BANK PERSONAL LOAN

By: Sohaib Elgenidy

SUPERVISOR: Dr: Doaa Mahmoud

Data science challenges

Reveal data secrets

Agenda

- Introduction
- Data Preprocessing
- Data Visualization
- Machine Learning Model
- Problem Statement

Data science challenges

Reveal data secrets

Introduction:

This case is about a bank which has a growing customer base. Majority of these customers are liability customers (depositors) with varying size of deposits.

The number of customers who are also borrowers (asset customers) is quite small, and the bank is interested in expanding this base rapidly to bring in more loan business and in the process.

earn more through the interest on loans. In particular, the management wants to explore ways of converting its liability customers to personal loan customers (while retaining them as depositors).

Data science challenges

Reveal data secrets

Age Customer's age

Experience Number of years of professional experience

Income Annual income of the customer

ZIPCode Home Address ZIP code

Family Family size of the customer

CCAvg Average spending on credit cards per month

Education Education Level. 1: Undergrad; 2: Graduate; 3: Advanced/Professional

Mortgage Value of house mortgage if any

Securities Account Does the customer have a securities account with the bank?

CD Account Does the customer have a certificate of deposit (CD) account with the bank?

Online Does the customer use internet banking facilities?

CreditCard Does the customer uses a credit card issued by UniversalBank?

Personal Loan Did this customer accept the personal loan offered in the last campaign?

Data science challenges
Reveal data secrets

Data Preprocessing

	count	mean	std	min	25%	50%	75%	max
ID	5000.000000	2500.500000	1443.520003	1.000000	1250.750000	2500.500000	3750.250000	5000.000000
Age	5000.000000	45.338400	11.463166	23.000000	35.000000	45.000000	55.000000	67.000000
Experience	5000.000000	20.104600	11.467954	-3.000000	10.000000	20.000000	30.000000	43.000000
Income	5000.000000	73.774200	46.033729	8.000000	39.000000	64.000000	98.000000	224.000000
ZIP Code	5000.000000	93152.503000	2121.852197	9307.000000	91911.000000	93437.000000	94608.000000	96651.000000
Family	5000.000000	2.396400	1.147663	1.000000	1.000000	2.000000	3.000000	4.000000
CCAvg	5000.000000	1.937938	1.747659	0.000000	0.700000	1.500000	2.500000	10.000000
Education	5000.000000	1.881000	0.839869	1.000000	1.000000	2.000000	3.000000	3.000000
Mortgage	5000.000000	56.498800	101.713802	0.000000	0.000000	0.000000	101.000000	635.000000
Personal Loan	5000.000000	0.096000	0.294621	0.000000	0.000000	0.000000	0.000000	1.000000
Securities Account	5000.000000	0.104400	0.305809	0.000000	0.000000	0.000000	0.000000	1.000000
CD Account	5000.000000	0.060400	0.238250	0.000000	0.000000	0.000000	0.000000	1.000000
Online	5000.000000	0.596800	0.490589	0.000000	0.000000	1.000000	1.000000	1.000000
CreditCard	5000.000000	0.294000	0.455637	0.000000	0.000000	0.000000	1.000000	1.000000

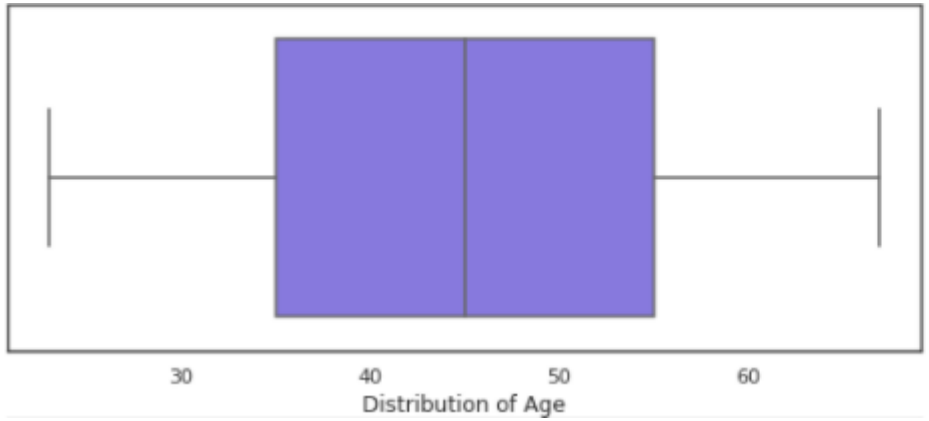
-The age column has a range of 23 to 67 years old. The mean age is 45 years old.

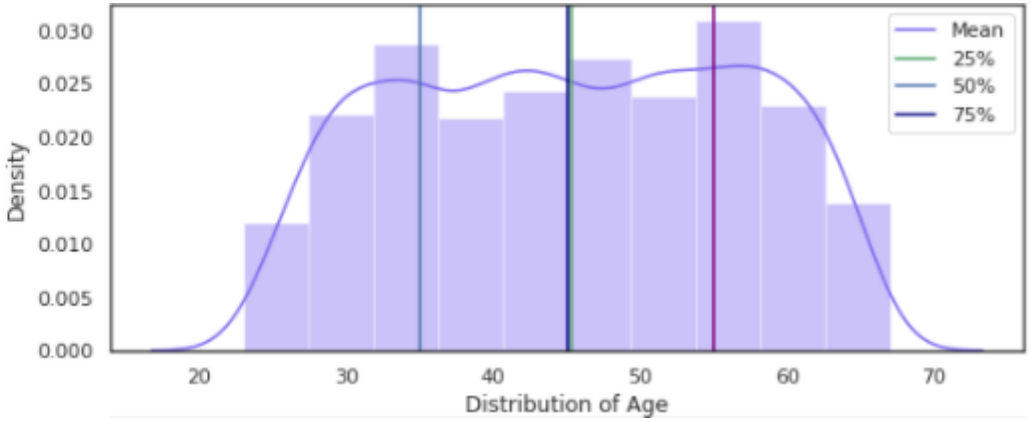
-The experience column has a min value of -3 and it is possible that there are many such values which need to be handled. But has a average value of 20 years of experience. (Further in our exploration, we will find that experience has a high correlation to age).

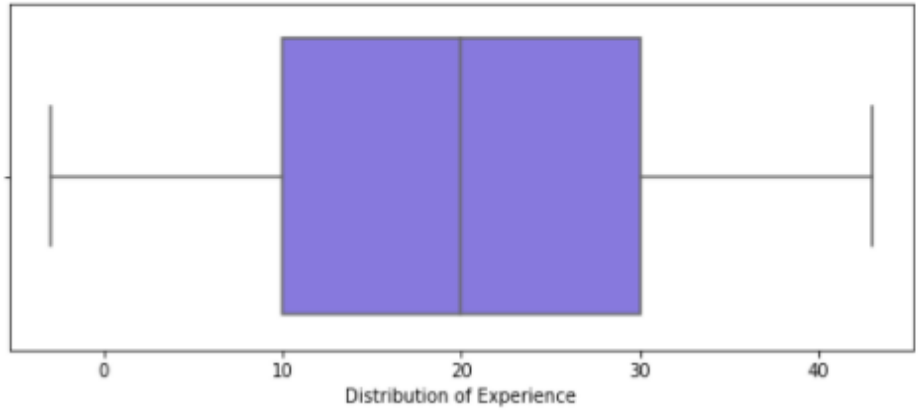
-The average income is 73.7 and the max value is more than 3x of the average income of 224.0.

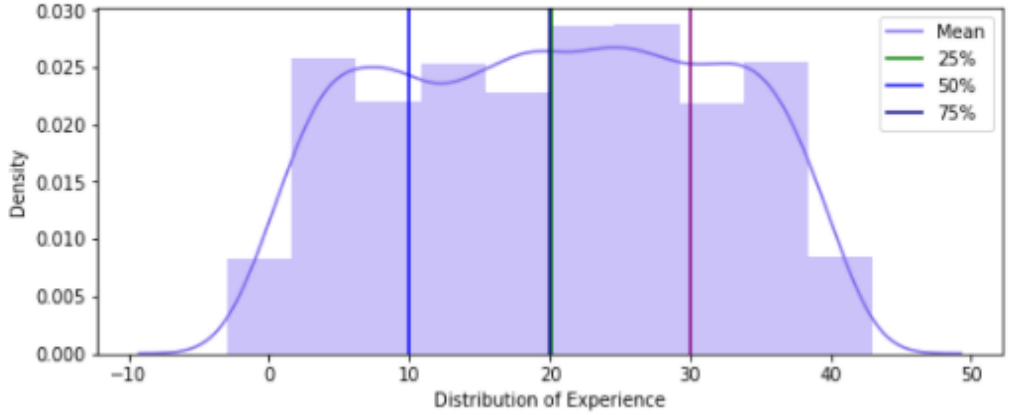
-The rest of columns don't tell us much about the data. We will explore them in depth in the following sections.

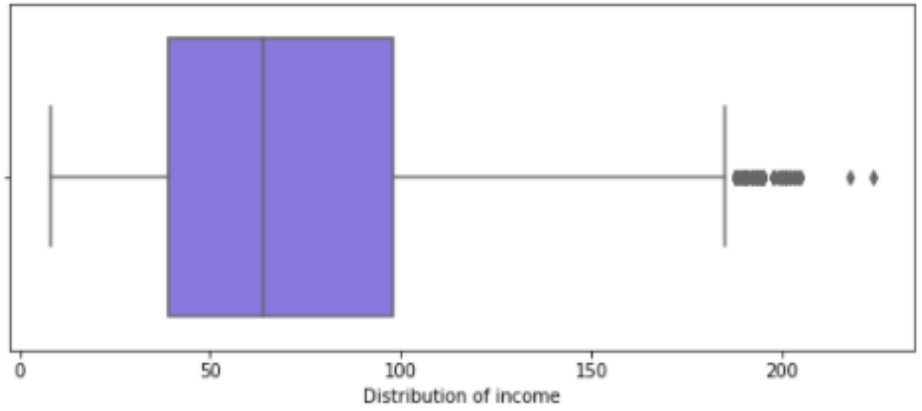
Data science challenges
Reveal data secrets

