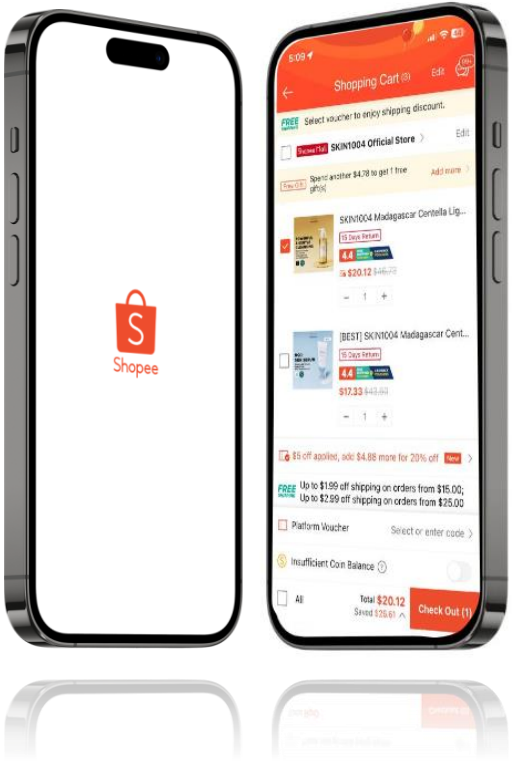
Team 6 Seminar Group 4 | Apr 2024

From Enigma to Engagement: Optimising Customer Retention Strategies for Shopee Thailand

Prepared by:

Andy Chan Yan Meng
Dhiya’ Diyana Binte Irwan
Gregory Goh Zong Jun
Lee Fang Hui, Jesslyn
Tan Yu Xiang (Javier)

BC2407 Analytics II Presentation



Our Agenda

1. Business Understanding
2. Proposed Solution
3. Data Preparation
4. Data Understanding
5. Data Modelling
6. Deployment of Solution
7. Evaluation of Solution

Dhiya' Diyana

Business Understanding

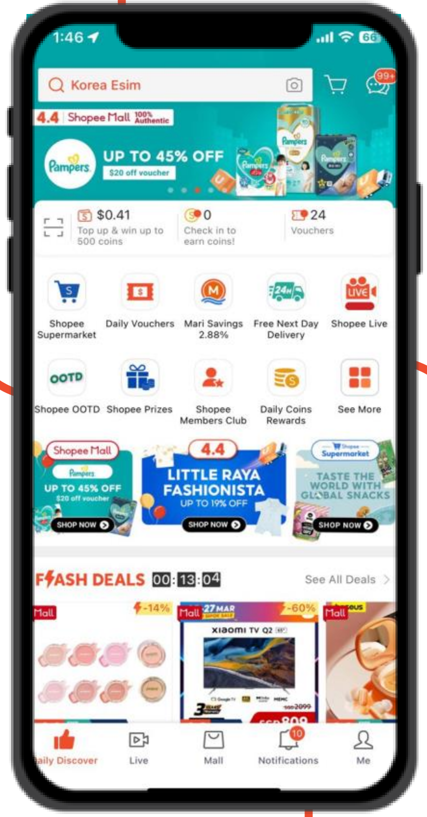
Dhiya' Diyana

1 Shopee Background

- First launched in 2015
- Today, Shopee is one of the prominent players in the e-commerce industry

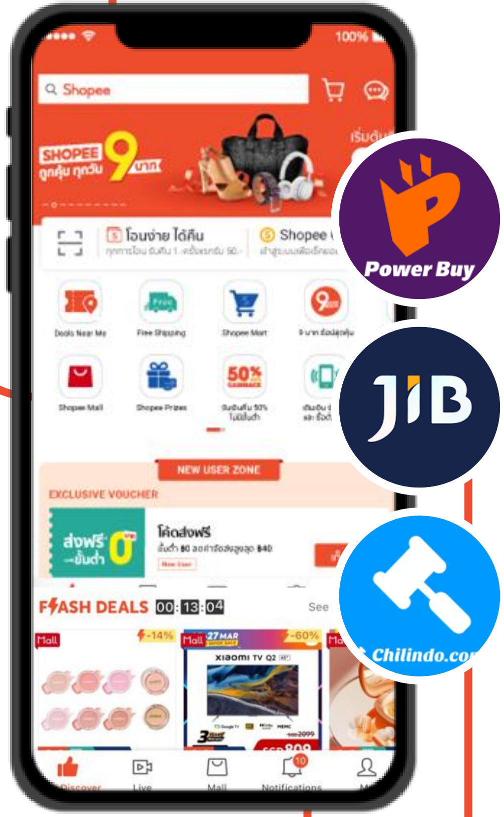
2 Southeast Asia (SEA) E-commerce Landscape

- Southeast Asia's e-commerce market boasts a population exceeding 600 million and a combined GDP of 3 trillion USD (Jaouadi & Chuidian, 2023)



