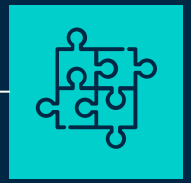
# BANK CREDIT CARD

THARUNESWAR J
NIKITA SHINTRE
RISHABH JAIN
RAVEESH M

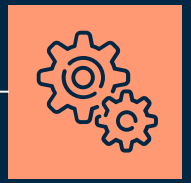
# TABLE OF CONTENTS



01

## INTRODUCTION

Understanding the dataset



02

## TARGET

The outcome expected at the end.

03

## PROCESS

Data visualization and Findings.

# INTRODUCTION

- This dataset contains 5000 unique individual data samples collected from 5000 individuals containing 14 different variables .

- Each unique variable contributes to an overall judgement of outcome if the user will buy credit card.

# The Features

- ID
- Age
- Experience
- Income
- ZIP Code
- Family
- CCAvg
- Education
- Mortgage
- Personal Loan
- Securities Account
- CD Account
- Online
- Credit Card

# The Target

- The target is to use logistic regression to train the algorithm to predict the Credit card purchase by individuals .

| Index | Age | Experience | Income | ZIP Code | Family | CCAvg | Education | Mortgage | Personal Loan | Securities Account | CD Account | Online | CreditCard |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | -0.00847259 | -0.00832576 | -0.0176947 | 0.0134315 | -0.0167972 | -0.0246752 | 0.0214632 | -0.0139199 | -0.0248012 | -0.0169723 | -0.0069094 | -0.00252841 | 0.0170282 |
| Age | 1 | 0.994215 | -0.0552686 | -0.0292163 | -0.0464177 | -0.0520122 | 0.0413344 | -0.0125386 | -0.00772562 | -0.000436242 | 0.00804255 | 0.0137024 | 0.00768104 |
| Experience | 0.994215 | 1 | -0.0465742 | -0.0286255 | -0.0525631 | -0.0500765 | 0.0131518 | -0.0105816 | -0.0074131 | -0.00123213 | 0.0103533 | 0.0138979 | 0.00896745 |
| Income | -0.0552686 | -0.0465742 | 1 | -0.0164098 | -0.157501 | 0.645984 | -0.187524 | 0.206806 | 0.502462 | -0.0026165 | 0.169738 | 0.0142059 | -0.00238501 |
| ZIP Code | -0.0292163 | -0.0286255 | -0.0164098 | 1 | 0.0117782 | -0.00406068 | -0.0173768 | 0.00738338 | 0.000107376 | 0.00470424 | 0.0199719 | 0.0169901 | 0.00769139 |
| Family | -0.0464177 | -0.0525631 | -0.157501 | 0.0117782 | 1 | -0.109275 | 0.0649289 | -0.0204449 | 0.061367 | 0.0199941 | 0.0141104 | 0.010354 | 0.0115881 |
| CCAvg | -0.0520122 | -0.0500765 | 0.645984 | -0.00406068 | -0.109275 | 1 | -0.136124 | 0.109905 | 0.366889 | 0.0150863 | 0.136534 | -0.00361101 | -0.00668949 |
| Education | 0.0413344 | 0.0131518 | -0.187524 | -0.0173768 | 0.0649289 | -0.136124 | 1 | -0.0333271 | 0.136722 | -0.010812 | 0.0139339 | -0.0150038 | -0.0110141 |
| Mortgage | -0.0125386 | -0.0105816 | 0.206806 | 0.00738338 | -0.0204449 | 0.109905 | -0.0333271 | 1 | 0.142095 | -0.00541097 | 0.0893111 | -0.0059949 | -0.00723092 |
| Personal Loan | -0.00772562 | -0.0074131 | 0.502462 | 0.000107376 | 0.061367 | 0.366889 | 0.136722 | 0.142095 | 1 | 0.0219539 | 0.316355 | 0.00627782 | 0.00280151 |
| Securities Account | -0.000436242 | -0.00123213 | -0.0026165 | 0.00470424 | 0.0199941 | 0.0150863 | -0.010812 | -0.00541097 | 0.0219539 | 1 | 0.317034 | 0.0126275 | -0.0150283 |
| CD Account | 0.00804255 | 0.0103533 | 0.169738 | 0.0199719 | 0.0141104 | 0.136534 | 0.0139339 | 0.0893111 | 0.316355 | 0.317034 | 1 | 0.17588 | 0.278644 |
| Online | 0.0137024 | 0.0138979 | 0.0142059 | 0.0169901 | 0.010354 | -0.00361101 | -0.0150038 | -0.0059949 | 0.00627782 | 0.0126275 | 0.17588 | 1 | 0.00420966 |
| CreditCard | 0.00768104 | 0.00896745 | -0.00238501 | 0.00769139 | 0.0115881 | -0.00668949 | -0.0110141 | -0.00723092 | 0.00280151 | -0.0150283 | 0.278644 | 0.00420966 | 1 |

