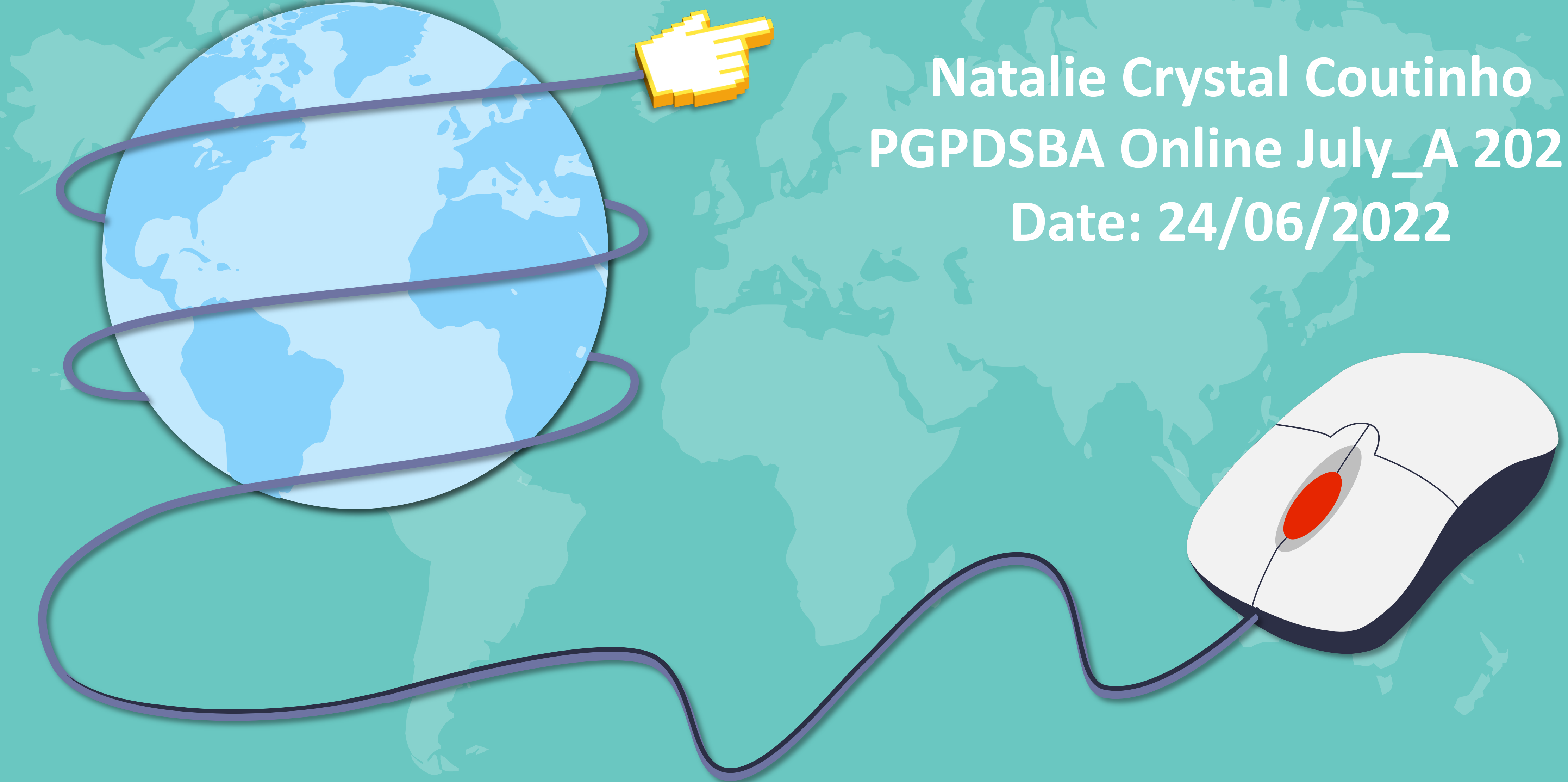
Customer Churn Analysis

Natalie Crystal Coutinho
PGPDSBA Online July_A 2021

Date: 24/06/2022



E-COMMERCE SHOPPING

CUSTOMER PURCHASE INTENTION

01 Problem Statement

- Business Problem, Constraints, Scope & Objective

