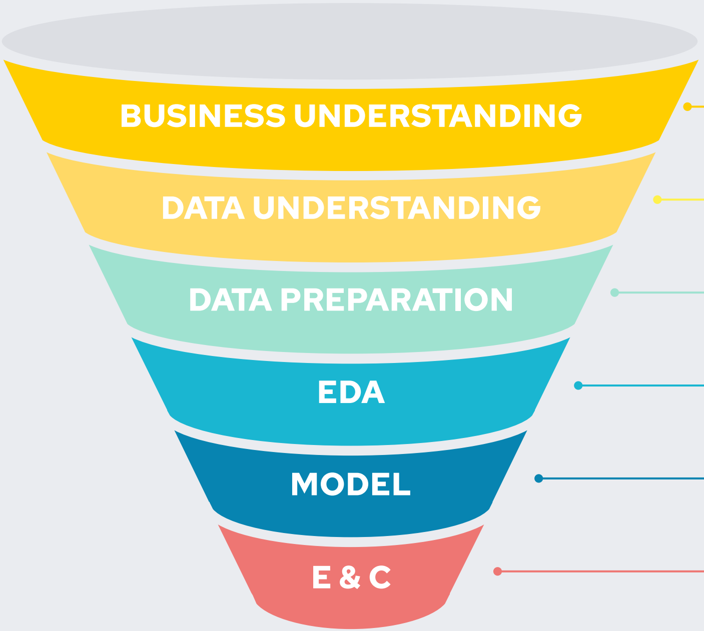
Marketing Campaign

Personal Loans

Table of Content



Business Understanding
Data Understanding
Data Preparation
Exploratory Data Analysis
Data Modelling
Evaluation & Conclusions

1 BUSINESS UNDERSTANDING

Our Understanding

ABC Bank - Marketing Campaign for Personal Loan

Existing Condition

- **Growing customer base**
  - Mostly depositor
- **Marketing campaign**
  - Targeting 5000 depositors
  - ~9% acquisition

Ideal Condition

- **Customer base**
  - Higher borrower to depositor ratio
- **Marketing campaign**
  - Lower budget/customer acquisition cost
  - Higher acquisition ratio

Problem Statement & Proposed Solution

ABC Bank

Problems

- **Customer base**
  - Better marketing target
- **Customer profile**
  - Significant variable(s)

Proposed Solutions

- **Machine learning model**
  - Give the probability of a customer taking personal loan
- **Market Segment**
  - Borrower's persona in general

2 DATA UNDERSTANDING

Dataset

The dataset is from ABC Bank **"Personal Loan".** The data collected in 2015 when bank ran a campaign for prospected liability customers (depositors) to convert into personal loan borrowers. This data set has **14 features** and **5000 rows** with demographic and transaction attributes.

Demographic Data

- ID
- Age
- Experience
- Income
- ZIP Code
- Family
- Education

Transaction Data

- CCAvg
- Mortgage
- Personal_Loan **(Y)**
- Securities_Account
- CD_Account
- Online
- CreditCard

Dataset

Demographic Data

- ID: Customer ID
- Age: Customer's age in completed years
- Experience: Years of professional experience
- Income: Annual income of the customer (in thousand dollars)
- ZIP Code: Home Address ZIP code.
- Family: Family size of the customer
- CCAvg: Average spending on credit cards per month (in thousand dollars)
- Education: Education Level. 1: Undergrad; 2: Graduate;3: Advanced/Professional

Transaction Data

- Mortgage: Value of house mortgage if any. (in thousand dollars)
- **Personal_Loan: Personal Loan acceptance on last campaign**
- Securities_Account: Securities Account ownership
- CD_Account: Certificate of deposit (CD) account ownership
- Online: Customer's using internet banking facilities?
- CreditCard: Credit card ownership issued by any other Bank (excluding ABC Bank)

3 DATA PREPARATION

Data Handling

Missing Value & Duplicate Checking

Checking through the dataset which have 'Nan' or null value and any duplicated rows. The result is there are no missing value or duplicates shown.

