



E-COMMERCE CHURN ANALYSIS

MARIA SOHAIL – JACK OF ALL TRADES
EUNICE WORIFAH – JACK OF ALL TRADES

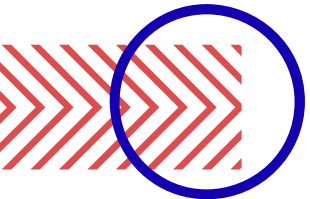
TABLE OF CONTENTS

01 BACKGROUND

02 HYPOTHESIS

03 DATA AND MODELLING

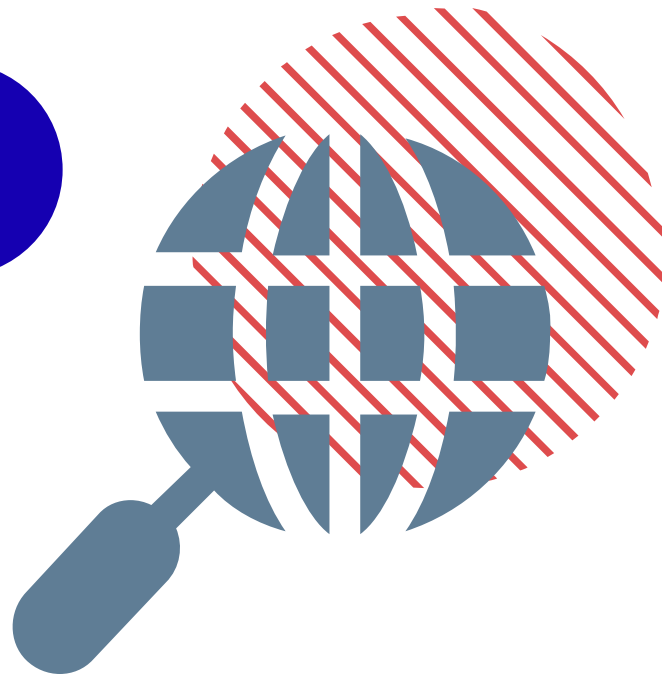
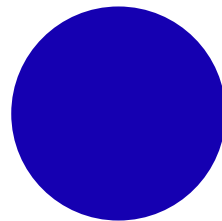
04 RESULTS AND
CONCLUSION

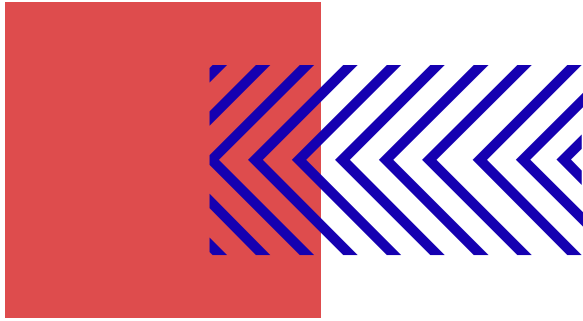


01



BACKGROUND



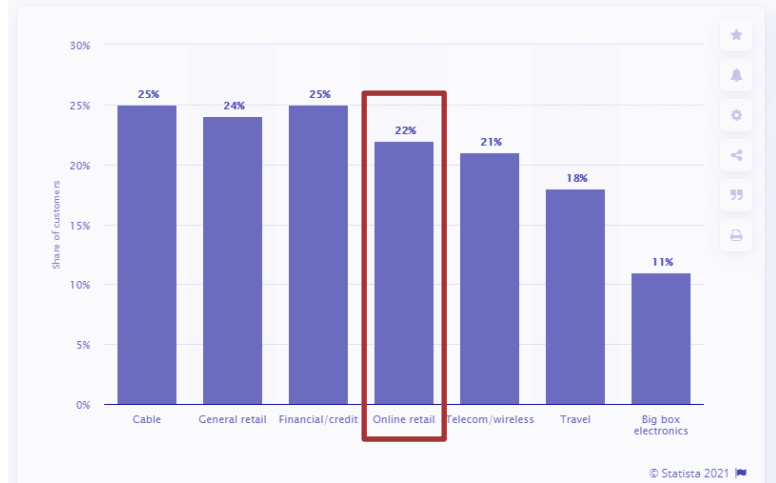


WHAT IS CHURN?

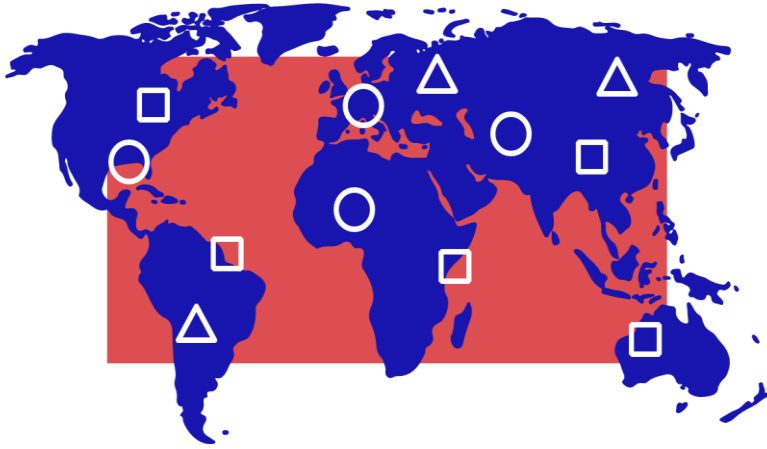
Customer churn is when customers stop doing business with an entity.

