Cross-Sell Insurance Analysis

Personal Protection Insurance(PPI)

Suprasanna Pradhan

Business Problem Understanding

Problem Statement:

A consumer bank wants to study and target the their customers who is not holding the personal protection insurance (PPI) product to those customers who have a secured or unsecured type of loan, but no PPI product as yet.

The provided sample data set from their customer portfolio containing various fields about their product ownership, credit standing, outstanding amounts, and whether they already have an insurance product and if not find out the reason from targeted variables

Benefits of Study:

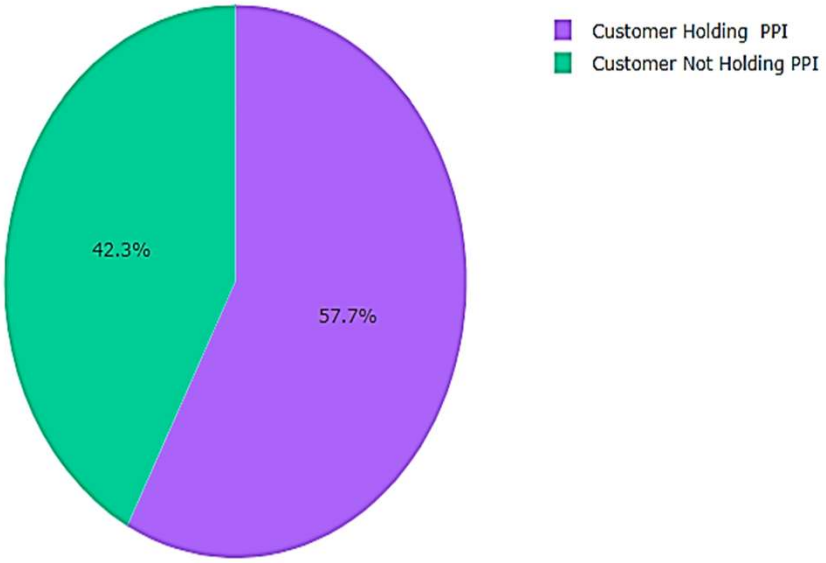
The bank would like to adopt analytics driven approach applied on this sample data for deciding:

- The quality of insights to be highlighted the major parameters (variable) having business influence to holding PPI or not holding PPI
- Who should they target from the pool of customers that currently do not have a PPI, and
- What type of PPI product they should be targeting them with..

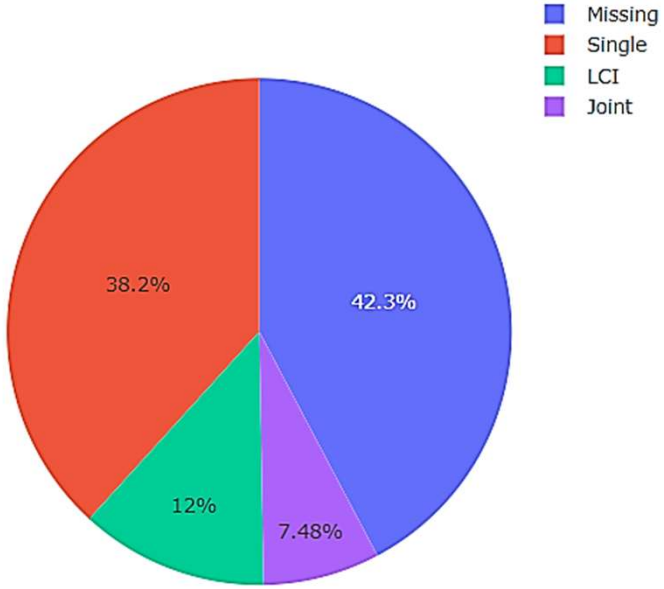
Insights from Analysis

- Almost 42% of the customers didn't hold a PPI product and for these customers category and insurance description was missing
- Single, LCI, Joint were the other values in the category variable with 38%, 12% and 7% share respectively

Distribution of PPI (Holding vs Non-holding Customers)