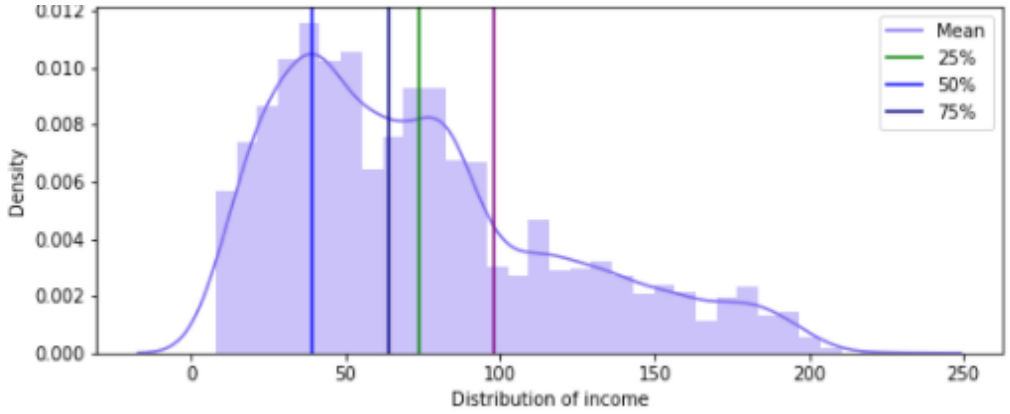






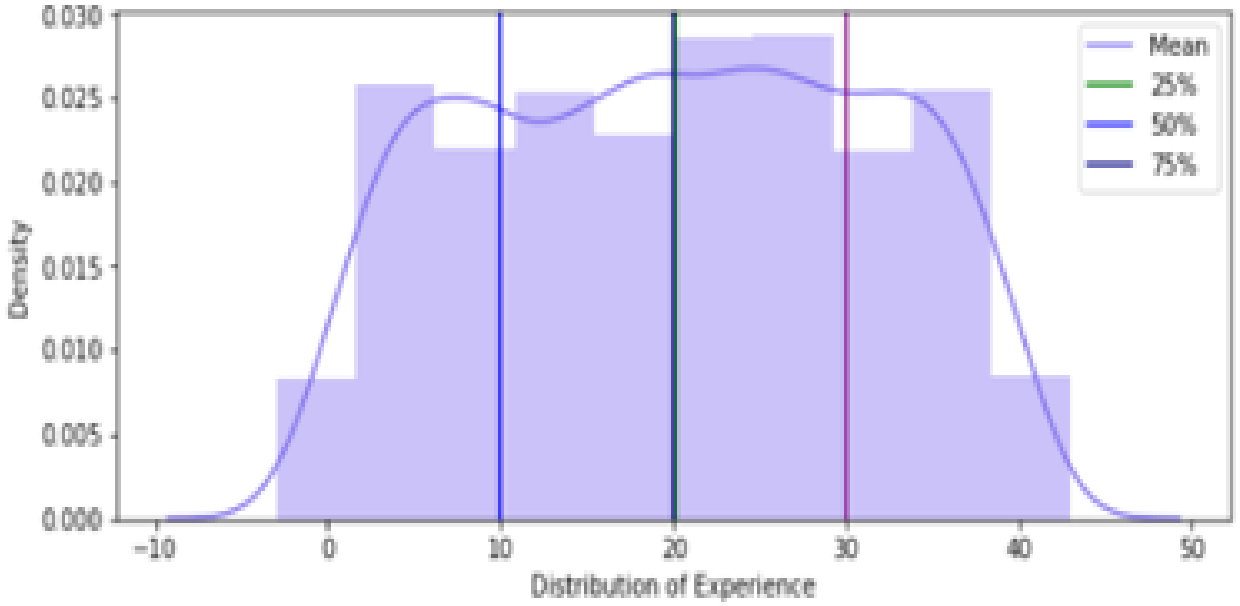




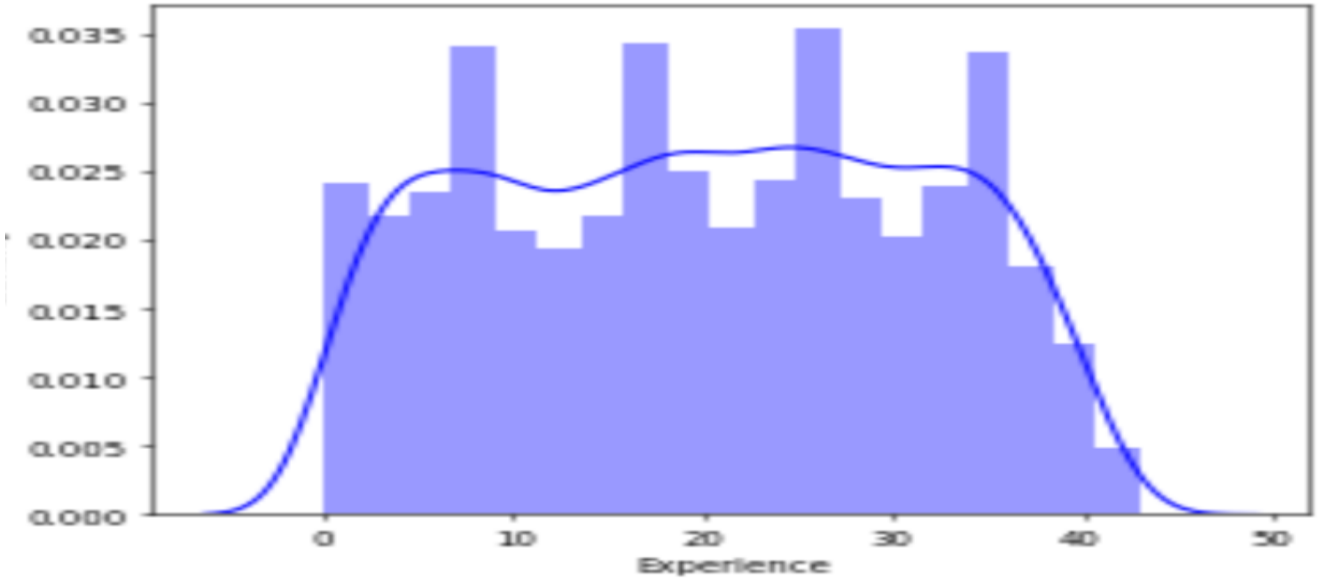


Data science challenges

Reveal data secrets



count	5000.000000
mean	20.104600
std	11.467954
min	-3.000000
25%	10.000000
50%	20.000000
75%	30.000000
max	43.000000



count	5000.000000
mean	20.134600
std	11.415189
min	0.000000
25%	10.000000
50%	20.000000
75%	30.000000
max	43.000000

Data science challenges

Reveal data secrets

-Income and CCAvg is moderately correlated.

