PERFORM PREDICTIVE ANALYTICS FOR CUSTOMERS' BEHAVIOUR IN MARKETING AND SALES

May 14, 2023

[58]:
```
install.packages('mvtnorm')
```

```
Installing package into 'C:/Users/JASHWANTH/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)


package 'mvtnorm' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\JASHWANTH\AppData\Local\Temp\RtmpYFfrBV\downloaded_packages
```

[59]:
```
# Load the mvtnorm package
library(mvtnorm)

# Set seed for reproducibility
set.seed(123)

# Define the desired means, standard deviations, and correlations
means <- c(Age = 40, Gender = 0.5, Income = 50000, Education = 2.5,␣
 ↪Bought_Product = 0.5)
sds <- c(Age = 10, Gender = 0.5, Income = 10000, Education = 1, Bought_Product␣
 ↪= 0.5)
cor_matrix <- matrix(c(1, -0.2, 0.5, -0.3, 0.8,
                       -0.2, 1, -0.1, 0.4, -0.3,
                       0.5, -0.1, 1, -0.2, 0.6,
                       -0.3, 0.4, -0.2, 1, -0.1,
                       0.8, -0.3, 0.6, -0.1, 1), nrow = 5, byrow = TRUE)

# Generate a dataset with the desired means, standard deviations, and␣
 ↪correlations
customers_data <- as.data.frame(rmvnorm(n = 500, mean = means, sigma =␣
 ↪diag(sds) %*% cor_matrix %*% diag(sds)))

# Set the column names
colnames(customers_data) <- c("Age", "Gender", "Income", "Education",␣
 ↪"Bought_Product")
```

```
# Write the dataframe to a csv file
write.csv(customers_data, "customers_data.csv", row.names = FALSE)
```

[17]:
```
library(corrplot)
```

```
corrplot 0.92 loaded
```

[3]:
```
library(caret)
```

```
Loading required package: ggplot2

Loading required package: lattice
```

[60]:
```
customer_data <- read.csv("customers_data.csv")
customer_data_clean <- na.omit(customer_data)
```

