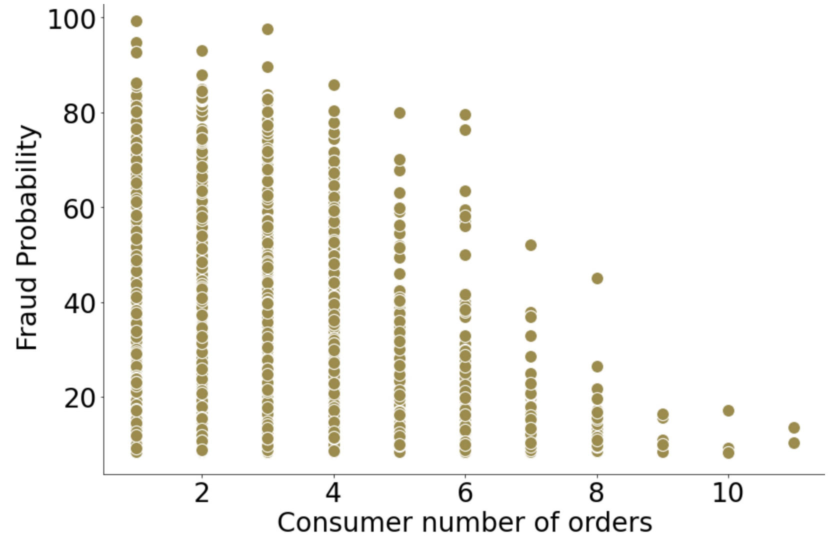
Insights & Findings

- Automation allows fraudsters to create tens of thousands of fraudulent consumer account with **stolen credentials**.
- These fake accounts can be monetized by making purchases without paying the installments, which incur **a loss to the BNPL firm.**
- Both daily consumer spending and number of orders by consumer show a **strong linear relationship** with **fraud probability**.
- Abnormally high consumer spending on one single day could indicate fraud.
- Consumer number of orders shows a negative correlation with fraud probability

Customer Daily Spending vs Fraud Rate



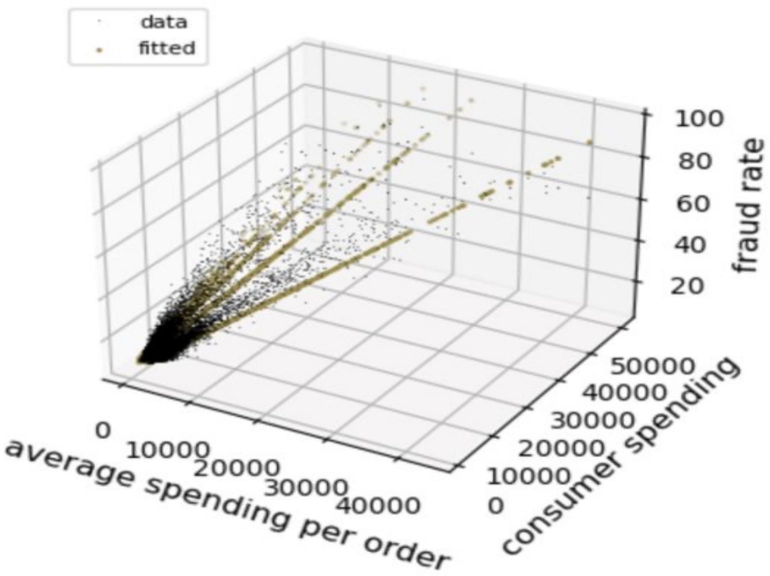
Customer Daily No. Order vs Fraud Rate



Fraudulent Transactions - Regression

Based on the linear relationship we found, a **linear regression model** is built to predict the fraud probability of a consumer on a given day. The model fits the data fairly well and **explains more than 80% of the fraud data**.

Model
Fraud Probability ~ consumer spending + average spending per order



average spending per order: the interaction term, equals consumer spending/consumer order

