

Buy Now Pay Later Project

Group 15

Our team

Keigo Warashina



Ider Byambadorj



Tugsgerel Lkhagvasuren



Beibei zhu



zihan Xu



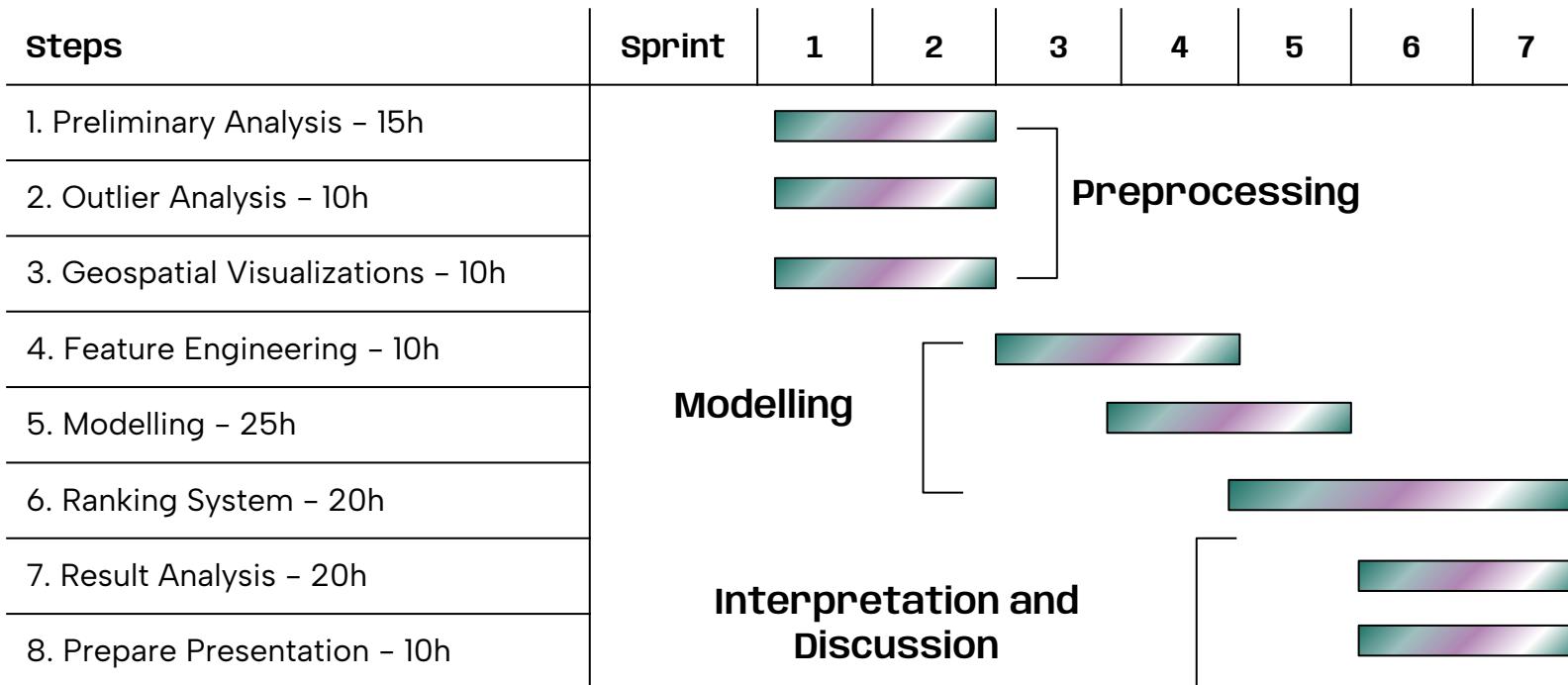
Project Recap

The Buy Now Pay Later (BNPL) Firm has begun offering a new “Pay in 5 Installments” feature and is going to onboard 100 merchants every year.

This project focuses on these tasks:

1. Assess the given and external datasets
2. Construct models to predict merchant's features
3. Find best 100 merchants to trade with

Project Timeline



Security and Privacy

Storage

All data is stored in a private Github repository

Access

Only group members accessed the data from their local, privately secured devices

Privacy

All data is confidential. No API keys were utilized.