02 Model Selection and Evaluation Parameter

03 Model Building Approach

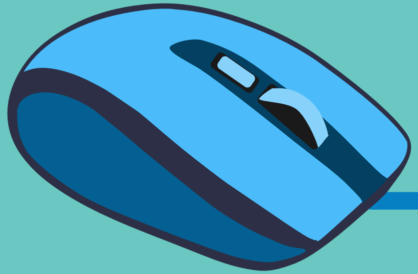
- K-Nearest Neighbors

04 Business insights

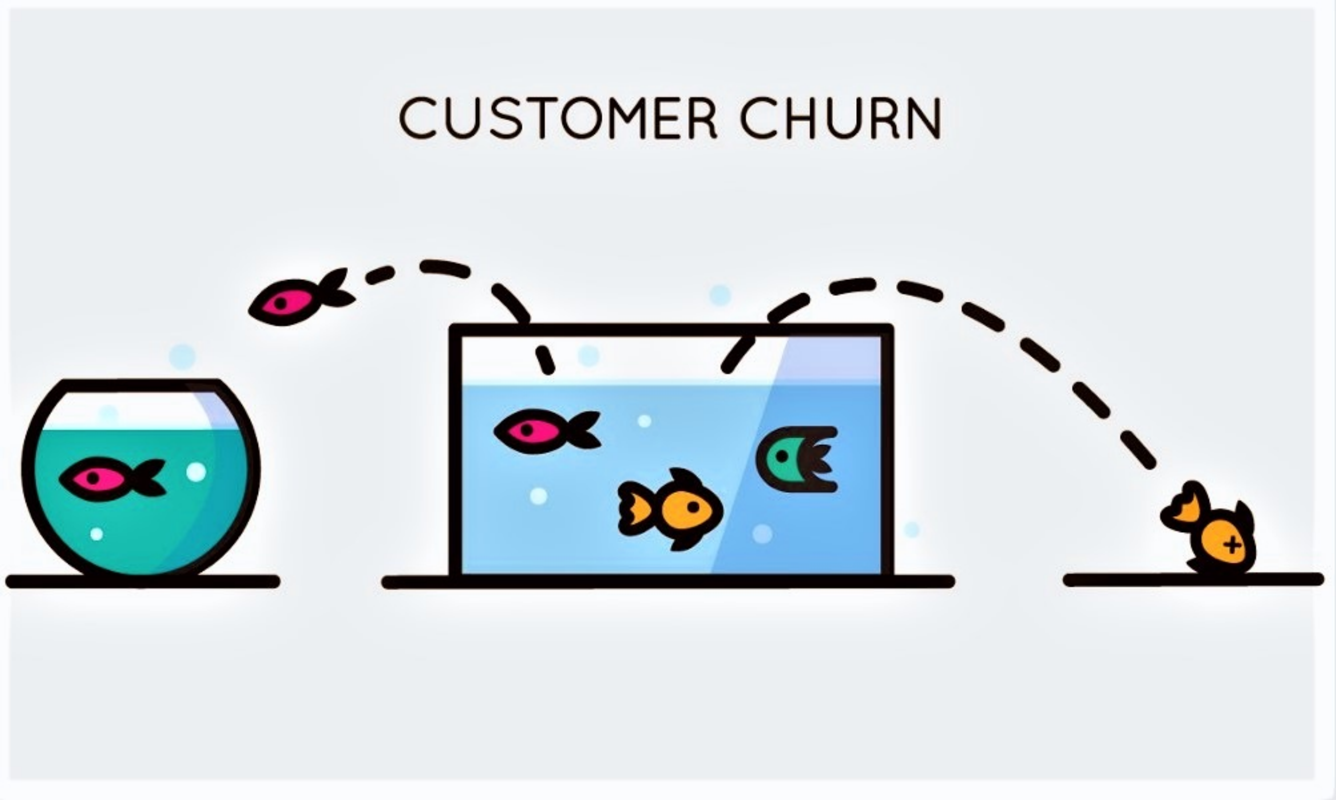
- Consumer Segmentation based on Behavior