- **Subscription** – customer does not renew their subscription -
- ***Non-Subscription** - customer gradually reduces their purchase frequency over time, or may all of a sudden never buy again

Customer churn rate in the United States in 2020, by industry



WHY DOES THIS MATTER?



E-COMMERCE IS SLOWLY BECOMING THE DOMINANT DISTRIBUTION CHANNEL

- Retail sales from worldwide e-commerce are forecast to grow to over 6.54 trillion in 2023.
- Revenue generated within the retail e-commerce market is expected to surpass 33 billion U.S. dollars by 2024, up from 25.4 billion in 2019.
- In late 2019, e-commerce retail trade sales amounted to almost 1.85 billion Canadian dollars, with approximately 28.1 million Canadians having made purchases online.

STAKEHOLDERS

COMPANY

- **Shareholders** - Revenue and Profit
- **Marketing** - Advertising and Customer Acquisition
- **Customer Service** - Customer Satisfaction
- **Logistics** - Product Delivery
- **Product Management** - UX/UI
- **Sales** - Customer Acquisition

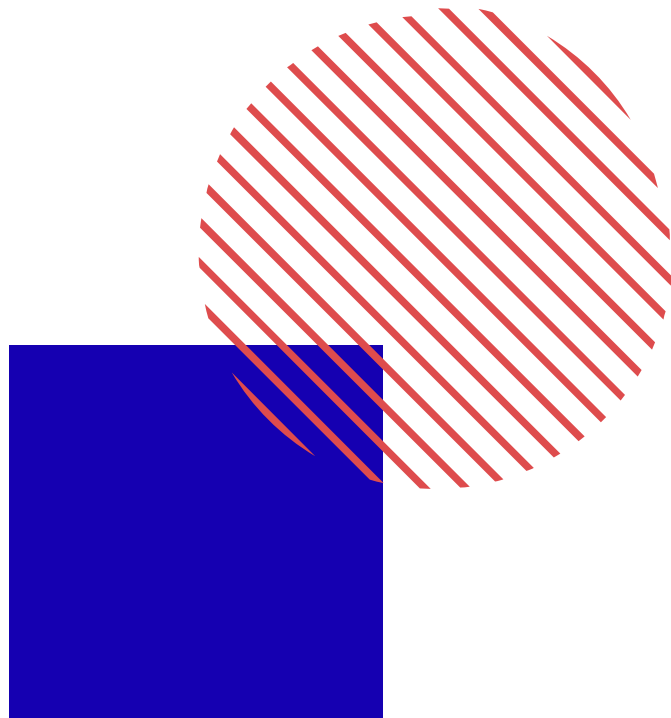
CUSTOMER

- Pleasant Experience and High Customer Satisfaction



OUR OBJECTIVES

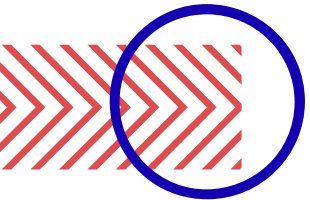
- Explore data and understand customer attributes that correlate to a higher risk of churn
- Build a model to predict customer churn





02

HYPOTHESIS



WE HYPOTHESIZE THAT...

TENURE

The longer the customer has been with the company, the less likely they are to churn.

COMPLAIN

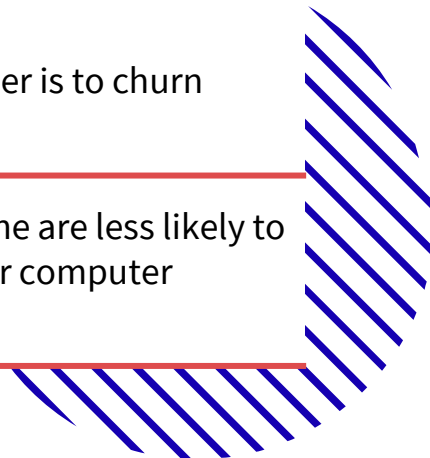
If a customer lodged a complaint within the last month, this increases their likelihood to churn.

