Prepared By: Pratik Sharma

Cross Sell

Personal Protection Insurance

Cross-sell the Personal Protection Insurance (PPI) product to customers who have a secured and unsecured loan, but no PPI product as yet.

Table of Contents

PROBLEM STATEMENT & OBJECTIVE !

Problem Statement

A consumer bank with a range of products would like to cross-sell insurance to its consumer base (that is, cross-sell the personal protection insurance (PPI) product to those customers who have a secured or unsecured type of loan, but no PPI product as yet).

Data is provided for their customer portfolio containing various fields about their product ownership, credit standing, outstanding amounts, and whether they already have an insurance product (called as PPI / personal protection insurance), if any, the type of PPI product they own.

Objective

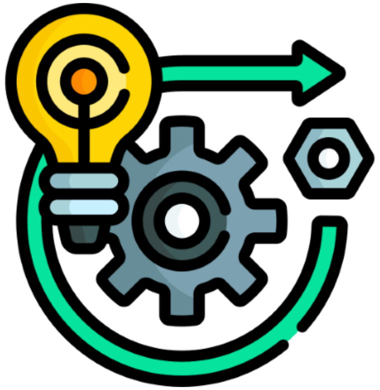
Analytics driven approach to identify:

- ✓ 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

APPROACH !

Exploratory Data Analysis

Data Cleaning, Univariate, Bivariate and Multivariate analysis; Identify insights for cross-selling the PPI products



Feature Reduction & Selection

Pearson Pairwise Correlation, Correlation with the target variable, Weight of Evidence and Information Value

Machine Learning Classification Model

Train, validate and test; Evaluate different classification models; Build a CatBoost classifier to predict Insurance Description

Market Basket Analysis

Apriori algorithm & Association rules to identify possible associations by number of antecedents and consequents; helps us in identify group of products that can be cross sold

Dataset Description !

- ✓ Shape of the data: **(16383, 59)**
- There were **no duplicates found** in the data
- There were **22 categorical** fields and **37 numerical** fields in the data
- **Numerical fields:**
  - **25** variables had **highly** skewed distribution
  - **4** variables had **moderately** skewed distribution
  - **8** variables had **fairly symmetrical** distribution
- **Categorical fields:**
  - More than **50% of the variables** had 2 unique values
  - Insurance_Description, prdt_desc, category variables had **~42% missing values**

- **List of numerical variables with highly skewed distribution:** 'Term', 'Net_Advance', 'Mosaic_Class', 'Time_at_Address', 'Number_of_Dependants', 'Time_in_Employment', 'Value_of_Property', 'Outstanding_Mortgage_Bal', 'Total_Outstanding_Balances', 'Bureau_Data__Monthly_Other_Co_R', 'Total_outstanding_balance__mortg', 'Total__Public_Info__CCJ___ban', 'Total_value__Public_Info__CCJ__', 'Time_since_most_recent_Public_In', 'Total_value__CAIS_8_9s', 'Worst_status_L6m', 'Worst_CUrrent_Status', '__of_status_3_s_L6m', 'Years_on_ER_for_SP', 'Total__outstanding_CCJ_s', 'Total_outstanding_balance__excl', 'Time_since_most_recent_outstandi', 'code', 'PPI_JOINT', 'PPI_LCI'

- **List of numerical variables with moderately skewed distribution:** 'APR', 'Income_Range', 'Time_with_Bank', 'Searches__Total__L6m'

- **List of numerical variables with fairly symmetrical distribution:** 'Ref', 'Credit_Score', 'Mosaic', 'Worst_History_CT', 'Age', 'Total__of_accounts', 'PPI', 'PPI_SINGLE'

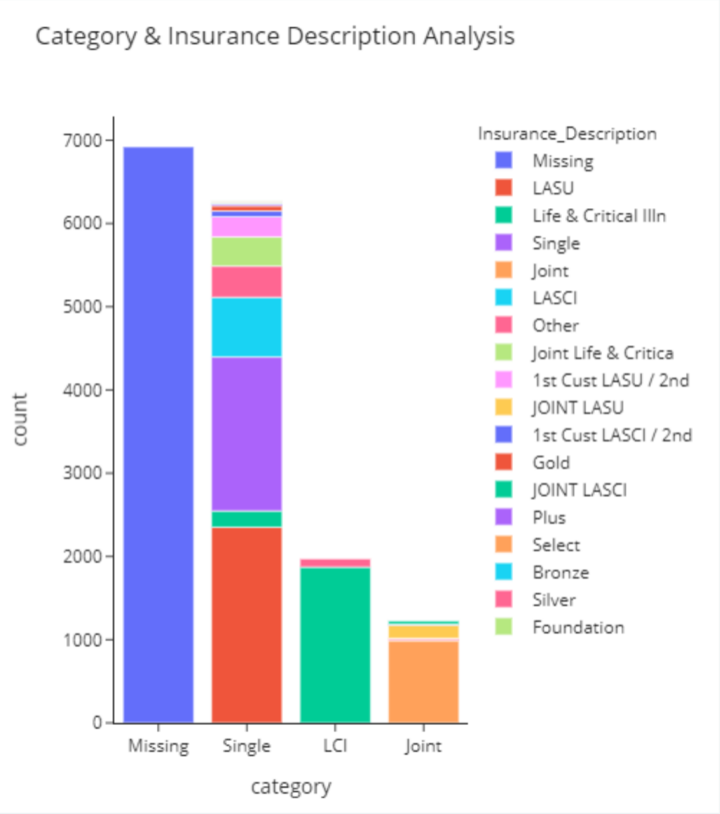
- **List of numerical variables with negative values:** 'Age', 'Total_outstanding_balance__mortg', 'Total__Public_Info__CCJ___ban', 'Total_value__Public_Info__CCJ__', 'Time_since_most_recent_Public_In', 'Total_value__CAIS_8_9s', '__of_status_3_s_L6m', 'Years_on_ER_for_SP', 'Total__outstanding_CCJ_s', 'Total_outstanding_balance__excl', 'Total__of_accounts', 'Time_since_most_recent_outstandi'