Customer Fraud Profile

- High daily spending
- High spending per order

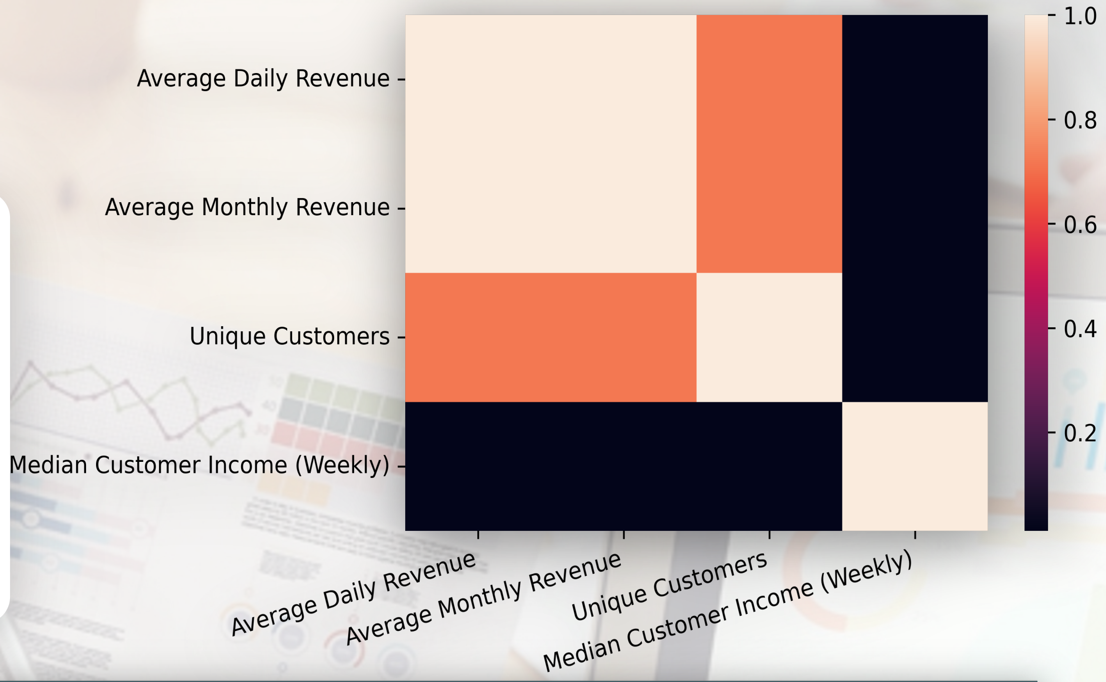
Metrics to Capture Fraud Info

- **Standard Deviation of Discounted Revenue:** captures the fluctuation of fraud rate discounted daily revenue
- **Overall Fraud Rate:** predicted overall proportion of fraudulent orders per merchant
- **Discounted average spending per order:** average spending per order discounted by predicted fraud rate

Metrics Summary

With over **27 different metrics** considered (e.g. Total orders, Total Revenue, Take Rate…), we narrowed down to the metrics we think are important.

We then conducted a **correlation analysis** and selected the metrics that are the least correlated to each other to **avoid overfitting** the data.



	Challenges Faced by BNPL			
	Exposure & Brand Familiarity	**Customer Demographic**	**Fraudulent Transactions**	**Merchants Sales Performance**
Final Chosen Metrics	• Number of sales region • Number of unique customers • Number of returning customers	• Median customer base weekly income	• Approximate fraud rate • Standard deviation daily discounted revenue • Discounted average value per order	• Average monthly commission • Average monthly order • Standard deviation monthly revenue

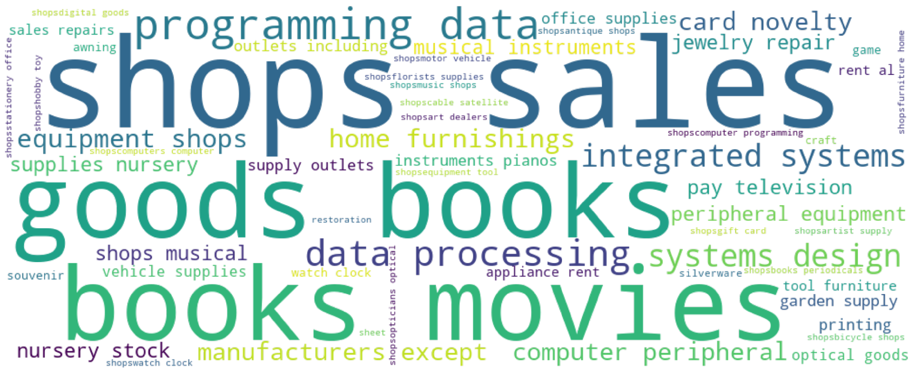
Merchants Segmentation

How do we segment merchants?
What characteristics does each segment generally have?

Segmentation by Merchants’ Tags

With over **50+ individual tags** in the merchants’ profile data, we need to segment the merchants and narrow down the segments that we want to focus on to have a meaningful analysis.