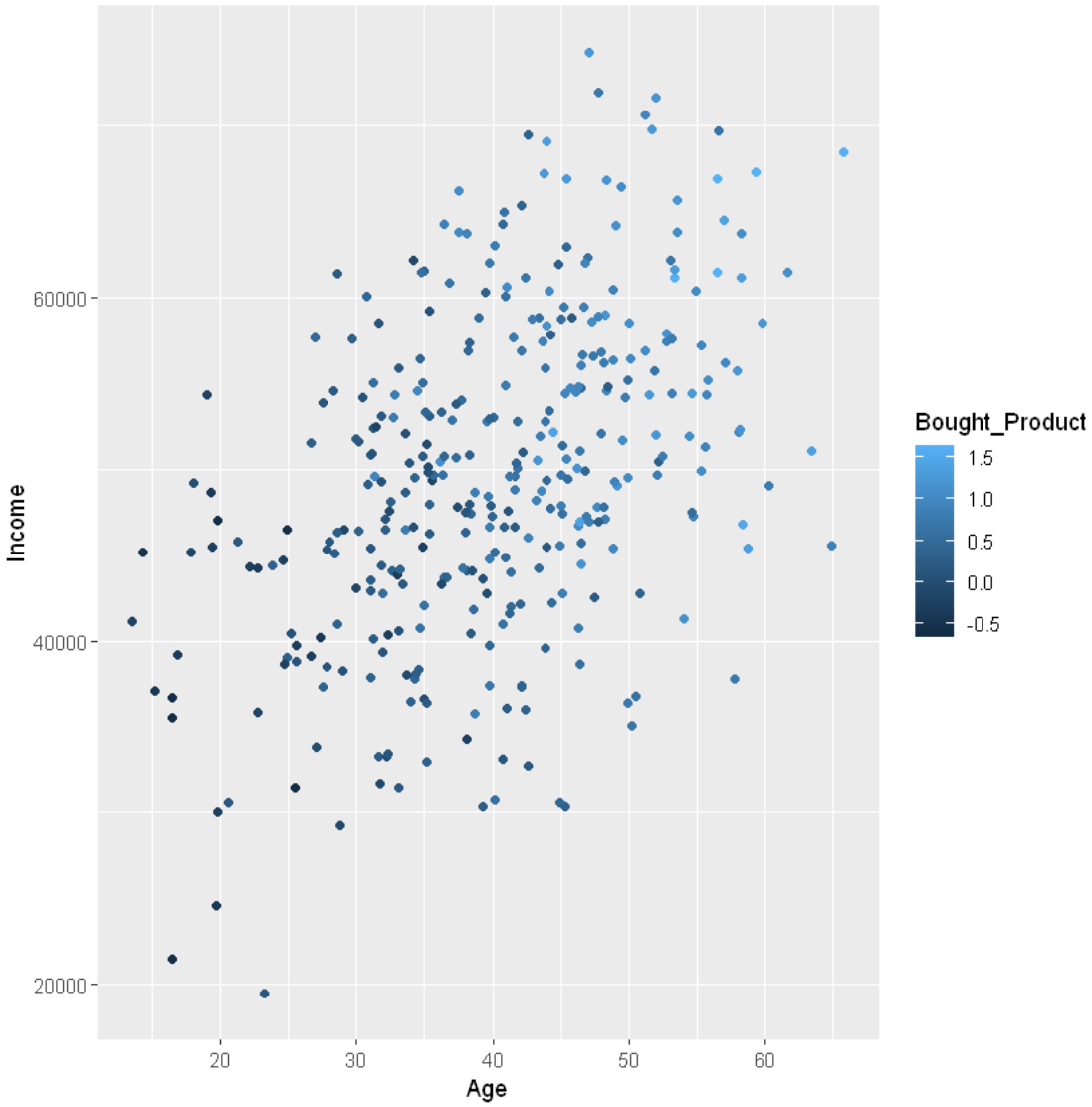
[61]:
```
customer_data
```

	Age <dbl>	Gender <dbl>	Income <dbl>	Education <dbl>	Bought_Product <dbl>
	42.97512	0.35789680	65584.32	2.351788	0.86347110
	48.50731	0.58915517	37357.94	1.773306	0.41060783
	52.38957	0.60314353	54013.62	2.278661	0.79868489
	45.32344	0.81171650	30342.45	3.215889	0.29815274
	25.62570	0.47877806	39734.59	2.171723	-0.30933511
	26.69328	0.79698909	51525.87	1.928524	0.28702216
	48.23308	0.34221164	58953.47	3.078032	1.18416998
	45.56523	0.69603203	49384.24	2.124040	0.51850306
	27.56754	0.71577009	37342.51	5.001991	0.40129836
	27.79242	0.52342786	45327.68	3.511104	0.09577713
	41.63069	0.65465512	49572.22	3.733681	0.58520297
	56.20081	-0.37415228	55853.83	2.002542	1.26434117
	41.57192	0.18960853	46669.73	1.392293	0.18412666
	43.24080	0.67307175	50531.97	3.484318	1.18591119
	40.90803	-0.64614708	60054.98	1.397498	0.57639345
	42.72339	0.37877150	37797.88	2.656614	0.41396006
	37.97150	0.79405123	46293.21	3.217448	0.34741521
	45.35207	0.85697497	54353.83	2.244713	0.94580340
	50.26333	0.50615362	52392.79	1.804153	1.14542300
	42.05224	1.53466502	65322.74	2.305678	0.36222312
	32.40401	0.70398414	47529.30	2.341222	-0.06901086
	31.67661	0.11222512	33320.75	2.442218	0.26067088
	27.04843	0.87251058	33818.44	3.000049	-0.04776540
	39.28742	0.51567339	43594.30	1.709284	0.05554558
	39.21316	-0.07774307	45095.67	2.328192	0.92754406
	34.90964	0.54136029	50776.51	1.735102	0.24557334
	52.18500	0.67354093	50418.99	1.716814	0.30861803
	52.79610	0.01642059	57404.38	3.649132	0.86048142
	38.10711	0.31044813	34281.90	1.099541	-0.28620010
A data.frame: 500 × 5	38.15316	0.16139585	56875.78	4.197981	0.43188661
	24.50515	0.89241682	51250.58	2.0152070	0.18152206
	44.28159	0.76574216	47745.43	2.1008852	0.50866646
	51.20214	0.76546106	65896.47	1.6519459	0.73395283
	38.18959	2.09601355	45713.89	4.0128733	0.21819058
	34.59908	0.39993475	47884.96	4.1268957	0.53118883
	53.93907	0.93987470	70719.22	2.7336809	0.86889902
	16.53836	0.16603166	21509.05	4.2775427	-0.59829181
	40.89690	1.01293406	44883.72	2.5619838	0.47035934
	45.47061	0.16755494	57309.13	2.7081126	0.78713677
	36.45880	0.94045502	64269.04	3.9119476	0.86773755
	33.20920	0.22074259	44121.06	2.3067620	0.38686918
	33.11502	-0.31361923	31395.55	2.5733773	0.16879746
	19.35696	0.60110024	50263.23	4.1325548	-0.54240647
	29.93418	0.14094249	51747.58	2.8853605	0.20076876
	23.79245	0.22946621	44387.64	2.7806920	0.34390237
	31.05091	0.42857094	42943.07	1.6578396	0.18738353
	31.12576	1.87696653	50920.21	3.0970512	0.04792246
	54.60396	0.60098969	47499.19	1.0779642	0.54764625
	32.54873	0.99841185	48106.59	2.1871777	0.31358165
	49.40038	0.13444773	45557.36	2.5470136	0.80853239