Table of contents

01 Datasets

02 Modelling

03 Ranking system

04 Result & Recommendation

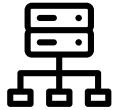


01

Datasets



Datasets



Transactions data

≈14,000,000 instances of transaction between merchants and their consumers, from Feb 2021 to Aug 2022



Merchant details

Total of ≈4,000 merchants, all having different take rate.



consumer details

≈500,000 consumers were registered. Attributes include consumers' postcode and gender.



Fraud data

≈4,000 and ≈80,000 data points indicating merchant and consumer fraud probabilities.



External datasets



Demographic data

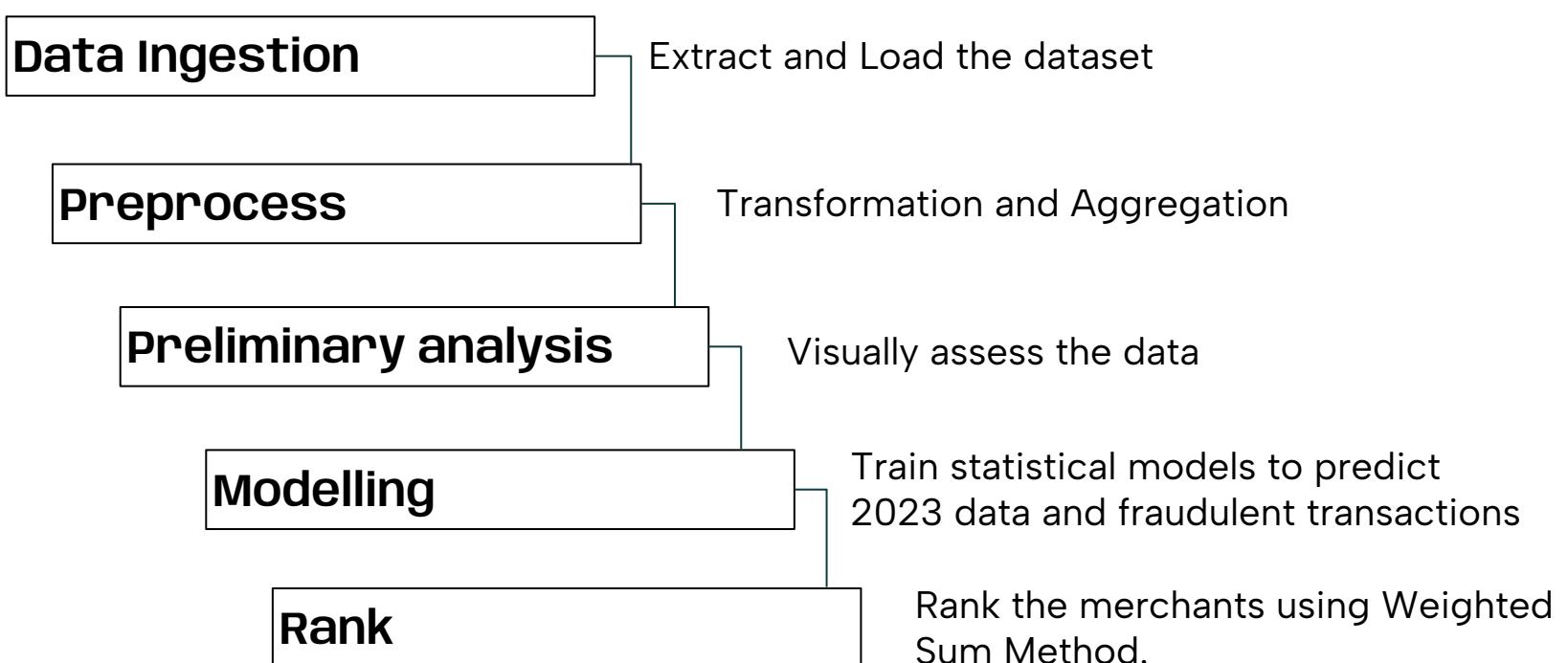
- Population
- Population density
- Median age