CASHBACK AMOUNT

The greater the cashback amount, the less likely a customer is to churn

ACCESSING APP BY PHONE

Customers who prefer to use the app on their mobile phone are less likely to churn compared to customers who prefer to log in on their computer because they have more/quicker access to the platform



03

DATA



SHAPE (5630, 20)

Data	Variable	Description
E Comm	CustomerID	Unique customer ID
E Comm	Churn	Churn Flag (0=customer retained, 1=customer churned)
E Comm	Tenure	Tenure of customer in organization
E Comm	PreferredLoginDevice	Preferred login device of customer
E Comm	CityTier	City tier
E Comm	WarehouseToHome	Distance in between warehouse to home of customer
E Comm	PreferredPaymentMode	Preferred payment method of customer
E Comm	Gender	Gender of customer
E Comm	HourSpendOnApp	Number of hours spent on mobile application or website
E Comm	NumberOfDevicesRegistered	Total number of devices registered on by a particular customer
E Comm	PreferredOrderCat	Preferred order category of customer in last month
E Comm	SatisfactionScore	Satisfactory score of customer on service (1-5)
E Comm	MaritalStatus	Marital status of customer
E Comm	NumberOfAddress	Total number of delivery addresses added on particular customer account
E Comm	Complain	Any complaints raised in last month
E Comm	OrderAmountHikeFromLastYear	Percentage increases in order from last year
E Comm	CouponsUsed	Total number of coupons used in the last month
E Comm	OrderCount	Total number of orders placed in last month
E Comm	DaysSinceLastOrder	Number of days since last order by customer
E Comm	CashbackAmount	Average cashback in the last month

SOURCE:

KAGGLE

Customer ID	Churn	Tenure	Preferred Login Device	CityTier	WarehouseToHome	Preferred Payment Mode	Gender	HourSpendOnApp	NumberOfDevicesRegistered	Preferred OrderCat	SatisfactionScore	MaritalStatus	NumberOfAddress	Complain	OrderAmountHikeFromlast Year	CouponUsed	OrderCount	DaySinceLastOrder	Cashback Amount
50001	1	4	Mobile Phone	3	6	Debit Card	Female	3	3	Laptop & Mobile	2	Single	9	1	11	1	1	5	160
50002	1		Phone	1	8	UPI	Male	3	4	Mobile	3	Single	7	1	15	0	1	0	121
50003	1		Phone	1	30	Debit Card	Male	2	4	Mobile	3	Single	6	1	14	0	1	3	120
50004	1	0	Phone	3	15	Debit Card	Male	2	4	Laptop & Mobile	5	Single	8	0	23	0	1	3	134
50005	1	0	Phone	1	12	CC	Male		3	Mobile	5	Single	3	0	11	1	1	3	130
50006	1	0	Computer	1	22	Debit Card	Female	3	5	Mobile Phone	5	Single	2	1	22	4	6	7	139
50007	1		Phone	3	11	Cash on Delivery	Male	2	3	Laptop & Mobile	2	Divorced	4	0	14	0	1	0	121
50008	1		Phone	1	6	CC	Male	3	3	Mobile	2	Divorced	3	1	16	2	2	0	123
50009	1	13	Phone	3	9	E wallet	Male		4	Mobile	3	Divorced	2	1	14	0	1	2	127
50010	1		Phone	1	31	Debit Card	Male	2	5	Mobile	3	Single	2	0	12	1	1	1	123
50011	1	4	Mobile Phone	1	18	Cash on Delivery	Female	2	3	Others	3	Divorced	2	0		9	15	8	295
50012	1	11	Mobile Phone	1	6	Debit Card	Male	3	4	Fashion	3	Single	10	1	13	0	1	0	154

DATA EXPLORATION

PANDAS PROFILING

Overview

Warnings 12

Reproduction

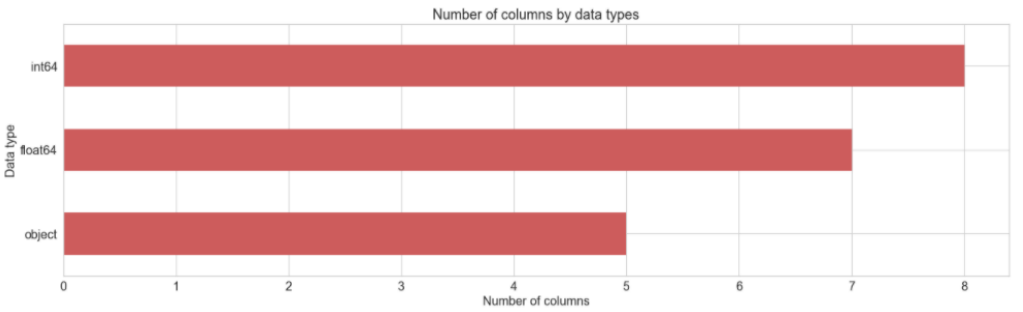
Dataset statistics

Number of variables	20
Number of observations	5630
Missing cells	1856
Missing cells (%)	1.6%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	879.8 KiB
Average record size in memory	160.0 B