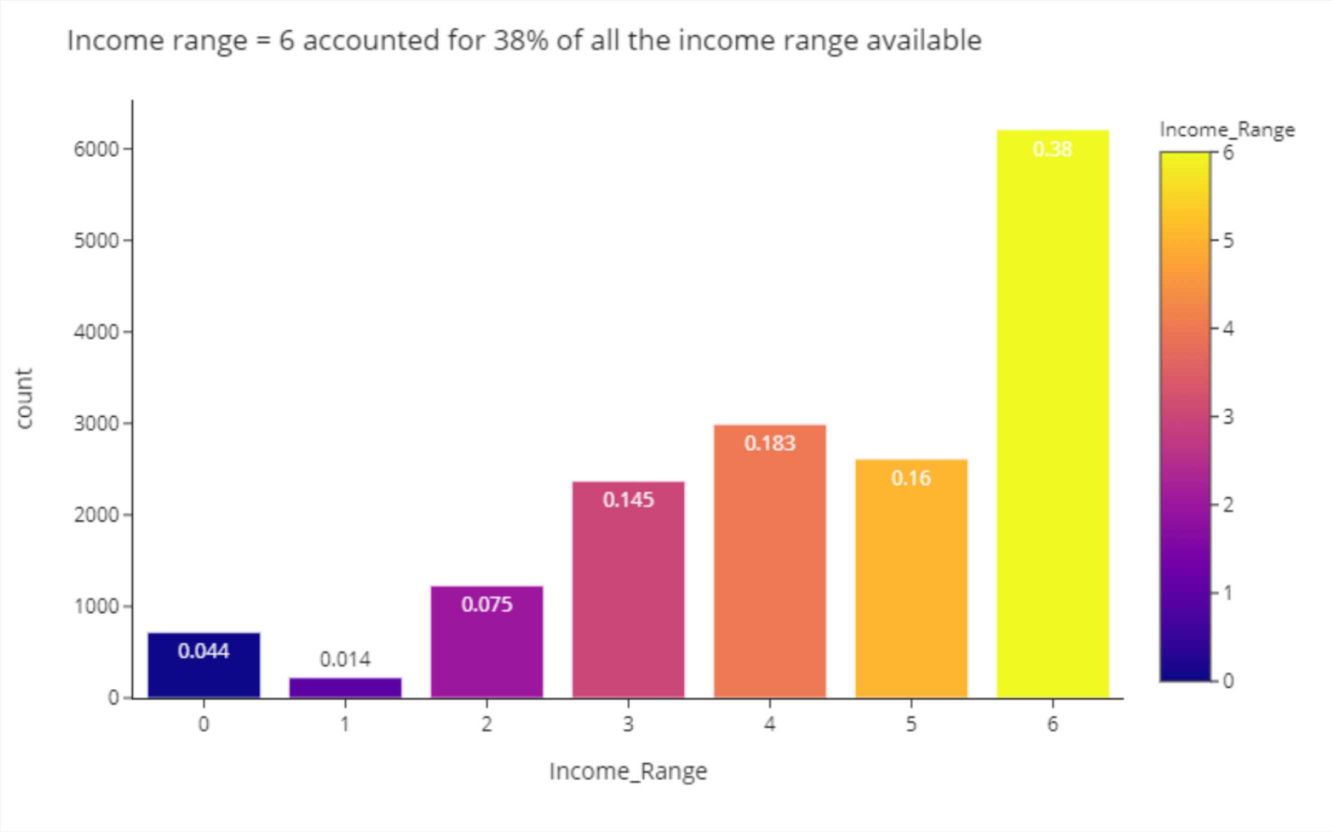
- **List of numerical variables with outliers:** 'Credit_Score', 'Term', 'Net_Advance', 'Mosaic', 'Mosaic_Class', 'Time_at_Address', 'Number_of_Dependants', 'Time_in_Employment', 'Time_with_Bank', 'Value_of_Property', 'Outstanding_Mortgage_Bal', 'Total_Outstanding_Balances', 'Bureau_Data__Monthly_Other_Co_R', 'Age', 'Total_outstanding_balance__mortg', 'Total__Public_Info__CCJ___ban', 'Total_value__Public_Info__CCJ__', 'Time_since_most_recent_Public_In', 'Total_value__CAIS_8_9s', 'Worst_status_L6m', 'Worst_CUrrent_Status', '__of_status_3_s_L6m', 'Searches__Total__L6m', 'Years_on_ER_for_SP', 'Total__outstanding_CCJ_s', 'Total_outstanding_balance__excl', 'Time_since_most_recent_outstandi', 'code', 'PPI_JOINT', 'PPI_LCI'

- **List of numerical variables without any outliers:** 'Ref', 'APR', 'Income_Range', 'Worst_History_CT', 'Total__of_accounts', 'PPI', 'PPI_SINGLE'

Some of the findings from the EDA ! (1/4)