Income data

- Median income
- Sum income

Data Pipeline

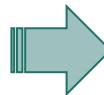


Preprocess: Missing values analysis

No details found for ≈ 400 merchants out of 4000 merchants

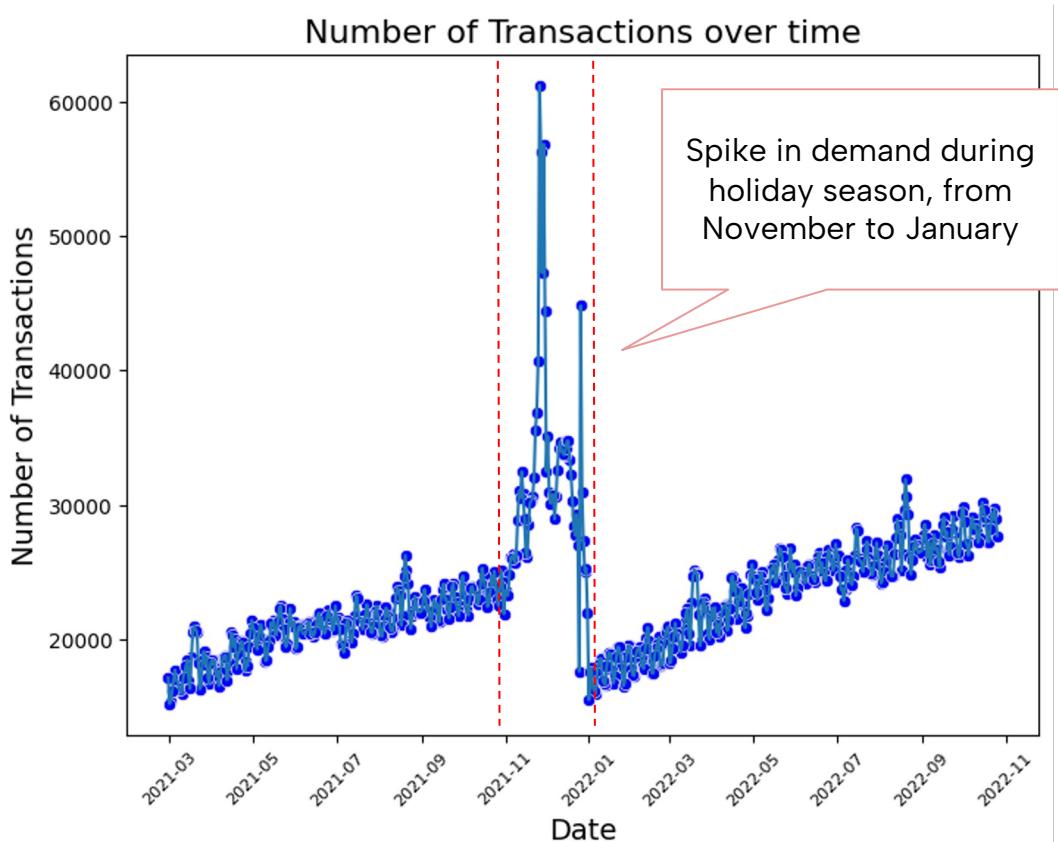


Likely to be not our clients



Removed those merchants and their corresponding transactions

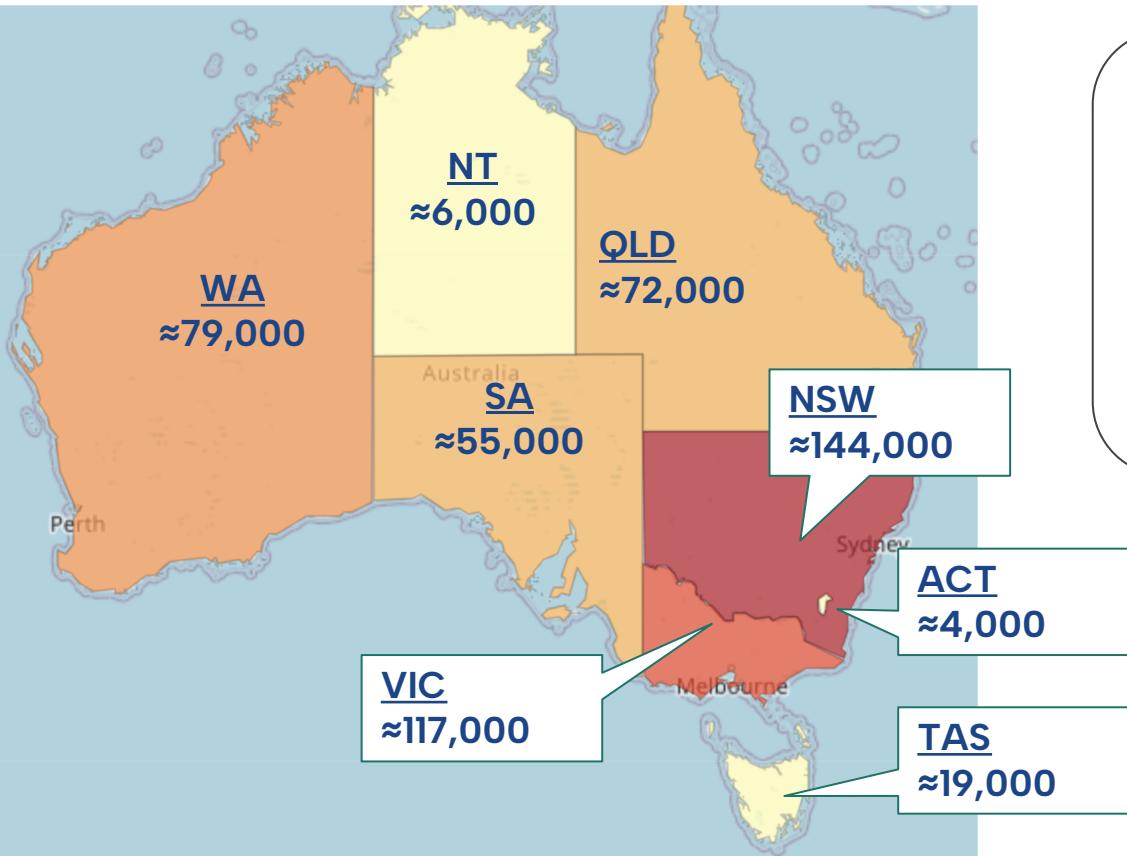
Preliminary analysis: transaction data



Observations:

1. Similar trend each year
1. High volume of transactions during Christmas and the Summer Holiday

Geospatial Analysis: Number of consumers by state



Observations:

- New South Wales and Victoria have the highest number of consumers