05 Recommendations

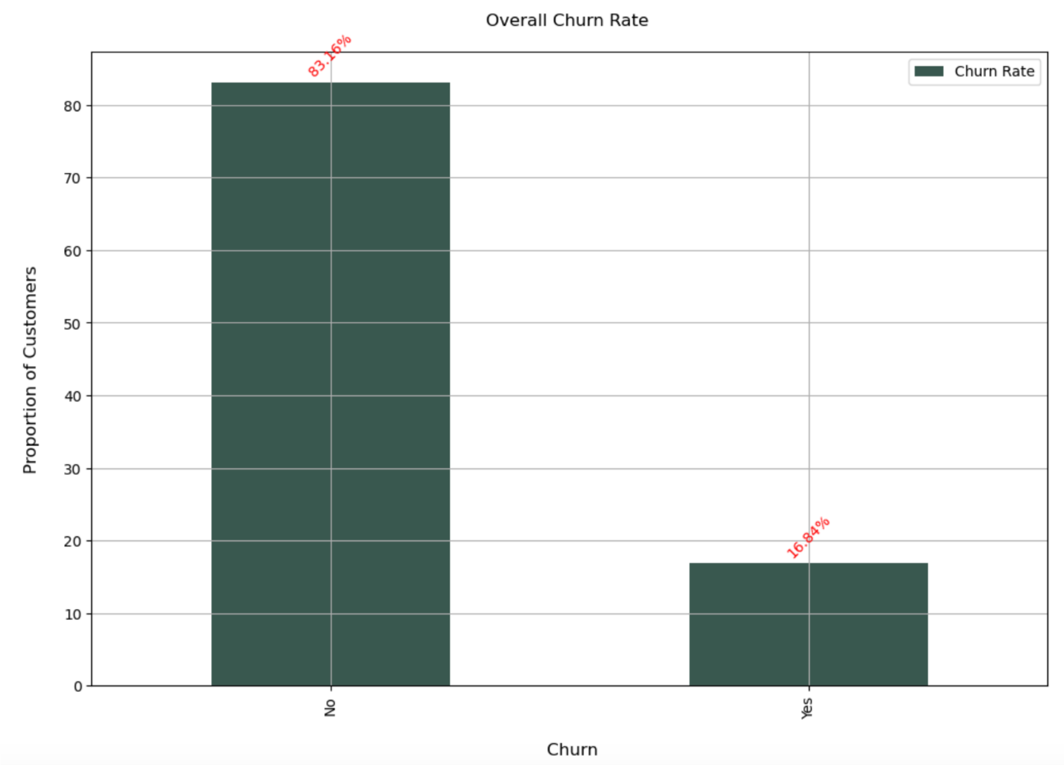
Suggestions & Lucrative Combos



Problem Statement:



A Closer Look



Predict and reduce the churn rate in the customer goods industry to industry standards (10%) to enhance income by 7%, utilizing data insights for strategic offer and promotional strategy development.

Data Overview

	AccountID	Churn	Tenure	City_Tier	CC_Contacted_LY	Payment	Gender	Service_Score	Account_user_count
0	20000	1	4.0	3.0	6.0	Debit Card	Female	3.0	3.0
1	20001	1	0.0	1.0	8.0	UPI	Male	3.0	4.0
2	20002	1	0.0	1.0	30.0	Debit Card	Male	2.0	4.0
3	20003	1	0.0	3.0	15.0	Debit Card	Male	2.0	4.0
4	20004	1	0.0	1.0	12.0	Credit Card	Male	2.0	3.0

account_segment	CC_Agent_Score	Marital_Status	rev_per_month	Complain_ly	rev_growth_yoy	coupon_used_for_payment	Day_Since_CC_connect	cashback	Login_device
Super	2.0	Single	9.0	1.0	11.0	1.0	5.0	159.93	Mobile
Regular Plus	3.0	Single	7.0	1.0	15.0	0.0	0.0	120.90	Mobile
Regular Plus	3.0	Single	6.0	1.0	14.0	0.0	3.0	NaN	Mobile
Super	5.0	Single	8.0	0.0	23.0	0.0	3.0	134.07	Mobile
Regular Plus	5.0	Single	3.0	0.0	11.0	1.0	3.0	129.60	Mobile

The dataset consists of 19 essential attributes representing the churn status. Churn labels are categorized as 0 = No Churn and 1 = Churn, offering a comprehensive grading system.