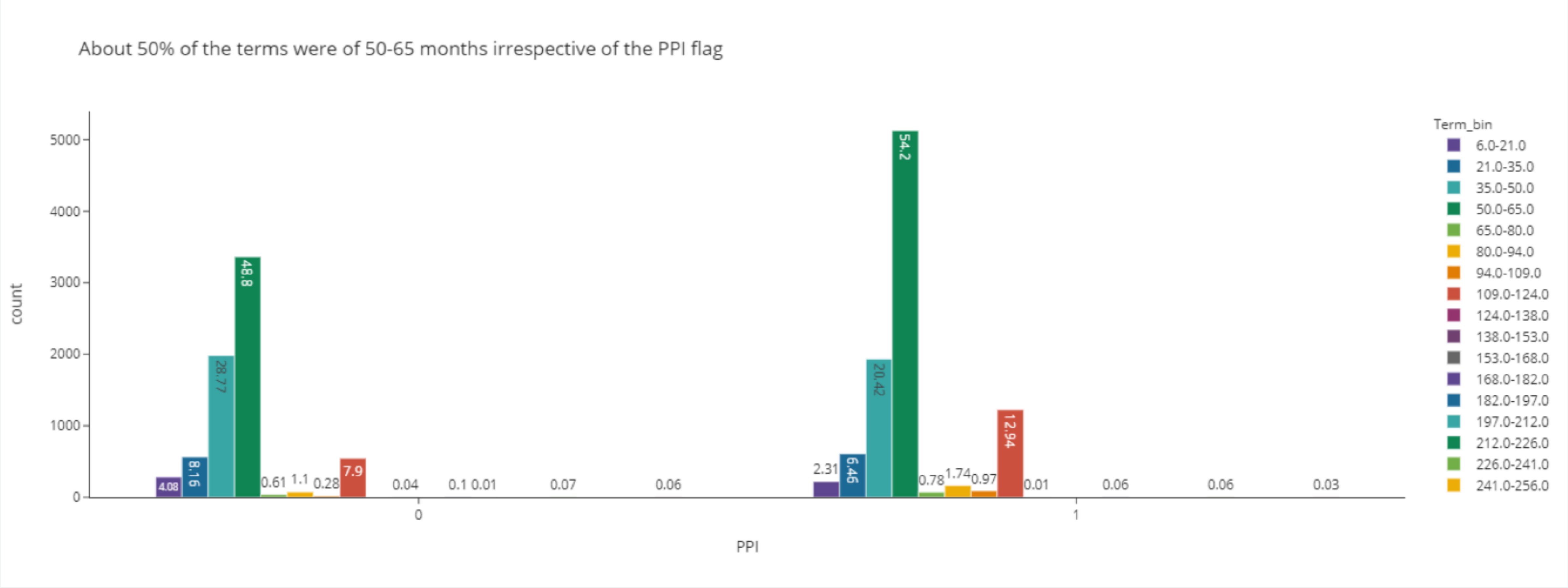
- For all categorical variables, **any value below 5% in proportion was considered as "Others"** provided that the column has more than 3 values
- For numerical variables, **binning was performed** for variables such as Age, Term, Total_outstanding_balance_mortg, Time_at_Address, Credit_Score and Value_of_Property
- Almost **42% of the customers didn't hold a PPI product** (our target group) and for these customers category and insurance description was missing
- **Bankruptcy was detected for 24 customers and CIFAS was detected for 83 customers** (exclusive sets)
- For target population i.e. **PPI=0, customers with no bankruptcy and no CIFAS detected were considered**





Some of the findings from the EDA ! (2/4)

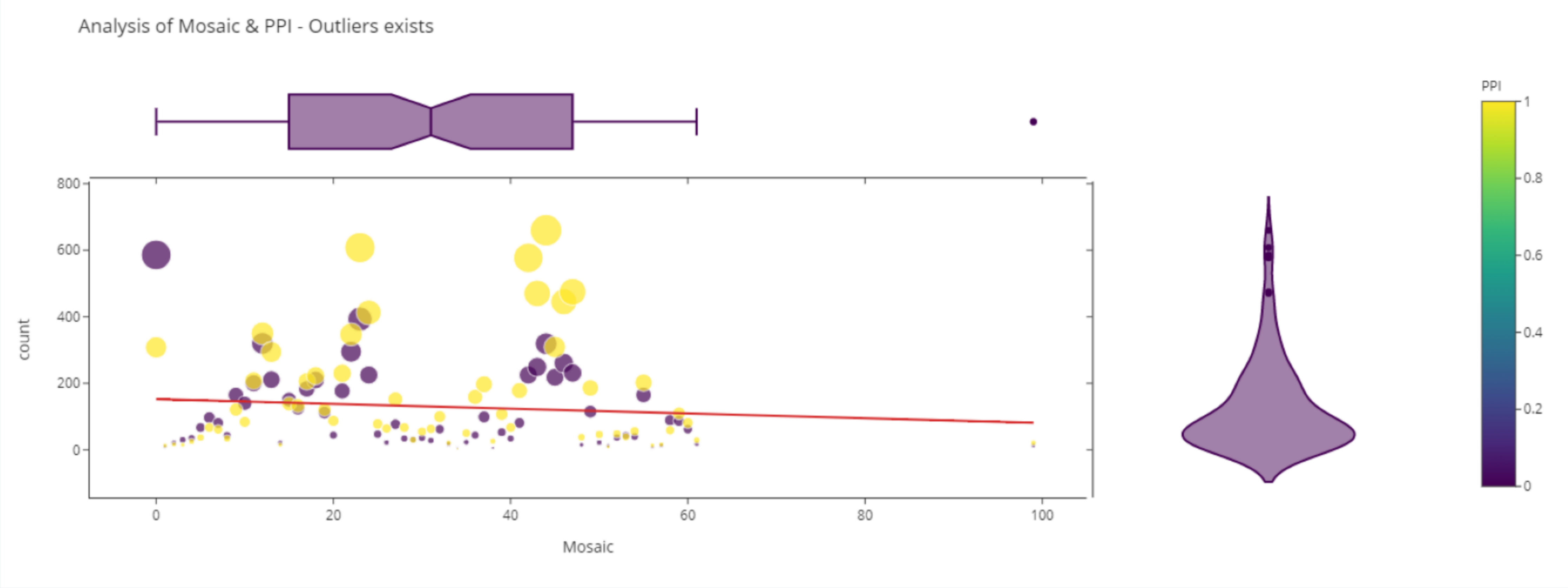
- Data had **higher number of customers with unsecured loans**; a lot of them were married and male customers
- About **80% of the customers were without any access cards**
- Customers with a PPI held **a cheque guarantee about 2% higher** than the customers without any PPI



Some of the findings from the EDA !