- Potential benefit to target these two states



Models 02

Models



Forecast model (linear model)



To predict:

- 2023 revenue
- 2023 # consumers
- 2023 # transactions,

for each merchant

(we only have 2021/02~2022/10 dataset)



Fraud detection model (random forest)



To predict:

- Merchant fraud
- Consumer fraud

for all the past transactions

from transaction history with fraud probabilities

Assumptions & Limitations

Assumptions

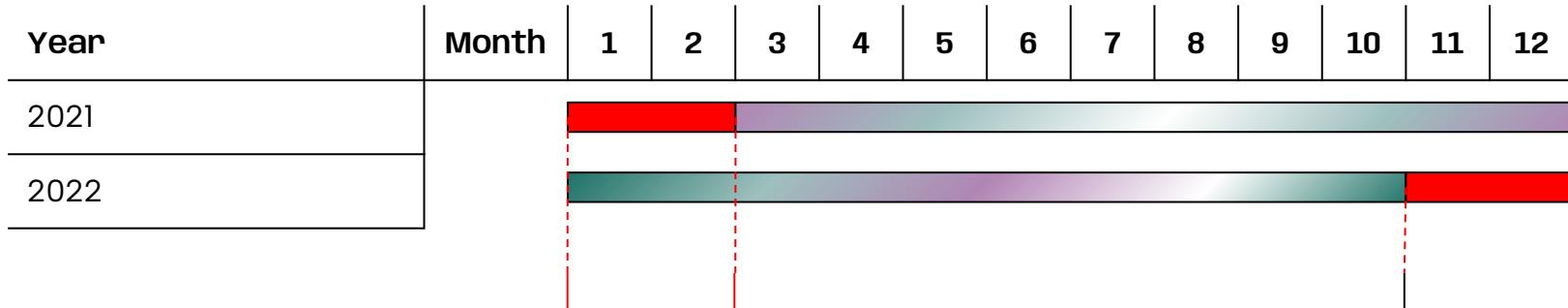
- Transactions follow similar pattern each year
 - Distribution of the fraud dataset is representative of the whole transaction dataset
-