-Age and Experience is highly correlated

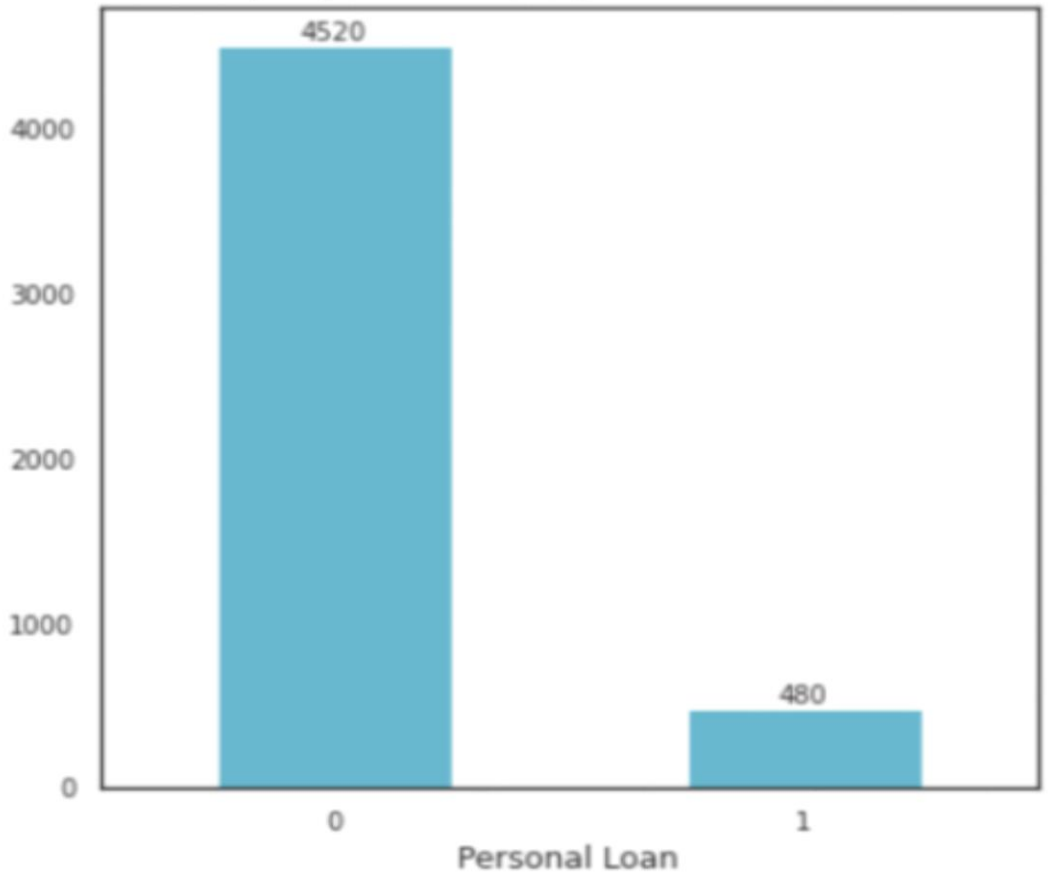
Data science challenges

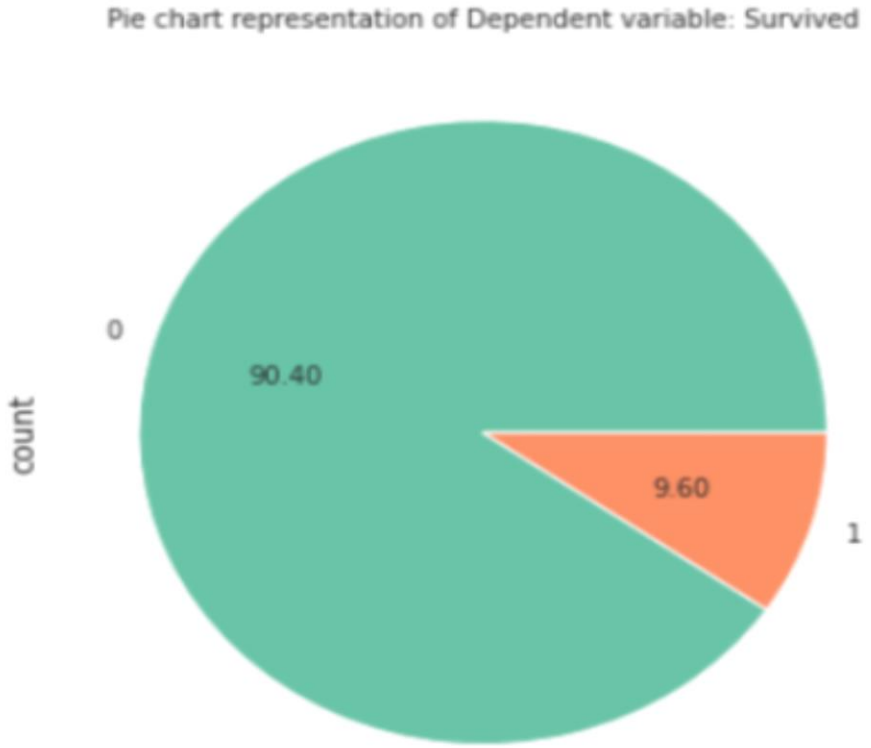
Reveal data secrets

Data Visualization

Data science challenges

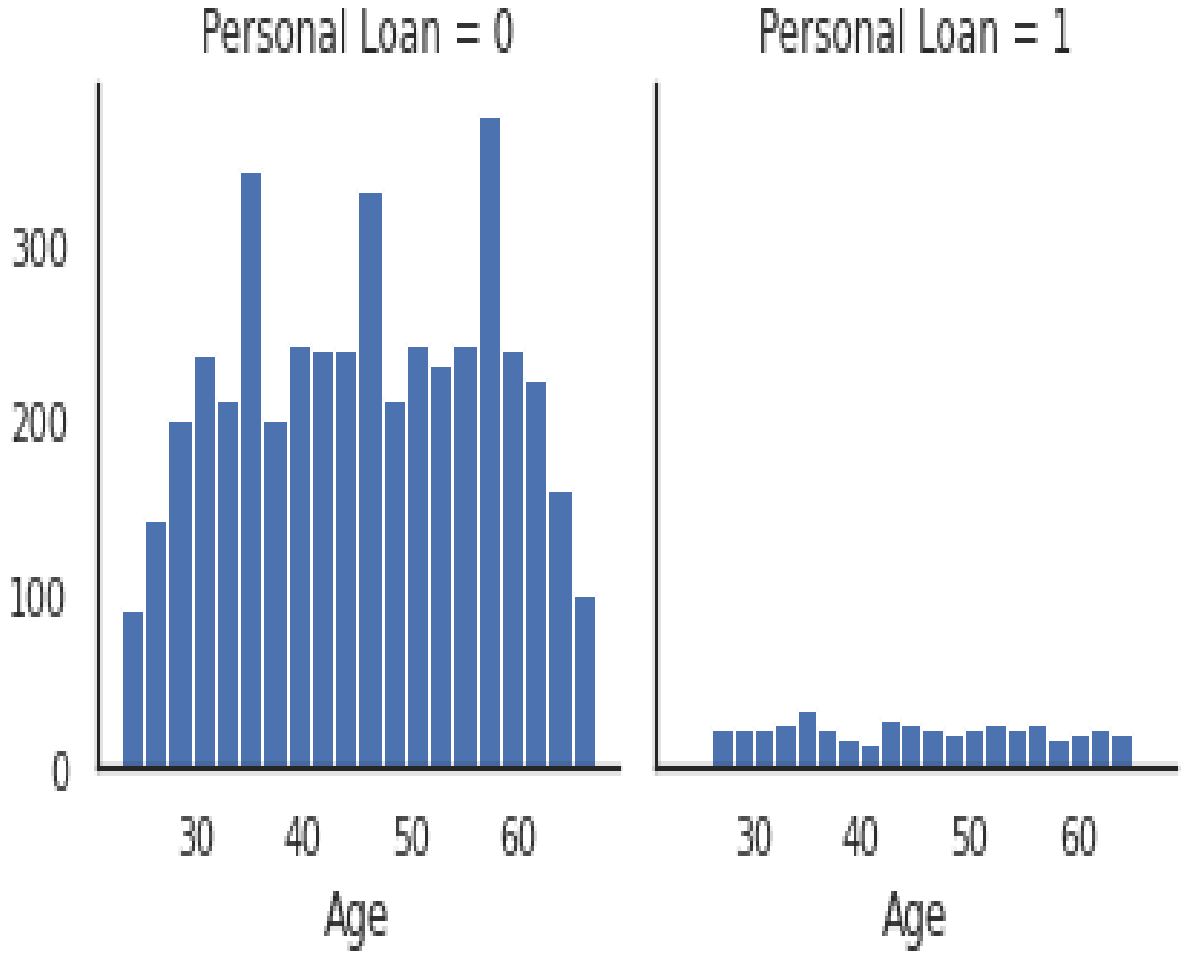
Reveal data secrets

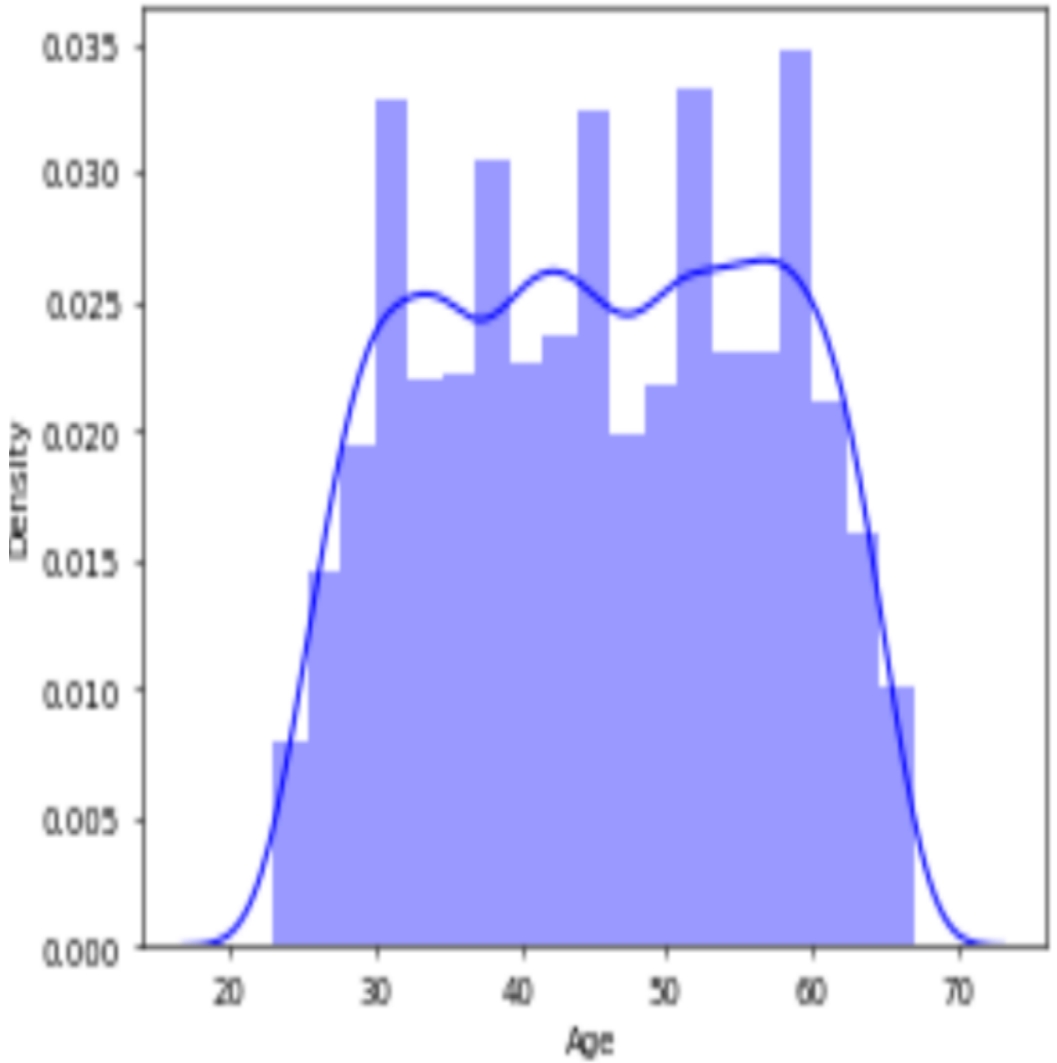




Dataset is likely to be imbalanced dataset.There ratio of Opted Personal Loan vs not opted is less than 90:10.We may need to find out a way to balance those 2 class.

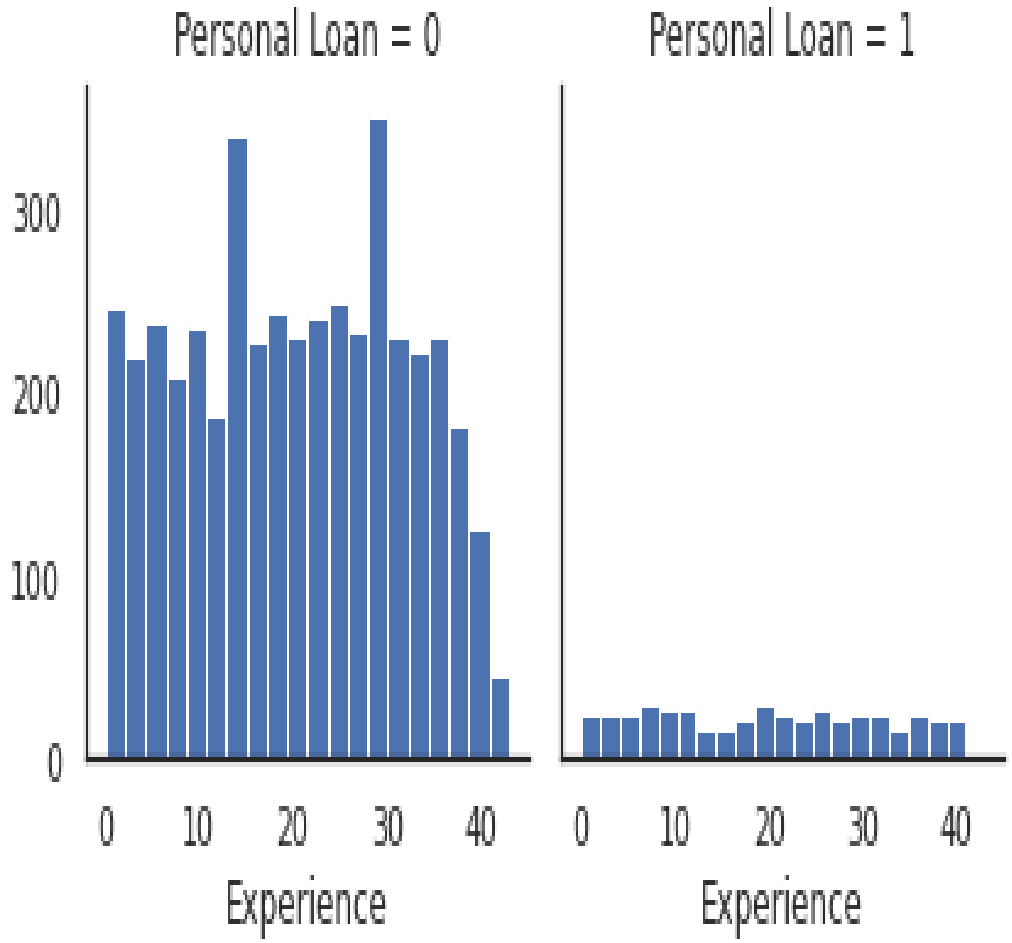
Data science challenges
Reveal data secrets