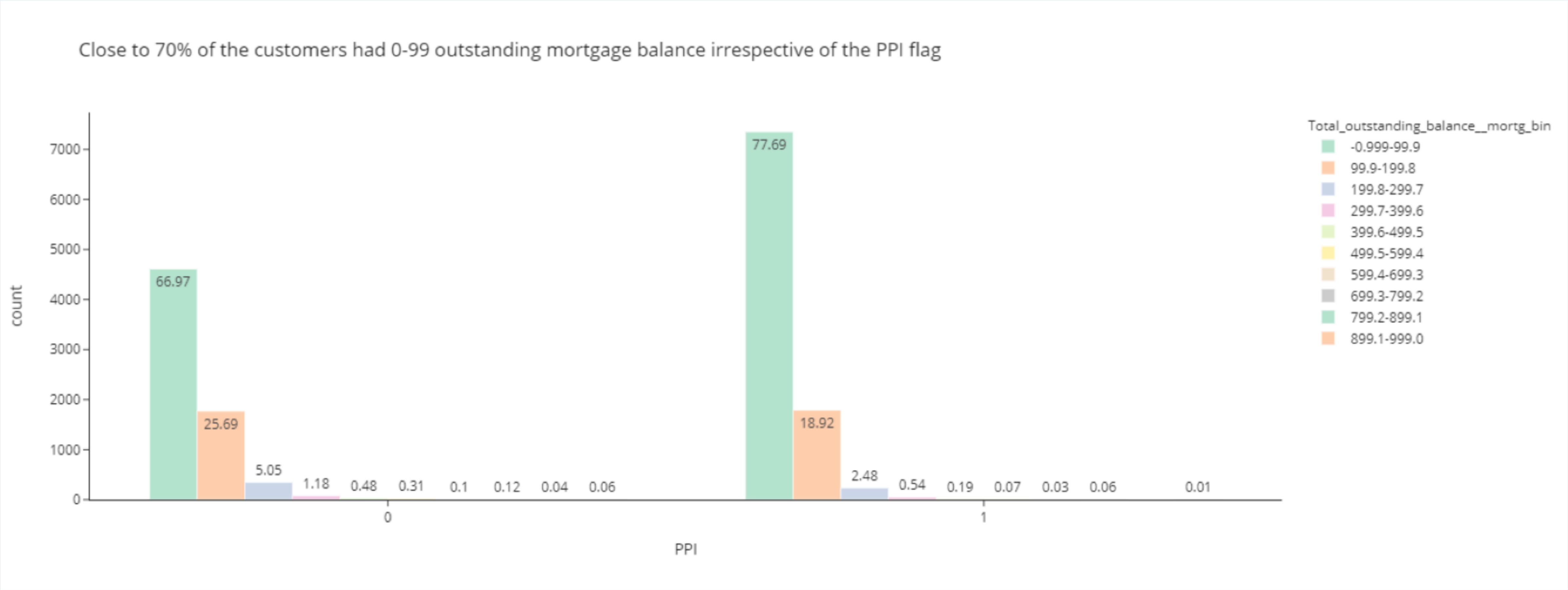
(3/4)

- Customers who were having PPI, **product description was available for 98% of them** whereas insurance description was available for all customers with a PPI
- Customers with **income range between 1and 5** are likely can be targetted for cross-selling the PPI products
- Customers with **total of accounts = 9** can be preferred while offering the PPI products



Some of the findings from the EDA ! (4/4)

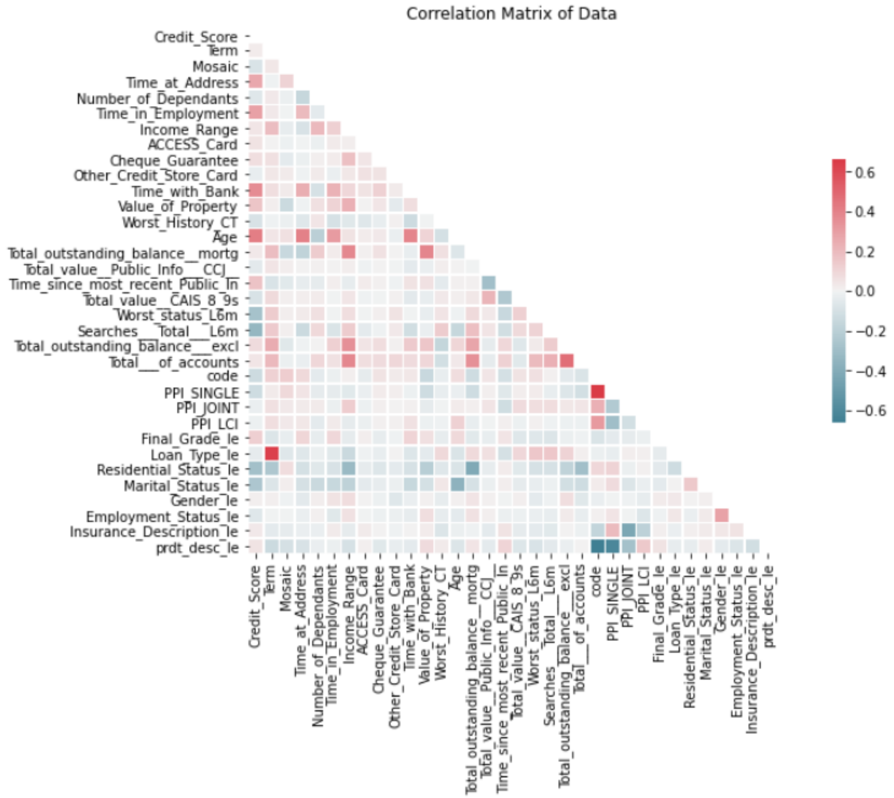
- **Average of total accounts for customers** holding a PPI product was highest in Joint insurance category/product
- Customers **staying at the address between 0-260 days can be priortized** while offering a PPI product
- PPI products can be **targeted to customers with a value of property b/w 115k and 120k, accounts for 25%-30% of the total customers** (with PPI & w/o a PPI product)