We managed to group these tags into **10 major industries.**

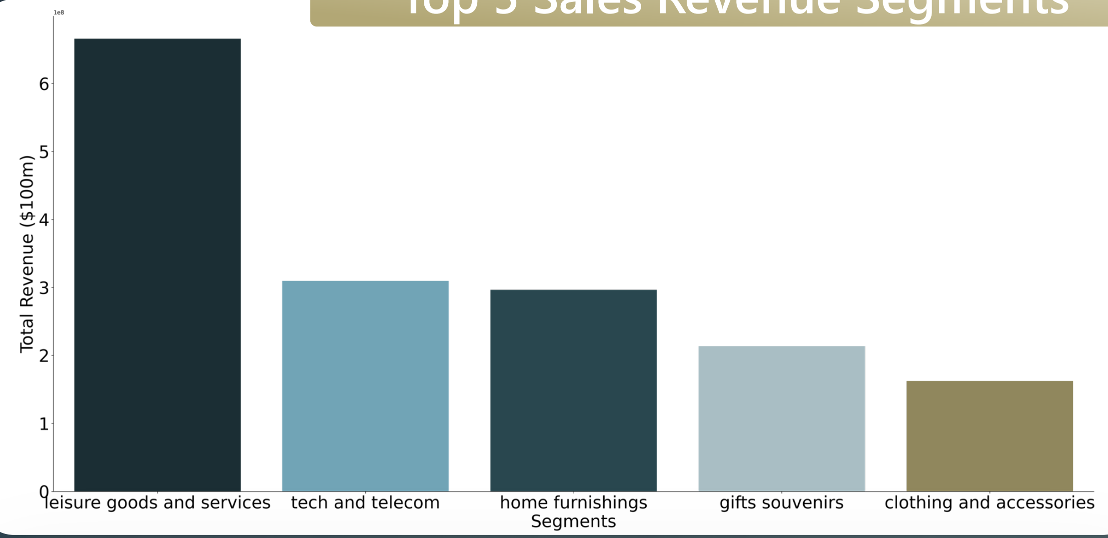


	Recession-Vulnerable						Recession-Proof			
	Luxury Goods	Leisure Goods	Home Furnishing	Gift & Souvenir	Clothing & Accessory	Office Equipment	Repair Services	Tech & Telecom	Motor & Bicycle	Health & Wellness
Example Tags under segment	Jewelry, Arts dealer gallery	Toy shop, Music, Movie	Furnitures, Lawn garden supply outlet	Gifts, Souvenir shops	Watch, Shoe shop	Office supply printing paper, Stationery	Repair, Restoration service	Computer, Data Processing, Cable	Motor vehicle supply part	Health beauty spa, Optician

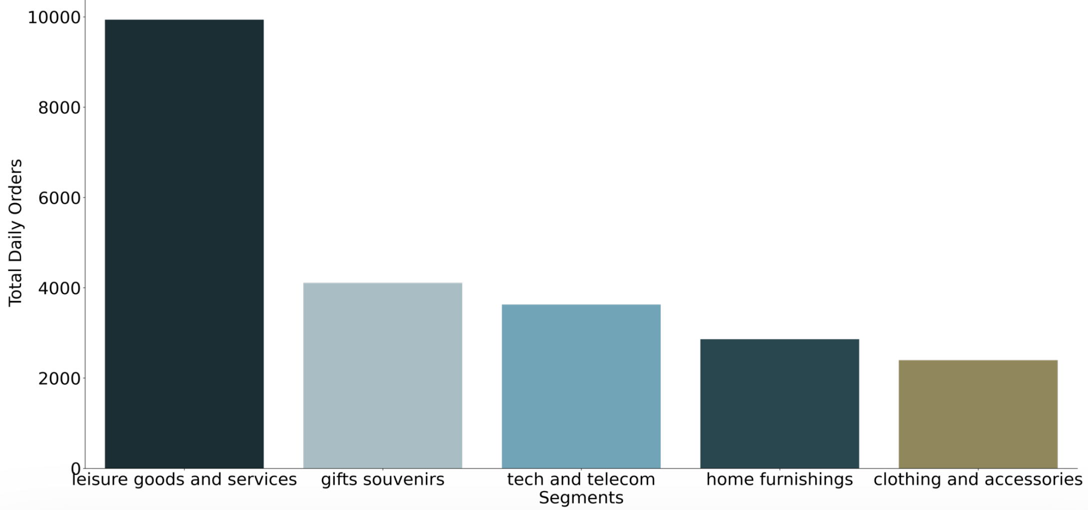
Credits: CNBC: Industries Hit Hardest by Recession, Forbes: What Industries Do Well In An L-Shaped Recession?

