```
#install.packages
library(caret)
library(ISLR)
library(gmodels)
library(ggplot2)
library(FNN)
library(dummies)
library(fastDummies)
library(e1071)
library(dplyr)

#install.packages("caret")
library(caret)
# install.packages("ISLR") # only install if needed
library(ISLR)
```

```
library(readxl)
uk <- read.csv("C:/Users/mnooo/Desktop/Datasets/UniversalBank.csv")
#View(uk)
```

```
# Clean up data and get it ready for KNN
library(dplyr)


uk$Education_1 = ifelse(uk$Education ==1, 1, 0)
uk$Education_2 = ifelse(uk$Education ==2, 1, 0)
uk$Education_3 = ifelse(uk$Education ==3, 1, 0)
uk$Personal_Loan = as.factor(uk$Personal.Loan)

View(uk)

uk1 <- uk[, c(2:4, 6:7, 9, 11:18)]   # Select a subset of variables, without ID and zipcode, ori
ginal variable Personal.Loan
head(uk1)
```

```
##   Age Experience Income Family CCAvg Mortgage Securities.Account CD.Account
## 1  25          1     49      4   1.6        0                  1          0
## 2  45         19     34      3   1.5        0                  1          0
## 3  39         15     11      1   1.0        0                  0          0
## 4  35          9    100      1   2.7        0                  0          0
## 5  35          8     45      4   1.0        0                  0          0
## 6  37         13     29      4   0.4      155                  0          0
##   Online CreditCard Education_1 Education_2 Education_3 Personal_Loan
## 1      0          0           1           0           0             0
## 2      0          0           1           0           0             0
## 3      0          0           1           0           0             0
## 4      0          0           0           1           0             0
## 5      0          1           0           1           0             0
## 6      1          0           0           1           0             0
```

```
str(uk1)
```

```
## 'data.frame':    5000 obs. of  14 variables:
##  $ Age               : int  25 45 39 35 35 37 53 50 35 34 ...
##  $ Experience        : int  1 19 15 9 8 13 27 24 10 9 ...
##  $ Income            : int  49 34 11 100 45 29 72 22 81 180 ...
##  $ Family            : int  4 3 1 1 4 4 2 1 3 1 ...
##  $ CCAvg             : num  1.6 1.5 1 2.7 1 0.4 1.5 0.3 0.6 8.9 ...
##  $ Mortgage          : int  0 0 0 0 0 155 0 0 104 0 ...
##  $ Securities.Account: int  1 1 0 0 0 0 0 0 0 0 ...
##  $ CD.Account        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Online            : int  0 0 0 0 0 1 1 0 1 0 ...
##  $ CreditCard        : int  0 0 0 0 1 0 0 1 0 0 ...
##  $ Education_1       : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ Education_2       : num  0 0 0 1 1 1 1 0 1 0 ...
##  $ Education_3       : num  0 0 0 0 0 0 0 1 0 1 ...
##  $ Personal_Loan     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 ...
```

```
set.seed(15)
Test_Index = createDataPartition(uk1$Age,p=0.2, list=FALSE) # 20% reserved for Test
Test_Data = uk1[Test_Index,]
TraVal_Data = uk1[-Test_Index,] # Validation and Training data is rest
Train_Index = createDataPartition(TraVal_Data$Age,p=0.75, list=FALSE) # 75% of remaining data as
training

# 75% of the 80% is 60% which is training

Train_Data = TraVal_Data[Train_Index,]
Validation_Data = TraVal_Data[-Train_Index,] # rest as validation
View(Train_Data)
summary(Train_Data)
```

```
##       Age           Experience        Income           Family
##  Min.   :23.00   Min.   :-3.00   Min.   :  8.0   Min.   :1.000
##  1st Qu.:35.00   1st Qu.:10.00   1st Qu.: 38.0   1st Qu.:1.000
##  Median :45.00   Median :20.00   Median : 63.0   Median :2.000
##  Mean   :45.37   Mean   :20.13   Mean   : 72.3   Mean   :2.401
##  3rd Qu.:55.00   3rd Qu.:30.00   3rd Qu.: 95.0   3rd Qu.:3.000
##  Max.   :67.00   Max.   :43.00   Max.   :218.0   Max.   :4.000
##      CCAvg           Mortgage      Securities.Account   CD.Account
##  Min.   :0.0000   Min.   :  0.00   Min.   :0.0000     Min.   :0.00000
##  1st Qu.:0.6925   1st Qu.:  0.00   1st Qu.:0.0000     1st Qu.:0.00000
##  Median :1.5000   Median :  0.00   Median :0.0000     Median :0.00000
##  Mean   :1.8996   Mean   : 55.77   Mean   :0.1037     Mean   :0.06033
##  3rd Qu.:2.5000   3rd Qu.: 98.00   3rd Qu.:0.0000     3rd Qu.:0.00000
##  Max.   :9.3000   Max.   :635.00   Max.   :1.0000     Max.   :1.00000
##      Online        CreditCard     Education_1     Education_2
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :1.000   Median :0.000   Median :0.000   Median :0.0000
##  Mean   :0.599   Mean   :0.301   Mean   :0.405   Mean   :0.2887
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.000   Max.   :1.000   Max.   :1.0000