Encoding 5 categorical variables using Label Encoding for Prediction Model.

	AccountID	Churn	Tenure	City_Tier	CC_Contacted_LY	Payment	Gender	Service_Score	Account_user_count
0	20000	1	4.00	2	6.00	2	0	3.0	3.00
1	20001	1	0.00	0	8.00	4	1	3.0	4.00
2	20002	1	0.00	0	30.00	2	1	2.0	4.00
3	20003	1	0.00	2	15.00	2	1	2.0	4.00
4	20004	1	0.00	0	12.00	1	1	2.0	3.00

account_segment	CC_Agent_Score	Marital_Status	rev_per_month	Complain_ly	rev_growth_yoy	coupon_used_for_payment	Day_Since_CC_connect	cashback	Login_device
3	2.0	2	9.00	1.0	11.00	1.00	5.00	159.93	1
2	3.0	2	7.00	1.0	15.00	0.00	0.00	120.90	1
2	3.0	2	6.00	1.0	14.00	0.00	3.00	NaN	1
3	5.0	2	8.00	0.0	23.00	0.00	3.00	134.07	1
2	5.0	2	3.00	0.0	11.00	1.00	3.00	129.60	1

Data Pre-Processing

Preparation for Model Building

The dataset was pre-processed and refined for optimal model performance through:

Split data into training and testing sets with 30% for testing. Use stratified sampling.

```
X_train: (13110, 17)
X_test: (3378, 17)
```

Addressing imbalance with SMOTE to improve model fairness and accuracy.