Industries Segments Analysis

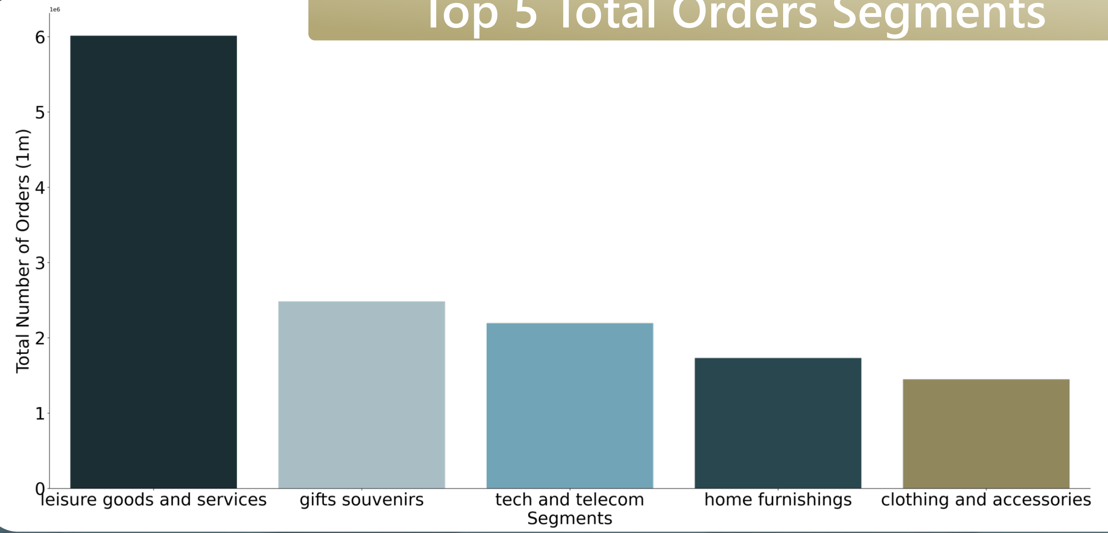
Top 5 Sales Revenue Segments



Top 5 Average Daily Orders Segments



Top 5 Total Orders Segments



Top 5 Sales Region Count Segment

Segments Selection Summary

	Recession-Vulnerable			Recession-Proof
	Leisure Goods	Home Furnishing	Gift & Souvenir	Tech & Telecom
Projected Annual Growth Rate (2022-2025)	• 16%	• 22.52%	• 6.7%	• **27%** for IT industry • **11%** for telecommunication industry.
Total Revenue ($)	• 660 mil	• 300 mil	• 210 mil	• 310 mil
Total Number of Orders	• 6 mil	• 1.7 mil	• 2.5 mil	• 2.2 mil
Average Number of Sales Region per Merchant	• 915	• 678	• 1087	• 872

Projected Annual Growth Rate (Last Accessed 9 Oct): Leisure Goods, Home Furnishings, Gift & Souvenir, Tech & Telecommunication

Ensemble Ranking System

How does it work?
Is it effective?

Average Normalized Feature Rank

Example Merchant	Feature 1 (LOWER IS BETTER)	Feature 2 (HIGHER IS BETTER)	Feature 1 (NORMALIZED)	Feature 2 (NORMALIZED)	Linear Combination	Final Rank
A	10	2	1	0.5	1.5	3
B	0	0	0	1	1	2
C	5	4	0.5	0	0.5	1
	1. Choose a subset of features and determine **ranking directions**.		2. Apply min-max **normalization**, accounting for ranking directions.		3. Calculate final rank from **linear combination** of normalized data.	

Normalization

- Makes the scale of features relative to each other the same (for even weighting).
- Preserves relative scale within the same feature.

Linear Combination

- **Flexible**: can adjust weights on features based on experience or input from executives.
- Initially evenly weights all features to remove the need for arbitrary weight generation.

Model Intuition and Benefits

The model evenly weighs the **opinion of several experts (metrics)** and ranks the merchants based on how well they excel on all metrics combined. This results in a selection of several jack-of-all-trades merchants.