Distribution of Category Variable Count



Insights from Analysis - I

- **A, X and B** were final grade were leading to most of the PPI products being bought by the customers
- Data has higher number of **unsecured loan** preferring customers
- Data had higher proportion of **Married** and **Male** customers
- Percentage of Female customers having a PPI product was 2% higher than Female customers not having it
- About 80% of the customers were without any access cards
- Customers without access cards in the total customers with **PPI products were ~3% higher** than customers without access cards and without PPI products
- Customers with a PPI hold a cheque guarantee about 2% higher than the customers without a PPI
- Almost a 2% difference in total number of customers holding other credit card store when compared against customers having a PPI product or not having it
- Customers who were having a PPI, **product description is available for 98%** of them and rest didn't had any product description
- All customers currently holding a **PPI product, had insurance description available**, hence this variable is likely to be given a priority above product description

Insights from Analysis -II

- Age group **33-39 & 48-55 hold 37% of PPI products** and age group 18-44 can be targeted the most for selling PPI products
- Customers **between 1 & 5 (inclusive) income range** are likely to prefer PPI products
- About 50% of the terms were of 50-65 months irrespective of the PPI flag
- Close to **70% of the customers had 0-99 outstanding mortgage balance** irrespective of the PPI flag
- Customers staying at the **address between 0 and 260 days** can be prioritized for cross-selling PPI products
- Customers with total number of accounts ranging from **1 to 6 (inclusive) are likely to prefer PPI products**
- Average of total accounts for customers holding PPI was highest in Joint insurance product
- Customers with total of accounts **= 9 / life & critical illness insurance product** can be preferred while offering PPI
- PPI products are more likely to be sold to customers with a credit score **between 792 and 918**
- PPI products can be targeted to customers with a value of property **b/w 115k and 120k**, accounts for **25%-30%** of the total customers (with PPI & w/o a PPI product)

Insights from Analysis -III

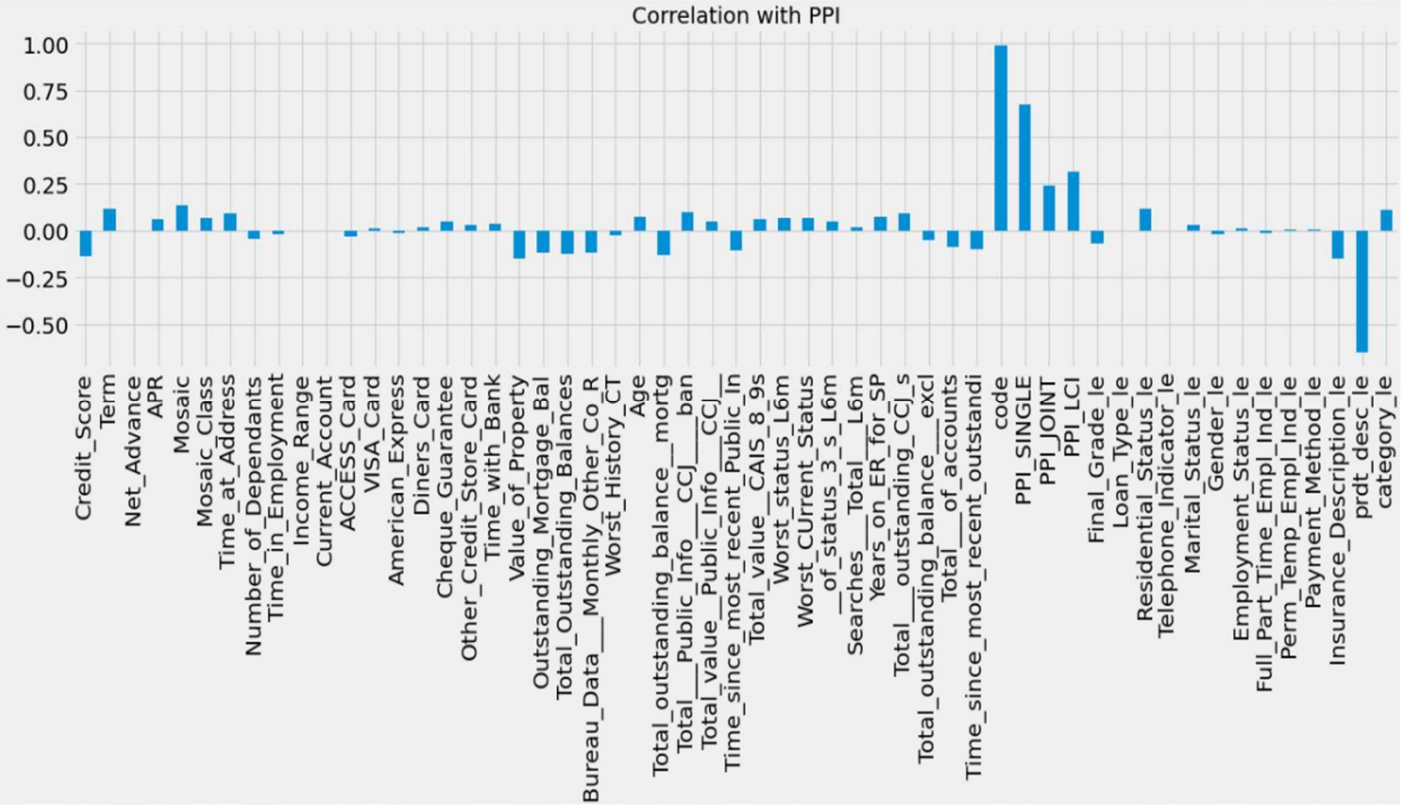
5 most Important Factors Influencing PPI

