FIRST CITY MONUMENT BANK

RETAIL BANKING GROUP

Early High Value Detection Model Class Review.

Feb 2020

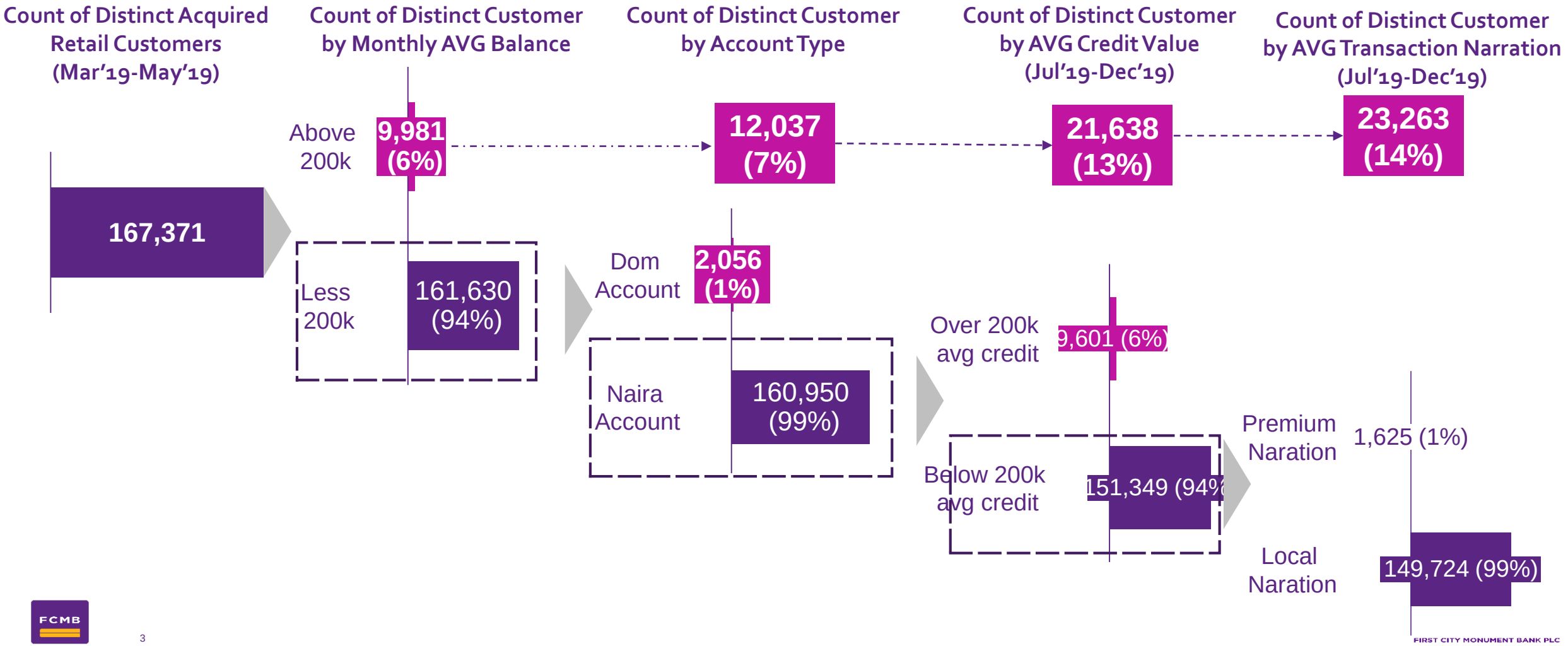
my bank and I

Project Objective

Project Objective	The project objectives is to demonstrate advanced analytics competence to FCMB by developing a Early High Value Detection Model.
Business Objective	To identify Retail customers any stage of their lifecycle who has the tendency of becoming an high value customer from the point of registration and possibly on their next customer induced transaction.
Model Objective	The model will classify newly enrolled customer has either premium customer of as regular transacting customer
Target Base	~167K customer who was enrolled between 1-mar-2019 and 31-May-2019 and transacted between July and December 2019 was use in training the model.

Early High Value Customer Segmentation Class

The parameters of the acquired high value customer between March 2019 and May 2019 with the respective cumulative percentage distribution of each parameter are Monthly Average Balance(6%), Account Type(7%), Average Credit Inflow(13%) and Average Transaction Narration(14%).
A total of 167,371 customers were acquired in which 14% qualified as high value customers based on the given parameters.



Typical transaction types common with Premium customers

A non-exclusive list of high value customer transaction narration.