Data Types

Data types is as the features meant to be, there is no need to change any data types.

```
bank.isnull().sum()

ID                    0
Age                   0
Experience            0
Income                0
ZIP Code              0
Family                0
CCAvg                 0
Education             0
Mortgage              0
Personal Loan         0
Securities Account    0
CD Account            0
Online                0
CreditCard            0
dtype: int64
```

```
bank.dtypes

ID                      int64
Age                     int64
Experience              int64
Income                  int64
ZIP Code                int64
Family                  int64
CCAvg                 float64
Education               int64
Mortgage                int64
Personal Loan           int64
Securities Account      int64
CD Account              int64
Online                  int64
CreditCard              int64
dtype: object
```

Data Handling

Data Scaling

Data scaling means transforming the data into specific scale (0 to 1), makes it easier for a model to learn and understand the problem. Any features with binary value doesn't need scaling. The features that need scaling with are:

- Age
- Income
- Zip Code
- Family
- CCavg
- Mortgage
- Education

Imbalance Data Handling

SMOTE (Synthetic Minority Oversampling Technique)

The dataset has imbalance examples between minority of people who took personal loan and majority of people who didn't take personal loan. To resolve this problem is to oversample the examples in minority class, by duplicating the data from minority class prior fitting a model.

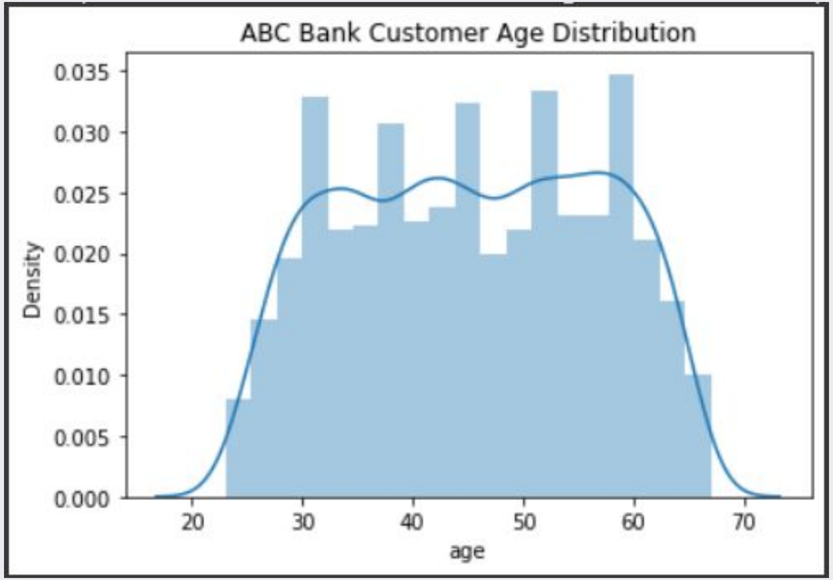
4 EXPLORATORY DATA ANALYSIS

Descriptive Analysis

	id	age	exp	income	zip_code	family	ccavg	edu	mortgage	personal_loan	securiti
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.0000
mean	2500.500000	45.338400	20.104600	73.774200	93152.503000	2.396400	1.937938	1.881000	56.498800	0.096000	0.1044
std	1443.520003	11.463166	11.467954	46.033729	2121.852197	1.147663	1.747659	0.839869	101.713802	0.294621	0.3058
min	1.000000	23.000000	-3.000000	8.000000	9307.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.0000
25%	1250.750000	35.000000	10.000000	39.000000	91911.000000	1.000000	0.700000	1.000000	0.000000	0.000000	0.0000
50%	2500.500000	45.000000	20.000000	64.000000	93437.000000	2.000000	1.500000	2.000000	0.000000	0.000000	0.0000
75%	3750.250000	55.000000	30.000000	98.000000	94608.000000	3.000000	2.500000	3.000000	101.000000	0.000000	0.0000
max	5000.000000	67.000000	43.000000	224.000000	96651.000000	4.000000	10.000000	3.000000	635.000000	1.000000	1.0000

- Average age of ABC customers is 45 years old, with the range of 23 until 67 years old.
- Average spending for credit card per person in a month is $1.937.
- Experience -3 years. More of investigation is needed to clean this data.
- Personal loan (target column) 0-75% data distribution value is 0 which indicated data imbalance.
- No duplicate and no null.