Most Positive Correlations

Index	Target
PPI_JOINT	0.24
PPI_LCI	0.32
PPI_SINGLE	0.67
code	0.99

Most Negative Correlations

Index	Target
prdt_desc_le	-0.65
Insurance_Description_le	-0.15
Value_of_Property	-0.15
Credit_Score	-0.14
Total Outstanding_balance__mortg	-0.13



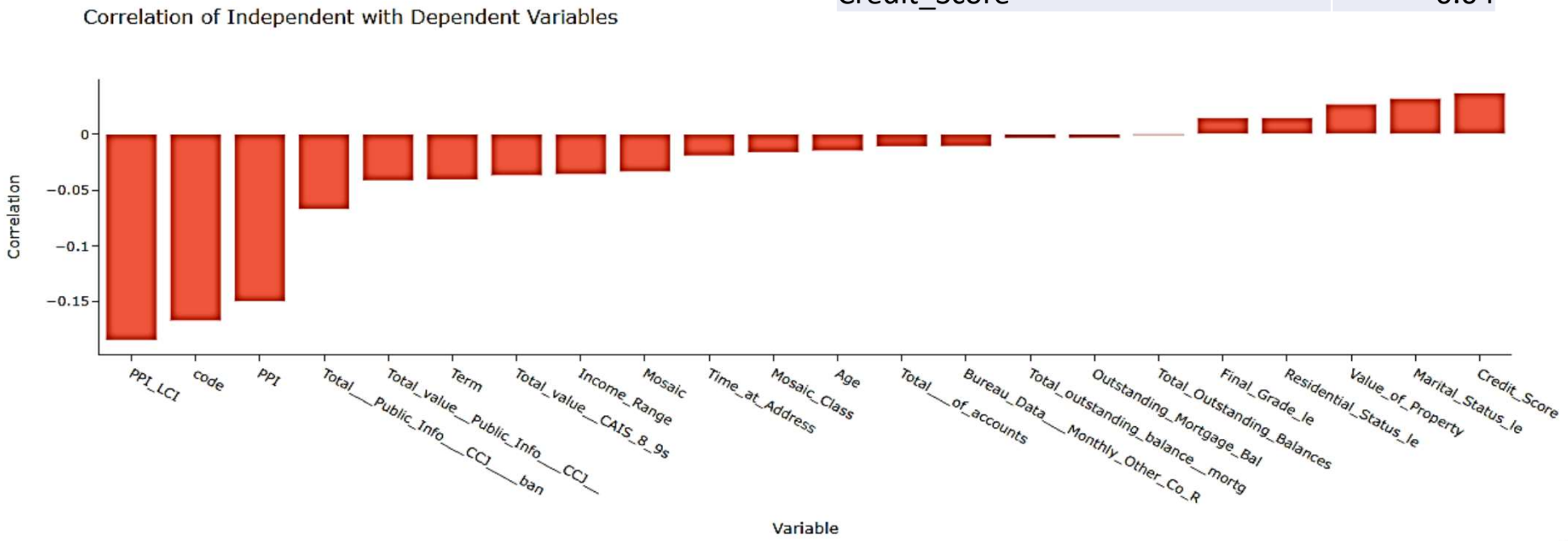
Method Used: Co-relation Bar plot

Insights from Analysis -IV

5 most Important Interpreter for Insurance Description

- Insurance_Description variable is used to find out the target for those customers not holding PPI

Most Positive Correlations	
Final_Grade_le	0.01
Residential_Status_le	0.01
Value_of_Property	0.03
Marital_Status_le	0.03
Credit_Score	0.04



Modelling Approach Used

- Feature Reduction:
- Correlations matrix
- Weight of Evidence and Information Value

- ANOVA Test, F-Test for Classification:
- P value to appropriate variable (Customer not holding PPI)

- Recommended Model :
- Logistic Regression
- Random tree Classifierr

- Market Basket Analysis:
- Along with Value_of_Property/ Income range

Model Output

- The model is prepared to predict for the Customers, those are not holding PPI (PPI=0) with targeted variable Insurance_Description (y).The performance of Random forest is occurred highest score comparing to other ML models

Model	Accuracy	Precision	Recall	F1 Score	ROC
Logistic Regression	0.937852	0.939676	0.997917	0.967921	0.500983
Stochastic Gradient Descent	0.936628	0.941336	0.994531	0.967203	0.515484
K-Nearest Neighbour	0.938341	0.94057	0.997396	0.96815	0.508819
Support Vector Machine	0.939564	0.939564	1	0.968841	0.5
Random tree Classifier	0.986787	0.986883	0.999219	0.993012	0.896371

- In Market Basket Analysis we have used two variables Value_of_Property and Income range to get the best association rules