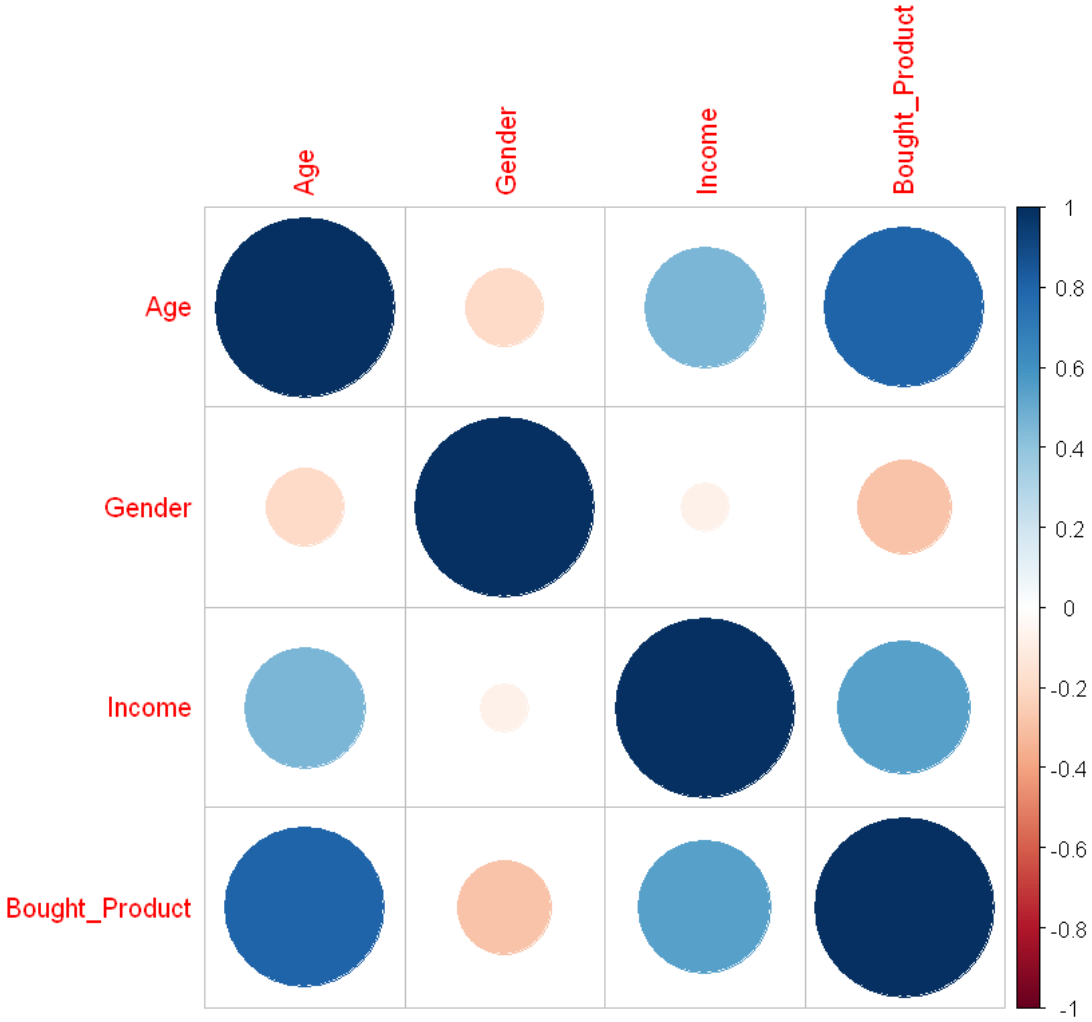
[62]:
```
set.seed(1234)
train_index <- createDataPartition(y = customer_data_clean$Bought_Product,p = 0.
 ↪7, list = FALSE)
train_data <- customer_data_clean[train_index,]
test_data <- customer_data_clean[-train_index,]
```