ABC Bank Customer Age Distribution

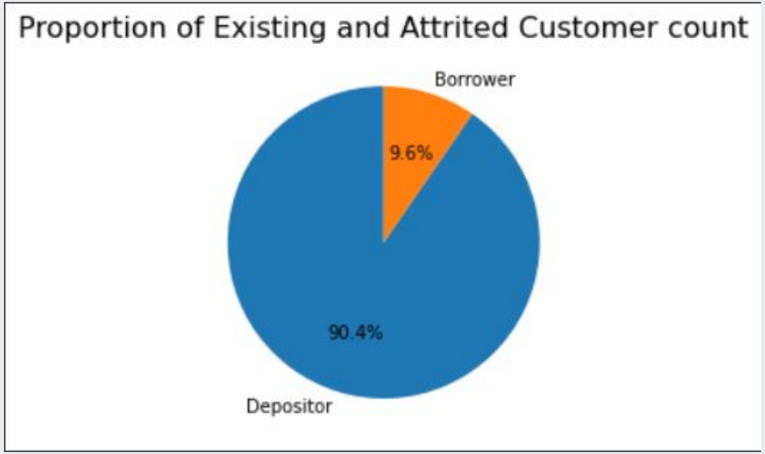


- Customer's age is almost uniformly distributed.
- Customer's age has fairly distributed in each 30s, 40s, 50s and 60s.

Imbalance Data

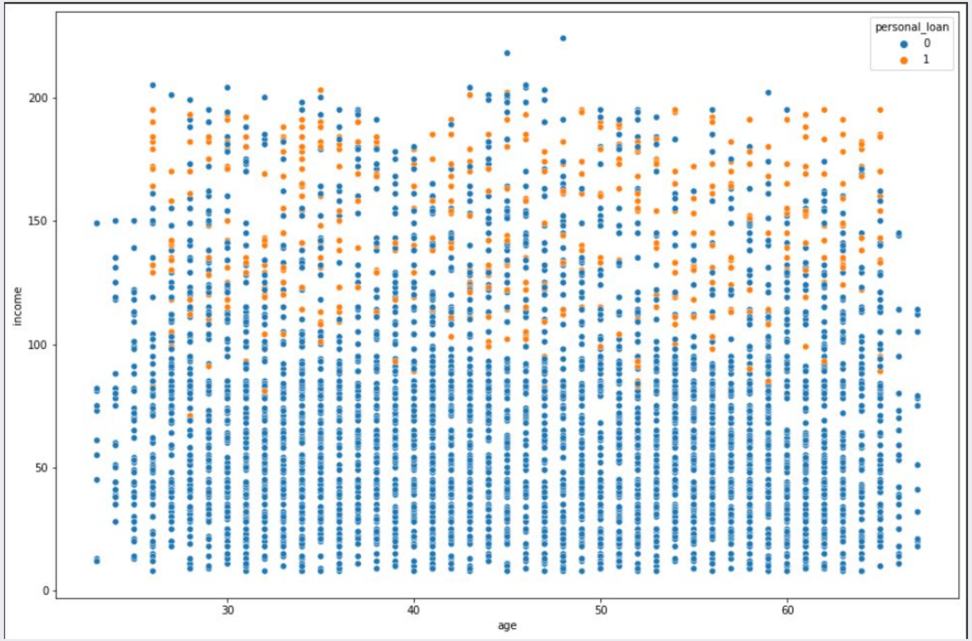
```
bank["personal_loan"].value_counts()
```

```
0    4520
1     480
Name: personal_loan, dtype: int64
```



The target variable personal_loan is highly imbalanced where only 9.6% of the customers have previously opted for a personal loan in the dataset. This can be handled using weight or SMOTE.