# Dataset Description

- The target is to use logistic regression to train the algorithm to predict the Credit card purchase by individuals .

| Index | ID | Age | Experience | Income | ZIP Code | Family | CCAvg | Education | Mortgage | Personal Loan | Securities Account | CD Account | Online | CreditCard |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| std | 1443.52 | 11.4632 | 11.468 | 46.0337 | 2121.85 | 1.14766 | 1.74766 | 0.839869 | 101.714 | 0.294621 | 0.305809 | 0.23825 | 0.490589 | 0.455637 |
| min | 1 | 23 | -3 | 8 | 9307 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| mean | 2500.5 | 45.3384 | 20.1046 | 73.7742 | 93152.5 | 2.3964 | 1.93794 | 1.881 | 56.4988 | 0.096 | 0.1044 | 0.0604 | 0.5968 | 0.294 |
| max | 5000 | 67 | 43 | 224 | 96651 | 4 | 10 | 3 | 635 | 1 | 1 | 1 | 1 | 1 |
| count | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 |
| 75% | 3750.25 | 55 | 30 | 98 | 94608 | 3 | 2.5 | 3 | 101 | 0 | 0 | 0 | 1 | 1 |
| 50% | 2500.5 | 45 | 20 | 64 | 93437 | 2 | 1.5 | 2 | 0 | 0 | 0 | 0 | 1 | 0 |
| 25% | 1250.75 | 35 | 10 | 39 | 91911 | 1 | 0.7 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

# Negative Values in Experience

The Min Value of Experience is shown as -3 and as experience cannot be a negative value the negative values are replaced with median values.

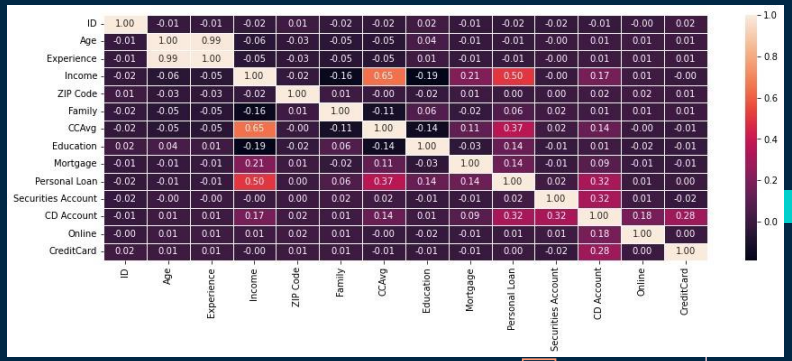
| Index | ID | Age | Experienc |
|---|---|---|---|
| std | 1443.52 | 11.4632 | 11.468 |
| min | 1 | 23 | -3 |
| mean | 2500.5 | 45.3384 | 20.1046 |
| max | 5000 | 67 | 43 |
| count | 5000 | 5000 | 5000 |
| 75% | 3750.25 | 55 | 30 |
| 50% | 2500.5 | 45 | 20 |
| 25% | 1250.75 | 35 | 10 |