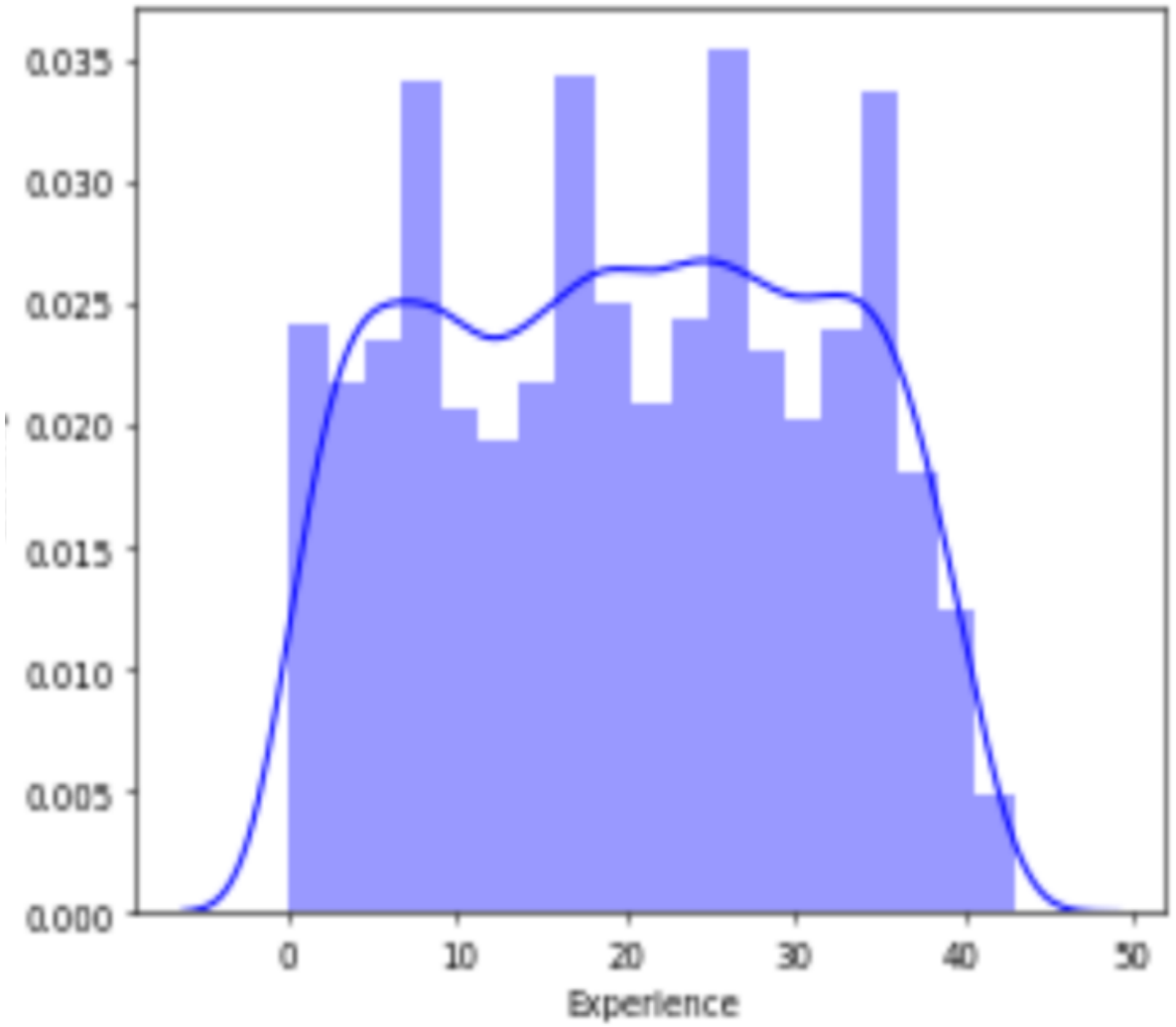




Age can be a vital factor in borrowinng loan, converting ages to bin to explore if there is any pattern.

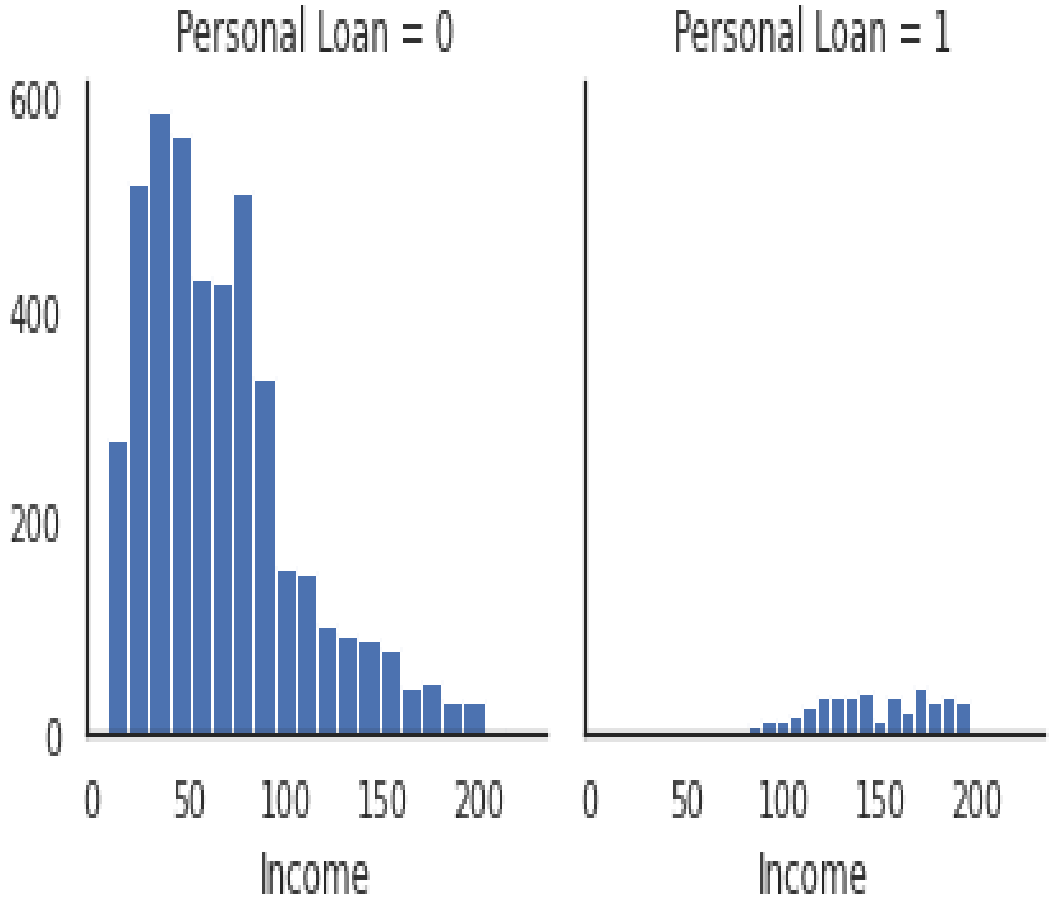
Data science challenges
Reveal data secrets

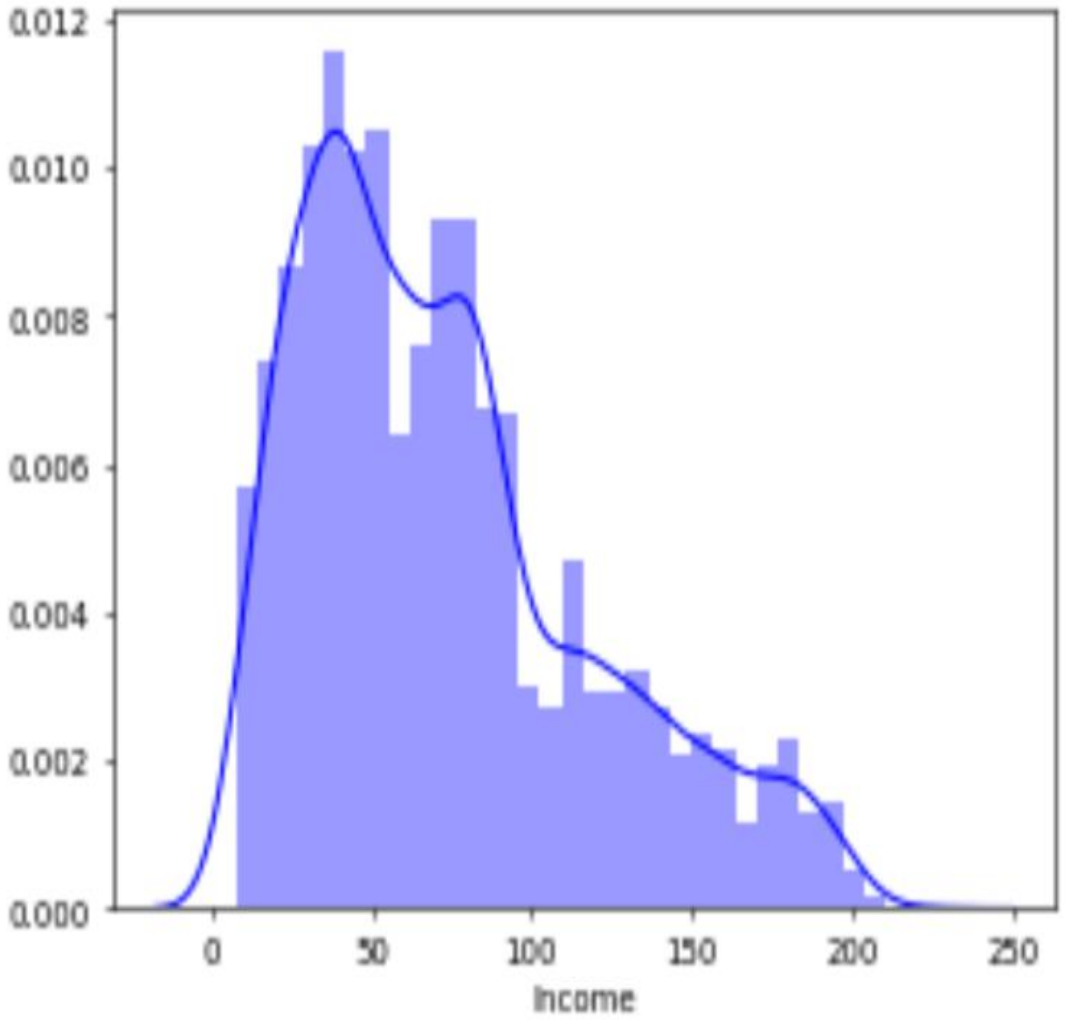




Experience is also normaly distributed.