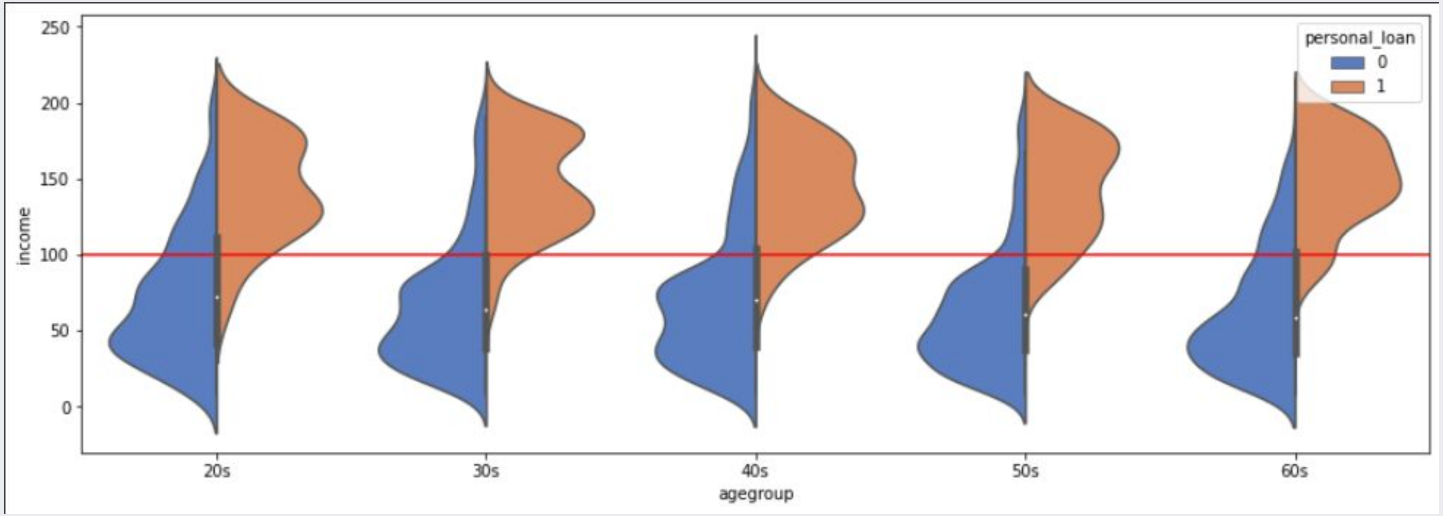
Age vs Income based on Personal Loan



Scatter plots are used to observe relationships between variables.

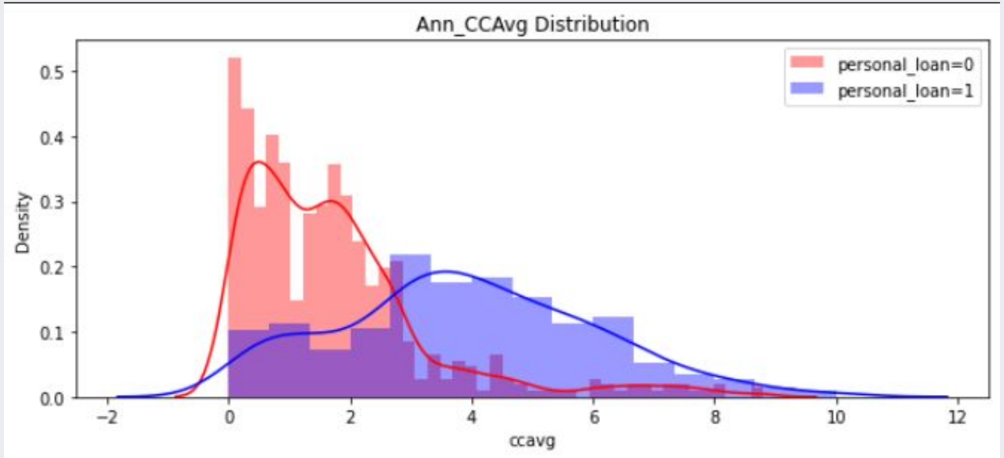
It can be seen from the graph that there is a **relationship** between **income** and customers who take personal loans.