Feature **Reduction !**

- Below table shows **pairwise pearson correlation above a threshold of 0.7**, some of the variables were highly correlated with each other and out of these following variables were dropped based on their correlation with the target variable (i.e. assumed to be PPI)
  'Total_Outstanding_Balances', 'Time_since_most_recent_outstandi', 'Total__outstanding_CCJ_s', 'Bureau_Data__Monthly_Other_Co_R', 'Total__Public_Info__CCJ___ban', 'Net_Advance', 'Years_on_ER_for_SP', 'Outstanding_Mortgage_Bal', '__of_status_3_s_L6m', 'Worst_CUrrent_Status', 'Mosaic_Class', 'APR', 'category_le'
- All categorical variables were label encoded, **correlation matrix after dropping the above variables**

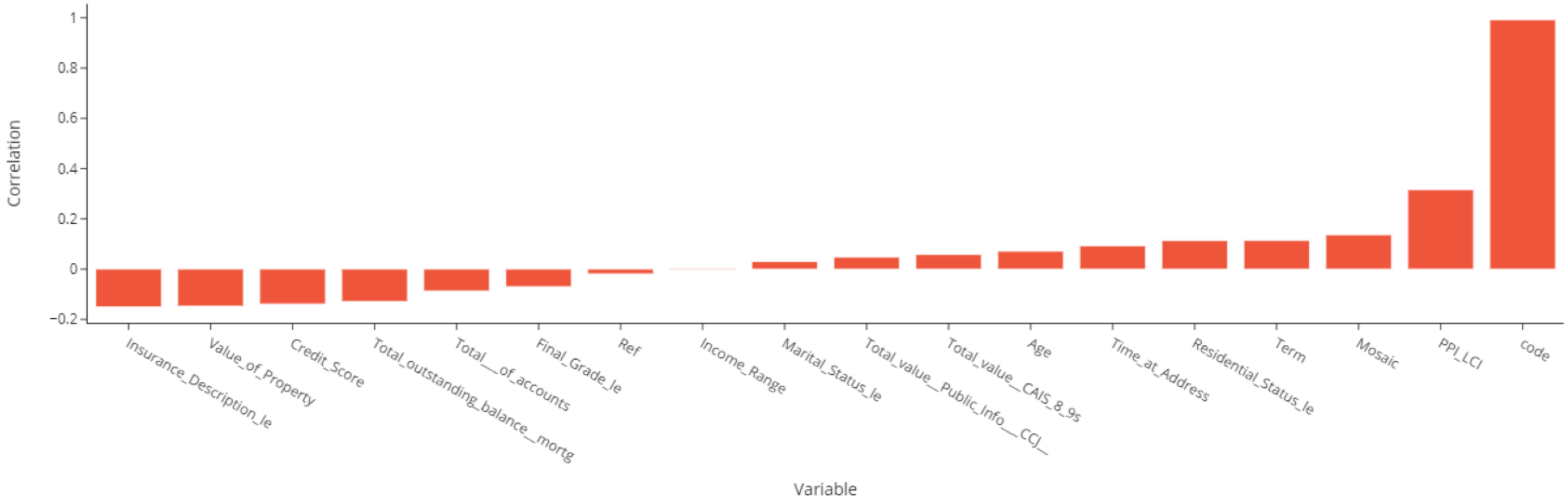
Feature 1	Feature 2	Pairwise Correlation
Total_Outstanding_Balances	Total_outstanding_balance__mortg	0.977
Total_outstanding_balance__mortg	Total_Outstanding_Balances	0.977
Time_since_most_recent_Public_In	Time_since_most_recent_outstandi	0.919
Time_since_most_recent_outstandi	Time_since_most_recent_Public_In	0.919
Total__outstanding_CCJ_s	Total__Public_Info__CCJ___ban	0.910
Total__Public_Info__CCJ___ban	Total__outstanding_CCJ_s	0.910
Bureau_Data__Monthly_Other_Co_R	Total_Outstanding_Balances	0.901
Total_Outstanding_Balances	Bureau_Data__Monthly_Other_Co_R	0.901
Bureau_Data__Monthly_Other_Co_R	Total_outstanding_balance__mortg	0.863
Total_outstanding_balance__mortg	Bureau_Data__Monthly_Other_Co_R	0.863
Time_since_most_recent_outstandi	Total__outstanding_CCJ_s	0.839
Total__outstanding_CCJ_s	Time_since_most_recent_outstandi	0.839
Total__Public_Info__CCJ___ban	Time_since_most_recent_Public_In	0.828
Time_since_most_recent_Public_In	Total__Public_Info__CCJ___ban	0.828
PPI_SINGLE	category_le	0.786
category_le	PPI_SINGLE	0.786
__of_status_3_s_L6m	Worst_CUrrent_Status	0.783
Worst_CUrrent_Status	__of_status_3_s_L6m	0.783
Worst_status_L6m	Worst_CUrrent_Status	0.778
Worst_CUrrent_Status	Worst_status_L6m	0.778
Time_since_most_recent_outstandi	Total__Public_Info__CCJ___ban	0.773
Total__Public_Info__CCJ___ban	Time_since_most_recent_outstandi	0.773
Time_since_most_recent_Public_In	Total__outstanding_CCJ_s	0.772
Total__outstanding_CCJ_s	Time_since_most_recent_Public_In	0.772
__of_status_3_s_L6m	Worst_status_L6m	0.768
Worst_status_L6m	__of_status_3_s_L6m	0.768
Time_at_Address	Years_on_ER_for_SP	0.750
Years_on_ER_for_SP	Time_at_Address	0.750
Term	Net_Advance	0.721
Net_Advance	Term	0.721
Total_outstanding_balance__mortg	Outstanding_Mortgage_Bal	0.713
Outstanding_Mortgage_Bal	Total_outstanding_balance__mortg	0.713
Total_Outstanding_Balances	Outstanding_Mortgage_Bal	0.703
Outstanding_Mortgage_Bal	Total_Outstanding_Balances	0.703



Feature Selection !

