**Mortgage Insurance Cross Sell Analysis**

▪ **Case Study:** A consumer bank with a range of products including mortgages, would like to cross sell its insurance products to its customer base based on customer portfolio data containing many fields

▪ **Table of Contents** 1. Introduction 2. Process flow Diagram 3. Hypothesis Generation – Customer segmentation for cross selling 4. Exploratory Data analysis

4.1 Data visualization using Python 5. Model Building – Generating Association Rules

5.1 Data Modelling using R 6. Metrics Evaluation 7. Sharing Insights and recommendations

▪ **Introduction** The sample data set from the customer portfolio data contains various fields about their product ownership, credit standing, have an outstanding mortgage, and insurance ownership (called as PPI / personal protection insurance),

The bank would like to adopt analytics driven approach for deciding: + who should they target for PPI, and + what type of PPI product they should be targeting them with

▪ **Process Flow Diagram** Kindly refer to attached PDF file

▪ **Hypothesis Generation** The problem is what to cross sell to , as well as whom to cross sell , Some key assumptions made were Ref denotes a customer id , the dataset is unique at this level, There are no repeated records of purchase of products , Hence we shall go ahead with Customer segment and calculating the product purchased per customer segment

▪ **Exploratory Data Analysis using python Library**

o **Importing the necessary libraries in python for performing analysis**

- *import pandas as pd*

**Mortgage Insurance Cross Sell Analysis**

- *import numpy as np* - *import seaborn as sns*

o **Loading the necessary Dataset into python**

- *Dataset = pd.read\_csv(‘C:\Data\Dataset.csv’)*

o **Analyzing the shape of the Data**

- *Dataset.shape()*(16386, 59) - ***The Dataset contains 59 features with 16386 observations***

o **Understand relationship among features using Correlation**

- Dataset.corr()

o **Key Observations from analyzing correlation HeatMap:**

- *Term and Net Advance shows a positive correlation with value (0.721)* - *Ref and Code shows a positive correlation(0.67)* - *There is a strong negative correlation observed between Total\_Outstanding\_CCJ*

*and Time\_since\_most\_recent\_outstanding (-0.841)*

**Mortgage Insurance Cross Sell Analysis**

- *A strong positive correlation was observed between Total\_value\_public\_info and*

*Time\_since\_most\_recent\_outstanding (0.91)*

o **Perform Univariate analysis on the Data**

- Univariate Analysis | **Categorical variable**(**Gender**) - Dataset['Gender'].value\_counts(normalize = **True**).plot.

Bar(figsize = (10,10),title = 'Gender') plt.xlabel('Gender') plt.ylabel('Applicants

It can be inferred from above bar plot that

➢ 60% of applicants in the Dataset are male ➢ 40% of applicants are female

**Mortgage Insurance Cross Sell Analysis**

- Independent Categorical variable**( Marital\_Status**) : - Dataset['Marital\_Status'].value\_counts(normalize =

**True**).plot.bar(figsize = (10,10),title = 'Marital\_status’)

It can be inferred from above bar plot that

➢ Around 58% of applicants are married ➢ Around 30% of applicants are separated ,close to 10% of them are divorced

- Independent Categorical variable –( **Bankruptcy**) - Dataset['Bankruptcy\_Detected\_\_SP\_'].value\_counts(normaliz

e = **True**).plot.bar(figsize=(10,10),title = 'Bankruptsy')

It can be inferred from above plot that,

➢ None of the applicants have history of Bankruptcy cases filed

**Independent Variable** | Categorical – (Number\_of\_Dependants)

- Dataset['Number\_of\_Dependants'].value\_counts(normalize = **True**).plot.bar(figsize=(20,6),title = 'Dependents')

**Mortgage Insurance Cross Sell Analysis**

It can be inferred from above plot that

➢ Around 70% of the applicants have no dependents ➢ Around 3% have dependents upto 5 members

- Independent Variable (**Employment Type**) - Dataset['Perm\_Temp\_Empl\_Ind'].value\_counts(normalize = **True**).plot.bar(figsize = (16,6),title = 'Employment type')

**Mortgage Insurance Cross Sell Analysis**

It is inferred from the above plot that

➢ All the applicants are mapped under permanent employment

➢ Around 80% of applicants are of low worst status wrt to outstanding mortgage

balance

o **Independent Variable | Numerical (Credit\_Score)**

- Dataset ['Credit\_Score'].hist

**Mortgage Insurance Cross Sell Analysis**

It is inferred that,

➢ The hist plot of Credit score confirms that data is normally distributed ➢ The box plot confirms the presence of outliers /extreme values .This can be

attributed to parameters in determining applicants score ➢ Let us categorize Credit score based on loan type ➢ Dataset.boxplot(column = 'Credit\_Score', by =

'Loan\_Type')

**Mortgage Insurance Cross Sell Analysis**

o **Outstanding Mortgage Balance**

- Dataset['Outstanding\_Mortgage\_Bal'].plot.box(figsize =

(16,5))

**Inferences:** - Few of the observations show extreme values which shows that many

applicants have outstanding payments to be made

**Mortgage Insurance Cross Sell Analysis**

▪ **Data Modelling Approach :**

1. The gist of problem is what to cross sell and whom to cross-sell , the first part we can use

Apriori algorithm and interpret the outcomes of the association rules

**Code Walkthrough:**

# Load required libraries

**library**(dplyr)

**library**(arules)

**library**(plotly)

**library**(arulesViz)

**library**(visNetwork)

**library**(igraph)

**library**(reshape2)

# Read the dataset

df <- **read.csv**("C:/Data/ Dataset.csv",

header = TRUE, stringsAsFactors = FALSE)

# Change the description to lowercase

df$Insurance\_Description <- **tolower**(df$Insurance\_Description)

# Specify columns to work with

required\_cols <- **c**("Category", "Insurance\_Description")

# Filter data and select only required columns. Filtering on PPI=1 as only

**Mortgage Insurance Cross Sell Analysis**

# they have product info.

required\_df <- df %>% **filter**(PPI == 1) %>% **select**(required\_cols)

# Wrange data to be usable as transaction data

products <- **aggregate**(Insurance\_Description ~ Category, required\_df,

c)

transactions <- **as**(products$Insurance\_Description, "transactions")

# Run apriori algorithm to get rules

rules <- **apriori**(transactions, parameter = **list**(supp = 0.006, conf = 0.25, minlen = 2))

# Sort the rules according to decreasing confidence

rules <- **sort**(rules, by = "confidence", decreasing = TRUE)

# Remove the redundant rules

rules <- rules[!**is.redundant**(rules)]

rules\_df <- **as**(rules, "data.frame")

# Output the rules into a csv file

**write.csv**(rules\_df, “C:/output/Output.csv")

**Insights and Outcomes :**

rules support confidence lift count

1 {bronze} => {1st cust-lasu/ 2nd c} 0.01666667 1 5.454545 1

2 {bronze} => {life & critical illn} 0.01666667 1 1.935484 1

3 {bronze} => {lasu} 0.01666667 1 1.500000 1

4 {bronze} => {joint} 0.01666667 1 1.333333 1

**Mortgage Insurance Cross Sell Analysis**

5 {bronze} => {life & ci} 0.01666667 1 1.153846 1

6 {bronze} => {single} 0.01666667 1 1.034483 1

The interpretation of these rules is that those who have purchase **bronze** are **1.935** times more likely to buy **life & critical illn**.