Age vs Income based on Personal Loan



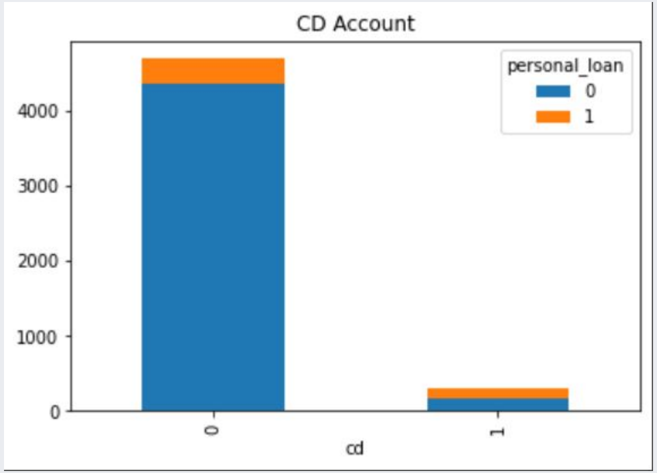
People with income above $100.000 are more likely take the Personal Loan despite of age.

CCAvg by Personal Loan



Customers who have taken personal loan have **higher credit card average** than those who did not take loan.

CD Account by Personal Loan

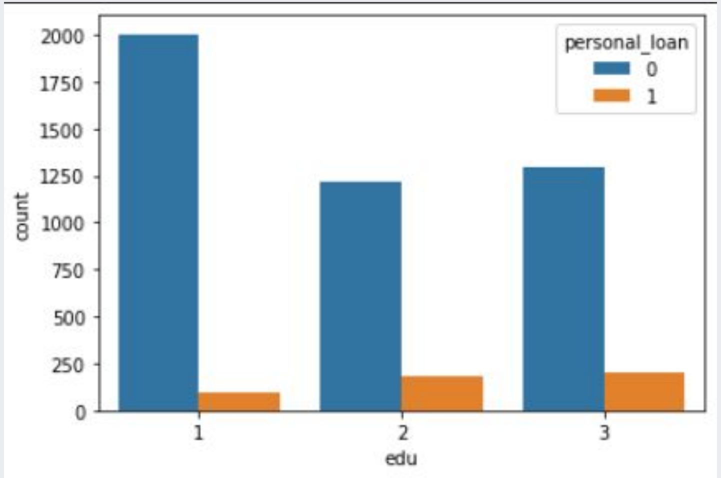



```
bank['cd'].value_counts()
```

```
0    4698
1     302
Name: cd, dtype: int64
```

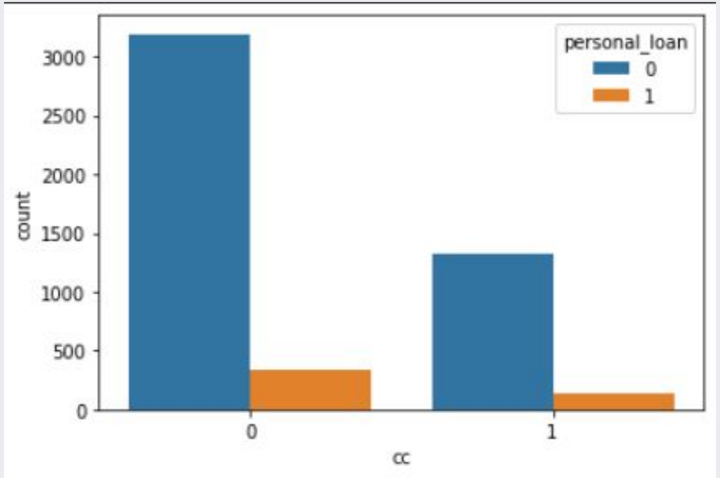
- Almost 50% of customers having Certified Deposit, had borrowed Personal Loan.
- However, 4358 customers out of 5000, do not have Certified Deposit Account and did not borrow Personal Loan.

