Logistic Regression Case Study

Overview

For the given data, build a Logistic Regression model to find out the key variables which significantly explains the **probability of 'Default_On_Payment'**. The Business Context of the question is as follows:

'The client, a financial service institution, want to increase revenue streams and intents to target a segment of their customers who are most likely to default on the loans/Credit taken.'

Approach Required:

1. **Business & Analytical Solution:** Prepare a slide on the business solution, which takes into account the business context.

2. **Hypothesis formulation:** Please frame a hypothesis table, and put down the hypothesis of independent variable wrt dependent var. You can follow the below template

Business Hypotheses

S.No	NAME	LABEL	VARIABLE TYPE	Degree of Impact	Sign of Effect	Hypothesis
1	AcctAge	Age of Oldest Account	Continuous	H	+	Greater the Age of Oldest Account, longer is association with Bank, hence likely to purchase
6	DirDep	Direct Deposit	Binary	H	+	Presence of the Direct Deposit, implies that the custome uses the a/c to pay all bills, hence likely to purchase
11	Sav	Saving Account	Binary	H	+	Presence of the Savings Deposit A/C, can drive the likelihood of purchase
12	SavBal	Saving Balance	Continuous	H	+	Greater of the Savings Deposit Bank, greater the likelihood of purchase
15	POS	Number Point of Sale	Continuous	H	+	Greater is the POS, greater is the likelihood to purchase
16	POSAmt	Amount Point of Sale	Continuous	H	+	Greater is the POS Amount, greater is the likelihood to purchase
17	CD	Certificate of Deposit	Binary	H	+	Presence of the Certificate of Deposit, can drive the likelihood of purchase
18	CDBal	CD Balance	Continuous	H	+	Greater the Certificate of Deposit, greater the likelihood of purchase
19	IRA	Retirement Account	Binary	H	+	Presence of the Retirement Account, can drive the likelihood of purchase
20	IRABal	IRA Balance	Continuous	H	+	Greater the IRA Balance, greater the likelihood of purchase
25	MM	Money Market	Binary	H	+	Presence of the Money Market Features, can drive the likelihood of purchase
26	MMBal	Money Market Balance	Continuous	H	+	Greater the Money Market Balance, greater the likelihood of purchase
30	CC	Credit Card	Binary	H	+	Presence of Credit card feature, implies the customer also has greater transcation relationships
31	CCBal	Credit Card Balance	Continuous	H	+	Greater the Credit Card Balance, greater the likelihood of purchase
33	SDB	Safety Deposit Box	Binary	H	+	A safety deposit box, is a risk averse feature, hence can be attributed to buy insurance
34	Income	Income	Continuous	H	+	Greater the income, greater the likelihood of purchase of insurance product
35	HMOwn	Owns Home	Binary	H	+	Presence of Home Feature, can increase in likelihood to purchase
37	HMVal	Home Value	Continuous	H	+	Greater the home value, greater is the likelihood to purchase
39	CRScore	Credit Score	Continuous	H	+	better the credit score, greater the likelihood to purchase
45	Dep	Checking Deposits	Continuous	H	+	Greater the checking deposits, greater is the likelihood to purchase
46	DepAmt	Amount Deposited	Continuous	H	+	Greater the amount deposited, greater is the likelihood to purchase
47	Inv	Investment	Binary	H	+	Presence of Investment feature in over banking relationships, implies the customer also has greater transcation relationships
	InvBal	Investment Balance	Continuous	H	+	Greater the Investment Balance, greater is the likelihood to purchase

Legend

H= High degree of impact on the dependent variable

+= Positive Impact on the dependant variable

3. **Data treatment:** Missing Values Treatment
4. **Data Exploration**
5. **Splitting** the data into train and test data
6. **Model Building and Refinement:** Train Data
7. **Model Building and Refinement:** Test Data

Deliverables Required

With given approach, you are required submit the following:

1. A **Power Point presentation** describing the Modelling Results and Interpretation (Logistic Regression Model Test, Results and Business Interpretation)
2. **R Code** for the Model
3. All the outputs of the model in a separate excel file.