Shopee

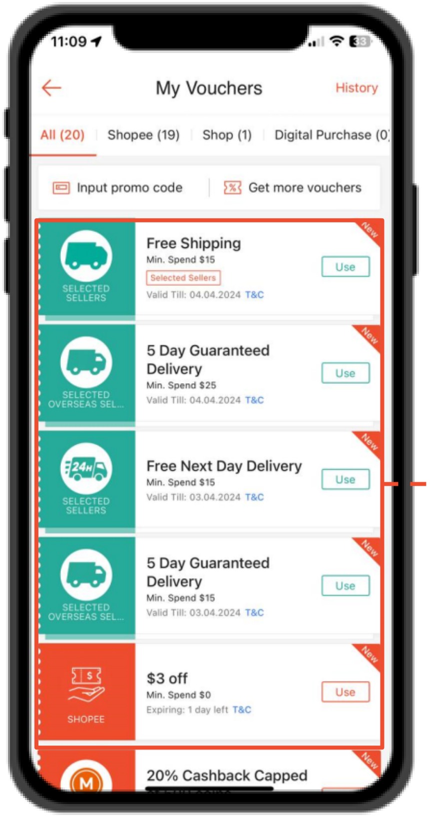
Dhiya’ Diyana



3 Business Problem

- Shopee’s customer retention in Thailand: low point of 37%
- Emergence of local e-commerce platforms offering niche Thai products dilutes Shopee’s market share
- 5% increase in customer retention can lead to a remarkable 95% boost in profits
- Shopee is paying 5-7 times more – in acquiring new customers compared to retaining existing ones
- Reduced customer lifetime value = negative impact on revenue and market competitiveness

Dhiya' Diyana



4 Current Strategies & Opportunity

- Broad, **one-size-fits-all** incentives strategy
- **Dilutes marketing efforts and financial resources** as rewards given to users who would have made purchases without incentives
- Lack of targeted incentives wastes resources and misses retention opportunities

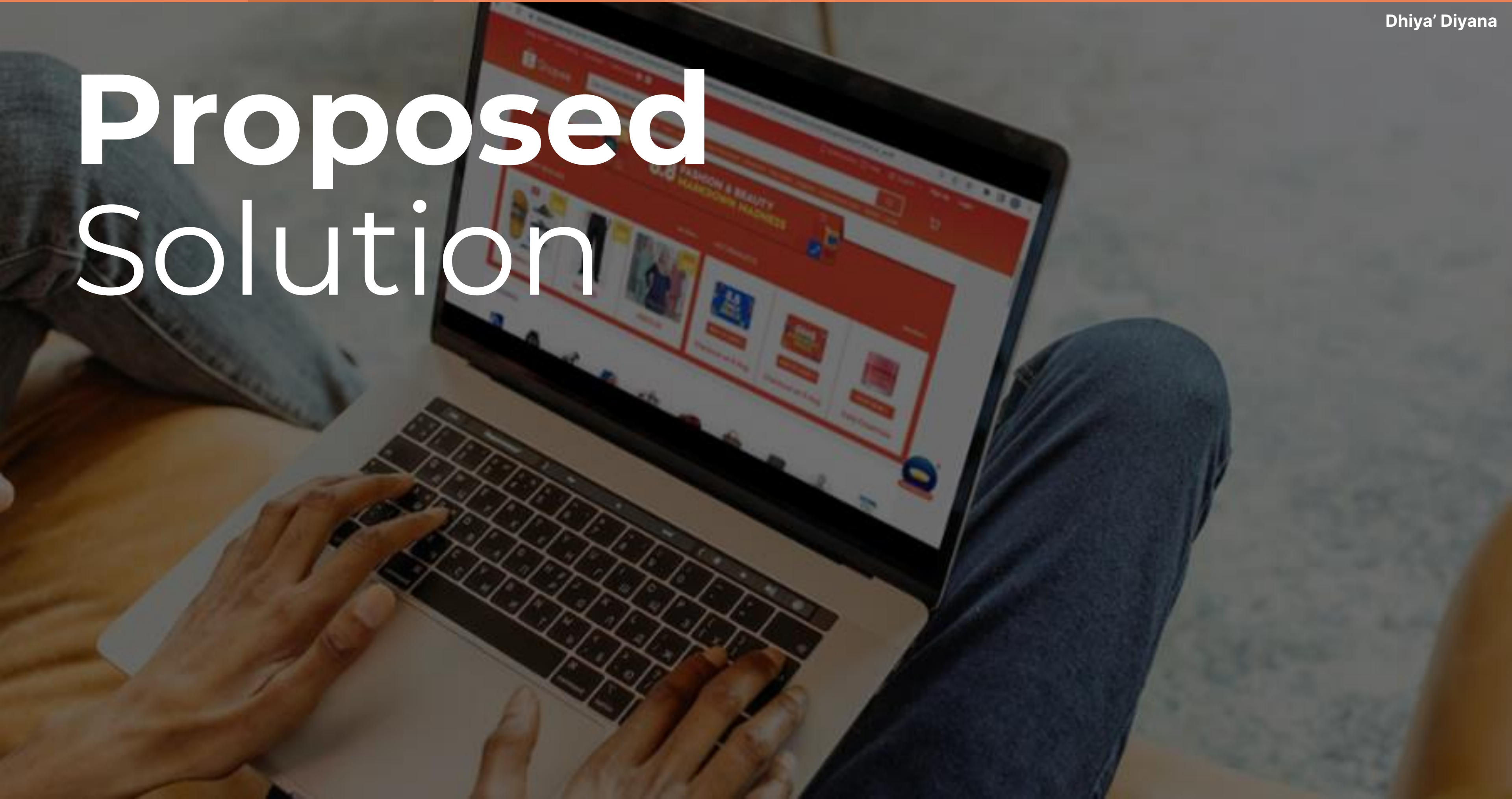
Opportunity: To **address and retain the segment truly at risk of disengagement and churn** to foster long-term retention