Limitations

- Limited amount of transaction and fraud data
- Conversion between SA2 \leftrightarrow postcode were approximated, therefore external features are also an approximate

1. Forecast model: Missing Data Imputation

Linear model requires two years of data for training and testing, but the data for the first 2 months of 2021 and last 2 months of 2022 are missing



As there is no prior data, these 2 months are imputed by the average of the next 3 months to preserve trends of the non-holiday data

Last 2 months are imputed with time series model (ARIMA)

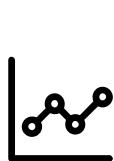
1. Forecast model: choice of model

LINEAR MODEL



- ✿ In contrast to time series model, it captures the effect of external features on the target variables
- ✿ Computationally fast
- ✿ Easy to interpret

1. Forecast model: Linear Model



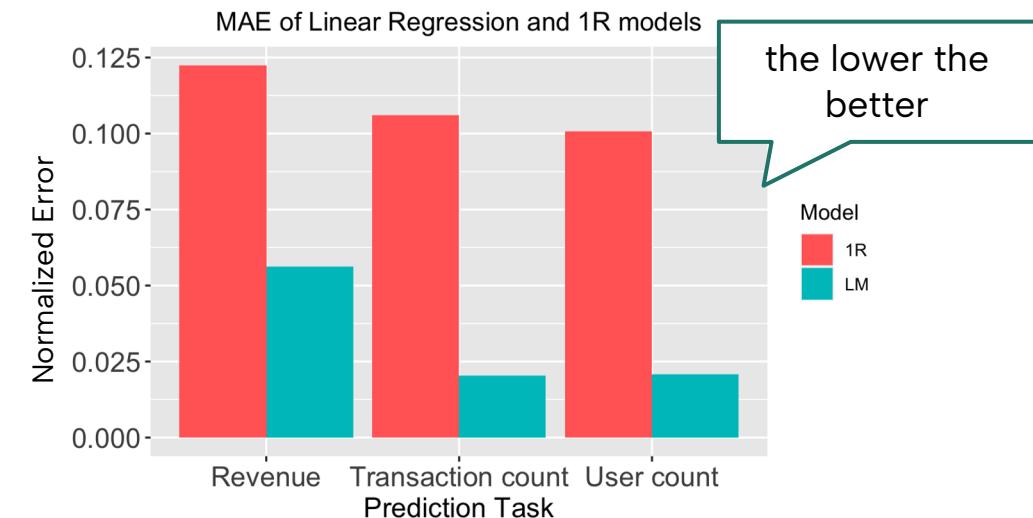
For each merchant, predict the total revenue, number of transactions and consumers during 2023, based on past data

* Due to lack of training data, time series model was not utilized to predict the 2023 data (12 months ahead)

Feature set:

Category of merchant

- Take rate
- Revenue level of merchant
- Demographic / Income features of the consumer base



2. Fraud detection model: Random Forest classifier



Classify whether the transaction is suspicious or not

* Suspicious transaction is defined to be the one with fraud probability over 50% to ensure sufficient sample size with true (suspicious) labels

Feature set:

- Datetime
- Merchant details
- Consumer details
- Demographic / Income features of consumer

	Precision	Recall
<u>Merchant</u> fraud detection model	0.97	0.64
<u>Consumer</u> fraud detection model	0.87	0.79

Precision is more important than accuracy and recall

03

Ranking system

what we found so far



Heuristics 1/2: Performance measures



1. Predicted Gain

Predicted Gain =
take rate \times predicted revenue



Measure of profitability. Most important measure.