| Index | ID | Age | xperienc | Income |
|---|---|---|---|---|
| count | 5000 | 5000 | 5000 | 5000 |
| mean | 2500.5 | 45.3384 | 20.3276 | 73.7742 |
| std | 1443.52 | 11.4632 | 11.253 | 46.0337 |
| min | 1 | 23 | 0 | 8 |
| 25% | 1250.75 | 35 | 11 | 39 |
| 50% | 2500.5 | 45 | 20 | 64 |
| 75% | 3750.25 | 55 | 30 | 98 |
| max | 5000 | 67 | 43 | 224 |

# The Process - CORRELATION BETWEEN ALL VARIABLES

- This can be done by relating the Values of Credit Card variable with the other values.
- Then the relationship between all the variables can be looked at with the help of correlation graph and Heatmap.

| Index | Age | Experience | Income | ZIP Code | Family | CCAvg | Education | Mortgage | Personal Loan | Securities Account | CD Account | Online | CreditCard |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | -0.00847259 | -0.00832576 | -0.0176947 | 0.0134315 | -0.0167972 | -0.0246752 | 0.0214632 | -0.0139199 | -0.0248012 | -0.0169723 | -0.0069094 | -0.00252841 | 0.0170282 |
| Age | 1 | 0.994215 | -0.0552686 | -0.0292163 | -0.0464177 | -0.0520122 | 0.0413344 | -0.0125386 | -0.00772562 | -0.000436242 | 0.00804255 | 0.0137024 | 0.00768104 |
| Experience | 0.994215 | 1 | -0.0465742 | -0.0286255 | -0.0525631 | -0.0500765 | 0.0131518 | -0.0105816 | -0.0074131 | -0.00123213 | 0.0103533 | 0.0138979 | 0.00896745 |
| Income | -0.0552686 | -0.0465742 | 1 | -0.0164098 | -0.157501 | 0.645984 | -0.187524 | 0.206806 | 0.502462 | -0.0026165 | 0.169738 | 0.0142059 | -0.00238501 |
| ZIP Code | -0.0292163 | -0.0286255 | -0.0164098 | 1 | 0.0117782 | -0.00406068 | -0.0173768 | 0.00738338 | 0.000107376 | 0.00470424 | 0.0199719 | 0.0169901 | 0.00769139 |
| Family | -0.0464177 | -0.0525631 | -0.157501 | 0.0117782 | 1 | -0.109275 | 0.0649289 | -0.0204449 | 0.061367 | 0.0199941 | 0.0141104 | 0.010354 | 0.0115881 |
| CCAvg | -0.0520122 | -0.0500765 | 0.645984 | -0.00406068 | -0.109275 | 1 | -0.136124 | 0.109905 | 0.366889 | 0.0150863 | 0.136534 | -0.00361101 | -0.00668949 |
| Education | 0.0413344 | 0.0131518 | -0.187524 | -0.0173768 | 0.0649289 | -0.136124 | 1 | -0.0333271 | 0.136722 | -0.010812 | 0.0139339 | -0.0150038 | -0.0110141 |
| Mortgage | -0.0125386 | -0.0105816 | 0.206806 | 0.00738338 | -0.0204449 | 0.109905 | -0.0333271 | 1 | 0.142095 | -0.00541097 | 0.0893111 | -0.0059949 | -0.00723092 |
| Personal Loan | -0.00772562 | -0.0074131 | 0.502462 | 0.000107376 | 0.061367 | 0.366889 | 0.136722 | 0.142095 | 1 | 0.0219539 | 0.316355 | 0.00627782 | 0.00280151 |
| Securities Account | -0.000436242 | -0.00123213 | -0.0026165 | 0.00470424 | 0.0199941 | 0.0150863 | -0.010812 | -0.00541097 | 0.0219539 | 1 | 0.317034 | 0.0126275 | -0.0150283 |
| CD Account | 0.00804255 | 0.0103533 | 0.169738 | 0.0199719 | 0.0141104 | 0.136534 | 0.0139339 | 0.0893111 | 0.316355 | 0.317034 | 1 | 0.17588 | 0.278644 |
| Online | 0.0137024 | 0.0138979 | 0.0142059 | 0.0169901 | 0.010354 | -0.00361101 | -0.0150038 | -0.0059949 | 0.00627782 | 0.0126275 | 0.17588 | 1 | 0.00420966 |
| CreditCard | 0.00768104 | 0.00896745 | -0.00238501 | 0.00769139 | 0.0115881 | -0.00668949 | -0.0110141 | -0.00723092 | 0.00280151 | -0.0150283 | 0.278644 | 0.00420966 | 1 |