##   Education_3     Personal_Loan
##  Min.   :0.0000   0:2718
##  1st Qu.:0.0000   1: 282
##  Median :0.0000
##  Mean   :0.3063
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
summary(Validation_Data)
```

```
##       Age           Experience        Income           Family
##  Min.   :23.00   Min.   :-2.00   Min.   :  8.00   Min.   :1.000
##  1st Qu.:35.00   1st Qu.:10.00   1st Qu.: 39.00   1st Qu.:1.000
##  Median :45.00   Median :20.00   Median : 65.00   Median :2.000
##  Mean   :45.19   Mean   :19.92   Mean   : 75.84   Mean   :2.396
##  3rd Qu.:55.00   3rd Qu.:29.00   3rd Qu.:102.50   3rd Qu.:3.000
##  Max.   :67.00   Max.   :43.00   Max.   :204.00   Max.   :4.000
##      CCAvg          Mortgage      Securities.Account   CD.Account
##  Min.   : 0.00   Min.   :  0.00   Min.   :0.0000     Min.   :0.00000
##  1st Qu.: 0.70   1st Qu.:  0.00   1st Qu.:0.0000     1st Qu.:0.00000
##  Median : 1.60   Median :  0.00   Median :0.0000     Median :0.00000
##  Mean   : 2.03   Mean   : 59.97   Mean   :0.0951     Mean   :0.06006
##  3rd Qu.: 2.60   3rd Qu.:110.00   3rd Qu.:0.0000     3rd Qu.:0.00000
##  Max.   :10.00   Max.   :617.00   Max.   :1.0000     Max.   :1.00000
##      Online         CreditCard      Education_1      Education_2
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.5696   Mean   :0.2803   Mean   :0.4454   Mean   :0.2633
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##   Education_3     Personal_Loan
##  Min.   :0.0000   0:902
##  1st Qu.:0.0000   1: 97
##  Median :0.0000
##  Mean   :0.2913
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
summary(Test_Data)
```

```
##       Age          Experience        Income           Family
##  Min.   :23.00   Min.   :-3.00   Min.   :  8.00   Min.   :1.000
##  1st Qu.:35.00   1st Qu.:10.00   1st Qu.: 39.00   1st Qu.:1.000
##  Median :45.00   Median :20.00   Median : 65.00   Median :2.000
##  Mean   :45.39   Mean   :20.21   Mean   : 76.14   Mean   :2.382
##  3rd Qu.:55.00   3rd Qu.:30.00   3rd Qu.:103.00   3rd Qu.:3.000
##  Max.   :67.00   Max.   :42.00   Max.   :224.00   Max.   :4.000
##      CCAvg           Mortgage      Securities.Account   CD.Account
##  Min.   : 0.000   Min.   :  0.00   Min.   :0.0000     Min.   :0.00000
##  1st Qu.: 0.700   1st Qu.:  0.00   1st Qu.:0.0000     1st Qu.:0.00000
##  Median : 1.600   Median :  0.00   Median :0.0000     Median :0.00000
##  Mean   : 1.961   Mean   : 55.23   Mean   :0.1159     Mean   :0.06094
##  3rd Qu.: 2.600   3rd Qu.:101.00   3rd Qu.:0.0000     3rd Qu.:0.00000
##  Max.   :10.000   Max.   :569.00   Max.   :1.0000     Max.   :1.00000
##      Online         CreditCard      Education_1      Education_2
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.6174   Mean   :0.2867   Mean   :0.4356   Mean   :0.2737
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##   Education_3     Personal_Loan
##  Min.   :0.0000   0:900
##  1st Qu.:0.0000   1:101
##  Median :0.0000
##  Mean   :0.2907
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
str(Train_Data)
```

```
## 'data.frame':    3000 obs. of  14 variables:
##  $ Age               : int  25 37 53 50 35 34 65 29 48 60 ...
##  $ Experience        : int  1 13 27 24 10 9 39 5 23 30 ...
##  $ Income            : int  49 29 72 22 81 180 105 45 114 22 ...
##  $ Family            : int  4 4 2 1 3 1 4 3 2 1 ...
##  $ CCAvg             : num  1.6 0.4 1.5 0.3 0.6 8.9 2.4 0.1 3.8 1.5 ...
##  $ Mortgage          : int  0 155 0 0 104 0 0 0 0 0 ...
##  $ Securities.Account: int  1 0 0 0 0 0 0 0 1 0 ...
##  $ CD.Account        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Online            : int  0 1 1 0 1 0 0 1 0 1 ...
##  $ CreditCard        : int  0 0 0 1 0 0 0 0 0 1 ...
##  $ Education_1       : num  1 0 0 0 0 0 0 0 0 0 ...
##  $ Education_2       : num  0 1 1 0 1 0 0 1 0 0 ...
##  $ Education_3       : num  0 0 0 1 0 1 1 0 1 1 ...
##  $ Personal_Loan     : Factor w/ 2 levels "0","1": 1 1 1 1 1 2 1 1 1 1 ...
```

```
str(Validation_Data)
```

```
## 'data.frame':    999 obs. of  14 variables:
##  $ Age               : int  35 67 46 46 56 40 48 51 42 57 ...
##  $ Experience        : int  8 41 21 20 30 16 24 25 18 32 ...
##  $ Income            : int  45 112 193 158 48 29 81 71 141 84 ...
##  $ Family            : int  4 1 2 1 1 1 3 1 3 3 ...
##  $ CCAvg             : num  1 2 8.1 2.4 2.2 2 0.7 1.4 5 1.6 ...
##  $ Mortgage          : int  0 0 0 0 0 0 0 198 0 0 ...
##  $ Securities.Account: int  0 1 0 0 0 0 0 0 1 1 ...
##  $ CD.Account        : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Online            : int  0 0 0 1 1 1 0 0 1 0 ...