Interpretability

Results are **highly interpretable**

Bias

Bias from individual metric is **greatly reduced**

Computation

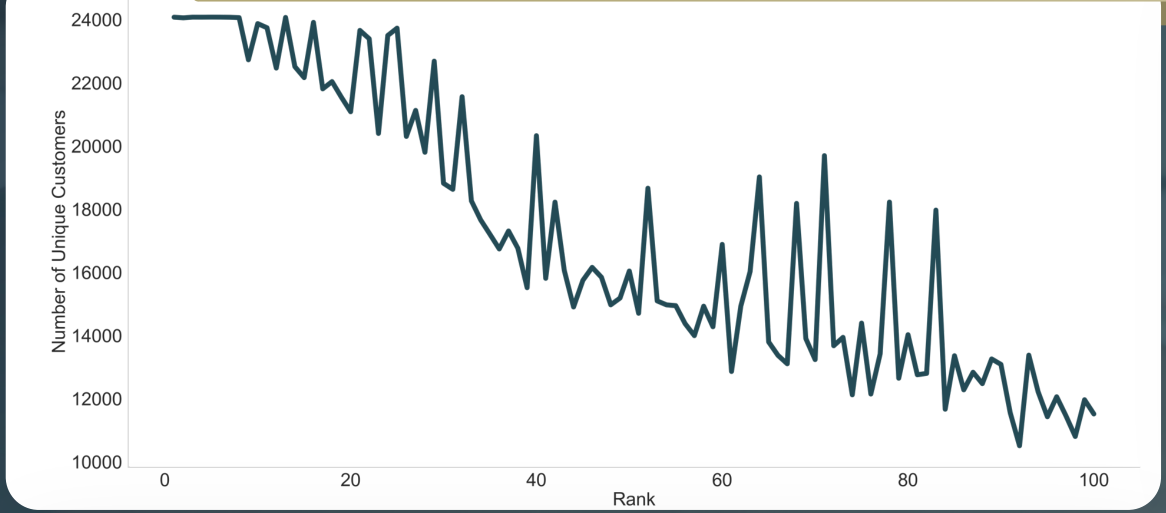
Computationally **cheap and quick** to output results

Flexibility

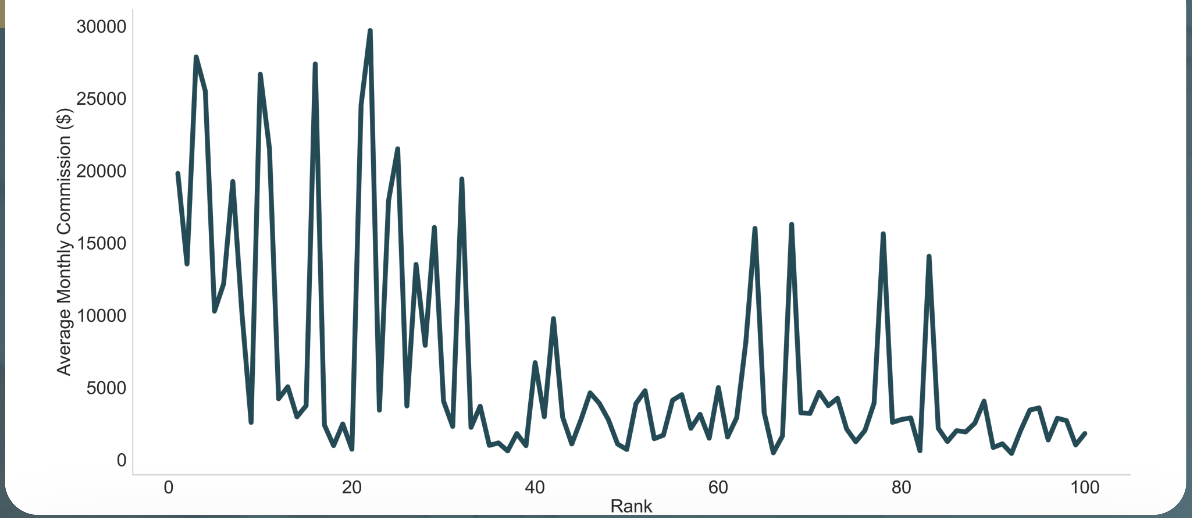
Weights can be **tailored** based on requirements from executives