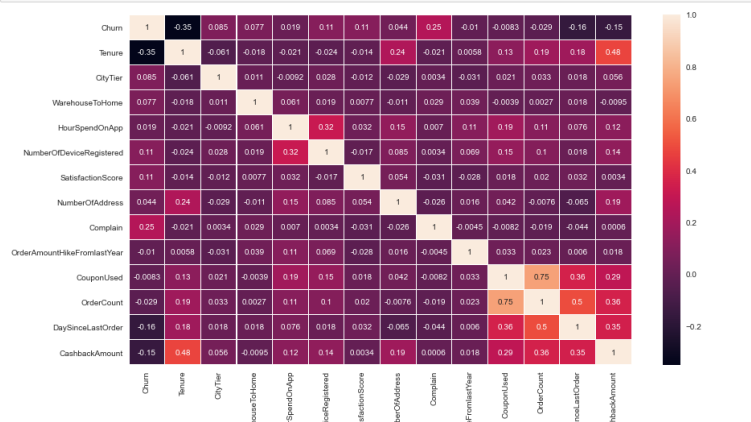
Variable types

Numeric
Categorical

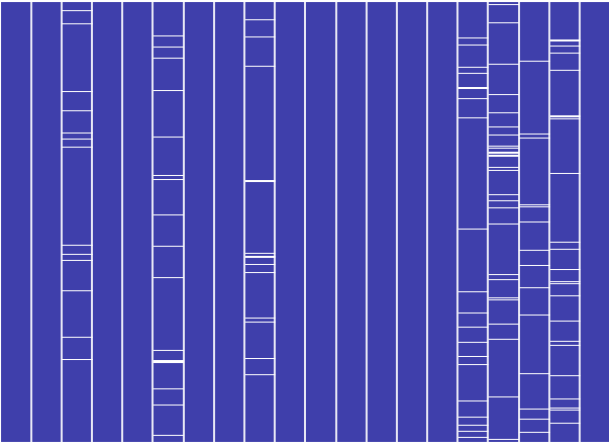
DATA TYPES



CORRELATION

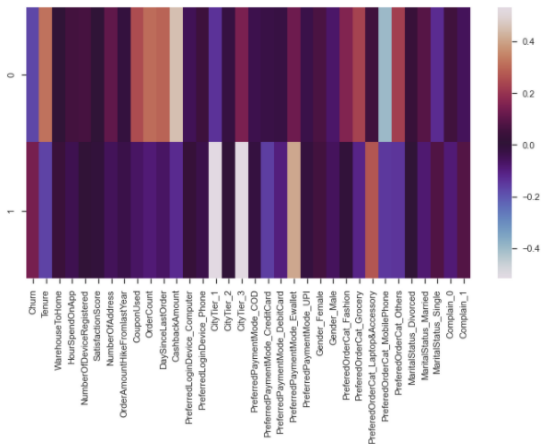
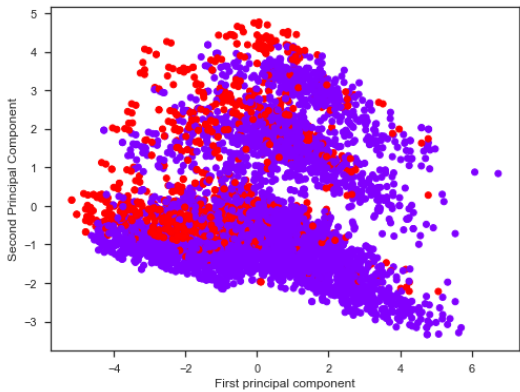