Education by Personal Loan



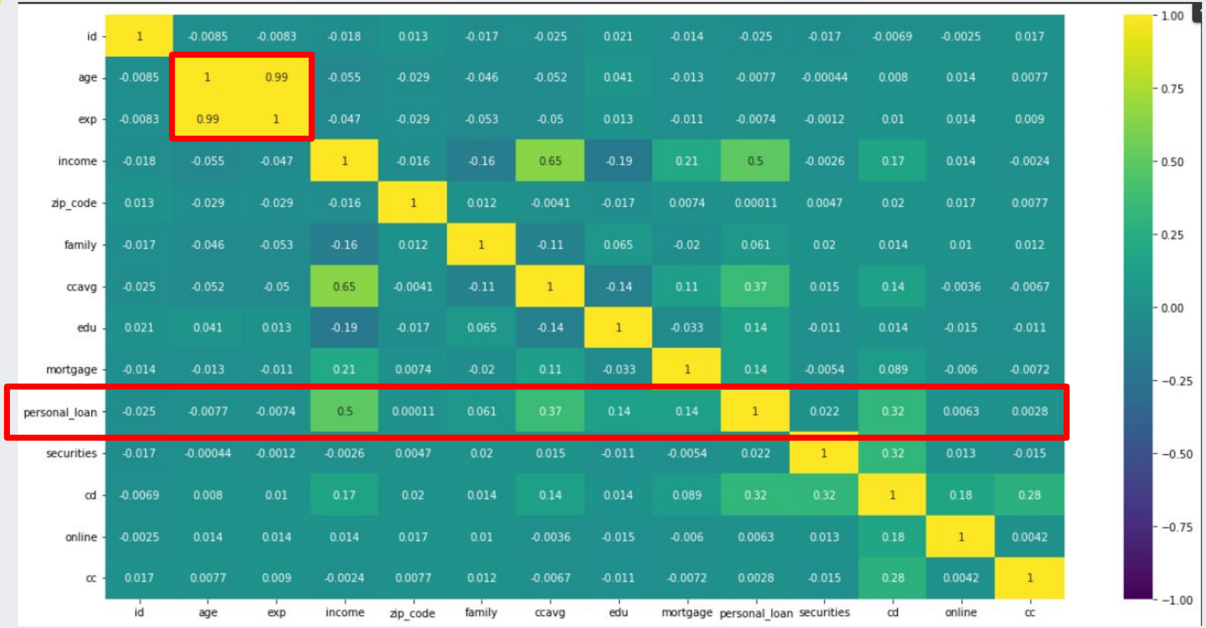
Based on education level, professionals are the most who took Personal Loan.
Education has low association with the 'Personal Loan'.

Credit Card by Personal Loan



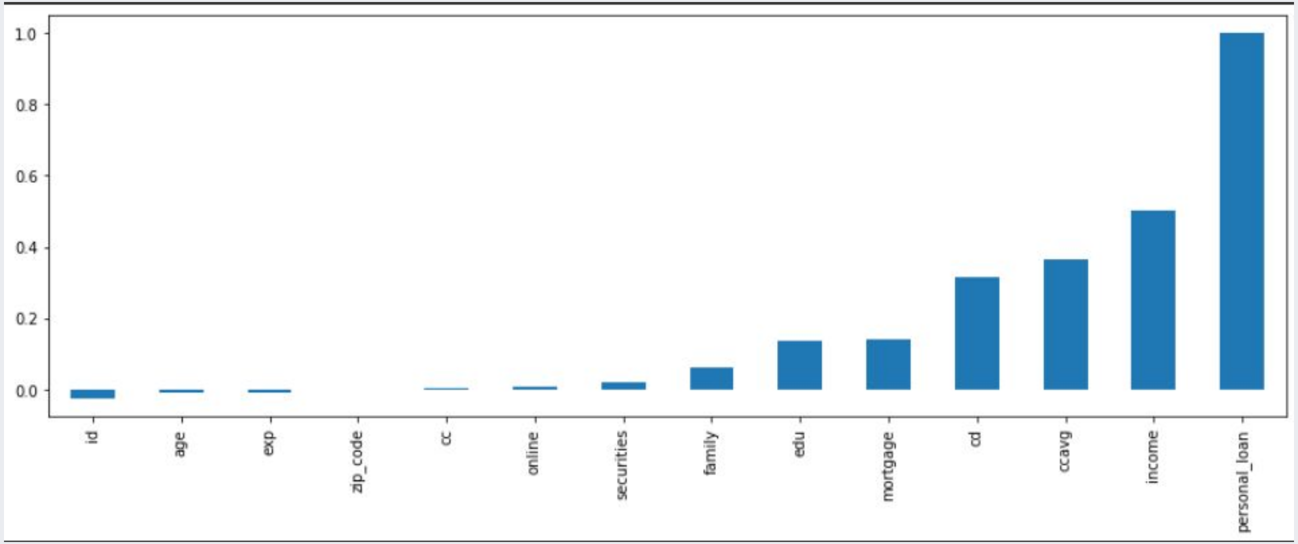
Majority customers who did have Personal Loan with the bank did not use Credit Card from other banks.