Transactions	Key word search
Inflow from offshore	Inflow
Treasury bills	T-bill/ custody fee
Purchase of flight ticket	Ticket / Airline/ travel
Hotel booking	Hotel
Purchase of property/ rent payment	Properties/ property/ rent
Stockbroking	Stockbrokers
Bonds	Bondmmdiscount
School fees	School/ Tuition/Form A
Transfers for investment	Securities/ Afriinvest/ CSL stockbrokers/ booking/ liquidation
Contribution	Tithe

Principal Component/ Feature Importance

Feature importance is a statistical concept in data science determines the correlation between the input variables and the target variable. Important Features also known as Principal components give some information about the goodness of fit of a variable in a model

KNIME has 2 primary techniques of deciding principal components for a model. These are Chi-square and R-square. The third is a combination of both. The idea is to determine the statistical p-value which is a measure of the degree of significance of an independent(input) variable to a dependent (target) variable.

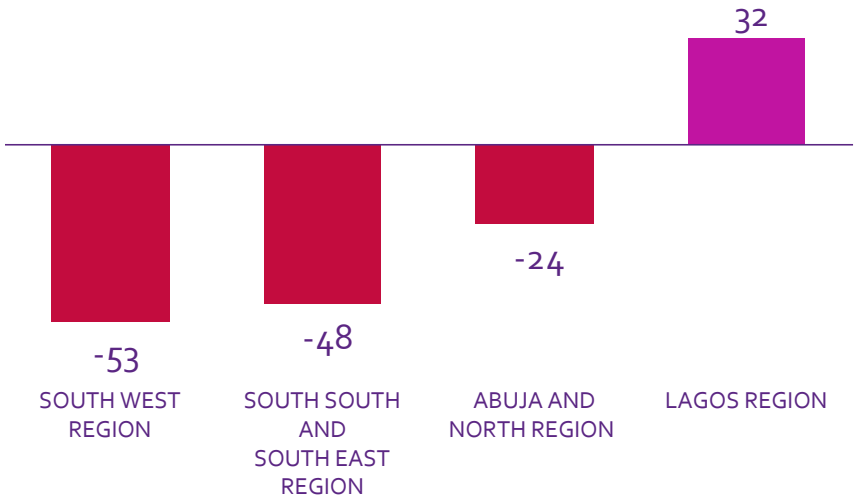
Chi-Square determines the degrees of freedom between an observed value(independent variable) and the expected value(dependent variable)
R-Square determines the correlation between an input variable or a set of input variables i.e. predictor(s) and a dependent variable. It indicates that whether a regression line perfectly fits the data being sampled. The R-square method removes variables that have large percentages of missing values, class variables that are based on the number of unique values, and variables in hierarchies. It also combines variables to assess how related they are to the target.

Data Robot on the other hand has a number of proprietary ways of determining the principal components of a model. In fact, with Data Robot, it appears as if the machine learning algorithm determine what features in its input dataset are the principal components. This means that the importance of a feature is a function of the techniques used in data modelling

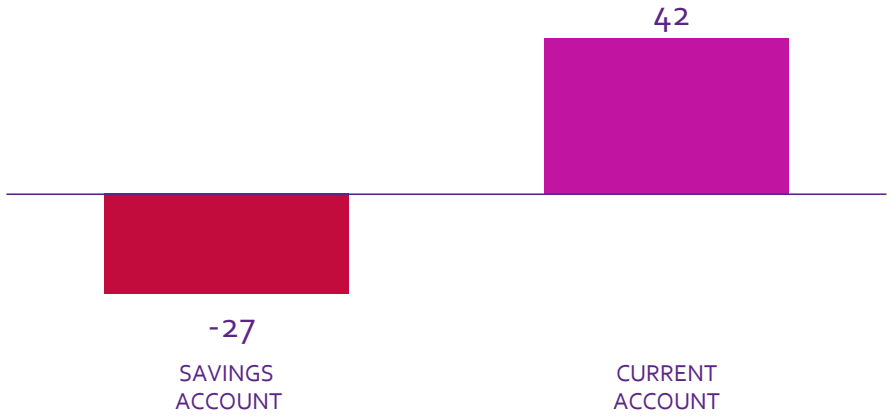
Early High Value Customer Segmentation Class

- South West/South-South and South-East/Abuja and North Region has marginal lower propensity of being a high value customer while Lagos Region has higher propensity of being a high value customer.