MISSING VALUES

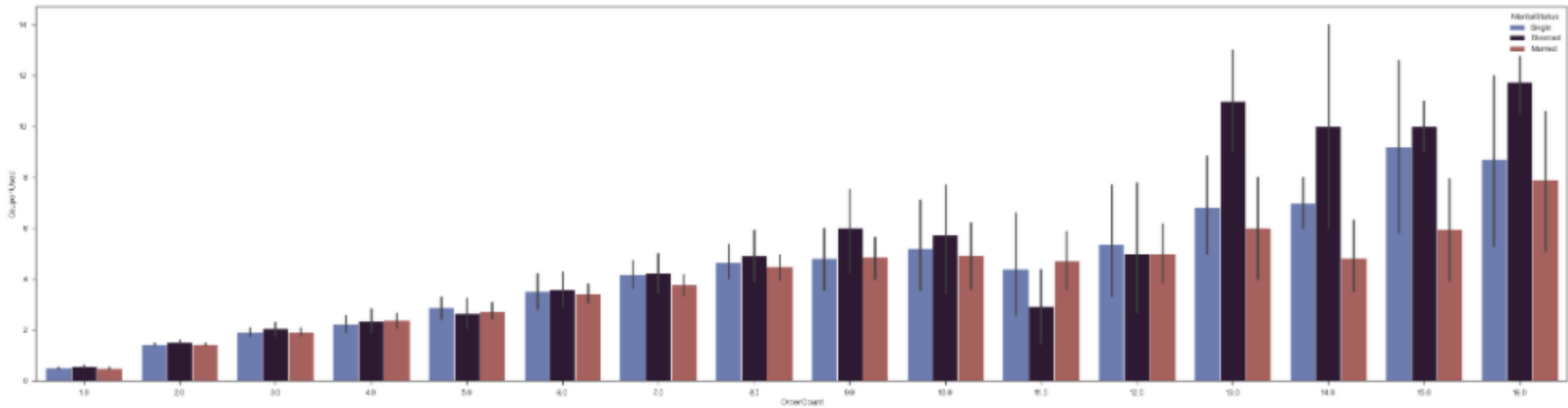


DATA EXPLORATION

PCA

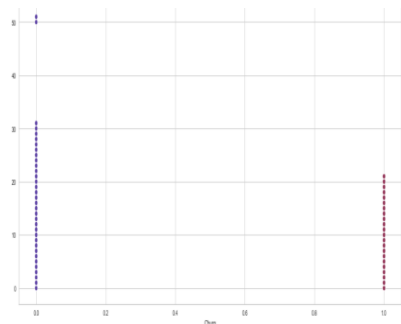


COUPONS VS ORDERCOUNT PER MARITAL STATUS

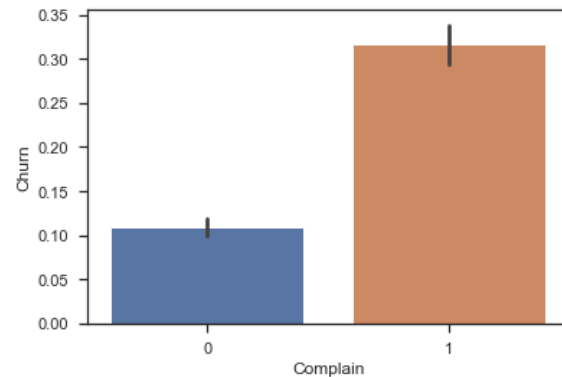


INSIGHTS (VERIFICATION OF HYPOTHESES)