##  $ CreditCard        : int  1 0 0 1 1 0 0 0 0 0 ...
##  $ Education_1       : num  0 1 0 1 0 0 1 0 0 0 ...
##  $ Education_2       : num  1 0 0 0 0 1 0 0 0 0 ...
##  $ Education_3       : num  0 0 1 0 1 0 0 1 1 1 ...
##  $ Personal_Loan     : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 2 1 ...
```

```
str(Test_Data)
```

```
## 'data.frame':    1001 obs. of  14 variables:
##  $ Age               : int  45 39 35 59 55 57 44 38 59 30 ...
##  $ Experience        : int  19 15 9 32 28 27 18 13 35 6 ...
##  $ Income            : int  34 11 100 40 21 63 43 119 35 18 ...
##  $ Family            : int  3 1 1 4 1 3 2 1 1 3 ...
##  $ CCAvg             : num  1.5 1 2.7 2.5 0.5 2 0.7 3.3 1.2 0.9 ...
##  $ Mortgage          : int  0 0 0 0 0 0 163 0 122 0 ...
##  $ Securities.Account: int  1 0 0 0 1 0 1 0 0 0 ...
##  $ CD.Account        : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ Online            : int  0 0 0 1 0 1 0 1 1 0 ...
##  $ CreditCard        : int  0 0 0 0 1 0 0 1 0 0 ...
##  $ Education_1       : num  1 1 0 0 0 0 1 0 0 0 ...
##  $ Education_2       : num  0 0 1 1 1 0 0 1 0 0 ...
##  $ Education_3       : num  0 0 0 0 0 1 0 0 1 1 ...
##  $ Personal_Loan     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 2 1 1 ...
```

```
#Normalization

# Copy the original data
train.norm.df <- Train_Data
valid.norm.df <- Validation_Data
traval.norm.df <- TraVal_Data
# use preProcess() from the caret package
norm.values <- preProcess(Train_Data[, 1:13], method=c("center", "scale"))
train.norm.df[, 1:13] <- predict(norm.values, Train_Data[, 1:13]) # Replace first two columns wi
th normalized values
valid.norm.df[, 1:13] <- predict(norm.values, Validation_Data[, 1:13])
traval.norm.df[, 1:13] <- predict(norm.values, traval.norm.df[, 1:13])
test.norm.df <- predict(norm.values, Test_Data[, 1:13])
summary(train.norm.df)
```

```
##       Age              Experience           Income            Family
##  Min.   :-1.94697   Min.   :-2.01326   Min.   :-1.4254   Min.   :-1.2145
##  1st Qu.:-0.90257   1st Qu.:-0.88168   1st Qu.:-0.7603   1st Qu.:-1.2145
##  Median :-0.03223   Median :-0.01123   Median :-0.2061   Median :-0.3478
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.83810   3rd Qu.: 0.85922   3rd Qu.: 0.5033   3rd Qu.: 0.5188
##  Max.   : 1.88251   Max.   : 1.99080   Max.   : 3.2301   Max.   : 1.3855
##      CCAvg            Mortgage       Securities.Account   CD.Account
##  Min.   :-1.1008   Min.   :-0.5545   Min.   :-0.34      Min.   :-0.2533
##  1st Qu.:-0.6995   1st Qu.:-0.5545   1st Qu.:-0.34      1st Qu.:-0.2533
##  Median :-0.2316   Median :-0.5545   Median :-0.34      Median :-0.2533
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00      Mean   : 0.0000
##  3rd Qu.: 0.3479   3rd Qu.: 0.4199   3rd Qu.:-0.34      3rd Qu.:-0.2533
##  Max.   : 4.2884   Max.   : 5.7589   Max.   : 2.94      Max.   : 3.9458
##      Online          CreditCard       Education_1       Education_2
##  Min.   :-1.2220   Min.   :-0.6561   Min.   :-0.8249   Min.   :-0.6369
##  1st Qu.:-1.2220   1st Qu.:-0.6561   1st Qu.:-0.8249   1st Qu.:-0.6369
##  Median : 0.8181   Median :-0.6561   Median :-0.8249   Median :-0.6369
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.8181   3rd Qu.: 1.5236   3rd Qu.: 1.2119   3rd Qu.: 1.5695
##  Max.   : 0.8181   Max.   : 1.5236   Max.   : 1.2119   Max.   : 1.5695
##   Education_3      Personal_Loan
##  Min.   :-0.6644   0:2718
##  1st Qu.:-0.6644   1: 282
##  Median :-0.6644
##  Mean   : 0.0000
##  3rd Qu.: 1.5045
##  Max.   : 1.5045
```

```
var(train.norm.df[, 1:13])
```

```
##                             Age   Experience      Income       Family
## Age                  1.000000000  0.994264125 -0.04770546 -0.070973950
## Experience           0.994264125  1.000000000 -0.03939227 -0.075446235
## Income              -0.047705457 -0.039392266  1.00000000 -0.151397484
## Family              -0.070973950 -0.075446235 -0.15139748  1.000000000
## CCAvg               -0.049316783 -0.047182664  0.65251394 -0.103109735
## Mortgage            -0.013781394 -0.012395885  0.21529220 -0.034950727
## Securities.Account  -0.005061513 -0.005532953  0.01228408  0.025768029
## CD.Account           0.003286609  0.005076368  0.17237914 -0.027474975
## Online               0.024303974  0.025353685  0.02449525 -0.036667200
## CreditCard           0.028250379  0.031096133 -0.02285159 -0.006553592
## Education_1         -0.025000360 -0.000752846  0.20725185 -0.105134433
## Education_2         -0.014004551 -0.015287490 -0.12034256  0.122708139
## Education_3          0.040389920  0.015829623 -0.10240570 -0.008665706
##                            CCAvg     Mortgage Securities.Account   CD.Account
## Age                  -0.049316783 -0.013781394       -0.005061513  0.003286609
## Experience           -0.047182664 -0.012395885       -0.005532953  