REGIONAL WEIGHT OF EVIDENCE

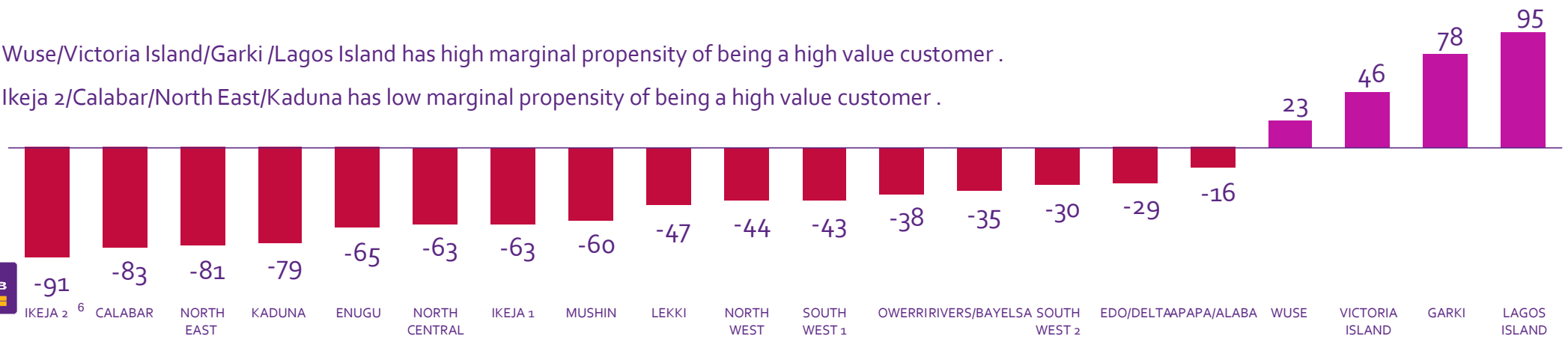


ACCOUNT TYPE WEIGHT OF EVIDENCE



BRANCH ZONE WEIGHT OF EVIDENCE

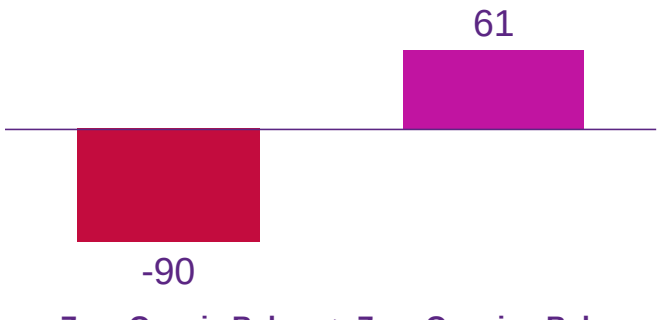
- Wuse/Victoria Island/Garki /Lagos Island has high marginal propensity of being a high value customer .
- Ikeja 2/Calabar/North East/Kaduna has low marginal propensity of being a high value customer .



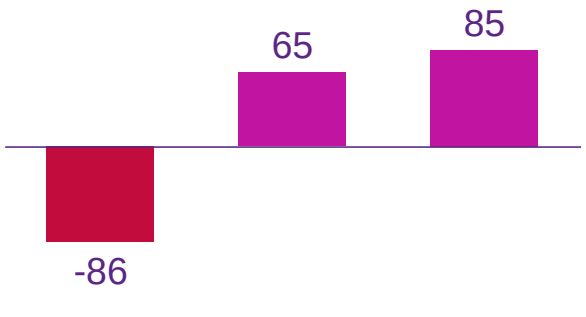
Early High Value Customer Segmentation Class

- Customers with Zero Opening Balance have low marginal propensity of being a high value customer while customers with Opening Balance >Zero have high marginal propensity of being a high value customer.
- Those who are active either in Inflow or Outflow have high marginal propensity of being a high value customer.
- Also customers who transact within 0-7 Days have high marginal propensity of being a high value customer.