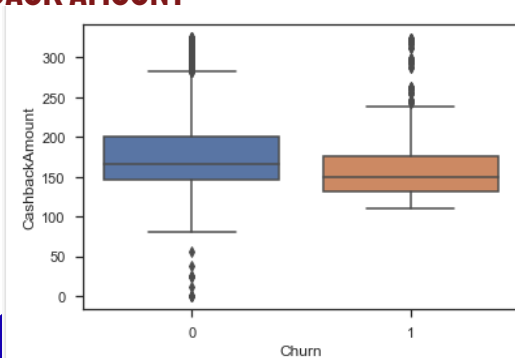
TENURE



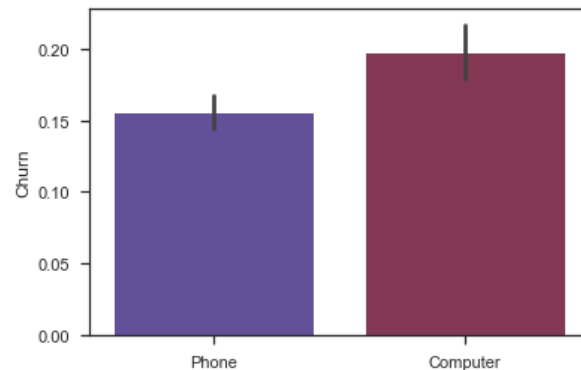
COMPLAIN



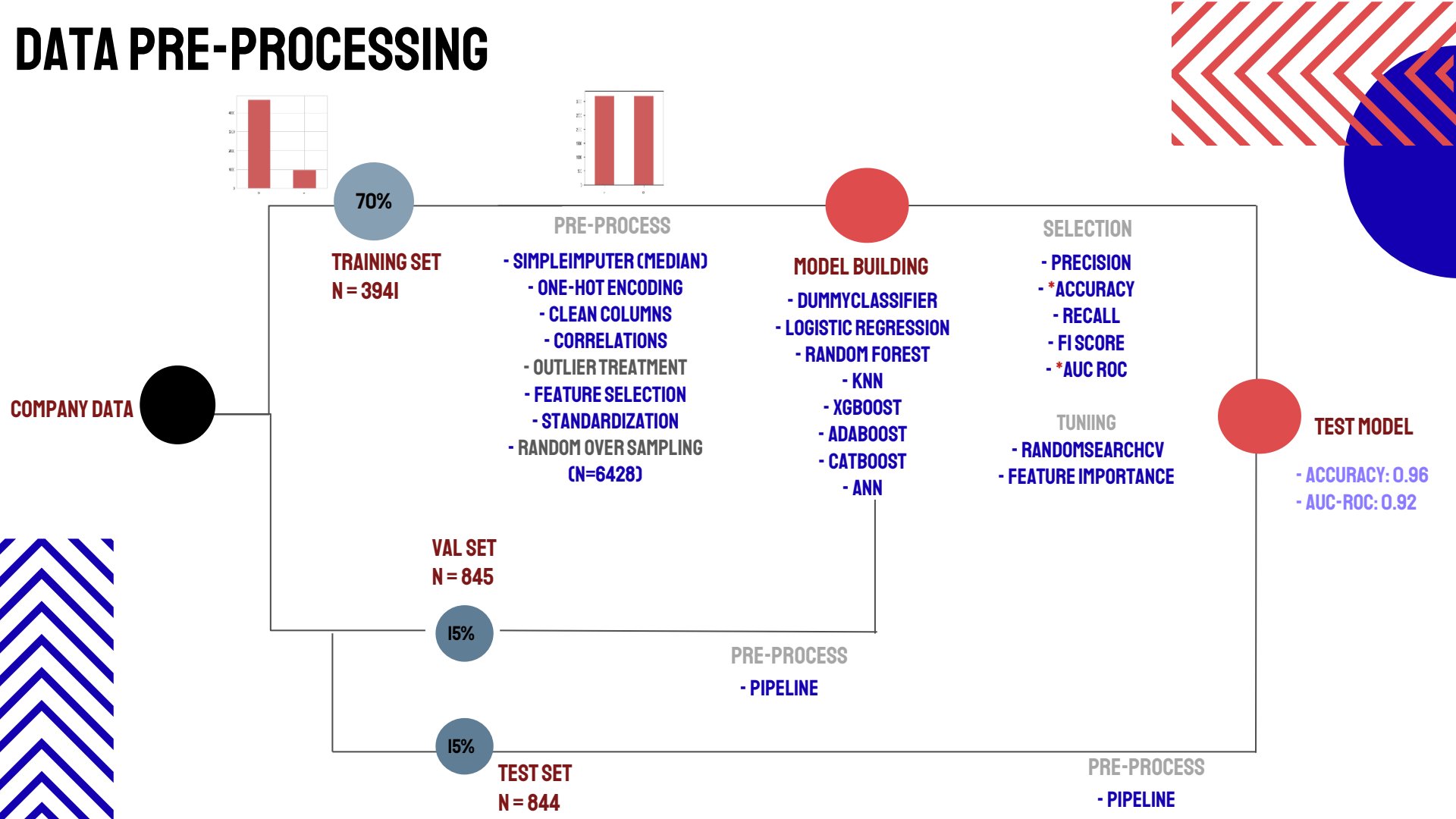
CASHBACK AMOUNT



ACCESS BY PHONE

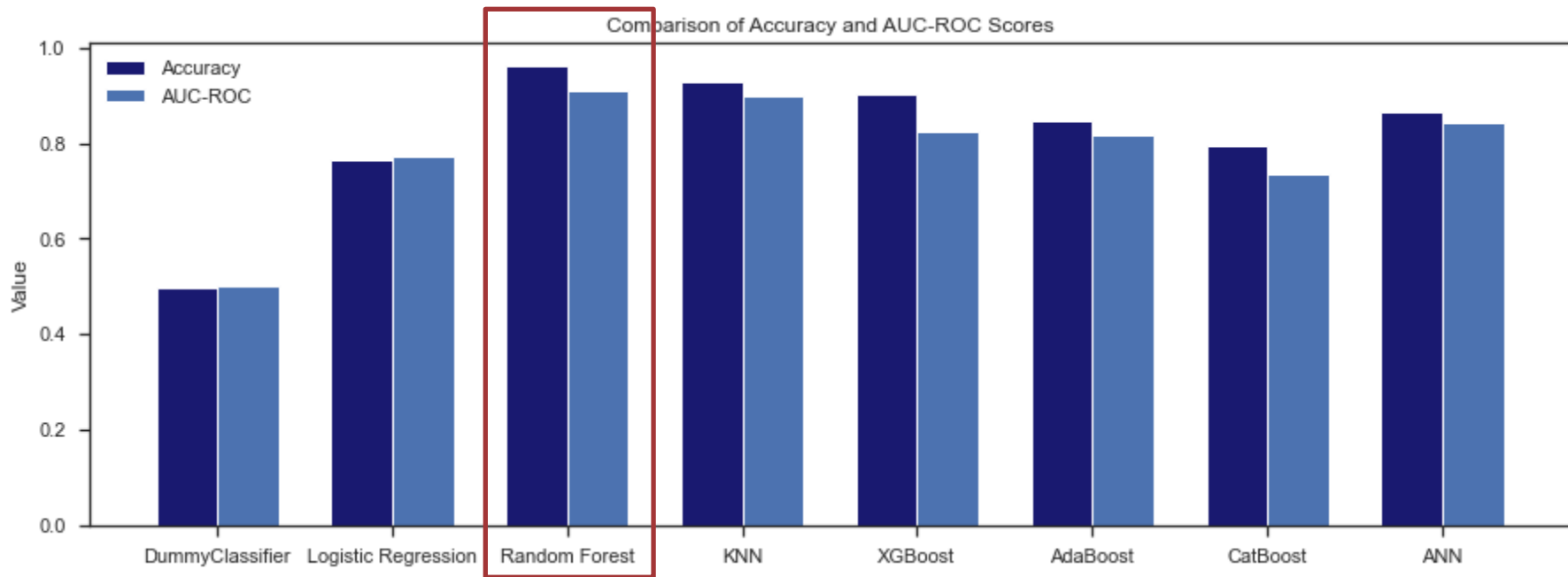


DATA PRE-PROCESSING



MODELLING

COMPARISON OF ACCURACY AND AUC-ROC SCORES

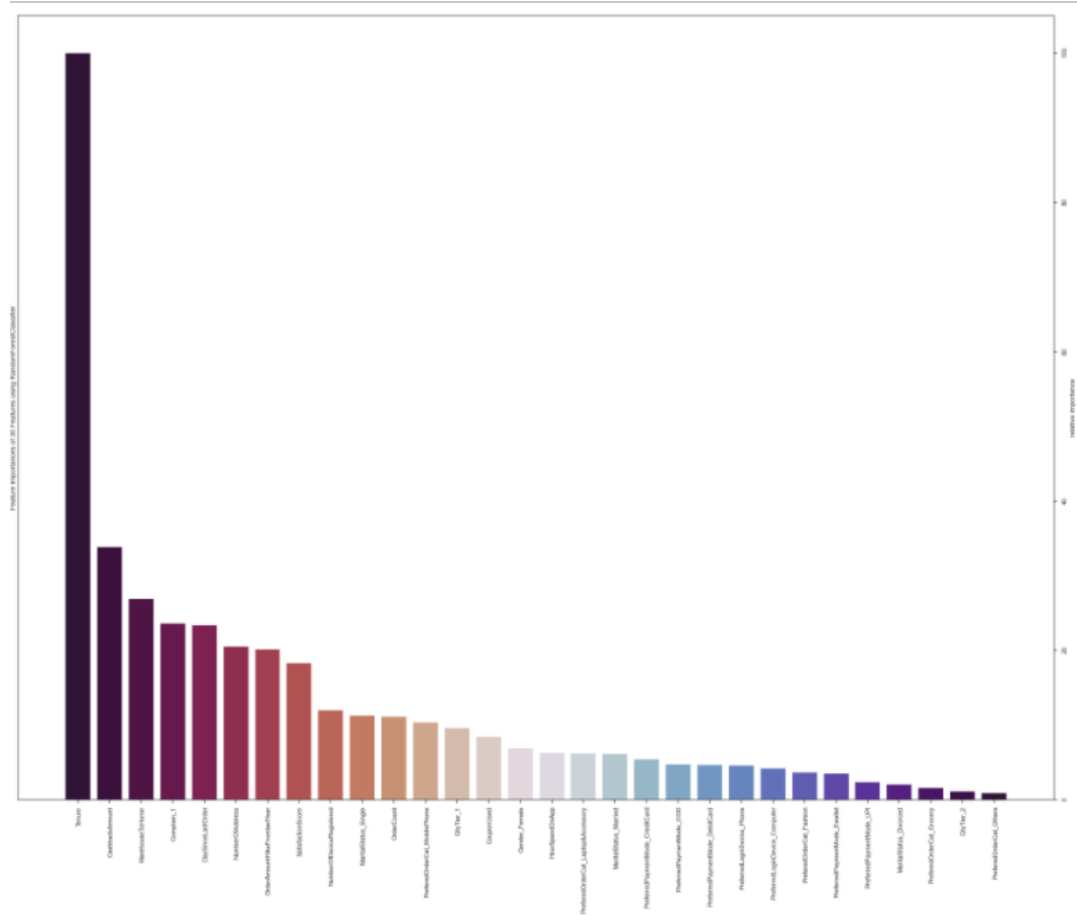


MODELLING

FEATURE IMPORTANCE

TOP 5 FEATURES:

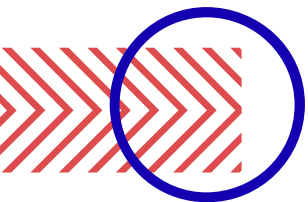
- Tenure
- CashbackAmount
- WarehouseToHome
- Complain
- DaysSinceLastOrder





04

RESULTS & CONCLUSION



USE CASE



CUSTOMER 1

- Gender: Female
- Login: Mobile Phone
- Payment: E-Wallet
- Marital Status: Divorced
- CityTier: 3
- Complain: 0
- Cashback Amount: \$160
- Hike from last year: 14%
- Hours on App: 2

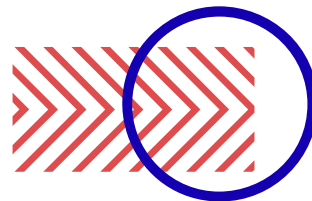
CHURN



CUSTOMER 2

- Gender: Male
- Login: Mobile Phone
- Payment: Credit Card
- Marital Status: Divorced
- CityTier: 1
- Complain: 1
- Cashback Amount: \$159
- Hike from last year: 13%
- Hours on App: 2