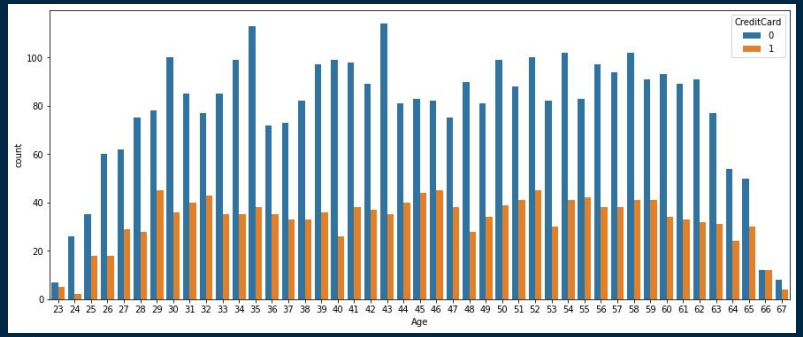


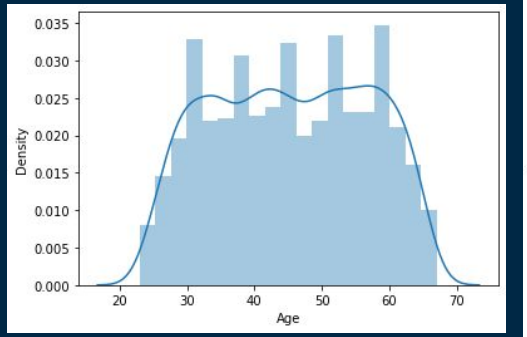
# The Process - CREDIT CARD VS OTHER VARIABLES

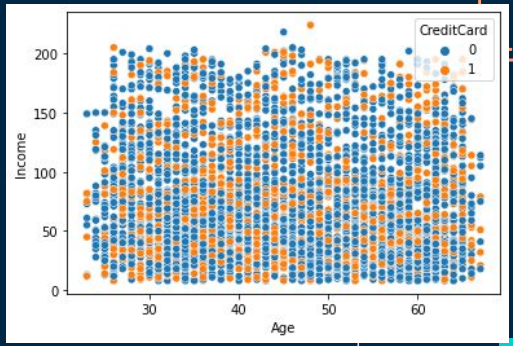
- Analysis on how the Credit card  and other variables values are related to each other variables prove to be an important tool in providing the weight to each variable in the regression model.
- Scatterplot with Regression lines is a better way to understand the relationship between credit Card values and the other variables .
- Representing them together makes it conventional  to determine the weightage of each variable in the regression model.

# Age

- Three small peaks can be indicating three values of age would be slightly more in number. However, the mean and median of the attribute is equal. So the distribution is Neutral.



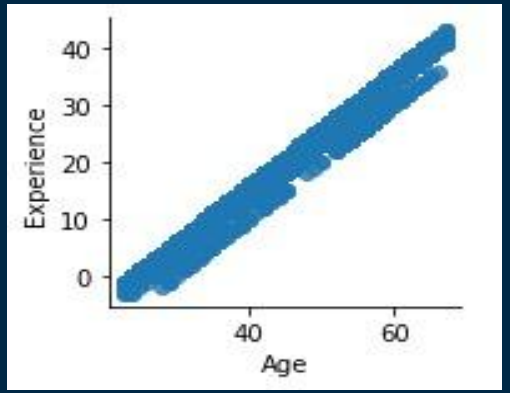



# Experience

"Experience" feature is also almost normally distributed and mean is also equal to median.But there are some negative values present which should be deleted, as Experience can not be negative. Age and Experience have similar distribution and the mean and median are almost equal. So experience can be dropped to prevent Multicollinearity issue.

| | ID | Age | Experience |
|---|---|---|---|
| ID | 1.00 | -0.01 | -0.01 |
| Age | -0.01 | 1.00 | 0.99 |
| Experience | -0.01 | 0.99 | 1.00 |
| Income | -0.02 | -0.06 | -0.05 |
| ZIP Code | 0.01 | -0.03 | -0.03 |
| Family | -0.02 | -0.05 | -0.05 |
| CCAvg | -0.02 | -0.05 | -0.05 |
| Education | 0.02 | 0.04 | 0.01 |
| Mortgage | -0.01 | -0.01 | -0.01 |
| Personal Loan | -0.02 | -0.01 | -0.01 |
| Securities Account | -0.02 | -0.00 | -0.00 |
| CD Account | -0.01 | 0.01 | 0.01 |
| Online | -0.00 | 0.01 | 0.01 |
| CreditCard | 0.02 | 0.01 | 0.01 |

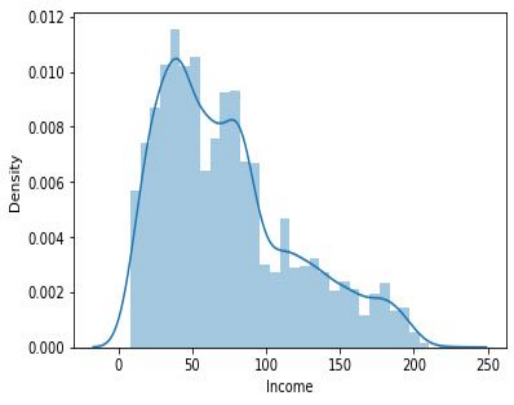


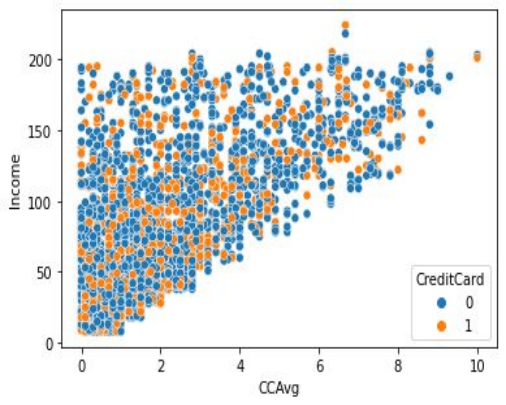

| Index | Age | xperienc |
|---|---|---|
| count | 5000 | 5000 |
| mean | 45.3384 | 20.3276 |
| std | 11.4632 | 11.253 |
| min | 23 | 0 |
| 25% | 35 | 11 |
| 50% | 45 | 20 |
| 75% | 55 | 30 |
| max | 67 | 43 |

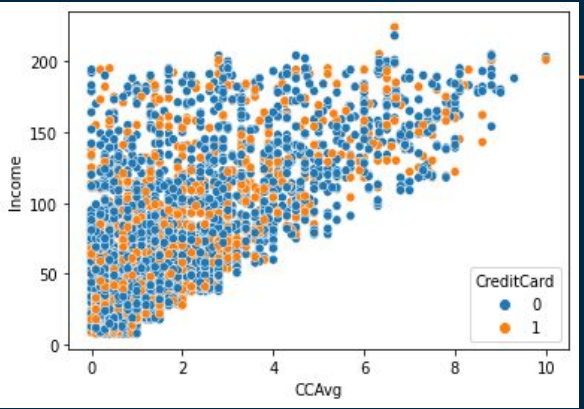
# Income

We can clearly see data is highly peaked on the left . Data for less income customers is more in the Dataset.