Problem Statement

“ How can we leverage predictive analytics and machine learning to optimise customer retention strategies for Shopee Thailand to improve customer retention and improve revenue? ”

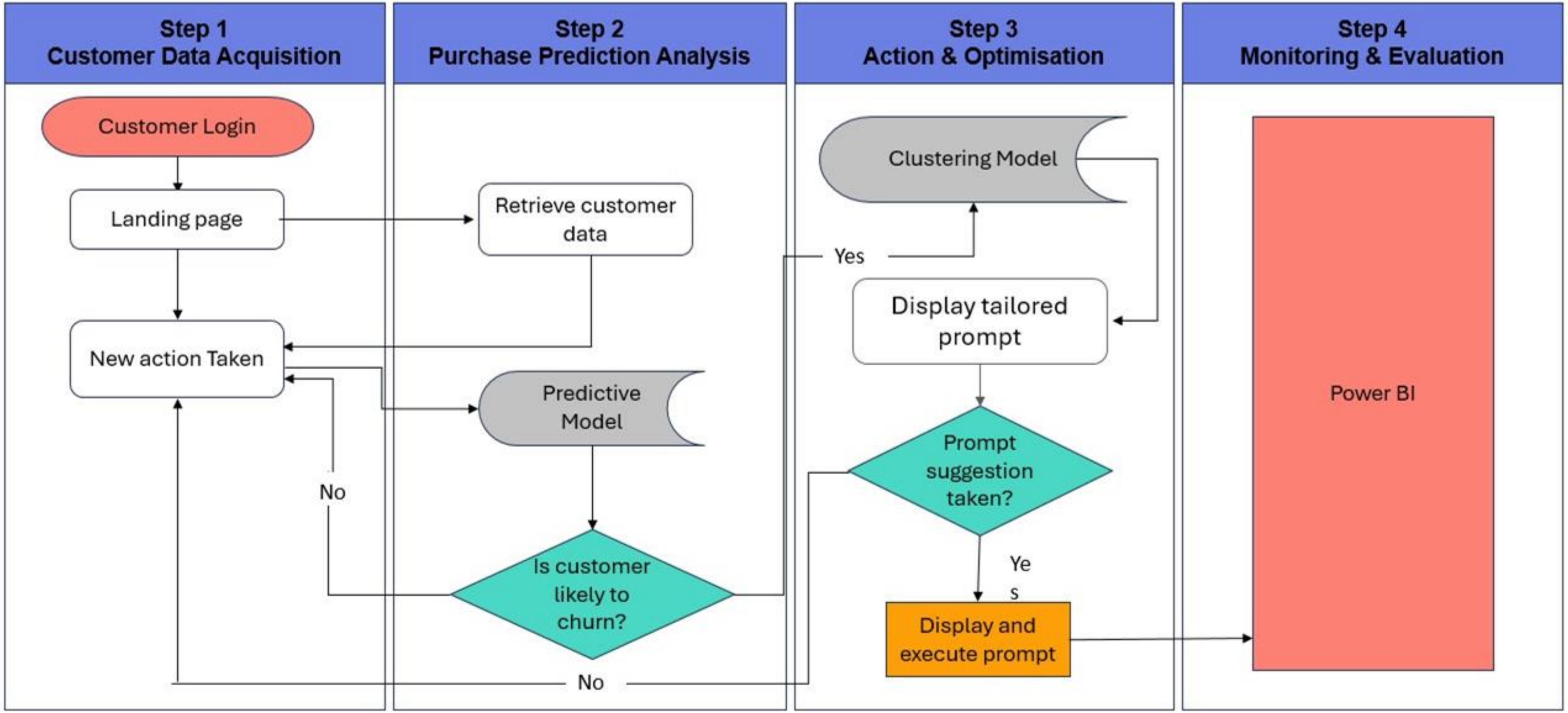
Dhiya' Diyana

Proposed Solution



Dhiya' Diyana

Our Proposed Solution



Dhiya' Diyana

Step 1

Collect real-time data on customers to Shopee Thailand

Step 2

Predict which customers are likely to churn using a predictive model

Decision

If customer is likely to churn, proceed to Step 3. Otherwise, return back to Step 1.

Step 3

Identify **unique demographics** for churning customer and **delivering targeted sales and marketing prompts**

Decision

Are targeted prompt suggestion taken? If yes, display and execute prompt (cart out). Otherwise, return back to Step 1.