RETAIN



THREATS TO VALIDITY



SMALL DATA SIZE AND IMBALANCED DATA

Training data had 3941 observations

- Collect more data
- Be aware of imbalance data

DATA COLLECTION AND LIMITED VARIABLES

Unaware of data collection process.
Time-series data would have been beneficial

- Selection Bias could be present (CityTier 3)
- Would be valuable to capture time series data to evaluate the fluctuations of customer behavior over time

CONCLUSION VALIDITY

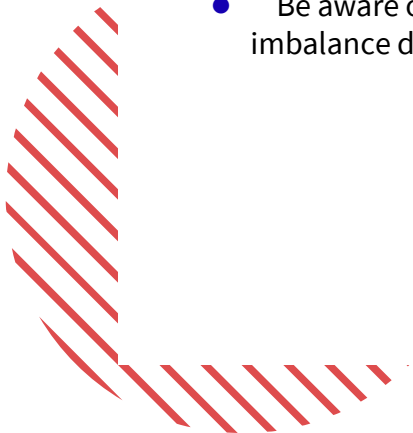
Need to statistically validate results from hypothesis tests

- In addition to the exploratory data task, it is important to conduct statistical analysis to accept or refute the hypotheses

UNCERTAINTIES

Exogeneous shocks (unforeseen events)

- Collect data from a wider time period to capture some market fluctuations and flexibility in user behaviour



RECOMMENDATIONS

BUSINESS/FINANCE

- Customer Churn → Revenue Churn. Understand spend (\$) per customer. Cost of retention for some customers may be less than the cost of acquisition of others.

MARKETING TEAM

- Target 'at-risk' customers. Use direct outreach, retention campaigns, discounts, loyalty programs and special offers to reel them back into your funnel.

DATA SCIENTISTS

- Classification → Regression. Instead of an absolute classification of churn, view it more as a scale [proba], and evaluate the health of the shopper

PRODUCT TEAM

- A/B Testing. Improve features on the laptop UI to reduce churn

- Improve Customer Service Team. Most people gave a rating of 3

- Customer Segmentation. High churn in CityTier3 and with Single customers.

- Do deeper analysis with clustering models

- Remove Cash-On-Delivery option, as this leads to the high churn.

- Customer Survey to understand reason behind customer churn: incidental vs deliberate churn

- Cohort Analysis – observe changes in the same group of customers who joined at the same time to see how long it takes on average to churn

Cohort	Month Since Purchase							
	0	1	2	3	4	5	6	7
1-Jan-2012	100%	81%	75%	71%	66%	65%	65%	63%
1-Feb-2012	100%	93%	92%	89%	81%	78%	76%	
1-Mar-2012	100%	86%	80%	73%	69%	67%		
1-Apr-2012	100%	95%	90%	90%	85%			
1-May-2012	100%	83%	77%	70%				
1-Jun-2012	100%	81%	75%					
1-Jul-2012	100%	84%						
1-Aug-2012	100%							



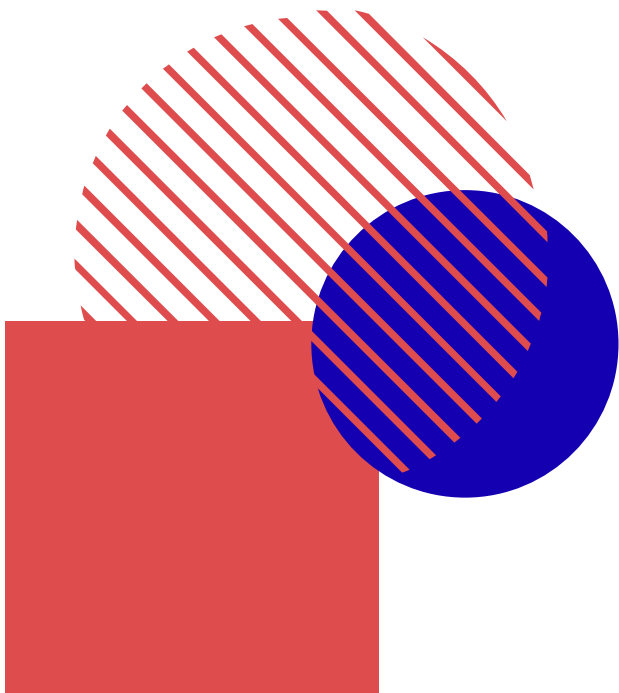
LESSONS LEARNED

- Importance of data pre-processing
- New visualization techniques
- How to identify best machine learning model
- Learned to perform different roles [project manager, data/business analyst, etc..]



NEXT STEPS

- Get a more encompassing time-series dataset that is more representative
 - Get industry specific data to find industry specific trends and patterns
 - Get customer feedback to gain a more in-depth analysis
 - Cohort Analysis
 - Customer Lifetime Value Analysis
 - Causal Inference
- 



THANKS!

<https://github.com/McGill-MMA-EnterpriseAnalytics/Churn-Analysis>