Mean is greater than median.Also we can confirm from this that majority of the customers have income between 30-60.



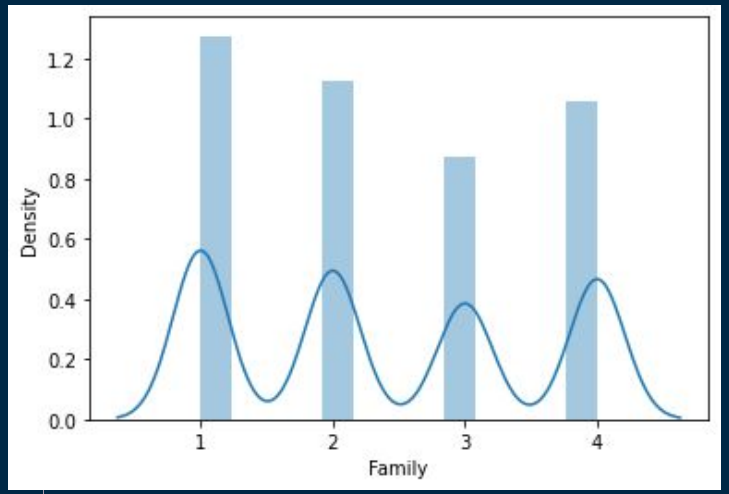



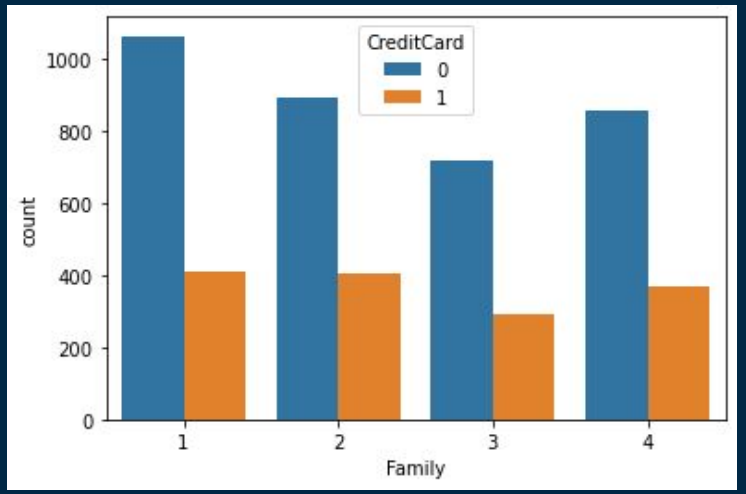




| reditCar | Income |
|---|---|
| 0 | 73.845 |
| 1 | 73.6041 |

# Family

It has 4 peaks(4 values) , families with least member is highest in the sample.

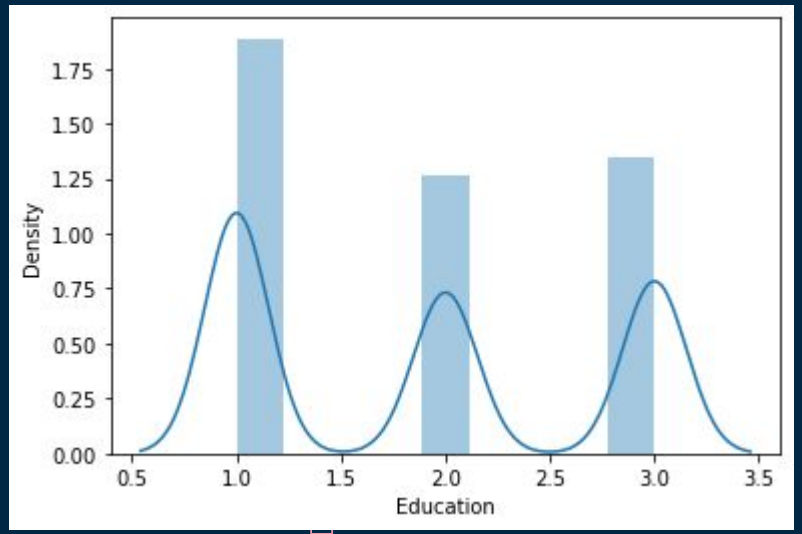



| Family | Income |
|---|---|
| 1 | 78.6223 |
| 2 | 84.2392 |
| 3 | 66.905 |
| 4 | 62.5131 |