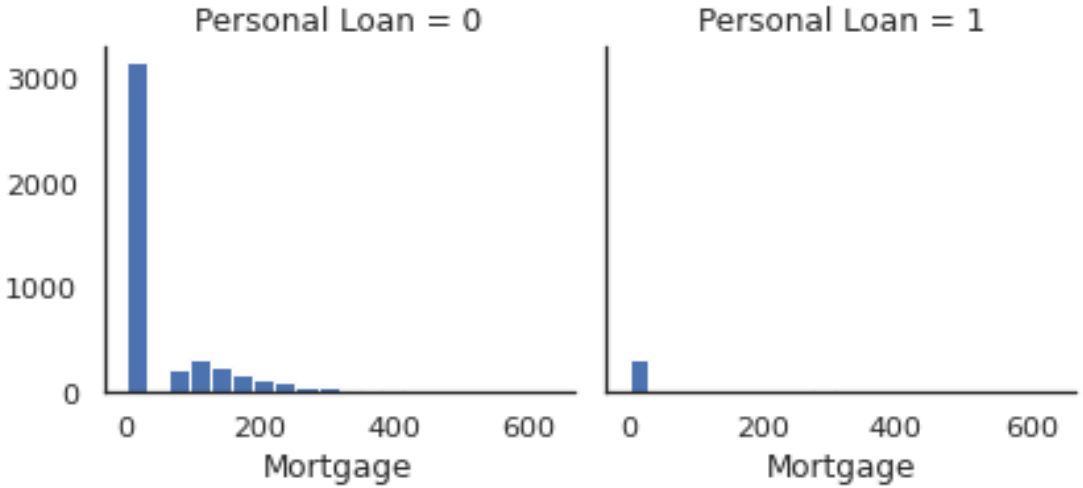
Data science challenges
Reveal data secrets

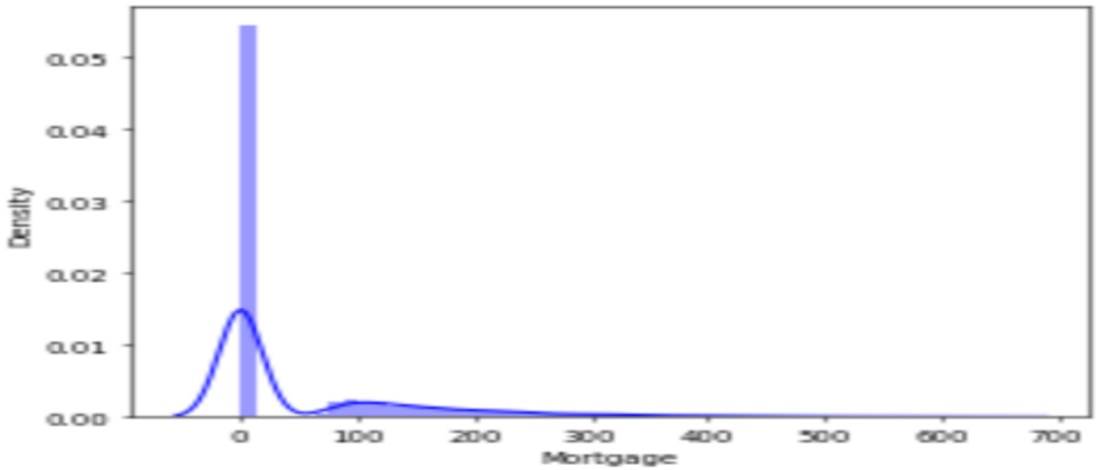


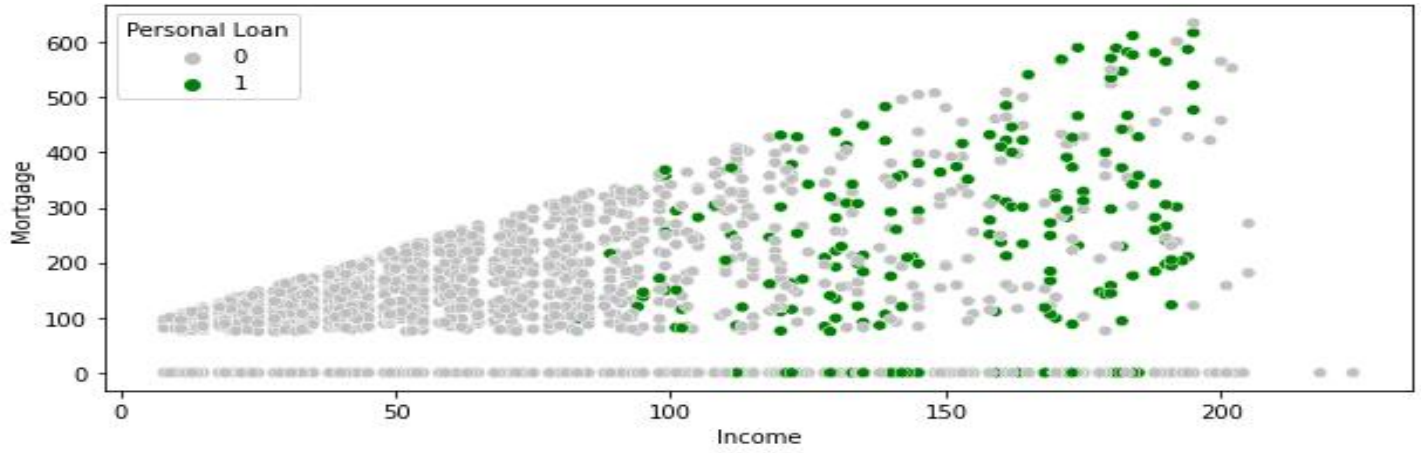


Value is slightly high and hence right skewed as shown in the plot above.

Data science challenges
Reveal data secrets







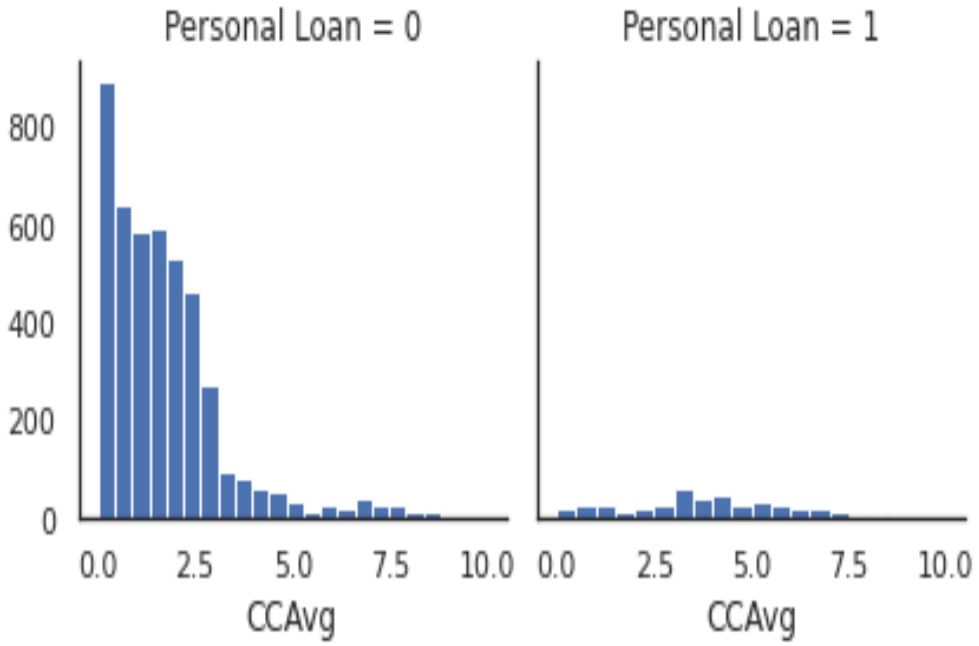
-Above distributionis is right skewed distribution because the tail goes to the right

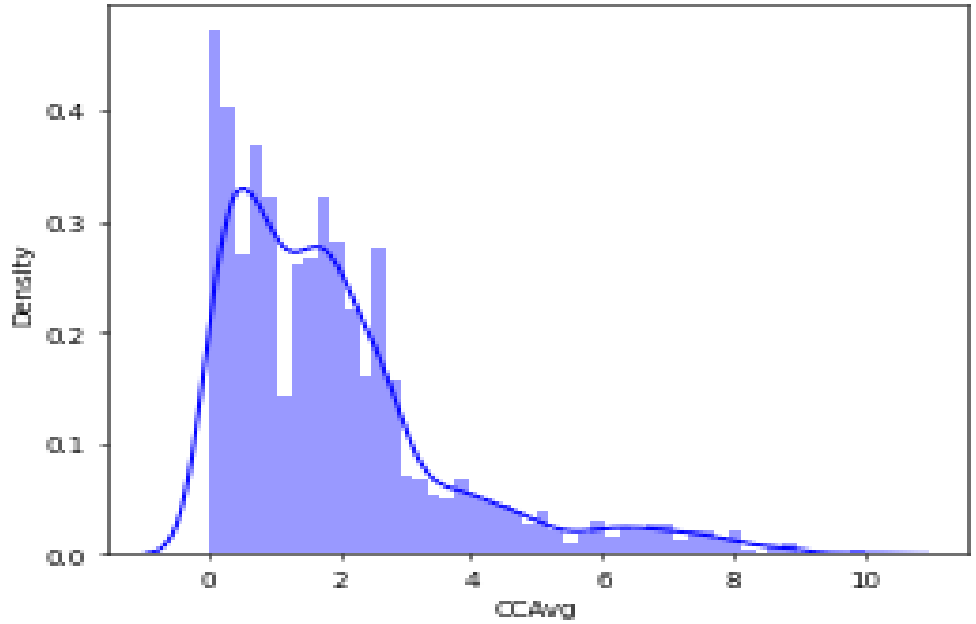
-Most of the customers do not have mortgage

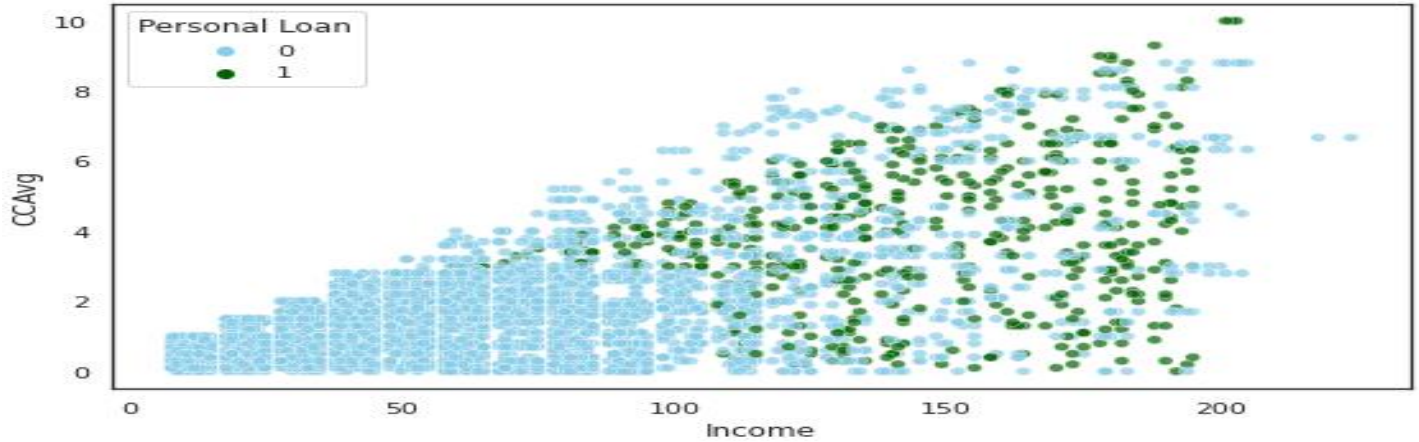
-Customer having 0 Mortgage had high rejection rate.