BEFORE

```
0     9364
1     1896
Name: Churn, dtype: int64
```

AFTER

```
0.00    6555
1.00    6555
Name: Churn, dtype: int64
```

Preparation for Model Building

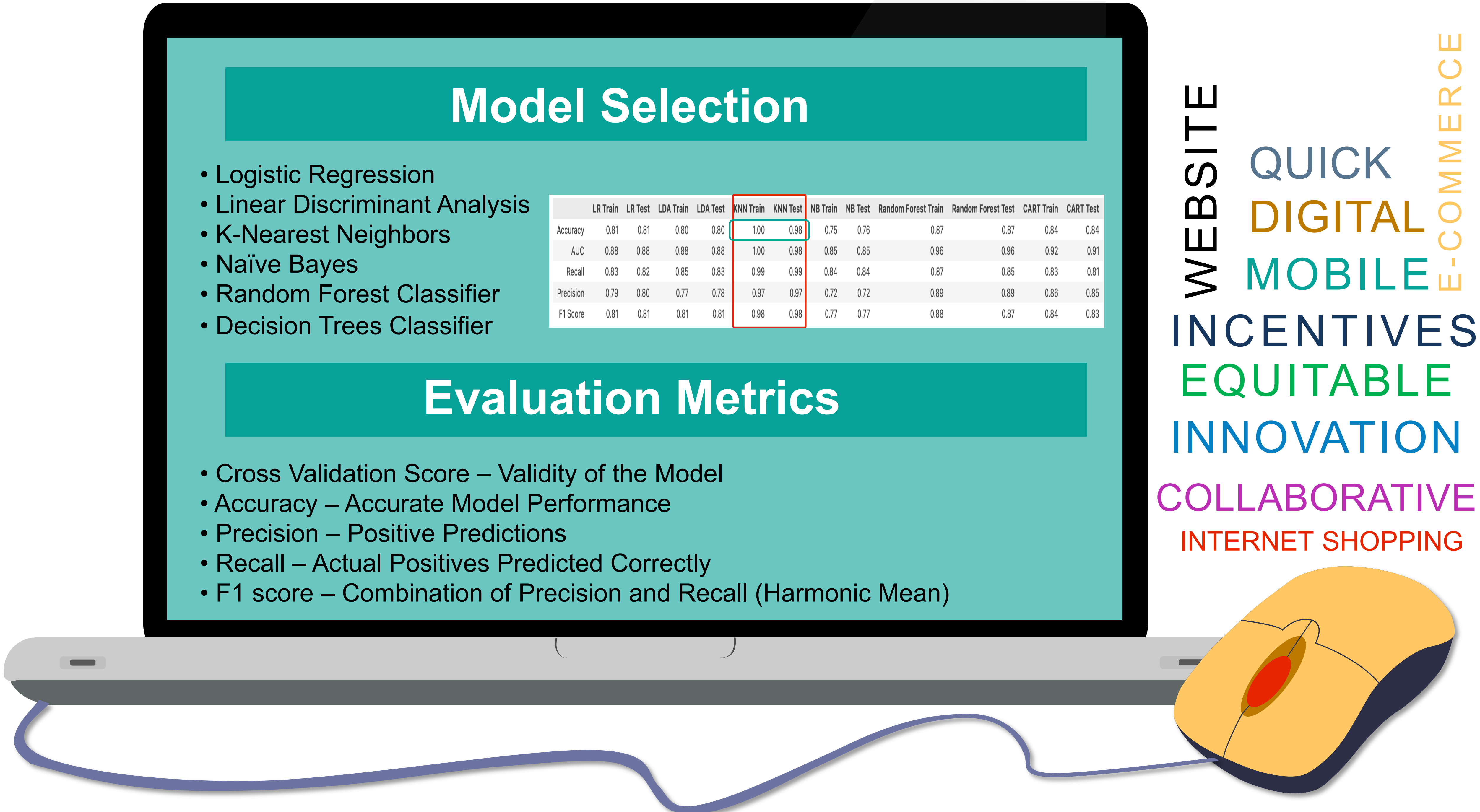
X_train

	Tenure	City_Tier	CC_Contacted_LY	Payment	Gender	Service_Score	Account_user_count	account_segment	CC_Agent_Score	Marital_Status	rev_per_month	Complain_ly	rev_growth_yoy	coupon_used_for_payment	Day_Since_CC_connect	cashback	Login_device
0	2.03	-0.79	-0.74	-1.84	0.77	0.13	-0.89	0.90	1.38	-0.42	-0.47	1.38	-1.09	0.51	1.48	-0.37	0.41
1	-0.29	1.40	1.58	1.37	0.77	0.13	-0.89	0.90	-1.68	-0.42	-1.51	-0.83	0.50	0.51	-0.63	-0.63	-1.52
2	-0.87	-0.79	0.77	-0.77	0.77	0.13	-0.89	-0.20	-0.91	-1.89	-0.47	-0.83	-1.36	-0.43	-0.63	-0.82	0.41
3	-0.06	1.40	-0.51	-0.77	0.77	0.13	-2.03	0.90	1.38	-0.42	0.57	-0.83	1.82	-0.43	1.48	-0.01	0.41
4	-0.52	1.40	0.19	1.90	0.77	0.13	-0.89	-1.86	-0.15	-0.42	1.26	-0.83	-1.09	1.94	0.88	0.13	0.41

X_test

	Tenure	City_Tier	CC_Contacted_LY	Payment	Gender	Service_Score	Account_user_count	account_segment	CC_Agent_Score	Marital_Status	rev_per_month	Complain_ly	rev_growth_yoy	coupon_used_for_payment	Day_Since_CC_connect	cashback	Login_device
0	0.06	-0.79	-1.20	0.30	0.77	-1.33	0.24	-1.86	0.62	-0.42	-0.81	-0.83	0.50	0.51	-0.03	0.72	0.41
1	1.80	-0.79	1.81	0.30	-1.32	-1.33	-0.89	2.01	1.38	-1.89	-0.12	1.38	-1.36	-1.38	-0.63	2.25	0.41
2	0.29	-0.79	-0.39	-1.84	-1.32	0.13	0.24	-1.86	-0.15	2.53	-0.81	-0.83	-0.56	1.94	1.48	1.26	0.41
3	-0.17	1.40	0.77	0.30	-1.32	-1.33	-0.89	0.90	-1.68	-0.42	-0.47	1.38	2.88	-1.38	-0.63	-0.40	-1.52
4	0.75	1.40	-0.85	1.37	0.77	1.60	1.37	-1.86	-1.68	-0.42	1.26	-0.83	-1.09	0.51	2.39	1.43	0.41