# Education

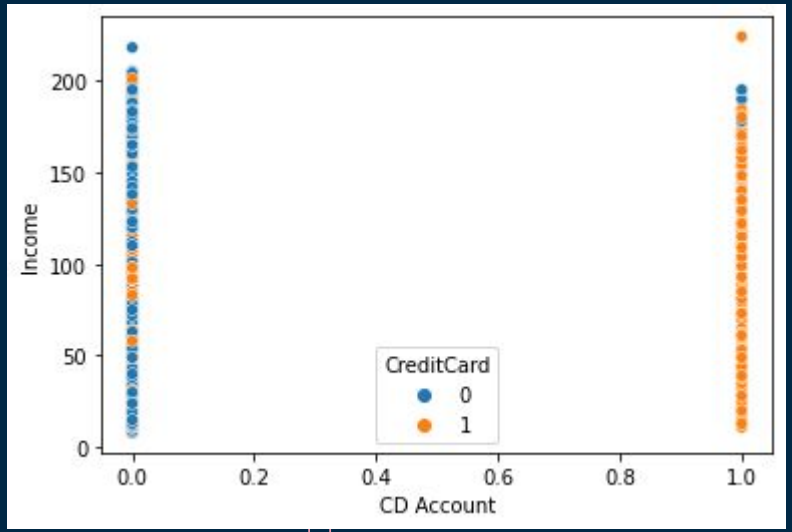
Mean and median is almost equal. Data is finely distributed. A few peaks shows different values dominance.

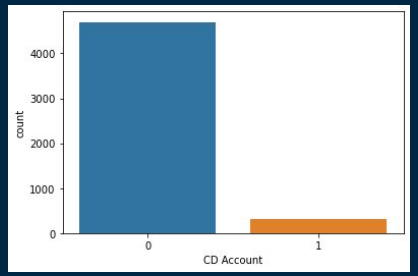



| Educatio | Income | CCAvg |
|---|---|---|
| 1 | 85.5864 | 2.26083 |
| 2 | 64.3136 | 1.68509 |
| 3 | 66.1226 | 1.72339 |

# CD Account

Most of the customers dont have CD accounts.
Customers who has CD Accounts Certainly have Credit Cards