Multicollinearity



- There are multicollinearity between age and experience with the value of 0.99.
- Income and 'CCavg' is moderately correlated.
- Exp column is dropped.

Correlation to Personal Loan



- Income has the highest correlation score towards personal loan (0.5) therefore we might want to focus on income variable.
- People with income above $100.000 most probably will take the loan.

5 DATA MODELLING

Predictive Approach

Classification model

From the historical data of campaign last year, we can find pattern how data are classified and analyze which variable was most significant as a base to predict potential borrower.

The build model will be used to predict potential borrower from ‘new data of customer”.

- Random Forest
- Support Vector Machine
- Logistics Regression
- K-Nearest Neighbor
- Decision Tree

Classification Model Result & Evaluation

	Precission		recall		f1-score		support		accuracy
	0	1	0	1	0	1	0	1	
Random Forest	0.98	0.87	0.99	0.85	0.99	0.86	452	48	0.97
Support Vector Machine	0.98	0.81	0.98	0.79	0.98	0.80	452	48	0.96
Logistics Regression	0.98	0.44	0.89	0.81	0.93	0.57	452	48	0.88
Decision Tree	0.98	0.89	0.99	0.85	0.99	0.87	452	48	0.98
K-Nearest Neighbor	0.97	0.74	0.97	0.71	0.97	0.72	452	48	0.95

Classification model

We decided to compare our model, not all variables entering the model we drop Experience variables because have multicollinearity with Age, in scaling we process only Age, Income, ZIP Code, Family, CCAvg, Mortgage and Education variables join scaling.

From the five models that we compare, the decision tree has the best performance with precision score 98% and 89%. For accuracy score decision tree has 98%.

Classification Model Result & Evaluation

Random Forest	0.026
Super Vector Machine	0.038
Logistics Regression	0.116
K-Nearest Neighbor	0.052
Decision Tree	0.024

Mean Absolute Error

We also conduct model evaluation to measure the error rate with mean absolute error (MAE). Based on the MAE value the decision tree has a low error value 2.4%

6 CONCLUSIONS

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

1. According to our analysis the variable that has a positive effect on personal loan is income, CCAVG and Credit Card.

1. The decision tree has the best performance with precision score 98% and 89%. For accuracy score decision tree has 98%

Recommendations

Based on evaluation and conclusion, we highly recommend some actions ABC should do in order to obtain more personal loan takers.

Bank managers should conduct marketing approach targeting depositor using machine learning. For example providing cashbacks and/or giving bonuses such as holiday fly tickets.

Depositors with income below $100k should be targeted as well with micro loans, longer installments and lower interest. By targeting both customer segment, we lower the customer acquisition cost.

Specific campaign on credit card users to encourage them try personal loan for specific needs, such as spending at restaurants with personal loan payment could get buy 2 get 2 free. Getting more merchants who accept personal loan payment will also boost higher rate of success of this campaign.