Step 4

Stakeholders at Shopee Thailand can utilise Power BI to continuously monitor and evaluate the effectiveness of tailored solutions

Boost purchase rates and strengthen customer relationships

Personalised engagement strategies to reinforce long-term customer loyalty, retention, & improve revenue generation

Gregory

Data Preparation

Gregory

Datasets Description

- Two datasets, "Ecommerce Customer Churn" and "Online Shoppers Purchasing Intention": 37 variables, categorised into categorical and continuous types
- Ecommerce Customer Churn": Approx. 1800 missing values exclusively within continuous variables
- Missing data primarily relate to customer demographic information
- Occurrence seems random, with many notable outliers

Gregory

Imputation of Data through rfImpute

- rfImpute → rough fix median imputation and the Random Forest algorithm
- Preference for Median imputation due to skewed distributions and outliers in continuous variables (Shiksha, 2023)
- Ensures quality, integrity, and reliability of modeling and analysis

Data Standardisation

- Abbreviations like "CC" substituted with full counterparts like "Credit Card" for uniformity
- Ensures **consistency** and **comparability** across the dataset

Merging Dataset

- Revenue = 0 merged with Churn = 1, indicating churn
- Excluded *'CustomerID'* → lack of predictive value
- Excluded *'Visitor_Type'* → concentrate on returning visitors, given their direct relevance to churn analysis
- Enables **simultaneous utilisation** of user behavior and demographic variables

Gregory

Data Understanding

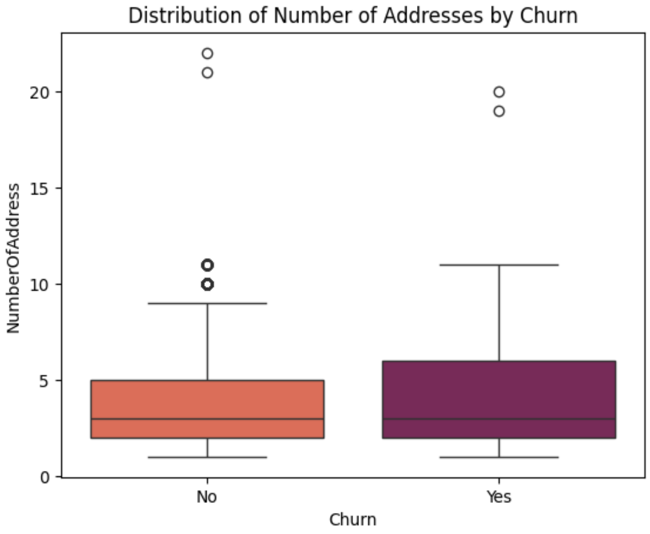
Gregory

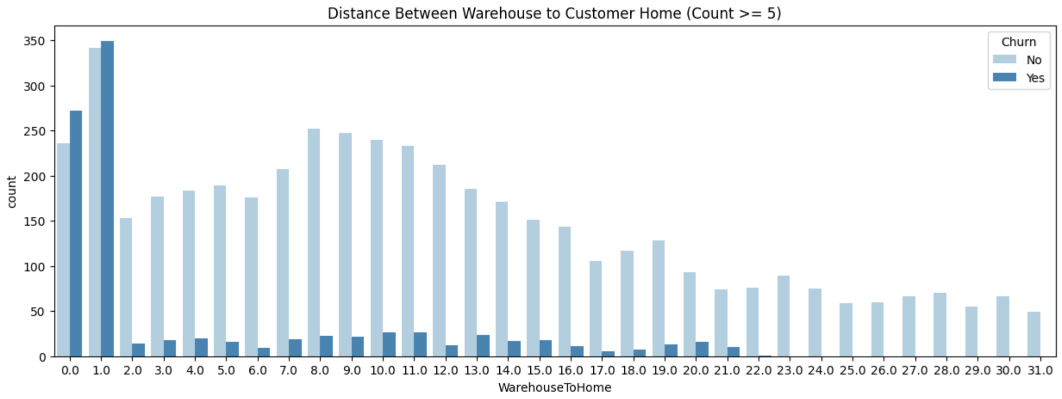
Categorisation of Variables

Category	Description
Demographic	These features represent basic characteristics of customers that might influence their loyalty and satisfaction.
Transactional	Transactional features relate to the customer's purchasing behaviour and preferences, which can signal their satisfaction and likelihood to continue using the service.
Engagement	These features are indicative of how engaged the customers are with the ecommerce platform, which can be critical for understanding churn.
Session	These features offer context about the session that might correlate with customer behaviour and preferences