Data science challenges

Reveal data secrets





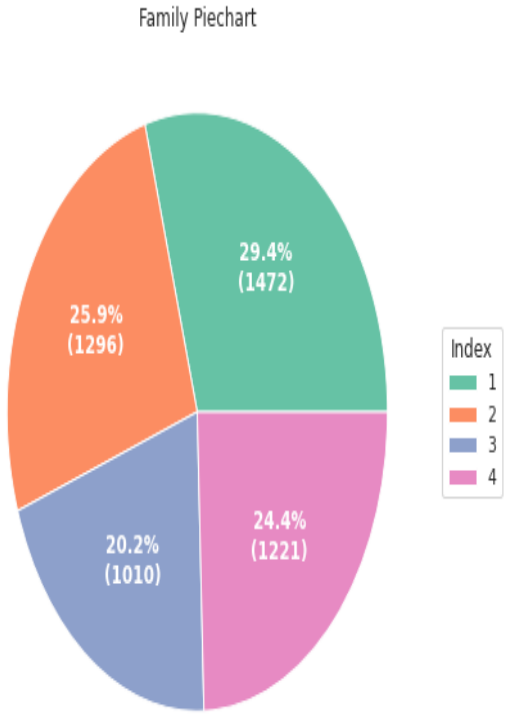


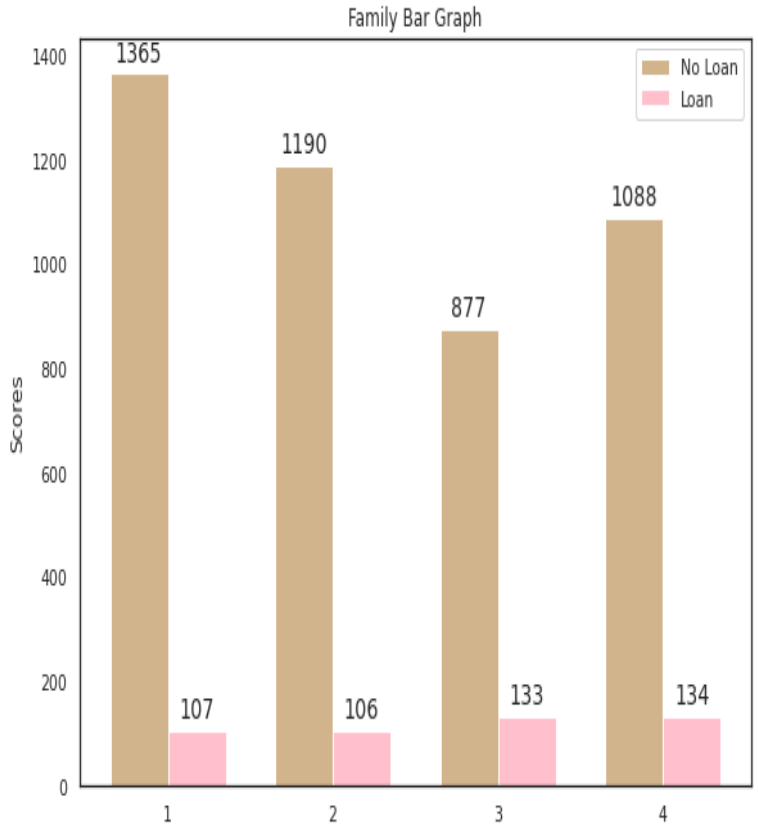
-The above distributionis is right skewed distribution because the tail goes to the right

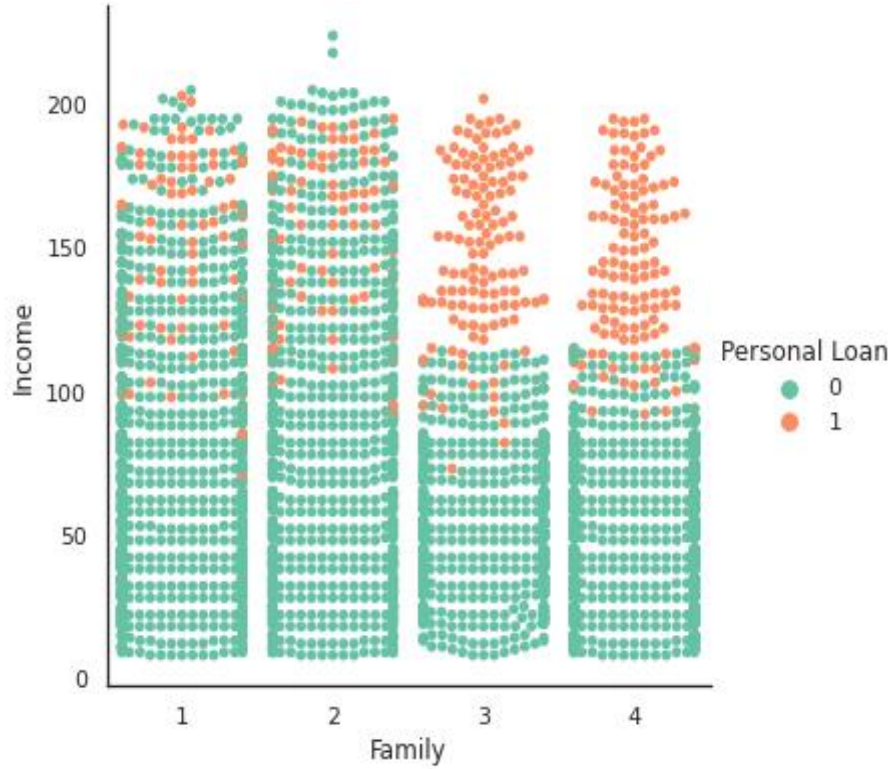
-Customer having low CCAvg mostly rejected Personal loan offer. Custer having CCAvg between 2.5 to 5 has higher rate of acceptance of the offer.

Data science challenges

Reveal data secrets



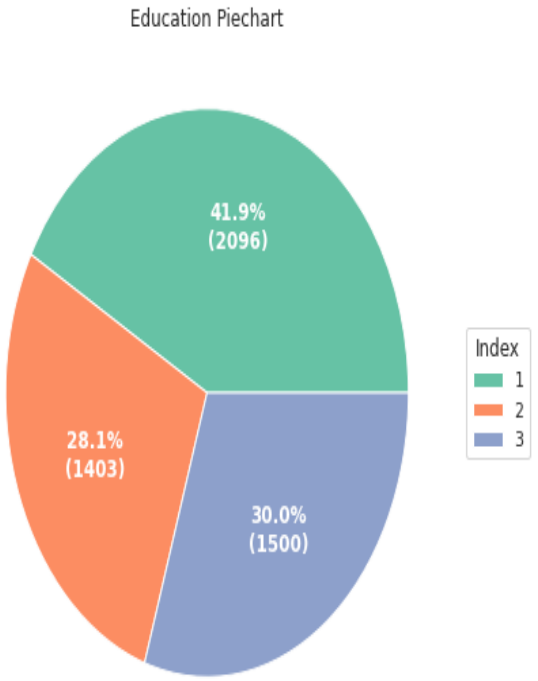


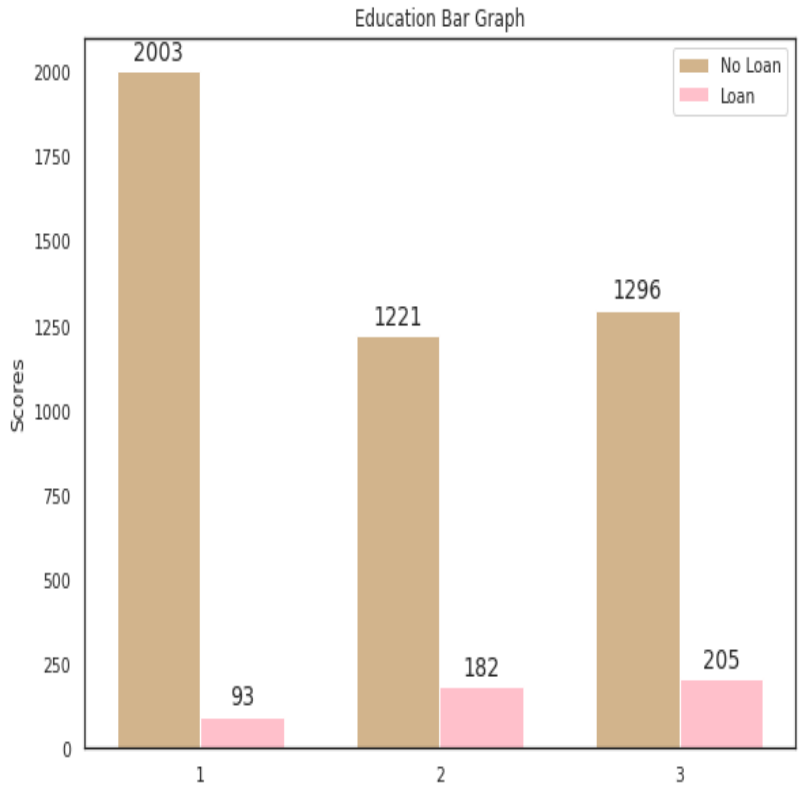


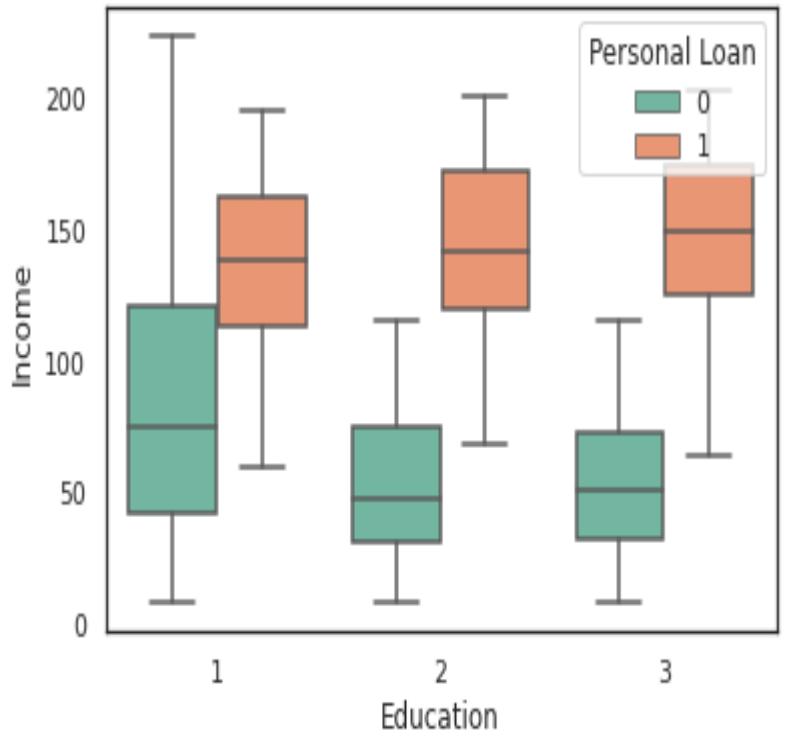
-Family size does not have any impact in personal loan. But it seems families with size of 3 and 4 are more likely to take loan.

- Customers who have family size 3 or greater with higher income between 100k to 200k are more likely to take loan.