StandardScaler normalizes feature data (X_train and X_test) and not the target variable data (y_train and y_test) to prevent information leakage. This enhances model performance and interpretability having mean = 0 and sd = 1.



	LR Train	LR Test	LDA Train	LDA Test	KNN Train	KNN Test	NB Train	NB Test	Random Forest Train	Random Forest Test	CART Train	CART Test
Accuracy	0.81	0.81	0.80	0.80	1.00	0.98	0.75	0.76	0.87	0.87	0.84	0.84
AUC	0.88	0.88	0.88	0.88	1.00	0.98	0.85	0.85	0.96	0.96	0.92	0.91
Recall	0.83	0.82	0.85	0.83	0.99	0.99	0.84	0.84	0.87	0.85	0.83	0.81
Precision	0.79	0.80	0.77	0.78	0.97	0.97	0.72	0.72	0.89	0.89	0.86	0.85
F1 Score	0.81	0.81	0.81	0.81	0.98	0.98	0.77	0.77	0.88	0.87	0.84	0.83

Model Building Approach

K-Nearest Neighbors

IMPORTANT to correctly identifying positive churn status to avoid business losses.

Balanced – Performance Metrics on Testing Data
AUC: 0.982

	precision	recall	f1-score	support
0.0	0.99	0.97	0.98	1967
1.0	0.97	0.99	0.98	1966
accuracy			0.98	3933
macro avg	0.98	0.98	0.98	3933
weighted avg	0.98	0.98	0.98	3933

Cross Validation Score: 0.9773346828129046

F-1 SCORE
- *All classes' scores are **close to 1.***

Balanced trade-off between precision and recall for predicting quality of churn status.

ACCURACY
- *Maintains high accuracy at 98%*

Reflecting strong generalization to new churn samples.

PRECISION
- *Class 0: Demonstrates **100%** precision, correctly identifying all customers who haven't churned.*
- *Class 1: Achieves **97%** precision in predicting churned samples.*

RECALL
- *Class 0: Successfully identifies 97% of customers who haven't churned.*
- *Class 1: Demonstrates perfect recall, capturing 100% of churned samples.*

Business Insights

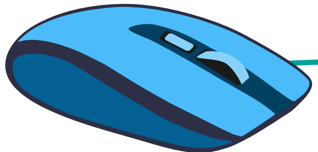
- Optimal clustering suggests three customer segments based on high, medium, and low tenure, aiding targeted promotional strategies.

- Urban customers in City_Tier 1, technologically adept, prefer online purchases and electronic payments, making them ideal targets for promotional campaigns.

- Price-driven incentives such as tie-ups, bundling, discounts, cash-backs, and EMI schemes attract budget-conscious consumers.

- Targeting male shoppers, known for their tech affinity, requires tailored strategies. Attracting female shoppers involves building trust and satisfaction through diverse options, incentivizing stress-free shopping across various categories.

- A comprehensive approach combining technology, tailored incentives, and a diverse product range is key to attracting and retaining customers.



Recommendations

- Leverage male shopping tendencies for electronic products with combo discounts and strategic product recommendations.

- Capitalize on festive seasons like Diwali and Christmas for targeted deals, enhancing sales and regional profits.

- Offer a diverse range of goods at affordable prices with free delivery, establishing the business as a one-stop-shop and partnering with other brands for profit accumulation.

- Utilize the internet as the primary access point, streamlining individual experiences through interactive portals and 24x7 online support, enhancing overall customer experience and attracting a broad consumer base.

THANK YOU