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
(Joint)	(Single)	0.317557	0.352672	0.233588	0.735577	2.085727	0.121594	2.448078
(Single)	(Joint)	0.352672	0.317557	0.233588	0.662338	2.085727	0.121594	2.02108
(LASU)	(Joint)	0.396947	0.317557	0.239695	0.603846	1.901535	0.113641	1.722671
(Joint)	(LASU)	0.317557	0.396947	0.239695	0.754808	1.901535	0.113641	2.459512
(Joint)	(Others)	0.317557	0.441221	0.258015	0.8125	1.841479	0.117902	2.980153
(Others)	(Joint)	0.441221	0.317557	0.258015	0.584775	1.841479	0.117902	1.64355
(Joint)	(Life & Critical Illn)	0.317557	0.459542	0.264122	0.831731	1.809912	0.118191	3.211865
(Life & Critical Illn)	(Joint)	0.459542	0.317557	0.264122	0.574751	1.809912	0.118191	1.604807
(Missing)	(Joint)	0.766412	0.317557	0.282443	0.368526	1.160502	0.039063	1.080714
(Joint)	(Missing)	0.317557	0.766412	0.282443	0.889423	1.160502	0.039063	2.112446

Communication Results

- **Insurance category almost 42 %** details are missing ,not recommended to leave blank .Missing says something is missed here.
- Payment_Method, Bankruptcy_Detected__SP_, CIFAS_detected variables had no impact in deciding **whether a PPI is bought or not**
- Also Telephone_Indicator, Full_Part_Time_Empl_Ind, Perm_Temp_Empl_Ind, Current_Account, VISA_Card, American_Express, Diners_Card **not playing any vital role to hold a PPI**
- Suggested, we will go with **Random Forest Classifier model showed as a higher F1 and ROC** score compared to other models.
- According to market basket analysis we found **Joint with Single and LASU with Joint goes well ,** Other word the customer Purchases **Joint and LASU** are likely to purchase **Single and Joint**

Source Code

Github: Suprasanna

Thank You