Gregory

1 Multivariate Analysis - Demographic Variables



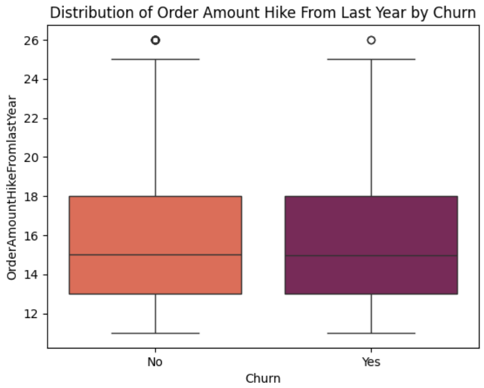


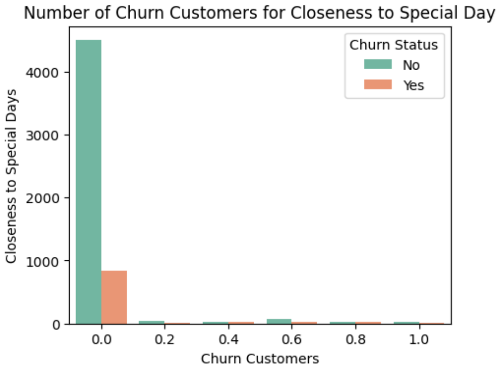
Distance of home from warehouse & Number of addresses not a significant predictor of churn

Gregory

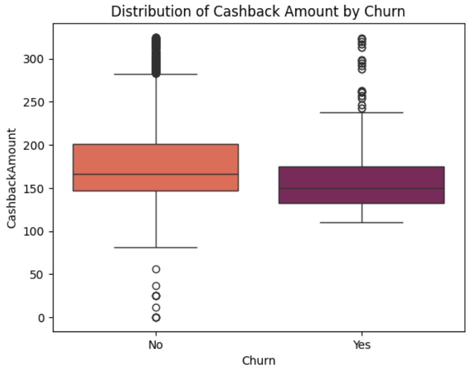
2 Multivariate Analysis - Transactional Variables

'OrderAmountHikeFromLastYear' does not significantly differ between churned and retained customers





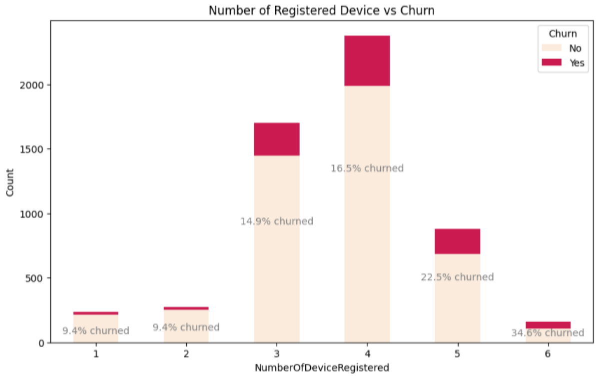
Lowest churn rates are observed during periods not closely aligned with special days



'CashbackAmount' distributions show a wider interquartile range for non-churned customers

Gregory

3 Multivariate Analysis - Engagement Variables



Churn rate increases with *'NumberOfRegisteredDevice'*

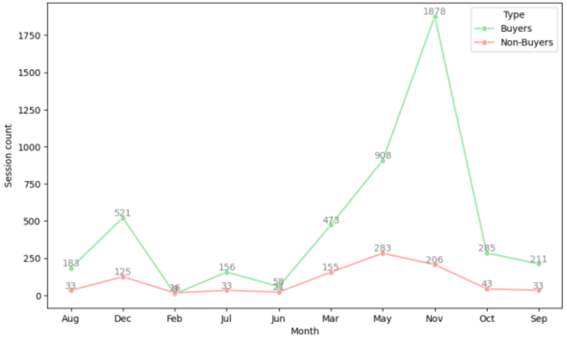
- Weak inverse relationship between *'SatisfactionScore'* and *'Complain'*
- High satisfaction scores show a correlation with increased churn rates
- Indicates a complex relationship

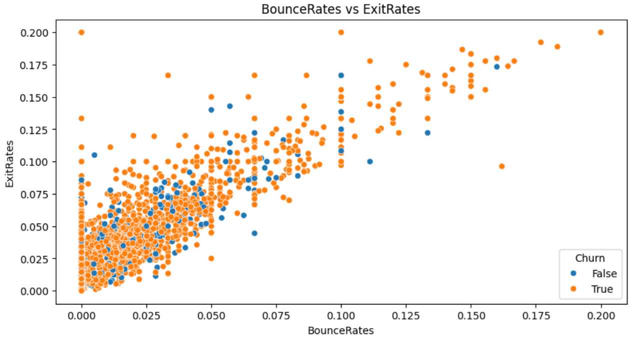
Gregory

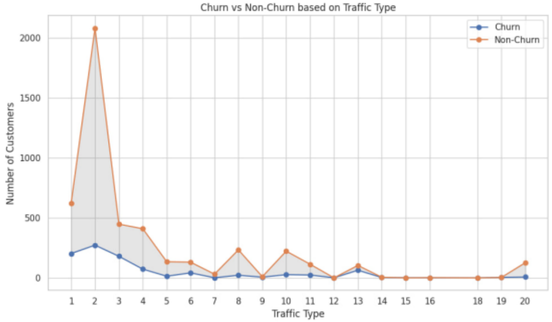
4 Multivariate Analysis - Session Variables

A time series plot of *'Month'* against session count shows a spike in retained customers in May and November

Sessions with increased exit and bounce rates are more inclined towards churning







'TrafficType' 13 stands out with an exceptionally high churn rate → significant risk factor and a potential predictor for customer attrition

Andy

Data Modelling

Churn
Prediction

Andy

Methodology

Objective

Accurately predict customers who are likely to churn

Approach

Employ supervised machine learning algorithms for classification.