- Feature selection was done based on the **Weight of Evidence (WOE) & Information Value (IV)**
- WOE tells the predictive power of an independent variable in relation to the dependent variable whereas IV helps us to rank variables on the basis of their importance.
- Before calculating the WOE & IV, **all negative values in the numerical features were replaced by medians**
- **Features with either a Strong, Medium and a Weak IV were considered**
- Shape of the data after EDA, feature reduction and then feature selection: (16346, 19) i.e. **from 59 variables initially we were now left with 19 variables**
- **Alternatively**, dimensionality reduction methods such as PCA / t-SNE could have been done to remove some of the highly correlated numerical features

Correlation between 18 Independent and a Dependent Features



Predicting Insurance Description for the Customers with PPI = 0 !

- Effort were towards estimating **what would be the Insurance Description for customers where PPI = 0** in the data provided
- **Various classifier models** (such as Logistic Regression, SVC, Decision Tree, Random Forest, Gaussian Naive Bayes, KNearest Neighbor, XGBoost, AdaBoost, Gradient Boosting and CatBoost) **were evaluated based on F1-score, Precision and Recall** to predict Insurance Description
- During the comparison, Gradient Boost emerged as the clear winner; however, the **CatBoost model was selected as the final model**
- CatBoost combines many binary decision trees (weak learners) to form a unified model (strong learner). It estimates target statistic to represent the categories of each high-cardinality categorical feature and subsequently clusters those category values into a small number of groups. This approach helps **handle very high cardinality without losing their associations with the target variable.**
- **CatBoost's weighed average recall and F1-score was 94% on the validation set**, which was 25% of the training set (i.e. the one with available Insurance Description)
- **No hyperparameter tuning** was performed since the focus was to find more cross-sell opportunities that being correct with each classification; yes, it might impact some of the cross-sell opportunity, however 94% was a good-enough value to proceed ahead

```
Weighted average recall and F1-score 94% on the validation set
              precision    recall  f1-score   support

           0       0.98      0.91      0.95       247
           1       0.88      0.99      0.93       588
           2       0.95      1.00      0.97       517
           4       0.95      0.79      0.86       551
           5       0.99      1.00      1.00       463

    accuracy                           0.94      2366
   macro avg       0.95      0.94      0.94      2366
weighted avg       0.94      0.94      0.94      2366
```

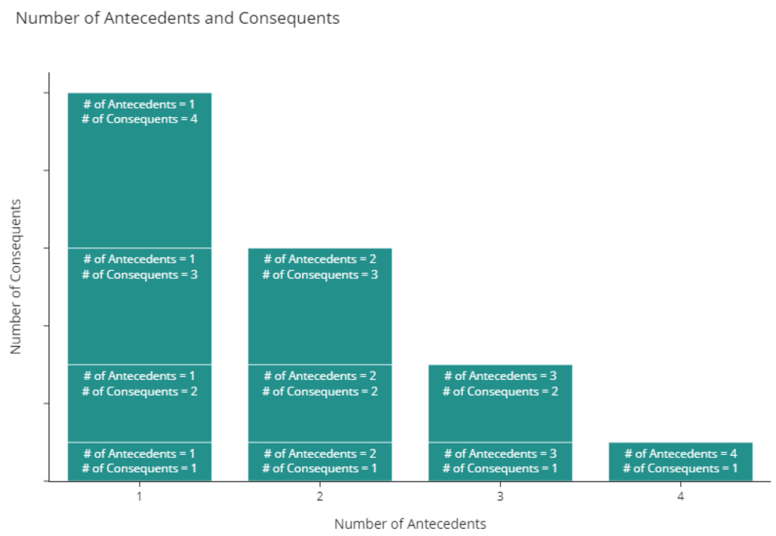
Note: {3: 'Missing', 2: 'Life & Critical Illn', 0: 'Joint', 5: 'Single', 4: 'Others', 1: 'LASU'}

Cross-Selling PPI Products - Market Basket Analysis !

- Assuming that predicted products (Insurance_Description) are the products customers are currently holding, used market basket analysis to find out what products customers tend to buy together, and then cross-sell different insurance products based on that information
- Some of the other applications of market basket analysis include: recommendation engine, cross-sell / bundle products, arranging items in the retail stores, credit card purchases of customers to build profiles for fraud detection purposes and cross-selling opportunities, telecom marketing efforts at customers, etc.