0.005076368
## Income                0.652513941  0.215292195        0.012284083  0.172379144
## Family               -0.103109735 -0.034950727        0.025768029 -0.027474975
## CCAvg                 1.000000000  0.120610504        0.022439170  0.131933222
## Mortgage              0.120610504  1.000000000       -0.027071056  0.076415649
## Securities.Account    0.022439170 -0.027071056        1.000000000  0.336345349
## CD.Account            0.131933222  0.076415649        0.336345349  1.000000000
## Online                0.002678438 -0.002206841        0.021667198  0.173046976
## CreditCard            0.002541498 -0.016104656       -0.010992531  0.273216272
## Education_1           0.163927388  0.056909450        0.017921092 -0.020832788
## Education_2          -0.089689490 -0.036875837        0.002955340  0.020857768
## Education_3          -0.086401572 -0.024353833       -0.021989532  0.001681470
##                             Online   CreditCard  Education_1 Education_2
## Age                   0.024303974  0.028250379 -0.025000360 -0.01400455
## Experience            0.025353685  0.031096133 -0.000752846 -0.01528749
## Income                0.024495249 -0.022851594  0.207251847 -0.12034256
## Family               -0.036667200 -0.006553592 -0.105134433  0.12270814
## CCAvg                 0.002678438  0.002541498  0.163927388 -0.08968949
## Mortgage             -0.002206841 -0.016104656  0.056909450 -0.03687584
## Securities.Account    0.021667198 -0.010992531  0.017921092  0.00295534
## CD.Account            0.173046976  0.273216272 -0.020832788  0.02085777
## Online                1.000000000 -0.005778322  0.023851417  0.01690939
## CreditCard           -0.005778322  1.000000000  0.022627489 -0.01870872
## Education_1           0.023851417  0.022627489  1.000000000 -0.52557024
## Education_2           0.016909392 -0.018708716 -0.525570242  1.00000000
## Education_3          -0.042021915 -0.005705227 -0.548265280 -0.42333430
##                      Education_3
## Age                   0.040389920
## Experience            0.015829623
## Income               -0.102405696
## Family               -0.008665706
## CCAvg                -0.086401572
## Mortgage             -0.024353833
## Securities.Account   -0.021989532
## CD.Account            0.001681470
## Online               -0.042021915
## CreditCard           -0.005705227
```

```
## Education_1        -0.548265280
## Education_2        -0.423334296
## Education_3         1.000000000
```

```
summary(valid.norm.df)
```

```
##       Age               Experience           Income             Family       
##  Min.   :-1.94697   Min.   :-1.92621   Min.   :-1.42542   Min.   :-1.214482  
##  1st Qu.:-0.90257   1st Qu.:-0.88168   1st Qu.:-0.73817   1st Qu.:-1.214482  
##  Median :-0.03223   Median :-0.01123   Median :-0.16177   Median :-0.347820  
##  Mean   :-0.01577   Mean   :-0.01785   Mean   : 0.07852   Mean   :-0.004279  
##  3rd Qu.: 0.83810   3rd Qu.: 0.77217   3rd Qu.: 0.66958   3rd Qu.: 0.518842  
##  Max.   : 1.88251   Max.   : 1.99080   Max.   : 2.91977   Max.   : 1.385504  
##      CCAvg              Mortgage        Securities.Account   CD.Account       
##  Min.   :-1.10078   Min.   :-0.55446   Min.   :-0.34003   Min.   :-0.253349  
##  1st Qu.:-0.69515   1st Qu.:-0.55446   1st Qu.:-0.34003   1st Qu.:-0.253349  
##  Median :-0.17361   Median :-0.55446   Median :-0.34003   Median :-0.253349  
##  Mean   : 0.07535   Mean   : 0.04176   Mean   :-0.02811   Mean   :-0.001148  
##  3rd Qu.: 0.40587   3rd Qu.: 0.53920   3rd Qu.:-0.34003   3rd Qu.:-0.253349  
##  Max.   : 4.69404   Max.   : 5.57998   Max.   : 2.93997   Max.   : 3.945807  
##      Online           CreditCard         Education_1        Education_2       
##  Min.   :-1.22199   Min.   :-0.65610   Min.   :-0.82489   Min.   :-0.63693  
##  1st Qu.:-1.22199   1st Qu.:-0.65610   1st Qu.:-0.82489   1st Qu.:-0.63693  
##  Median : 0.81806   Median :-0.65610   Median :-0.82489   Median :-0.63693  
##  Mean   :-0.06004   Mean   :-0.04516   Mean   : 0.08238   Mean   :-0.05605  
##  3rd Qu.: 0.81806   3rd Qu.: 1.52364   3rd Qu.: 1.21188   3rd Qu.: 1.56952  
##  Max.   : 0.81806   Max.   : 1.52364   Max.   : 1.21188   Max.   : 1.56952  
##   Education_3       Personal_Loan
##  Min.   :-0.66443   0:902        
##  1st Qu.:-0.66443   1: 97        
##  Median :-0.66443                
##  Mean   :-0.03263                
##  3rd Qu.: 1.50455                
##  Max.   : 1.50455
```

```
var(valid.norm.df[, 1:13])
```

```