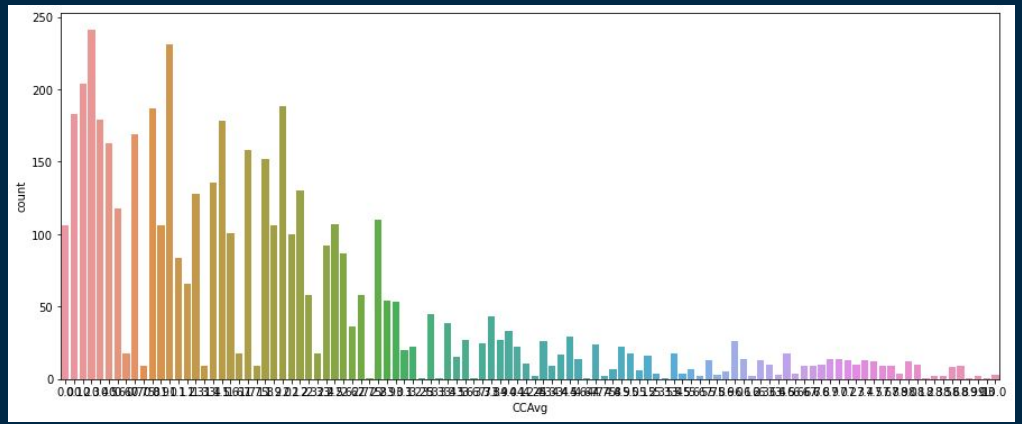



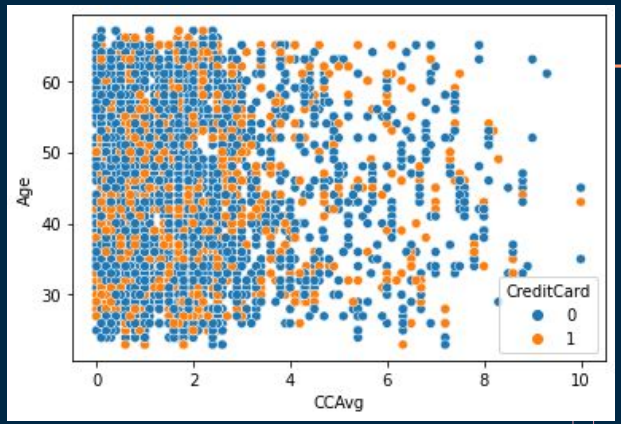

| CreditCar | CD Account |
|---|---|
| 0 | 0.0175637 |
| 1 | 0.163265 |

# CC Avg

The data is more biased to the left . There are more people with less CC Average and the Credit Card holders are evenly distributed here.

| reditCard | CCAvg |
|---|---|
| 0 | 1.94548 |
| 1 | 1.91982 |



# The Process - PREPARING TRAIN AND TEST VALUES

- The dataset is divided into test and train datasets in with 80% of the entries are used to train the data and the rest 20% is used to test the data .
- The Values that are then plotted as X and Y in which Y is the BMI dataset and X is a combination of the other variables in the dataset.
- X train and Y train values are used to train the model and the outcome is then plotted with Y test values to find the accuracy of the model.

# The Variables in order of Weightage to Credit card

1. CD account
2. Family
3. Education
4. Age
5. Securities account
6. Mortgage
7. CC average
8. Online
9. Personal Loan
10. income

# Model

Features = All Features Except - ID, ZIP Code , Experience

```
------------------------
------------------------
LOGISTIC REGRESSION
------------------------
------------------------
--------------------
| Confusion Matrix |
--------------------
[[697   7]
 [245  51]]
------------------------
| Classifiction Report |
------------------------
              precision    recall  f1-score   support

           0       0.74      0.99      0.85       704
           1       0.88      0.17      0.29       296

    accuracy                           0.75      1000
   macro avg       0.81      0.58      0.57      1000
weighted avg       0.78      0.75      0.68      1000

------------------------
| Accuracy Score |
------------------------
0.748
------------------------
| Cross Validation Score |
------------------------
0.7422
```

```
------------------------
------------------------
KNN CLASSIFIER
------------------------
------------------------
--------------------
| Confusion Matrix |
--------------------
[[697   7]
 [245  51]]
------------------------
| Classifiction Report |
------------------------
              precision    recall  f1-score   support

           0       0.74      0.99      0.85       704
           1       0.88      0.17      0.29       296

    accuracy                           0.75      1000
   macro avg       0.81      0.58      0.57      1000
weighted avg       0.78      0.75      0.68      1000

------------------------
| Accuracy Score |
------------------------
0.748
------------------------
| Cross Validation Score |
------------------------
0.6195999999999999
```

```
------------------------
------------------------
RANDOM FORST CLASSIFIER
------------------------
------------------------
--------------------
| Confusion Matrix |
--------------------
[[698   6]
 [247  49]]
------------------------
| Classifiction Report |
------------------------
              precision    recall  f1-score   support

           0       0.74      0.99      0.85       704
           1       0.89      0.17      0.28       296

    accuracy                           0.75      1000
   macro avg       0.81      0.58      0.56      1000
weighted avg       0.78      0.75      0.68      1000

------------------------
| Accuracy Score |
------------------------
0.747
------------------------
| Cross Validation Score |
------------------------
0.7414
```