- Approach:
    - Frequent itemsets: Find insurance products that are purchased together atleast 5% of the times
    - Association rules by confidence: Form rules based on confidence assuming that an Insurance Product Y is purchased 20% of the times given that Insurance Product X has already been purchased
    - Filtering results by Lift: Sort the rules in descending order based on Lift. Assuming Lift >= 1 (positive correlation within the itemset) means that Insurance Product Y is as likely to be purchased as Insurance Product X provided that the latter has already been purchased
    - Analyze & Recommend: Analyze 3 scenarios with 2 minimum consequents and provide recommendations based on them

- Three scenarios with consequent a minimum of 2:
    - Scenario 1: Customer buys 1 Insurance Product and bank cross-sells a minimum of 2 Insurance Product
    - Scenario 2: Customer buys 2 Insurance Product and bank cross-sells a minimum of 2 Insurance Product
    - Scenario 3: Customer buys 3 Insurance Product and bank cross-sells a minimum of 2 Insurance Product



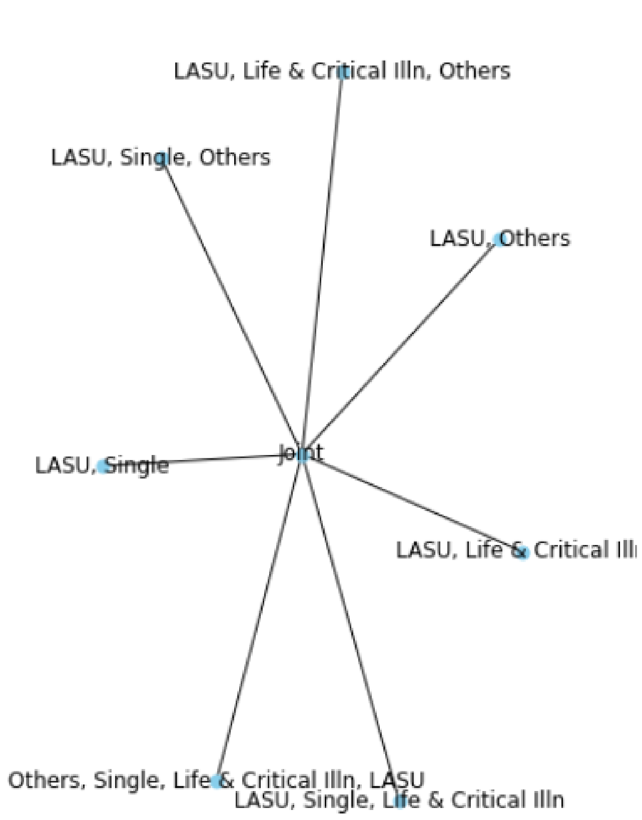
Cross-Selling PPI Products - Recommendations ! (1/3)

- Scenario 1: Customer buys 1 Insurance Product and bank cross-sells a minimum of 2 Insurance Product
  - For the frequent itemset, association rules by confidence and filtering of Lift > 1 approach mentioned earlier,
    - Single, Life & Critical Illness, Others insurance product will be bought given that either a LASU or a Joint insurance product is being bought

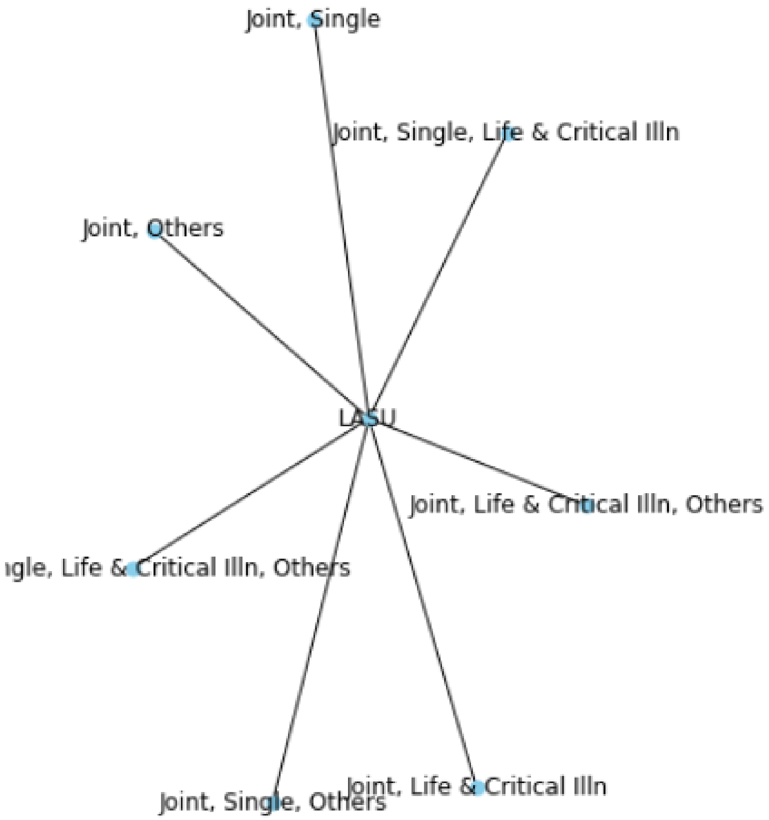
All possible combinations for the above scenario

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	Number of Antecedents	Number of Consequents
LASU	Joint, Single, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	3
Joint	LASU, Single, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	3
LASU	Joint, Life & Critical Illn, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	3
Joint	LASU, Single, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	3
LASU	Joint, Single, Life & Critical Illn, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	4
Joint	Others, Single, Life & Critical Illn, LASU	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	4
LASU	Joint, Single, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	3
Joint	LASU, Life & Critical Illn, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	3
LASU	Joint, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	2
Joint	LASU, Single	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	2
Joint	LASU, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	2
LASU	Joint, Single	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	2
LASU	Joint, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	2
Joint	LASU, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	1	2