##                             Age   Experience      Income        Family
## Age                  0.988640890  0.983620705 -0.05324144 -0.0291677024
## Experience           0.983620705  0.991389432 -0.04230932 -0.0416402612
## Income              -0.053241441 -0.042309325  1.05976315 -0.1719799311
## Family              -0.029167702 -0.041640261 -0.17197993  0.9716375312
## CCAvg               -0.089843407 -0.089038093  0.71890864 -0.1310234800
## Mortgage             0.013309492  0.015918091  0.23962745 -0.0084001235
## Securities.Account   0.003440234  0.005214413 -0.02263717  0.0009751088
## CD.Account          -0.006720266 -0.005653175  0.12990686  0.0664262003
## Online              -0.012213226 -0.016674470 -0.01292899  0.0752044456
## CreditCard          -0.013491343 -0.017623450  0.01409637  0.0378748178
## Education_1          0.007248919  0.037990081  0.23978179 -0.1298181906
## Education_2         -0.034973547 -0.035985475 -0.14510972  0.1336417328
## Education_3          0.026660240 -0.005081626 -0.11270059  0.0068723665
##                           CCAvg     Mortgage Securities.Account   CD.Account
## Age                  -0.089843407  0.013309492       0.0034402338 -0.006720266
## Experience           -0.089038093  0.015918091       0.0052144127 -0.005653175
## Income                0.718908643  0.239627446      -0.0226371651  0.129906860
## Family               -0.131023480 -0.008400124       0.0009751088  0.066426200
## CCAvg                 1.142585679  0.126223438      -0.0173574250  0.117260515
## Mortgage              0.126223438  1.108702229       0.0184666092  0.136166383
## Securities.Account   -0.017357425  0.018466609       0.9267071962  0.321480264
## CD.Account            0.117260515  0.136166383       0.3214802639  0.996425614
## Online               -0.007905135  0.005950674       0.0059732302  0.187345897
## CreditCard           -0.040087709  0.023675047      -0.0044890799  0.295165953
## Education_1           0.158685238  0.031931752       0.0246517999  0.002341910
## Education_2          -0.108116750 -0.037682693      -0.0435823481 -0.025955515
## Education_3          -0.062704782  0.003038356       0.0165903389  0.023020864
##                            Online  CreditCard  Education_1  Education_2
## Age                  -0.012213226 -0.01349134  0.007248919 -0.034973547
## Experience           -0.016674470 -0.01762345  0.037990081 -0.035985475
## Income               -0.012928991  0.01409637  0.239781791 -0.145109725
## Family                0.075204446  0.03787482 -0.129818191  0.133641733
## CCAvg                -0.007905135 -0.04008771  0.158685238 -0.108116750
## Mortgage              0.005950674  0.02367505  0.031931752 -0.037682693
## Securities.Account    0.005973230 -0.00448908  0.024651800 -0.043582348
## CD.Account            0.187345897  0.29516595  0.002341910 -0.025955515
## Online                1.021336161  0.03350928 -0.026889452 -0.003593782
## CreditCard            0.033509283  0.95940600 -0.025466628  0.011017893
## Education_1          -0.026889452 -0.02546663  1.025786474 -0.527538363
## Education_2          -0.003593782  0.01101789 -0.527538363  0.945198332
## Education_3           0.032167642  0.01628889 -0.573790886 -0.367367285
##                       Education_3
## Age                   0.026660240
## Experience           -0.005081626
## Income               -0.112700587
## Family                0.006872366
## CCAvg                -0.062704782
## Mortgage              0.003038356
## Securities.Account    0.016590339
## CD.Account            0.023020864
## Online                0.032167642
## CreditCard            0.016288892
```

```
## Education_1        -0.573790886
## Education_2        -0.367367285
## Education_3         0.972165816
```

```
dim(train.norm.df[, 1:13]) # Train is 13 numeric variable
```

```
## [1] 3000   13
```

```
dim(test.norm.df)         # Test is 13 numeric variable
```

```
## [1] 1001   13
```

```
table(train.norm.df[, 14])  # DV is logic (Y/N) variable
```

```
##
##    0    1
## 2718  282
```

```
library(FNN)
nn <- knn(train = train.norm.df[, 1:13], test = test.norm.df,
          cl = train.norm.df[, 14], k = 1, prob=TRUE) # We use k = #Q1

customer_df<- data.frame("Age" =40, "Experience"=10, "Income"=84, "Family"=2, "CCAvg"=2, "Mortga
ge"=0,  "Securities Account"=0, "CD Account"=0,  "Online" =1, "Credit Card"=1, "Education_1" = 0
, "Education_2" = 1, "Education_3" = 0)

customerClass <- knn(train.norm.df[, 1:13], customer_df,  cl = train.norm.df[, 14], k = 1, prob
 = 0.5)

paste0("The customer is classified as: ", summary(customerClass))
```

```
## [1] "The customer is classified as: 1"
```

```
# print(nn) # uncomment for more output
# row.names(Train_Data)[attr(nn, "nn.index")]

#table(nn)
#table(train.norm.df[, 14])
#table(valid.norm.df[, 14])

#nn <- knn(train.norm.df[, 1:13], valid.norm.df[, 1:13], cl = train.norm.df[, 14], k = 1)

#confusionMatrix(nn, valid.norm.df[, 14])$overall[1]
```

Q1 Answer: The customer is classified as: 1 (loan acceptance)

```
#Q2
# initialize a data frame with two columns: k, and accuracy.
library(caret)
accuracy.df <- data.frame(k = seq(1, 14, 1), accuracy = rep(0, 14))