2. Predicted number of consumers



Measure of future potential. Also relates to usage rate of BNPL service.



3. Predicted number of transactions



Measure of demand. Frequent transactions imply that there is a high demand towards the merchant's business.

Heuristics 2/2: Safety measures



4. Reliability of the merchant



How reliable is the merchant concerning fraudulent transactions.



5. Reliability of consumer base



How reliable is the consumer base concerning fraudulent transactions.

$$* \text{Reliability} = 1 - \frac{\# \text{ of suspicious transactions by a merchant or a consumer}}{\# \text{ of all transactions by a merchant or a consumer}}$$

Our ranking system



Steps:

1. Normalise each feature values
2. Calculate the weighted sum for each merchant
3. Sort merchants by the weighted sum



Merchant rating = Σ weight \times normalized features

Distribution of weights

40%



25%



25%



5%



5%



* Order of priority was determined upon our research,
weight values can be adjusted with specific needs or business goals

04

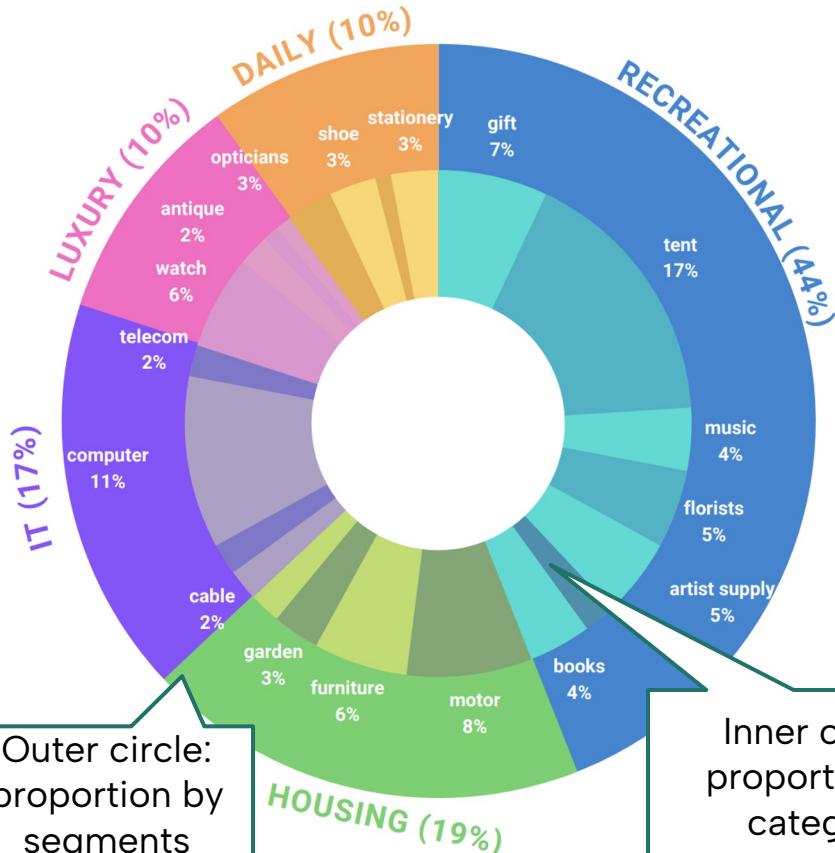
Result & Recommendation



Results: top 10 out of 100 selected merchants

rank	merchant name	segment	merchant rating
1	Leo In Consulting	luxury	0.95
2	Lacus Consulting	recreational	0.87
3	Est Nunc Consulting	recreational	0.82
4	Non Vestibulum Industries	recreational	0.79
5	Erat Vitae LLP	recreational	0.76
6	Lorem Ipsum Sodales Industries	recreational	0.71
7	Pede Nonummy Corp.	recreational	0.69
8	Mauris Non Institute	IT	0.68
9	Lobortis Ultrices Company	recreational	0.66
10	Orci In Consequat Corporation	recreational	0.65