Joint being the antecedent



LASU being the antecedent



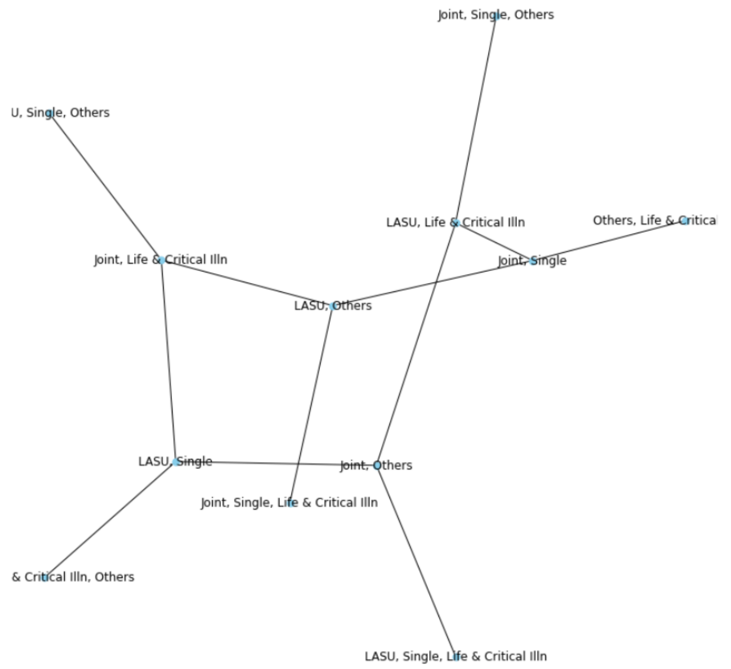
Cross-Selling PPI Products - Recommendations ! (2/3)

- Scenario 2: Customer buys 2 Insurance Product and bank cross-sells a minimum of 2 Insurance Product
  - For the frequent itemset, association rules by confidence and filtering of Lift > 1 approach mentioned earlier,
    - LASU, Single=>Joint, Others;Joint, Single=>LASU, Others;LASU, Life & Critical Illn=>Joint, Single;LASU, Single=>Joint, Life & Critical Illn;Joint, Life & Critical Illn=>LASU, Single;Joint, Single=>LASU, Life & Critical Illn;LASU, Others=>Joint, Life & Critical Illn;LASU, Life & Critical Illn=>Joint, Others;Joint, Others=>LASU, Life & Critical Illn;Joint, Life & Critical Illn=>LASU, Others;Joint, Others=>LASU, Single;LASU, Others=>Joint, Single;LASU, Single=>Joint, Life & Critical Illn, Others;LASU, Others=>Joint, Single, Life & Critical Illn;LASU, Life & Critical Illn=>Joint, Single, Others;Joint, Single=>Others, Life & Critical Illn, LASU;Joint, Others=>LASU, Single, Life & Critical Illn;Joint, Life & Critical Illn=>LASU, Single, Others are the possible antecedent and consequent recommendations for the above scenario

All possible combinations for the above scenario

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	Number of Antecedents	Number of Consequents
LASU, Single	Joint, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
Joint, Single	LASU, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
LASU, Life & Critical Illn	Joint, Single	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
LASU, Single	Joint, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
Joint, Life & Critical Illn	LASU, Single	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
Joint, Single	LASU, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
LASU, Others	Joint, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
LASU, Life & Critical Illn	Joint, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
Joint, Others	LASU, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
Joint, Life & Critical Illn	LASU, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
Joint, Others	LASU, Single	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
LASU, Others	Joint, Single	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	2
LASU, Single	Joint, Life & Critical Illn, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	3
LASU, Others	Joint, Single, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	3
LASU, Life & Critical Illn	Joint, Single, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	3
Joint, Single	Others, Life & Critical Illn, LASU	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	3
Joint, Others	LASU, Single, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	3
Joint, Life & Critical Illn	LASU, Single, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	2	3

Antecedent and Consequent for the above scenario



Cross-Selling PPI Products - Recommendations ! (3/3)

- Scenario 3: Customer buys 3 Insurance Product and bank cross-sells a minimum of 2 Insurance Product
    - For the frequent itemset, association rules by confidence and filtering of Lift > 1 approach mentioned earlier,
        - LASU, Single, Others => Joint, Life & Critical Illn; LASU, Single, Life & Critical Illn => Joint, Others; Others, Life & Critical Illn, LASU => Joint, Single; Joint, Single, Others => LASU, Life & Critical Illn; Joint, Single, Life & Critical Illn => LASU, Others; Joint, Life & Critical Illn, Others => LASU, Single are the possible antecedent and consequent recommendations for the above scenario

All possible combinations for the above scenario

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	Number of Antecedents	Number of Consequents
	LASU, Single, Others	Joint, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	3	2
	LASU, Single, Life & Critical Illn	Joint, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	3	2
	Others, Life & Critical Illn, LASU	Joint, Single	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	3	2
	Joint, Single, Others	LASU, Life & Critical Illn	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	3	2
	Joint, Single, Life & Critical Illn	LASU, Others	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	3	2
	Joint, Life & Critical Illn, Others	LASU, Single	0.857143	0.857143	0.857143	1.0	1.166667	0.122449	inf	3	2

Antecedent and Consequent for the above scenario

Conclusion !

- Here we used an Machine Learning driven approach to suggest cross sell of Insurance Products for the consumer bank with a range of products.
- EDA, helped us understand in detail about the data provided at a features that are being provided and derive data driven insights. Based on the EDA, we also saw some of the features highly contributing to whether a PPI product will be purchased or not.
- Feature reduction and selection, helped us focus on the features with strong, medium and weak Information Value and avoid our focus on suspicious (too good to be true) features; further pairwise correlations and correlation analysis with the target variable (PPI) allowed us to remove independent features that were highly correlated with each other.
- Machine Learning classifier model, helped us in predicting the insurance description for the ones currently not holding a PPI product with the bank and thus widening the scope of customers that can be targetted in the cross-selling.
- Market Basket Analysis, allowed us to determine/identify cross-selling Insurance Product opportunities by identifying/suggesting frequently bought together Insurance Products.
- Recommendations provided using Market Basket Analysis were categorized in three scenarios using the approach mentioned earlier. This should reduce improve conversion rates, reduce unsolicited sales connects and avoid mis-sellings.
- Improvements using this approach can be measured by identifying conversion rates and monthly sales improvements using scenarios suggested.