Data science challenges
Reveal data secrets

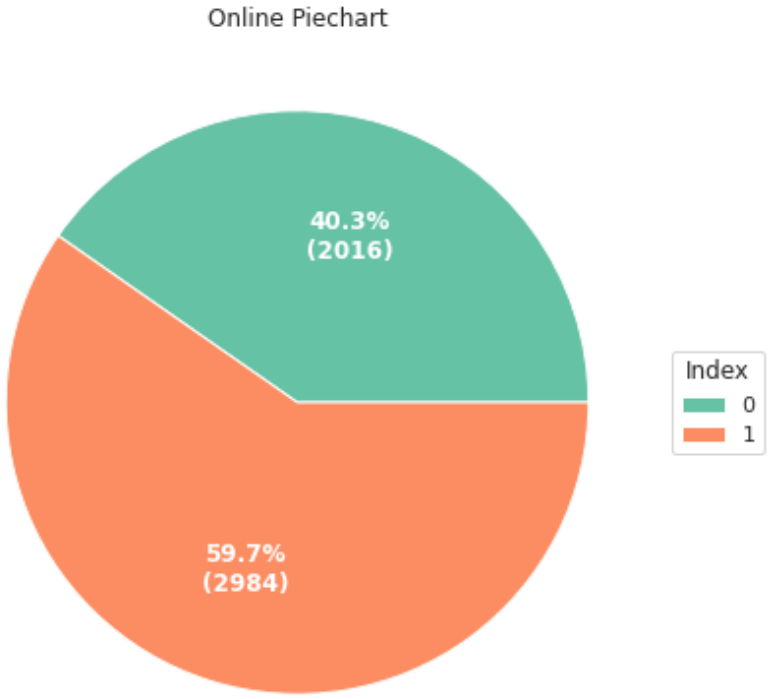


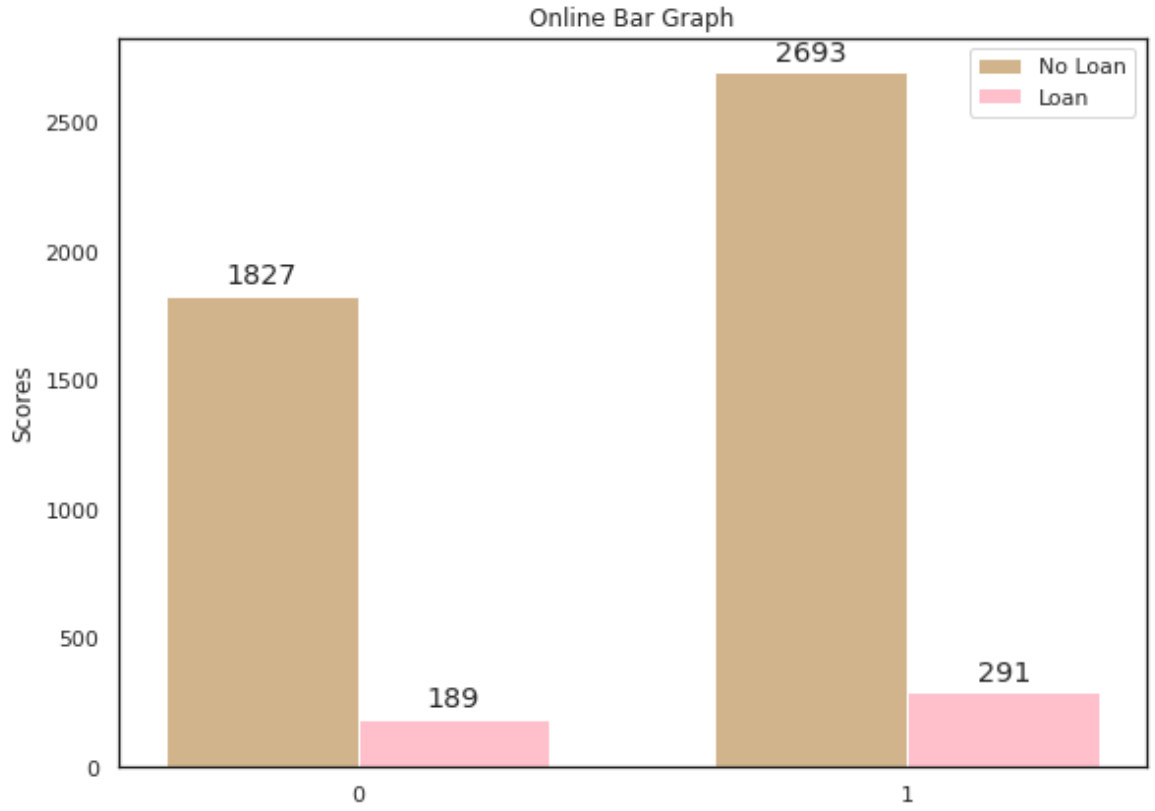




From above we can say that customers with undergraduate level of education and family greater than 3 are good customers who took loan. Customer who took loan have same income range irrespective of education level. Education of Graduate and above have more chance to take loan.

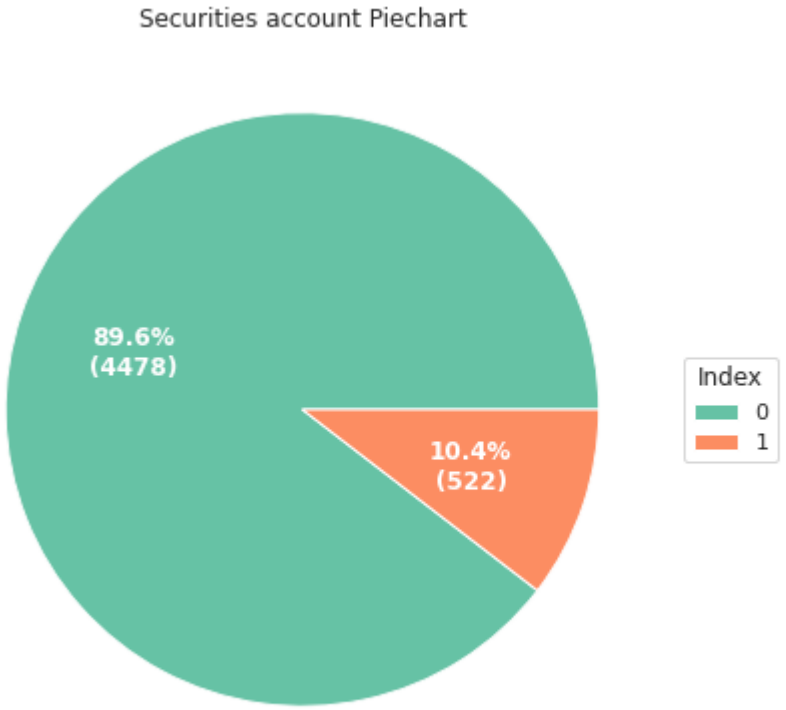
Data science challenges
Reveal data secrets

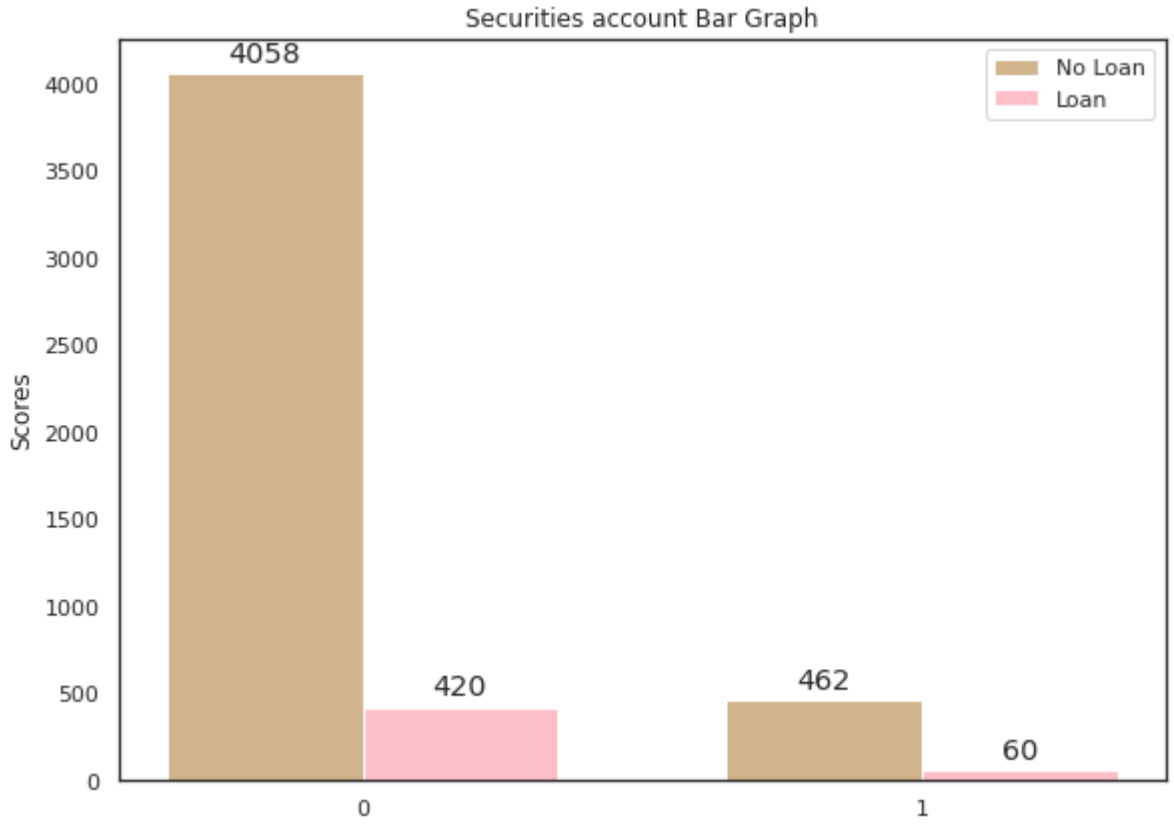




Number of customers who have Online accout is geater than the number of customers who do not have online account .