# compute knn for different k on validation.
for(i in 1:14) {
  knn.pred <- knn(train.norm.df[, 1:13], valid.norm.df[, 1:13],
                  cl = train.norm.df[, 14], k = i)
  accuracy.df[i, 2] <- confusionMatrix(knn.pred, valid.norm.df[, 14])$overall[1]
}
accuracy.df
```

```
##     k  accuracy
## 1   1 0.9489489
## 2   2 0.9499499
## 3   3 0.9549550
## 4   4 0.9529530
## 5   5 0.9529530
## 6   6 0.9519520
## 7   7 0.9539540
## 8   8 0.9499499
## 9   9 0.9529530
## 10 10 0.9439439
## 11 11 0.9469469
## 12 12 0.9439439
## 13 13 0.9439439
## 14 14 0.9419419
```

The value of k that provides the best performance is k = 3. k=3 gives the highest accuracy percentage of 95.5% however it should be noted that this classification model is built based on actual converted customers percentage of only 9.8% of the whole dataset which might be causing overfitting in the model

```
#Q3 - Confusion Matrix

knn.pred <- knn(train.norm.df[, 1:13], valid.norm.df[, 1:13],
                  cl = train.norm.df[, 14], k = 3)
confusionMatrix(knn.pred, valid.norm.df[, 14])$overall[1]
```

```
## Accuracy
## 0.954955
```

```
confusionMatrix(knn.pred, valid.norm.df[, 14])
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 897  40
##          1   5  57
##
##                Accuracy : 0.955
##                  95% CI : (0.9402, 0.967)
##     No Information Rate : 0.9029
##     P-Value [Acc > NIR] : 6.550e-10
##
##                   Kappa : 0.6938
##
##  Mcnemar's Test P-Value : 4.011e-07
##
##             Sensitivity : 0.9945
##             Specificity : 0.5876
##          Pos Pred Value : 0.9573
##          Neg Pred Value : 0.9194
##              Prevalence : 0.9029
##          Detection Rate : 0.8979
##    Detection Prevalence : 0.9379
##       Balanced Accuracy : 0.7910
##
##        'Positive' Class : 0
##
```

```
#Q4
customerClass <- knn(train.norm.df[, 1:13], customer_df,  cl = train.norm.df[, 14], k = 3, prob
 = 0.5)

paste0("The customer is classified as: ", summary(customerClass))
```

```
## [1] "The customer is classified as: 1"
```

The customer is classified as: 1 (loan acceptance)

```
#Q5
set.seed(15)
Test_Index = createDataPartition(uk1$Age,p=0.2, list=FALSE) # 20% reserved for Test
Test_Data = uk1[Test_Index,]
TraVal_Data = uk1[-Test_Index,] # Validation and Training data is rest
Train_Index = createDataPartition(TraVal_Data$Age, p=0.625, list=FALSE) # 62.5% of remaining dat
a as training

# 62.5% of the 80% is 50% which is training

Train_Data = TraVal_Data[Train_Index,]
Validation_Data = TraVal_Data[-Train_Index,] # rest as validation
```

Normalization

```
# Copy the original data
train.norm.df <- Train_Data
valid.norm.df <- Validation_Data
traval.norm.df <- TraVal_Data
# use preProcess() from the caret package to normalize Sales and Age.
norm.values <- preProcess(Train_Data[, 1:13], method=c("center", "scale"))
train.norm.df[, 1:13] <- predict(norm.values, Train_Data[, 1:13]) # Replace first two columns wi
th normalized values
valid.norm.df[, 1:13] <- predict(norm.values, Validation_Data[, 1:13])
traval.norm.df[, 1:13] <- predict(norm.values, traval.norm.df[, 1:13])
test.norm.df <- predict(norm.values, Test_Data[, 1:13])
summary(train.norm.df)
```

```
##       Age              Experience            Income            Family
##  Min.   :-1.95829   Min.   :-2.024044   Min.   :-1.4303   Min.   :-1.2228
##  1st Qu.:-0.90687   1st Qu.:-0.884555   1st Qu.:-0.7662   1st Qu.:-1.2228
##  Median :-0.03069   Median :-0.008026   Median :-0.2128   Median :-0.3559
##  Mean   : 0.00000   Mean   : 0.000000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.84549   3rd Qu.: 0.868504   3rd Qu.: 0.4955   3rd Qu.: 1.3777
##  Max.   : 1.89691   Max.   : 2.007992   Max.   : 3.2181   Max.   : 1.3777
##      CCAvg            Mortgage       Securities.Account   CD.Account
##  Min.   :-1.1015   Min.   :-0.5593   Min.   :-0.3413    Min.   :-0.2552
##  1st Qu.:-0.7026   1st Qu.:-0.5593   1st Qu.:-0.3413    1st Qu.:-0.2552
##  Median :-0.2466   Median :-0.5593   Median :-0.3413    Median :-0.2552
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000    Mean   : 0.0000
##  3rd Qu.: 0.3803   3rd Qu.: 0.4566   3rd Qu.:-0.3413    3rd Qu.:-0.2552
##  Max.   : 4.1989   Max.   : 5.8919   Max.   : 2.9290    Max.   : 3.9167
##      Online          CreditCard       Education_1       Education_2
##  Min.   :-1.1860   Min.   :-0.6531   Min.   :-0.8256   Min.   :-0.6307
##  1st Qu.:-1.1860   1st Qu.:-0.6531   1st Qu.:-0.8256   1st Qu.:-0.6307
##  Median : 0.8428   Median :-0.6531   Median :-0.8256   Median :-0.6307
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.8428   3rd Qu.: 1.5306   3rd Qu.: 1.2107   3rd Qu.: 1.5848
##  Max.   : 0.8428   Max.   : 1.5306   Max.   : 1.2107   Max.   : 1.5848
##   Education_3     Personal_Loan
##  Min.   :-0.670   0:2264
##  1st Qu.:-0.670   1: 237
##  Median :-0.670
##  Mean   : 0.000
##  3rd Qu.: 1.492
##  Max.   : 1.492
```

```
var(train.norm.df[, 1:13])
```

```
##                              Age    Experience      Income        Family
## Age                 1.0000000000  0.9940225594 -0.05087467 -0.0448921532
## Experience          0.9940225594  1.0000000000 -0.04161191 -0.0511499669
## Income             -0.0508746670 -0.0416119125  1.00000000 -0.1540511082
## Family             -0.0448921532 -0.0511499669 -0.15405111  1.0000000000
## CCAvg              -0.0499342272 -0.0477227657  0.65470526 -0.1094696043
## Mortgage           -0.0061419578 -0.0030797005  0.17857854 -0.0291903291
## Securities.Account  0.0110696326  0.0094137497 -0.00604324  0.0349508776
## CD.Account         -0.0095900712 -0.0080462575  0.14013474 -0.0026431396
## Online             -0.0007167343 -0.0009863074  0.03006638 -0.0213492131
## CreditCard          0.0029851469  0.0056280970 -0.02879987  0.0029107587
## Education_1        -0.0142139861  0.0107205722  0.20863503 -0.1012372986
## Education_2        -0.0227033797 -0.0228528694 -0.10242607  0.1111009110
## Education_3         0.0372456641  0.0109184348 -0.12155815 -0.0009315549
##                          CCAvg     Mortgage Securities.Account   CD.Account
## Age                 -0.04993423 -0.006141958       0.0110696326 -0.009590071
## Experience          -0.04772277 -0.003079701       0.0094137497 -0.008046258
## Income               0.65470526  0.178578543      -0.0060432404  0.140134735
## Family              -0.10946960 -0.029190329       0.0349508776 -0.002643140
## CCAvg                1.00000000  0.083809670       0.0138911748  0.121924323
## Mortgage             0.08380967  1.000000000      -0.0258171796  0.051877435
## Securities.Account   0.01389117 -0.025817180       1.0000000000  0.365817547
## CD.Account           0.12192432  0.051877435       0.3658175470  1.000000000
## Online               0.01061331  0.005725267       0.0117521686  0.177952540
## CreditCard          -0.02317236 -0.026694132       0.0055415949  0.270532380
## Education_1          0.16482286  0.055649900       0.0004814151 -0.010303193
## Education_2         -0.08409744 -0.015173448       0.0020197981  0.012729568
## Education_3         -0.09292824 -0.044276991      -0.0024820985 -0.001482952
##                           Online   CreditCard   Education_1  Education_2
## Age                 -0.0007167343  0.002985147 -0.0142139861 -0.022703380
## Experience          -0.0009863074  0.005628097  0.0107205722 -0.022852869
## Income               0.0300663793 -0.028799867  0.2086350284 -0.102426065
## Family              -0.0213492131  0.002910759 -0.1012372986  0.111100911
## CCAvg                0.0106133117 -0.023172363  0.1648228643 -0.084097439
## Mortgage             0.0057252667 -0.026694132  0.0556499000 -0.015173448
## Securities.Account   0.0117521686  0.005541595  0.0004814151  0.002019798
## CD.Account           0.1779525401  0.270532380 -0.0103031929  0.012729568
## Online               1.0000000000  0.006635667  0.0136334716  0.021196286
## CreditCard           0.0066356667  1.000000000  0.0173110016 -0.023116442
## Education_1          0.0136334716  0.017311002  1.0000000000 -0.520952732
## Education_2          0.0211962862 -0.023116442 -0.5209527318  1.000000000
## Education_3         -0.0351586636  0.004178555 -0.5533423992 -0.422732154
##                      Education_3
## Age                 0.0372456641
## Experience          0.0109184348
## Income             -0.1215581517
## Family             -0.0009315549
## CCAvg              -0.0929282409
## Mortgage           -0.0442769911
## Securities.Account -0.0024820985
## CD.Account         -0.0014829520
## Online             -0.0351586636
## CreditCard          0.0041785553
```

```
## Education_1        -0.5533423992
## Education_2        -0.4227321536
## Education_3         1.0000000000
```

```
summary(valid.norm.df)
```

```
##       Age               Experience            Income             Family
##  Min.   :-1.958287   Min.   :-1.936391   Min.   :-1.43027   Min.   :-1.22276
##  1st Qu.:-0.906870   1st Qu.:-0.884555   1st Qu.:-0.72195   1st Qu.:-1.22276
##  Median :-0.030689   Median :-0.008026   Median :-0.19071   Median :-0.35594
##  Mean   :-0.005889   Mean   :-0.003228   Mean   : 0.03345   Mean   :-0.02438
##  3rd Qu.: 0.845492   3rd Qu.: 0.868504   3rd Qu.: 0.56188   3rd Qu.: 0.51087
##  Max.   : 1.896908   Max.   : 2.007992   Max.   : 2.90820   Max.   : 1.37768
##      CCAvg              Mortgage        Securities.Account   CD.Account
##  Min.   :-1.101522   Min.   :-0.55930   Min.   :-0.34128   Min.   :-0.25522
##  1st Qu.:-0.702567   1st Qu.:-0.55930   1st Qu.:-0.34128   1st Qu.:-0.25522
##  Median :-0.246618   Median :-0.55930   Median :-0.34128   Median :-0.25522
##  Mean   :-0.000963   Mean   : 0.04784   Mean   :-0.02473   Mean   :-0.01014
##  3rd Qu.: 0.323319   3rd Qu.: 0.47442   3rd Qu.:-0.34128   3rd Qu.:-0.25522
##  Max.   : 4.597842   Max.   : 5.42457   Max.   : 2.92898   Max.   : 3.91667
##      Online           CreditCard        Education_1        Education_2
##  Min.   :-1.18598   Min.   :-0.65309   Min.   :-0.82561   Min.   :-0.63074
##  1st Qu.:-1.18598   1st Qu.:-0.65309   1st Qu.:-0.82561   1st Qu.:-0.63074
##  Median : 0.84284   Median :-0.65309   Median :-0.82561   Median :-0.63074
##  Mean   : 0.03836   Mean   :-0.01898   Mean   : 0.05255   Mean   :-0.01399
##  3rd Qu.: 0.84284   3rd Qu.: 1.53057   3rd Qu.: 1.21074   3rd Qu.: 1.58481
##  Max.   : 0.84284   Max.   : 1.53057   Max.   : 1.21074   Max.   : 1.58481
##   Education_3       Personal_Loan
##  Min.   :-0.66995   0:1356
##  1st Qu.:-0.66995   1: 142
##  Median :-0.66995
##  Mean   :-0.04213
##  3rd Qu.: 1.49205
##  Max.   : 1.49205
```

```
var(valid.norm.df[, 1:13])
```

```
##                              Age   Experience      Income       Family
## Age                 1.028410613  1.024109453 -0.04727932 -0.087829557
## Experience          1.024109453  1.031783061 -0.03884993 -0.094776283
## Income             -0.047279315 -0.038849927  1.03384575 -0.159771772
## Family             -0.087829557 -0.094776283 -0.15977177  0.981654164
## CCAvg              -0.074329772 -0.073375939  0.66390120 -0.106480198
## Mortgage           -0.009178939 -0.009777290  0.30546794 -0.027895739
## Securities.Account -0.026230479 -0.023157273  0.01898459 -0.006559310
## CD.Account          0.018091560  0.019830856  0.19461122 -0.006427784
## Online              0.042407374  0.041995080 -0.01318926  0.013170172
## CreditCard          0.043310936  0.041961680  0.01032662  0.007091008
## Education_1        -0.022269432  0.005441586  0.22775413 -0.127464047
## Education_2        -0.013619599 -0.016582389 -0.16956079  0.150729617
## Education_3         0.036933939  0.010404233 -0.07634492 -0.011757385
##                           CCAvg     Mortgage Securities.Account   CD.Account
## Age                -0.074329772 -0.009178939       -0.026230479  0.018091560
## Experience         -0.073375939 -0.009777290       -0.023157273  0.019830856
## Income              0.663901201  0.305467938        0.018984593  0.194611222
## Family             -0.106480198 -0.027895739       -0.006559310 -0.006427784
## CCAvg               1.007459332  0.189048851        0.008481167  0.130988547
## Mortgage            0.189048851  1.192971189        0.000547136  0.161211890
## Securities.Account  0.008481167  0.000547136        0.935615396  0.268688738
## CD.Account          0.130988547  0.161211890        0.268688738  0.963013440
## Online             -0.019822694 -0.012446249        0.028793633  0.169023601
## CreditCard          0.015692529  0.027311519       -0.033898442  0.288731606
## Education_1         0.154784243  0.045472017        0.051024068 -0.022411918
## Education_2        -0.110284479 -0.077018583       -0.025960708  0.003107789
## Education_3        -0.056716117  0.026879206       -0.028839203  0.020762138
##                           Online   CreditCard  Education_1  Education_2
## Age                 0.042407374  0.043310936 -0.022269432 -0.013619599
## Experience          0.041995080  0.041961680  0.005441586 -0.016582389
## Income             -0.013189264  0.010326618  0.227754126 -0.169560790
## Family              0.013170172  0.007091008 -0.127464047  0.150729617
## CCAvg              -0.019822694  0.015692529  0.154784243 -0.110284479
## Mortgage           -0.012446249  0.027311519  0.045472017 -0.077018583
## Securities.Account  0.028793633 -0.033898442  0.051024068 -0.025960708
## CD.Account          0.169023601  0.288731606 -0.022411918  0.003107789
## Online              0.985625752  0.001450083  0.003194576 -0.001944309
## CreditCard          0.001450083  0.983238995 -0.001752896  0.009399846
## Education_1         0.003194576 -0.001752896  1.017755073 -0.541962242
## Education_2        -0.001944309  0.009399846 -0.541962242  0.986714480
## Education_3        -0.001494379 -0.007311591 -0.551691322 -0.387461883
##                    Education_3
## Age                 0.036933939
## Experience          0.010404233
## Income             -0.076344922
## Family             -0.011757385
## CCAvg              -0.056716117
## Mortgage            0.026879206
## Securities.Account -0.028839203
## CD.Account          0.020762138
## Online             -0.001494379
## CreditCard         -0.007311591
```

```
## Education_1        -0.551691322
## Education_2        -0.387461883
## Education_3         0.963829277
```

KNN

```
library(FNN)
nn <- knn(train = train.norm.df[, 1:13], test = test.norm.df,
          cl = train.norm.df[, 14], k = 3, prob=TRUE)
```

Confusion matrix

```
knn.pred.valid <- knn(train.norm.df[, 1:13], valid.norm.df[, 1:13],
                  cl = train.norm.df[, 14], k = 3)
confusionMatrix(knn.pred.valid, valid.norm.df[, 14])
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1348   56
##          1    8   86
##
##                Accuracy : 0.9573
##                  95% CI : (0.9458, 0.9669)
##     No Information Rate : 0.9052
##     P-Value [Acc > NIR] : 1.836e-14
##
##                   Kappa : 0.7067
##
##  Mcnemar's Test P-Value : 4.228e-09
##
##             Sensitivity : 0.9941
##             Specificity : 0.6056
##          Pos Pred Value : 0.9601
##          Neg Pred Value : 0.9149
##              Prevalence : 0.9052
##          Detection Rate : 0.8999
##    Detection Prevalence : 0.9372
##       Balanced Accuracy : 0.7999
##
##        'Positive' Class : 0
##
```

```
knn.pred.train <- knn(train.norm.df[, 1:13], train.norm.df[, 1:13],
                  cl = train.norm.df[, 14], k = 3)
confusionMatrix(knn.pred.train, train.norm.df[, 14])
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 2263   58
##          1    1  179
##
##                Accuracy : 0.9764
##                  95% CI : (0.9697, 0.982)
##     No Information Rate : 0.9052
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8459
##
##  Mcnemar's Test P-Value : 3.086e-13
##
##             Sensitivity : 0.9996
##             Specificity : 0.7553
##          Pos Pred Value : 0.9750
##          Neg Pred Value : 0.9944
##              Prevalence : 0.9052
##          Detection Rate : 0.9048
##    Detection Prevalence : 0.9280
##       Balanced Accuracy : 0.8774
##
##        'Positive' Class : 0
##
```