Top 100 merchants: by Segments and Categories



Outer circle:
proportion by
segments

Inner circle:
proportion by
category

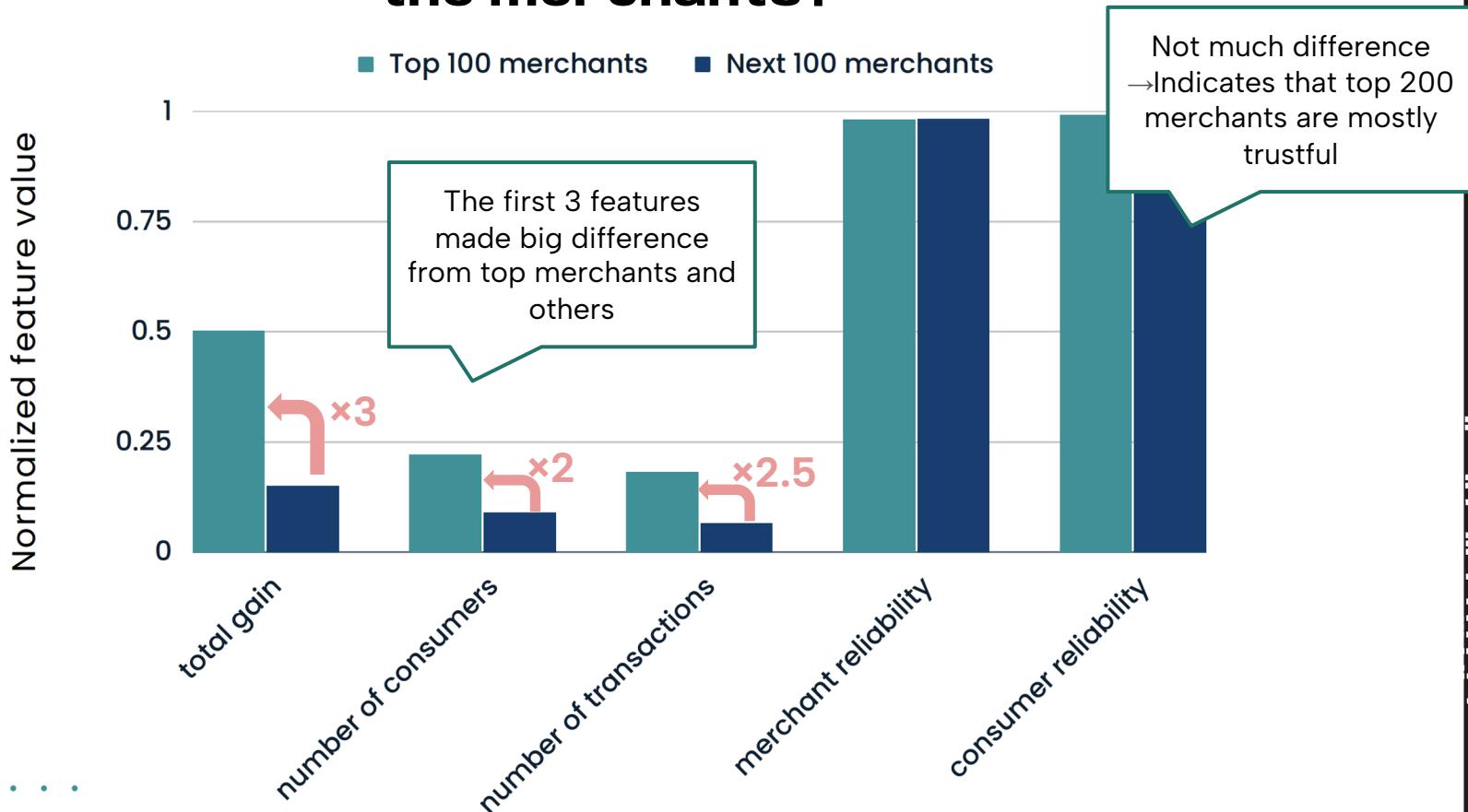
Top Segments:

1. Recreation (44%): The most common segment in the top 100 merchants
2. Housing (19%): The second most common segment, frequent transactions are being made

Top categories:

1. Tent (17%): large amount of transactions
2. Computer (11%): high average gain for single transaction

To what extent did the features differentiate the merchants?