Top 100 Merchants Results

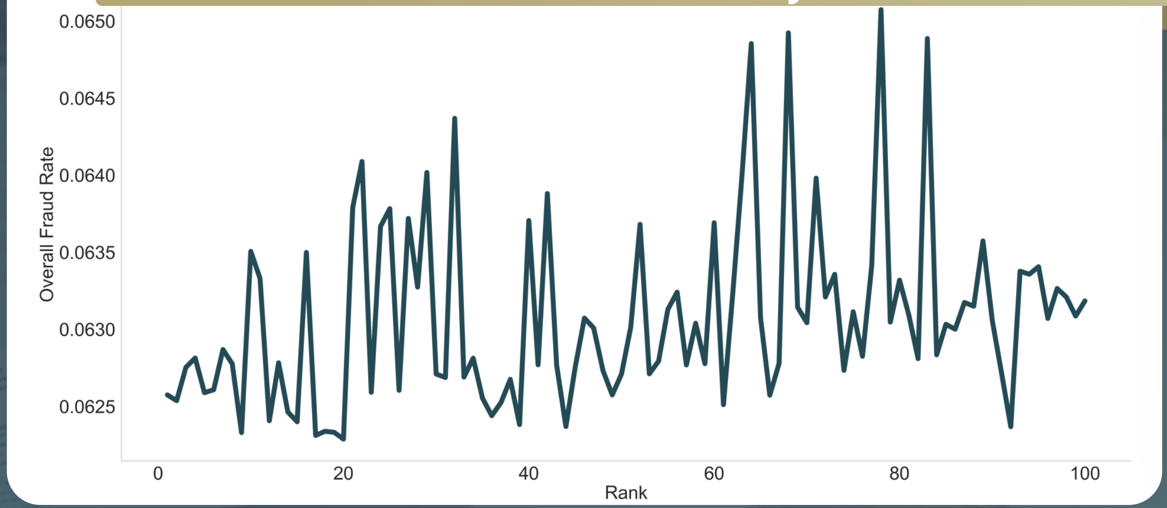
Number of Unique Customers by Rank



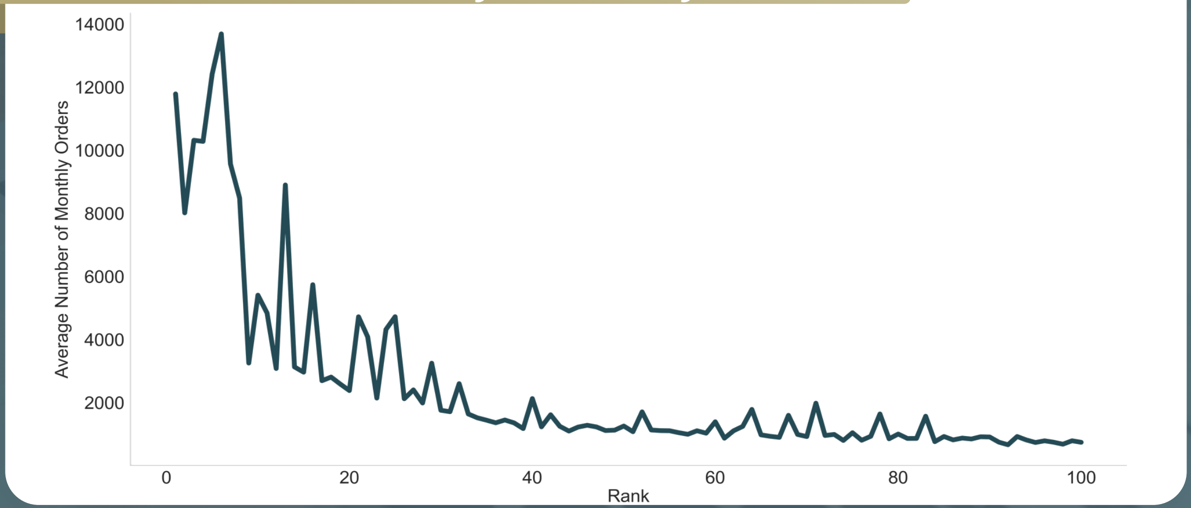
Average Monthly Commission by Rank



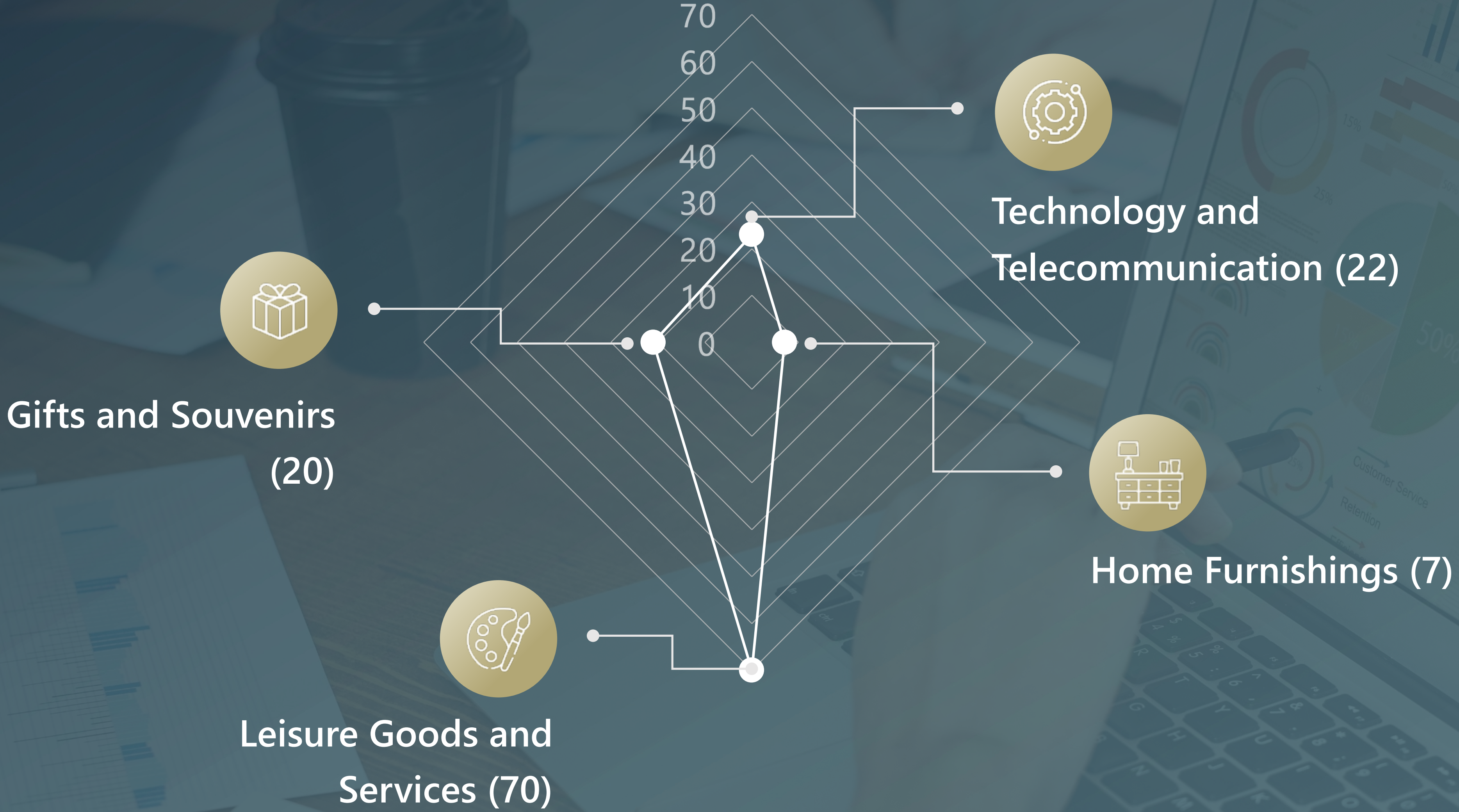
Overall Fraud Rate by Rank



Number of Monthly Order by Rank

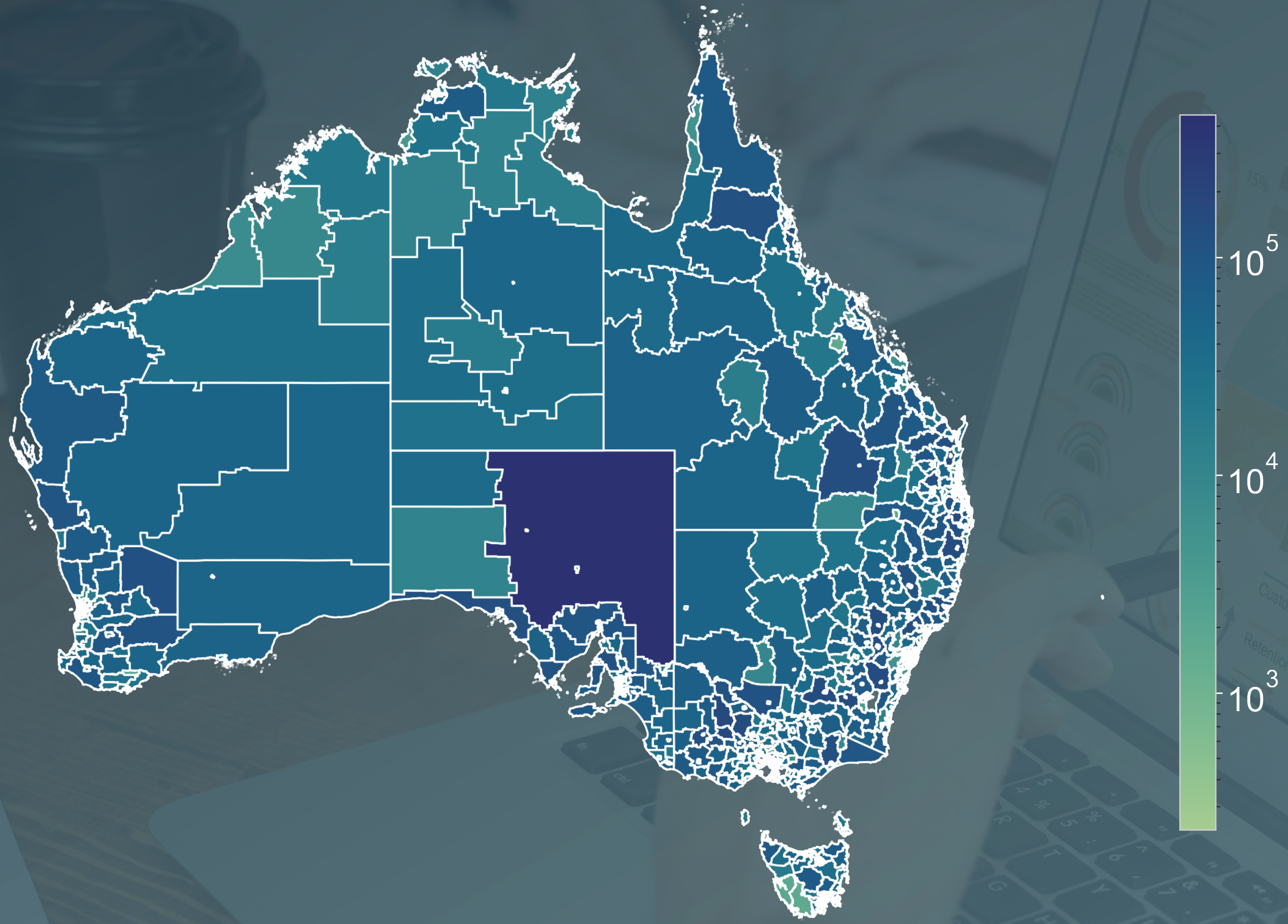


Industry Breakdown – Top 100 Merchants



Note: Some merchants have tags that belong to multiple segments

Potential Regions Reach – Top 100 Merchants



Total Sales Orders by Regions from the Top 100 Merchants' Based on Historical Transactions

Top 10 Merchants Per Segment

Merchants from certain segments will display features to solve different challenges our BNPL firm faces. The top 10 merchants per segment are therefore determined with weightings toward the segment's strengths.

Gifts & Souvenirs

Prioritize **financial performance** metrics.

Home & Furnishing

Prioritize **risk of default payment** metrics.

Leisure Goods

Tech & Telecom

Prioritize **exposure** metrics.

For each metric that we prioritize, double the weight (from 0.1 to 0.2). For non-priority metrics, evenly distribute the remaining weight (to sum to 1)