Models Used

Random Forest

CART

Logistic Regression

Target Variable

Churn

1: Customer has churned

0: Customer retained

Andy

Features Selection

Original Feature Set

35
Variables

Engagement & Session Feature Set

22
Variables

Andy

Features Selection

Excluded Features

Variables deemed less informative: 'HourSpendOnApp', 'OperatingSystems', 'Browser', 'Region', 'Weekend'.

Variables with marginal predictive value: 'NumberOfRegisteredDevice', 'SatisfactionScore'.

Variables with high correlation: 'OrderCount', 'Administrative', 'Informational', and 'ProductRelated' excluded in favour of 'DaySinceLastOrder', 'Administrative_Duration', 'Informational_Duration', 'ProductRelated_Duration'.

Final Features

Tenure, Complain, DaySinceLastOrder, Administrative_Duration, Informational_Duration, ProductRelated_Duration, BounceRates, ExitRates, PageValues, Month, TrafficType

Andy

Results of Predictive Models

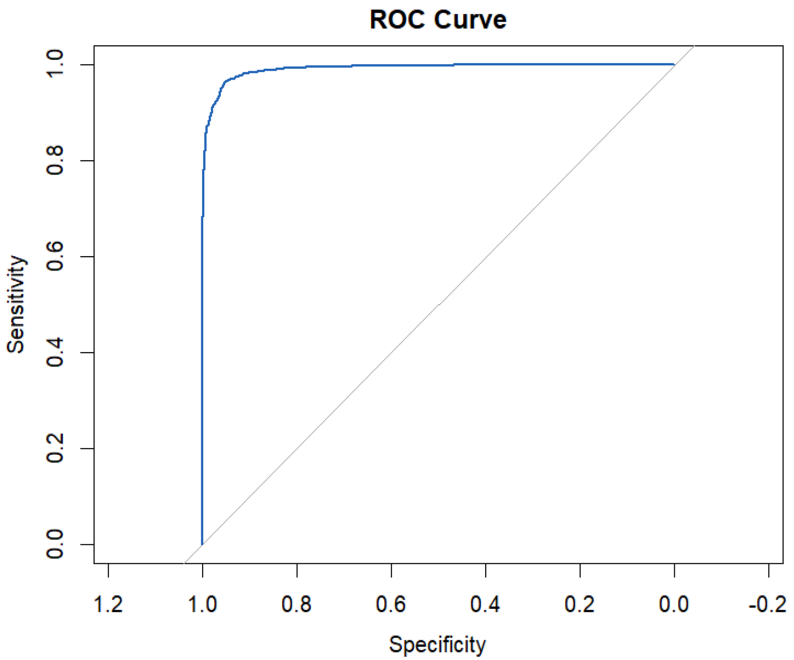
Model	Accuracy	Precision	Recall	F1-Score
Random Forest Model	94.67%	82.27%	86.93%	84.54%
CART Model	92.36%	73.48%	85.16%	78.89%
Logistic Regression Model	89.69%	63.59%	87.28%	73.58%

Andy

Selected Model Evaluation: Random Forest

```
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 3

        OOB estimate of  error rate: 4.35%
Confusion matrix:
     1    0 class.error
1 2520  136  0.05120482
0  122 3155  0.03722917
```



Andy

Clustering
Algorithm

Methodology

Objective

Segment the at-risk customers into meaningful groups for tailored marketing interventions

Approach

Employ unsupervised machine learning technique

Model Used

K-Prototypes Clustering

Rationale

Iterative optimisation approach with the ability to handle mixed data types effectively and clusters customers based on both numerical and categorical attributes

Andy

Features Selection

Original Feature Set

35 Variables

Demographic & Transactional Feature Set

12 Variables

Andy

Feature Selection

Excluded Features

Variables deemed less informative: 'OrderAmountHikeFromLastYear', 'NumberOfAddress'.

Variables with marginal clustering value: 'Gender', 'SpecialDay'.

Strategic Inclusion

Although they are Engagement features, 'OrderCount' and 'DaySinceLastOrder' included for their strategic value.