Top 100 merchants under our solution ...

Predicted total gain for next year:

\$ 23,000,000

Predicted total number of consumers:

> 24,000

Predicted total number of transactions:

> 3,700,000

With all merchants having reliability score greater than 0.99.

Worth pursuing?

Multi-criteria ranking system

Rather than using single measure, our ranking system takes into consideration about multiple important features for the company's sustainable growth and safety.



Feasibility

Both the forecast model and the fraud detection model runs under feasible amount of time. (i.e. < 15 minutes)



Ease of customization

Depending on the company goals / market situation, the feature weights are easily customizable.

Conclusion



- Choice of merchants is crucial for the growth of the firm
- Our solution provides a framework indicating merchants with greatest profitability and safety
- The models can be further refined upon the accumulation of data over years



Thanks!

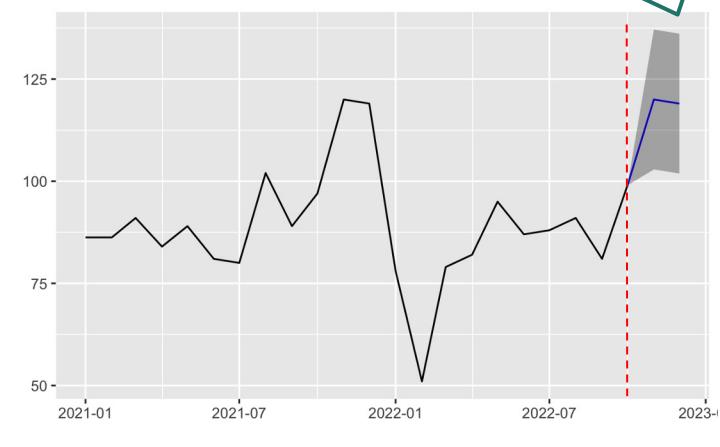
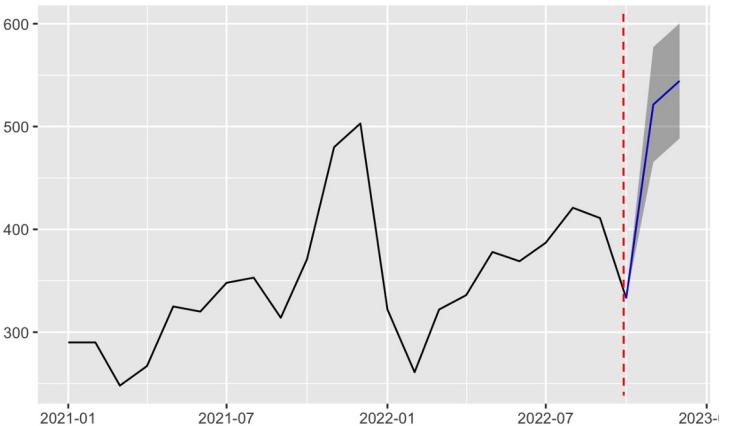
Do you have any questions?

Appendix 1: Merchant segments and categories

	Segment	Category
1	Recreational	Hobby, books, tent, gift, music, florists, artist supply
2	Housing	Equipment, furniture, garden, bicycle, motor
3	IT	Computer, telecom, cable, digital
4	Luxury	Antique, art, jewellery, watch
5	Daily	Health, shoe, stationery, opticians

Appendix 2: missing months imputation with ARIMA Model

Ex: total user count imputation for 2 merchants



Data from
2022/11~12 has
been imputed

Appendix 3: Linear Model performance

Model	MAE	Adj R-squared	P-value
Revenue prediction	17606.85	0.9975	< 2.2e-16 ***
User count prediction	38.76	0.9999	< 2.2e-16 ***
Transaction count prediction	40.25	0.9999	< 2.2e-16 ***