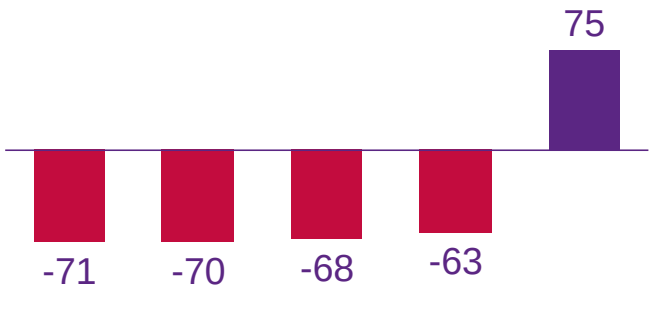
OPENING BAL. WEIGHT OF EVIDENCE



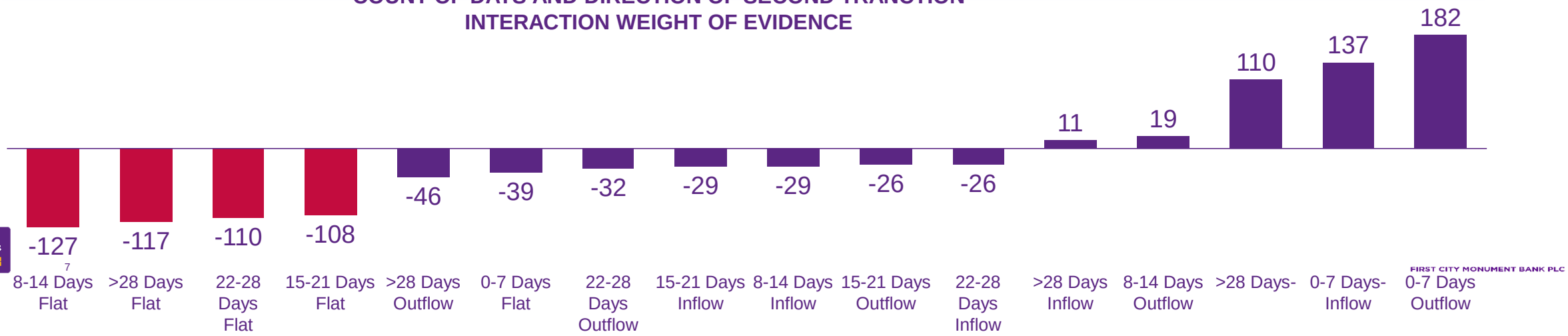
CUSTOMERS SECOND TRANSACTION WEIGHT OF EVIDENCE



NUMBER OF DAYS TO SECOND TRANSACTION WEIGHT OF EVIDENCE

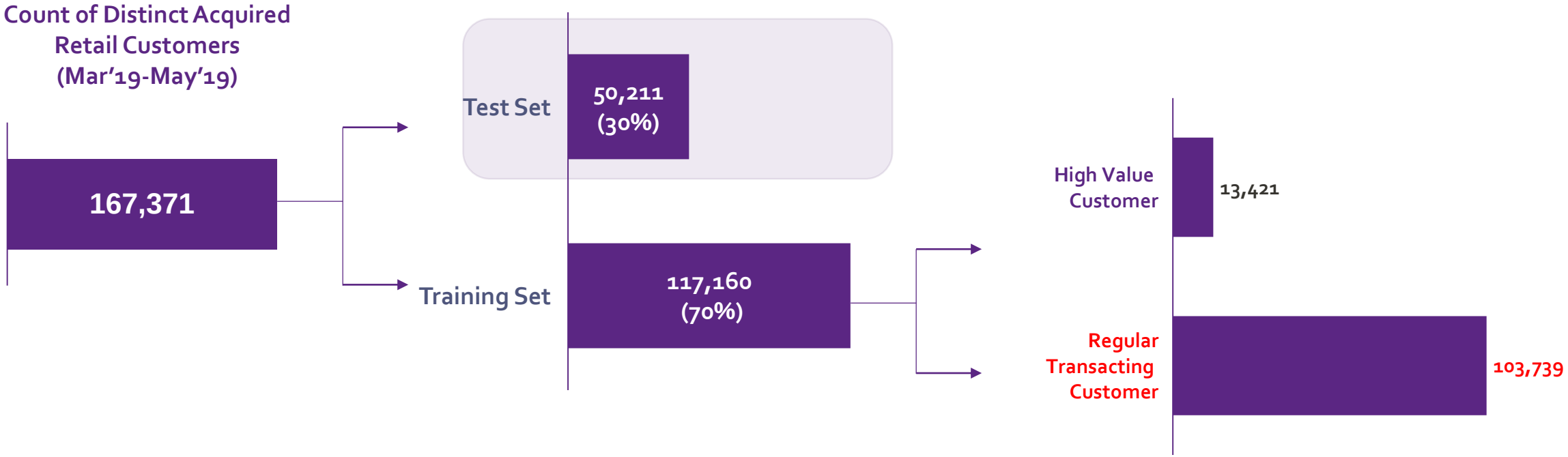


COUNT OF DAYS AND DIRECTION OF SECOND TRANCTION INTERACTION WEIGHT OF EVIDENCE



Early High Value Detection Model

Train-Test split proportion is 70:30, 91% accuracy was observed on the test set wit and average of 91% sensitivity.



MODEL ACCURACY STATISTICS

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specifity	F-meas...	Accuracy	Cohen'...
LV	55601	7519	59207	4214	0.93	0.881	0.93	0.887	0.905	?	?
HV	59207	4214	55601	7519	0.887	0.934	0.887	0.93	0.91	?	?
Overall	?	?	?	?	?	?	?	?	?	0.907	0.815

Thank you