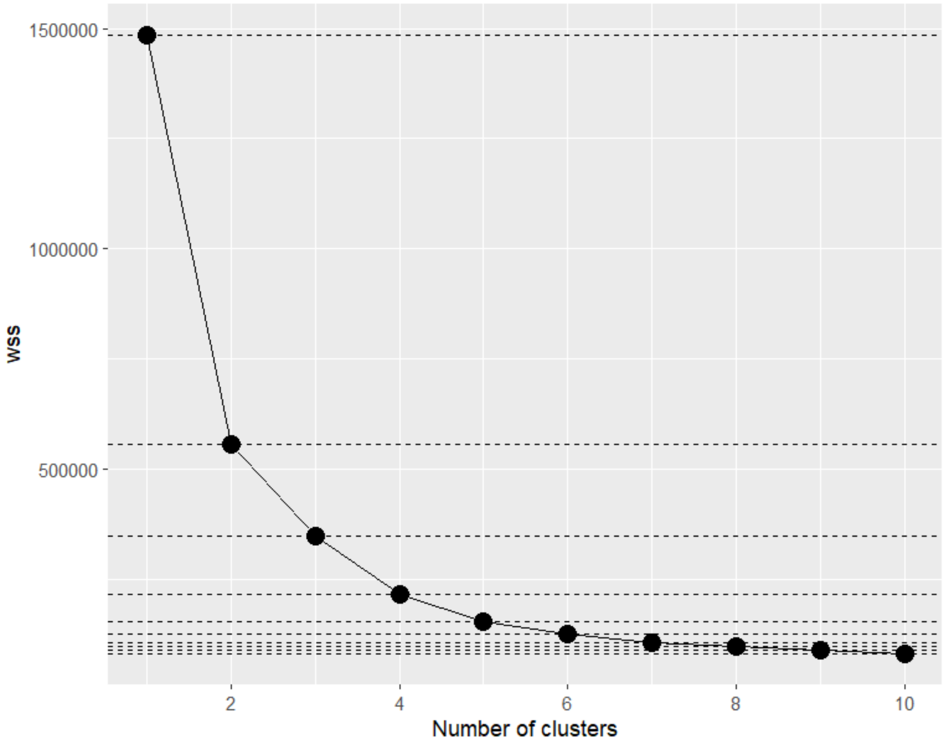
Combined with 'CouponUsed', these features offer a comprehensive view of customer behavior.

Final Features

PreferredOrderCat, PreferredLoginDevice, PreferredPaymentMode, CityTier, WarehouseToHome, MaritalStatus, CashbackAmount, CouponUsed, OrderCount, DaySinceLastOrder

Andy

Visualising the Elbow Method



Andy

Evaluation of Clusters

```
> print(centers)
  PreferedOrderCat PreferredLoginDevice CityTier WarehouseToHome PreferredPaymentMode MaritalStatus
1          Fashion         Mobile Phone        3        19.14535           Debit Card       Married
2     Mobile Phone         Mobile Phone        1        17.54977           Debit Card        Single
3           Others         Mobile Phone        3        14.20000           Debit Card        Single
4     Mobile Phone         Mobile Phone        1        14.81579           Debit Card        Single
  CouponUsed OrderCount DaySinceLastOrder CashbackAmount Cluster
1  2.7732558   4.279070          4.587209       208.1279       1
2  1.7307692   2.705882          3.307692       156.3032       2
3  3.2000000   7.133333          8.600000       291.0667       3
4  0.9342105   1.697368          1.796053       126.3618       4
```

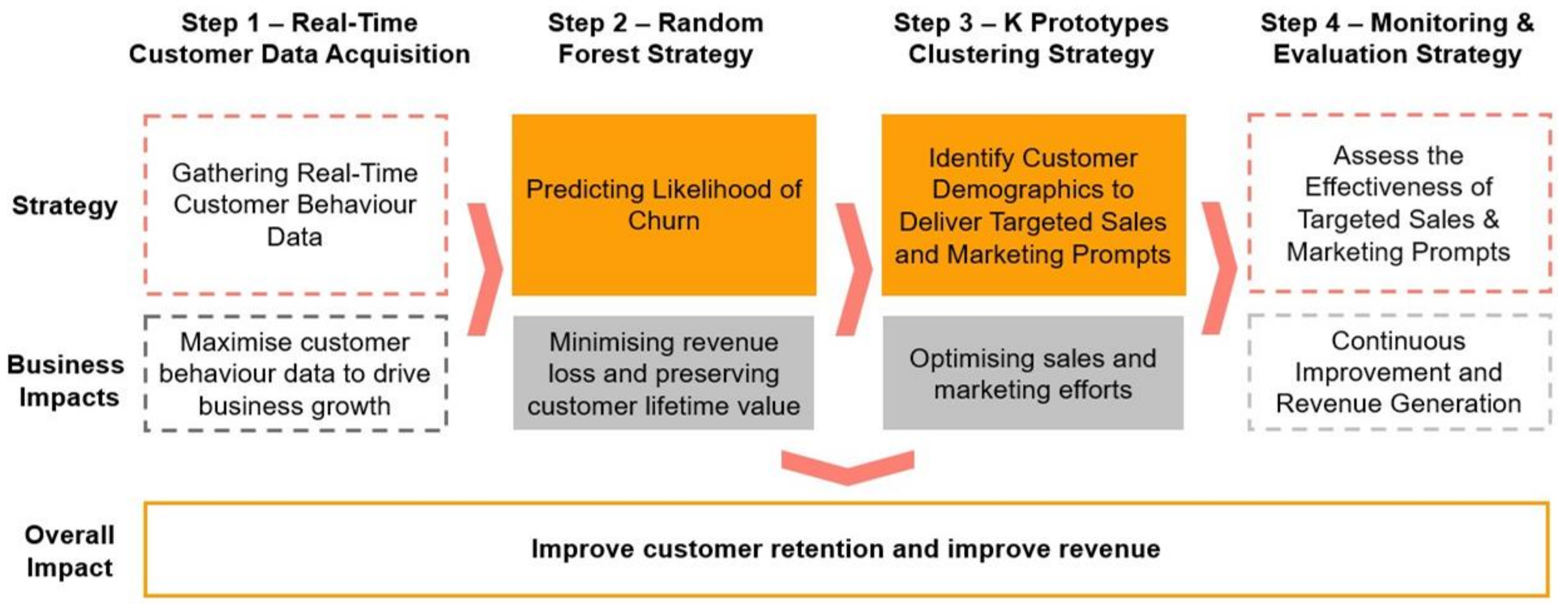
Distinct patterns observed in these clusters can then direct us to **tailor specific marketing interventions**