Top 10 Merchants Per Segment (with overall ranks for comparison)

Home Furnishing		Gifts/Souvenirs		Leisure		Tech/Telecom	
6	Erat Vitae LLP	3	Lacus Consulting	1	Non Vestibulum Industries	8	Placerat Eget Venenatis Limited
7	Lorem Ipsum Sodales Industries	6	Erat Vitae LLP	3	Lacus Consulting	16	Mauris Non Institute
39	Orci Corp.	7	Lorem Ipsum Sodales Industries	2	Vehicula Pellentesque Corporation	23	Vel Est Tempor LLP
40	Eget Laoreet Posuere PC	22	Orci In Consequat Corporation	4	Est Nunc Consulting	71	Amet Consulting
68	Purus Gravida Sagittis Ltd	21	Dictum Phasellus In Institute	5	Pede Nonummy Corp.	64	Eleifend PC
78	Eu Inc.	13	Ipsum Dolor Sit Corporation	13	Ipsum Dolor Sit Corporation	34	At Sem Corp.
83	Semper Corp.	25	Phasellus At Company	10	Lobortis Ultrices Company	42	Posuere Cubilia Curae Corporation
103	A Scelerisque Foundation	27	Faucibus Leo Ltd	11	Nullam Consulting	43	Suspendisse Incorporated
132	Volutpat Nulla Incorporated	24	Ultricies Dignissim LLP	21	Dictum Phasellus In Institute	45	Quisque Fringilla Limited
148	Cursus Non Egestas Foundation	39	Orci Corp.	25	Phasellus At Company	278	Arcu Sed Eu Incorporated

Limitations and Future Works

What are the assumptions, limitations and possible future works?

Assumptions, Limitations & Future Works

Assumptions

- Features to maximize in merchants are assumed to prioritize the **prosperity and exposure** of the BNPL company.
- Commission calculations assume **100%** of merchants' consumers use our BNPL platform for payments.

Limitations

- Our ranking does not account for pillars of **sustainability** (no focus on social or environmental sustainability in metrics).
- Lack of ground truth ranking labels limits the set of approaches in ranking merchants.

Future Works

- Which metrics to be **weighted higher** should be discussed prior to running the ranking algorithm with the business partners or based on company's future direction.
- Collect **actual commissions earned** based on these selected merchants and analyze which metrics have high impact on the actual commissions earned.

Thank You

Check out curated python notebook for details on ETL and Modelling Processes