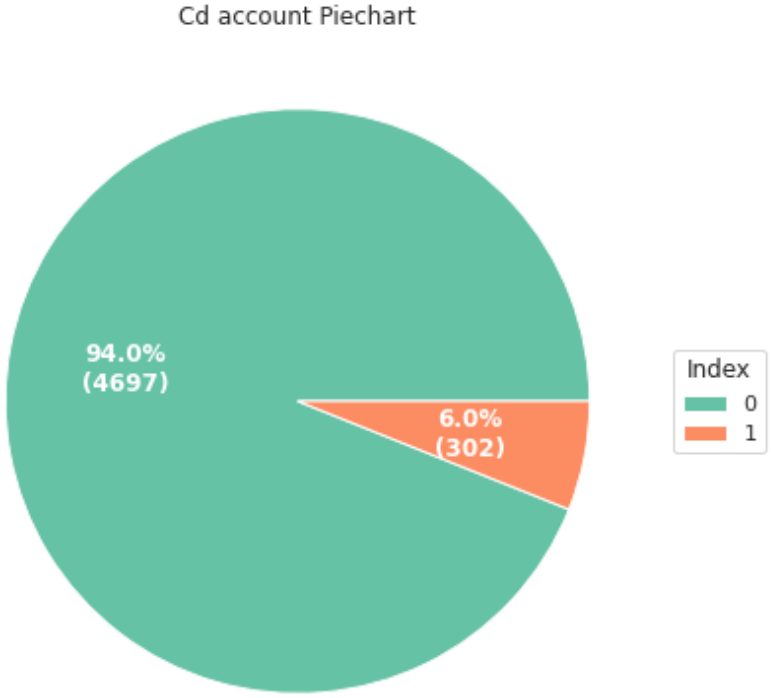
Data science challenges
Reveal data secrets

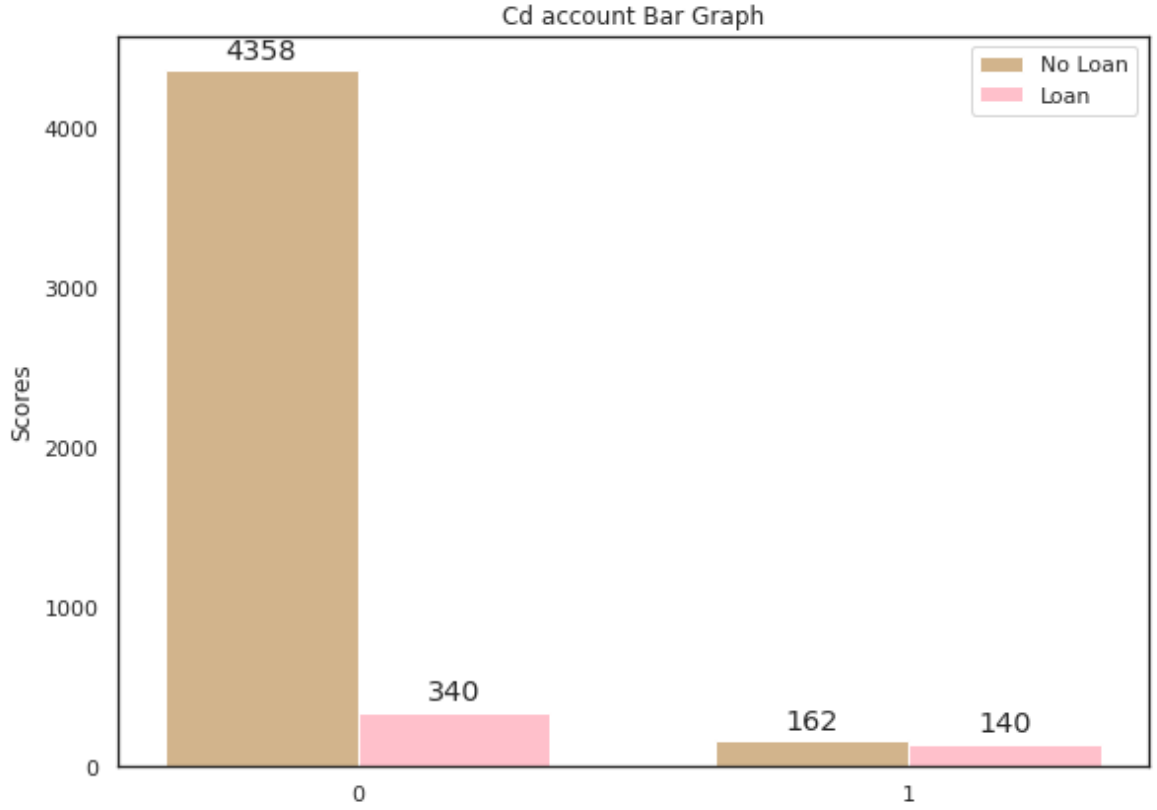




Customers who has securies account are more likly to take loan. Majority of customers who does not have loan do not have securities account.

Data science challenges
Reveal data secrets

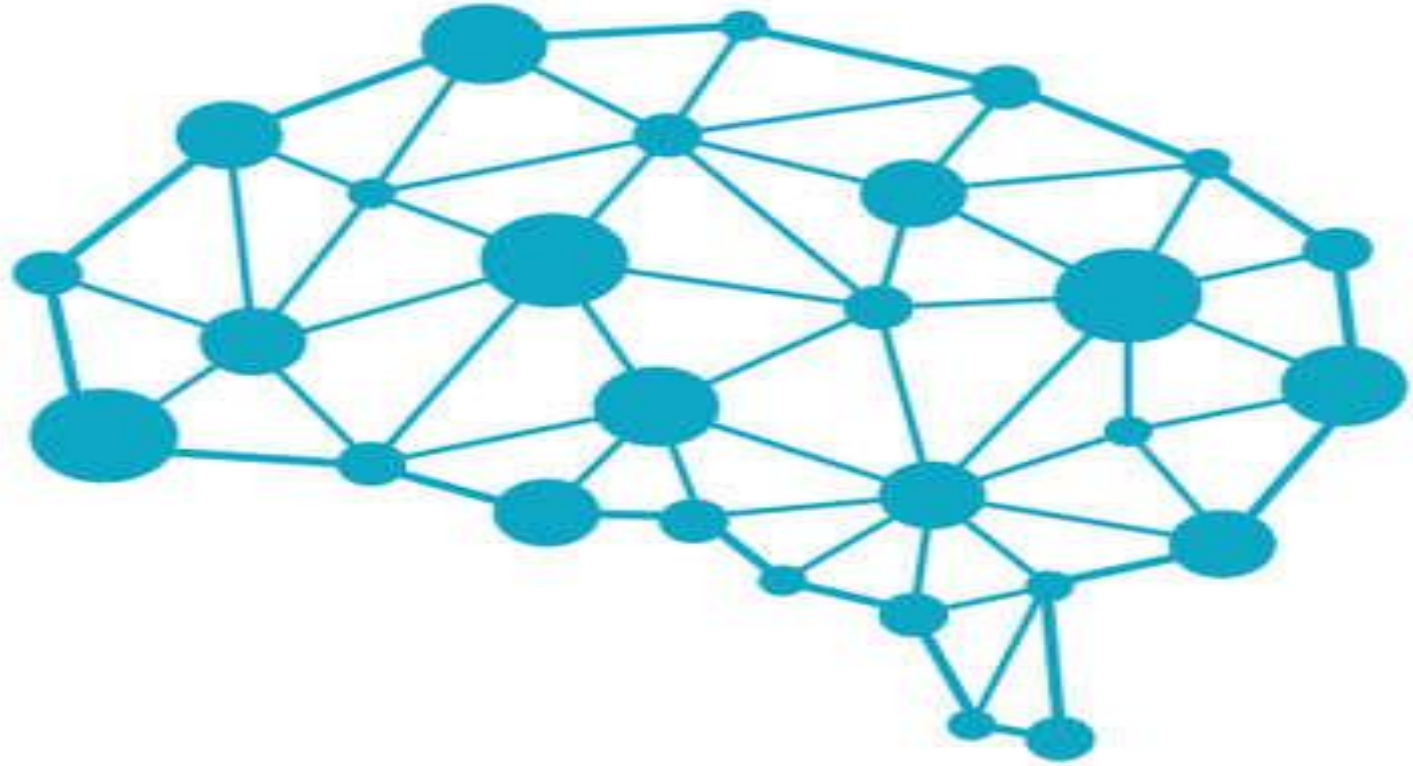




Customers who does not have CD account , does not have loan as well.
This seems to be majority. But almost all customers who has CD account has loan as well

Data science challenges
Reveal data secrets

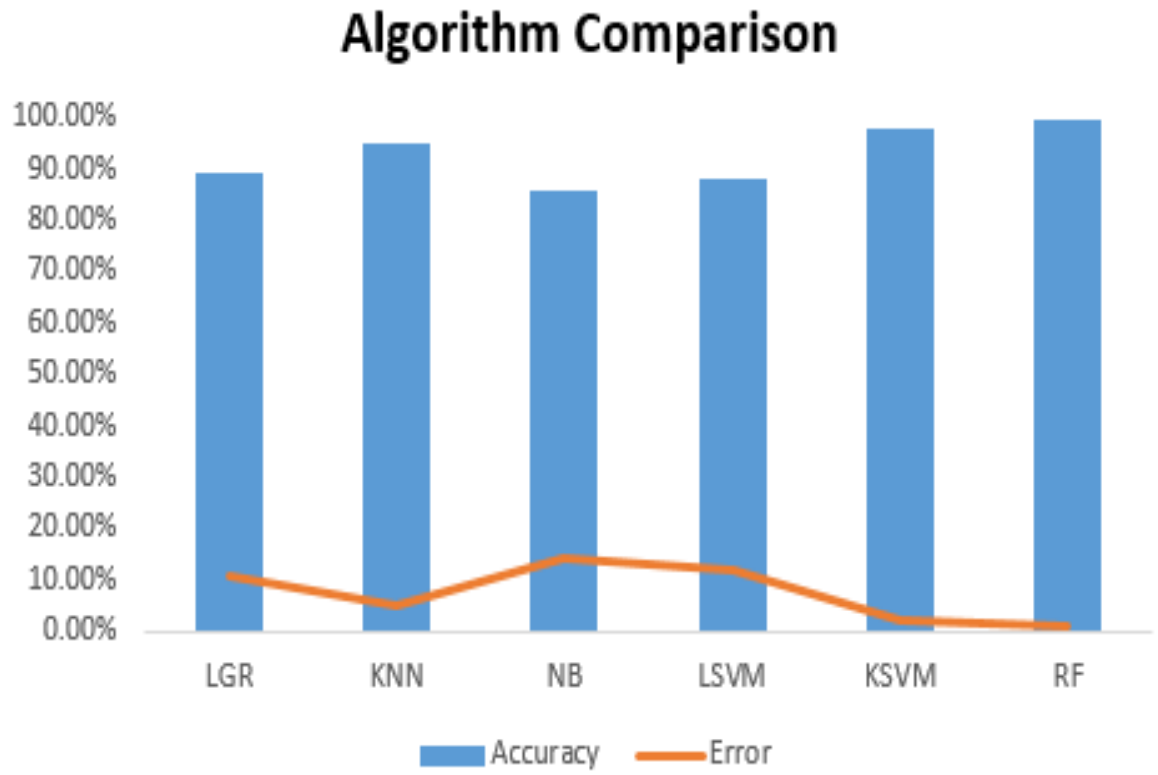
Machine Learning Model



Data science challenges

Reveal data secrets

	Accuracy	Error
Logistic Regression	: 89.38%	10.62%
KNN	: 94.87%	5.125%
Naive Bayes	: 85.73%	14.27%
Linear SVC	: 88.2%	11.8%
Kernel SVM	: 97.82%	2.176%
Random Forest	: 99.26%	0.7375 %



Data science challenges

Reveal data secrets

Problem Statement:

Data science challenges

Reveal data secrets

Based on the previous analysis of the data, it was found that the percentage of loan acceptance by the customer is for those whose income is more than or equal to 100k although the analysis showed a number of those whose income is less than 100 k.

So the following must be done:

1- Restructuring the strategic plan in the Marketing Department and providing support for those whose income is less than 100 K.

2- The marketing department coordinates between all customer service departments until it is confirmed that the set plan will be implemented.

3- After that the role of the sales department is but before that the bank management must ensure that all the sales department employees possess communication skills and emotional intelligence.

4- The sales department communicates with the customer in several ways which he determines according to the situation and according to the situation of the customer.

Data science challenges
Reveal data secrets

Thank you

Data science challenges

Reveal data secrets