Yu Xiang (Javier)

Deployment of Solution

Yu Xiang (Javier)

Our Proposed Framework



Yu Xiang (Javier)

1 Real-Time Data Acquisition

- Leverages on the customer data to monitor and analyse customer behaviour data in real time
- Collect and analyse customer behaviour to gain insights to current consumer sentiments
- Leverage its sophisticated, existing real-time customer data acquisition system by harnessing its robust web scraping and data extraction capabilities (Shukla, 2023).

Yu Xiang (Javier)

2 Predicting Likelihood of Churn

Step 2 – Random Forest Strategy

Strategy	Predicting Likelihood of Churn
Business Impacts	Minimising revenue loss and preserving customer lifetime value

- Analyze real-time customer data to identify churn probability
- Error Rate: 4.62%, indicating feasibility and relatively high accuracy
- Develop user-friendly API for seamless data input to the model
- Utilise in-house platform and Google Cloud Platform for reliability and scalability and automate responses to potential churn risks

Yu Xiang (Javier)

3 Identify Customer Demographics for Targeted Prompts

- **Firstly,** run the K-Prototypes algorithm on Shopee Thailand's historical data to create clusters representing different customer profiles.
- Save these centroids

- **Secondly**, model measures the proximity between newly identified potential churn customer and established cluster centroids based on the customer's data and assigned to the nearest centroid

- **Lastly,** cluster will represent the customer profile that most closely matches the customer's current behaviour and characteristics → Implement the associated retention strategy linked to this specific cluster

Step 3 – K Prototypes Clustering Strategy

Strategy	Identify Customer Demographics to Deliver Targeted Sales and Marketing Prompts
Business Impacts	Optimising sales and marketing efforts

Yu Xiang (Javier)

4 Monitoring & Evaluation of Tailored Initiatives

- Assess the impact of Phase 2 and 3 strategies on churn rates
- A real-time dashboard will be introduced to provide stakeholders immediate visibility into key performance metrics
- Empowers stakeholders to promptly identify and address gaps in customer satisfaction, enhancing retention efforts swiftly
- Proactive monitoring ensures continuous optimisation of targeted efforts, dynamically responding to customer feedback and behaviors

Step 4 – Monitoring & Evaluation Strategy

Strategy	Assess the Effectiveness of Targeted Sales & Marketing Prompts
Business Impacts	Continuous Improvement and Revenue Generation

Yu Xiang (Javier)

Power BI

Real-Time Dashboard Demo

Jesslyn

Evaluation of Proposed Solution

Jesslyn

Benefits of Solution

Reducing Churn, Foster Better Relations

- Prediction of user's churn risk gives Shopee the opportunity to intervene
- Allow Shopee to tailor personalised engagement strategies, fostering stronger customer relations

Increase Retention Rate & Revenue Generation

- Satisfied customers are more likely to increase purchase frequency
- Loyal customers will decrease churn rates and improve profitability, improving competitiveness edge within the e-commerce landscape

Jesslyn

Limitations

1 Reliance on Real-Time Customer Data

- Our proposed framework only involves one layer of algorithm to predict churn
- Any inaccuracies can lead to misguided predictions → wasting resources or missed opportunities

2 Complexities of Customer Behaviour

Essential to consider other factors in providing deeper and more accurate insights

Jesslyn

Future Considerations

1. **Adapt to Evolving Customer Behaviour & Market Dynamics**

2. **Incorporate New Data & Adjusting Model Parameters**

3. **Incorporate New Features to Improve Prediction of Churn**

Conclusion

1. Potential of Integrating Random Forest and K-Prototypes clustering for decreasing churn rates and enhancing customer retention

2. Strategic advantage of real-time monitoring of customer behavior for providing stakeholders with actionable insights

3. Personalisation of customer experience and optimization of marketing efforts to streamline cost and increase revenue

4. Utilization of sophisticated machine learning models and advanced analytics to align with the business objective of minimizing churn and increased customer loyalty

THANK YOU

Any Questions?

Shopee