[63]:
```
summary(train_data)
ggplot(train_data, aes(x = Age, y = Income, color = Bought_Product)) +
  geom_point()
```

```
      Age             Gender             Income         Education
 Min.   :13.60   Min.   :-0.6461   Min.   :19521   Min.   :-0.03918
 1st Qu.:33.12   1st Qu.: 0.1803   1st Qu.:43922   1st Qu.: 1.95024
 Median :40.45   Median : 0.4944   Median :49540   Median : 2.56768
 Mean   :40.05   Mean   : 0.4968   Mean   :49579   Mean   : 2.58520
 3rd Qu.:46.77   3rd Qu.: 0.7899   3rd Qu.:56049   3rd Qu.: 3.24836
 Max.   :65.73   Max.   : 2.2548   Max.   :74242   Max.   : 5.66863
 Bought_Product
 Min.   :-0.6627
 1st Qu.: 0.1996
 Median : 0.4962
 Mean   : 0.5083
 3rd Qu.: 0.8335
 Max.   : 1.6310
```



[64]:
```
cor_matrix <- cor(train_data[,c("Age","Gender","Income", "Bought_Product")])
corrplot(cor_matrix)
```



[66]:
```
train_data_selected <- train_data[,c("Age","Gender","Income", "Bought_Product")]
test_data_selected <- test_data[,c("Age","Gender","Income", "Bought_Product")]
```

[69]:
```
model <- train(Bought_Product ~ ., data = train_data_selected,method = "glm",␣
 ↪family = "poisson")
predictions <- predict(model, test_data_selected[,3:5], type = "prob")
confusionMatrix(predictions, test_data_selected$Bought_Product)
```

```
Warning message:
"model fit failed for Resample01: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
```

```
Warning message:
"model fit failed for Resample02: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample03: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample04: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample05: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample06: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample07: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample08: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample09: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample10: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample11: parameter=none Error in
eval(family$initialize) :
```

```
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample12: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample13: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample14: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample15: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample16: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample17: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample18: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample19: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample20: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
```

```
"model fit failed for Resample21: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample22: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample23: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample24: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message:
"model fit failed for Resample25: parameter=none Error in
eval(family$initialize) :
  negative values not allowed for the 'Poisson' family
"
Warning message in nominalTrainWorkflow(x = x, y = y, wts = weights, info =
trainInfo, :
"There were missing values in resampled performance measures."
```

```
Something is wrong; all the RMSE metric values are missing:
      RMSE        Rsquared        MAE
 Min.   : NA   Min.   : NA   Min.   : NA
 1st Qu.: NA   1st Qu.: NA   1st Qu.: NA
 Median : NA   Median : NA   Median : NA
 Mean   :NaN   Mean   :NaN   Mean   :NaN
 3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA
 Max.   : NA   Max.   : NA   Max.   : NA
 NA's   :1     NA's   :1     NA's   :1
```

```
Error: Stopping
Traceback:

1. train(Bought_Product ~ ., data = train_data_selected, method = "glm",
 .     family = "poisson")
2. train.formula(Bought_Product ~ ., data = train_data_selected,
 .     method = "glm", family = "poisson")
3. train(x, y, weights = w, …)
4. train.default(x, y, weights = w, …)
```

```
5. stop("Stopping", call. = FALSE)
```

[70]:
```
new_customer <- data.frame(Age = 30, Gender = 0.6789, Income = 50000)
predict(model, new_customer, type = "prob")
```

```
Error in eval(expr, envir, enclos): object 'model' not found
Traceback:

1. predict(model, new_customer, type = "prob")
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: