Kyaw_Exercise_4

May Kyaw

2022-12-07

Question 1:

```
bankchurn <- read.csv('~/Downloads/bankchurn_ex4.csv', header = TRUE)
```

Number of rows equal 9900 and columns 14

```
nrow(bankchurn) # this is for the number of rows

## [1] 9900

ncol(bankchurn) # this is for the number of columns

## [1] 14

dim(bankchurn) # this is for the dimensions (rows, columns)

## [1] 9900   14
```

The variables are:RowNumbers, CustomersId, Surnames, CreditScores, Geographies Gender, Age, Tenure, Balances, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary and Exited

```
ls(bankchurn)

##  [1] "Age"             "Balances"        "CreditScores"    "CustomersId"
##  [5] "EstimatedSalary" "Exited"          "Gender"          "Geographies"
##  [9] "HasCrCard"       "IsActiveMember"  "NumOfProducts"   "RowNumbers"
## [13] "Surnames"        "Tenure"
```

Which values are NA

```
which(is.na(bankchurn))

## integer(0)
```

There are no NA values Question 2:

```
max(bankchurn$CreditScores) # this is for the max

## [1] 850

min(bankchurn$CreditScores) # this is for the min

## [1] 350

mean(bankchurn$CreditScores) # this is for the mean
```

```
## [1] 650.3768
```

CreditScores: Max:850 Min:350 Mean:650.3768

Question 3:Two variables that I belive to be plausibly associated with credit score is HasCrCard and IsActiveMember. It is important that one has a credit card to have a credit score. While being an active member is also a factor into ones credit score as using the credit card is correlated ones score.

Question 4:

```
max(bankchurn$Age) # this is for the max

## [1] 92

min(bankchurn$Age) # this is for the min

## [1] 18

mean(bankchurn$Age) # this is for the mean

## [1] 38.92657
```

Age: Max:92 Min:18 Mean:38.92657

Question 5:

```
cor.test(bankchurn$Age,bankchurn$CreditScores)

##
##  Pearson's product-moment correlation
##
## data:  bankchurn$Age and bankchurn$CreditScores
## t = -0.23796, df = 9898, p-value = 0.8119
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.02208958  0.01730782
## sample estimates:
##          cor
## -0.002391807
```

From the results we can are unable to conclude that a significant relationship between exists as the p-value is large that .05 Question 6:

```
mod1 <- lm(bankchurn$CreditScores ~ bankchurn$Age + bankchurn$Geographies + bankchurn$Gender)
summary(mod1)

##
## Call:
## lm(formula = bankchurn$CreditScores ~ bankchurn$Age + bankchurn$Geographies +
##     bankchurn$Gender)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -301.211  -66.949    1.362   66.930  201.761
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  650.64975    3.99790 162.748   <2e-16 ***
## bankchurn$Age                 -0.02595    0.09269  -0.280    0.780
## bankchurn$GeographiesGermany   1.88265    2.38030   0.791    0.429
## bankchurn$GeographiesSpain     1.74725    2.38543   0.732    0.464
## bankchurn$GenderMale          -0.30887    1.95272  -0.158    0.874
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 96.68 on 9895 degrees of freedom
## Multiple R-squared:  9.677e-05,  Adjusted R-squared:  -0.0003074
## F-statistic: 0.2394 on 4 and 9895 DF,  p-value: 0.9161
```

From the results we can are unable to conclude that a significant relationship between age and credit score, after controlling for customer “Geography” and “Gender” exists as the p-value is large that .05

Question 7: I personally would feel confident in using these predictive model to make business decisions as these models can assist businesses predict various consumer trends.It can also help business predict employee productivity shifts, which will help drive supply and marketing decisions and improve efficiency. I believe that when utilizing these predictive model it will overall help the businesses examine the future of their comapny and maximize the full potential of the business.