Part2ba

Arnav Ushnik Harshita

November 23, 2016

Source Code

AAA - visitors coming from the AAA travel. AAA is the American Automobile Association. The AAA is a federation of motor clubs throughout North America. AAA is a non-profit member service organization; with 54 million members in the United States and Canada. AAA provides services to its members, including roadside assistance and others

WEB - visitors that are staying overnight at a hotel affiliated with “The Casino”. In addition to the money spent in gambling, WEB customers expend additional money in overnight accommodations.

WALK - are walking-ins with to prior reservations on the system. “The Casino” has little information about this type of customer.

Reads the Casino File and shows the head and structure of the data frame Casino

```
casino<-read.csv("/Users/arnavsomani/Desktop/NYU COURSE/sem 3/ba/r programming csv file
s/Casino.csv")
head(casino,2)
```

```
##        Player Total.Spend Gender Age Source
## 1 Player 2690    13.01958 Female  37    AAA
## 2 Player 3688    16.04045 Female  27   WALK
```

```
str(casino)
```

```
## 'data.frame':    5000 obs. of  5 variables:
##  $ Player     : Factor w/ 5000 levels "Player 1","Player 10",..: 1880 2988 463 4199 2
171 3155 1606 4556 1053 1346 ...
##  $ Total.Spend: num  13 16 17.9 24.6 27.5 ...
##  $ Gender     : Factor w/ 2 levels "Female","Male": 1 1 1 1 1 1 1 2 1 2 ...
##  $ Age        : int  37 27 55 42 26 60 48 33 65 67 ...
##  $ Source     : Factor w/ 3 levels "AAA","WALK","WEB": 1 2 3 3 1 3 3 3 3 2 ...
```

```
sd(casino$Age)
```

```
## [1] 12.3771
```

```
sd(casino$Total.Spend)
```

```
## [1] 2226.21
```

Creates a dummy variable and gives numeric values to Male and Female respectively

```
casino$gen<-as.numeric(0)

for (i in 1:nrow(casino)){
  if (casino$Gender[i]=="Male")
    casino$gen[i]<-as.numeric(1)
  else
    casino$gen[i]<-as.numeric(0)
}
```

Creates a dummy variable and assigns numeric values to each Source (AAA = 0, WALK = 1, and WEB = 2)

```
casino$Sou<-as.numeric(0)
for (i in 1:nrow(casino)){
  if (casino$Source[i]=="AAA")
    casino$Sou[i]<-as.numeric(0)
  else if
  (casino$Source[i]=="WALK")
    casino$Sou[i]<-as.numeric(1)
  else
    casino$Sou[i]<-as.numeric(2)
      }
```

Creates a dummy variable and gives numeric values to Age groups of 20-40, 41 - 65, and 65 and above

```
casino$Agegrp<-as.numeric(0)
for (i in 1:nrow(casino)){
  if (casino$Age[i]>20 & casino$Age[i]<41)
    casino$Agegrp[i]<-as.numeric(0)
  else if
  (casino$Age[i]>40 & casino$Age[i]<66)
    casino$Agegrp[i]<-as.numeric(1)
  else
    casino$Agegrp[i]<-as.numeric(2)
    }
```

Shows the correlation between Male and all Sources (AAA, WALK and WEB)

```
aaa<-cor(casino$gen==1,casino$Sou==0)
aaa
```

```
## [1] -0.001185616
```

```
walk<-cor(casino$gen==1,casino$Sou==1)
walk
```

```
## [1] 0.01311349
```

```
web<-cor(casino$gen==1,casino$Sou==2)
web
```

```
## [1] -0.01031758
```

Shows the correlation between Female and all Source (AAA, WALK and WEB)

```
aaa1<-cor(casino$gen==0,casino$Sou==0)
aaa1
```

```
## [1] 0.001185616
```

```
walk1<-cor(casino$gen==0,casino$Sou==1)
walk1
```

```
## [1] -0.01311349
```

```
web1<-cor(casino$gen==0,casino$Sou==2)
web1
```

```
## [1] 0.01031758
```

Creates a data frame that returns the total amount spent by Female visitors coming from AAA.

```
aaafemale1<-subset(casino,subset=casino$gen==0 & casino$Sou==0 & casino$Agegrp==0)
head(aaafemale1,2)
```

```
##        Player Total.Spend Gender Age Source gen Sou Agegrp
## 1 Player 2690    13.01958 Female  37    AAA   0   0      0
## 5 Player 2952    27.49134 Female  26    AAA   0   0      0
```

```
sum(aaafemale1$Total.Spend)
```

```
## [1] 53395.24
```

```
aaafemale2<-subset(casino,subset=casino$gen==0 & casino$Sou==0 & casino$Agegrp==1)
head(aaafemale2,2)
```

```
##          Player Total.Spend Gender Age Source gen Sou Agegrp
## 80    Player 60    59.88830 Female  55    AAA   0   0      1
## 120 Player 1952    69.28833 Female  53    AAA   0   0      1
```

```
sum(aaafemale2$Total.Spend)
```

```
## [1] 426655
```

```
aaafemale3<-subset(casino,subset=casino$gen==0 & casino$Sou==0 & casino$Agegrp==2)
head(aaafemale3,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 32 Player 4803    46.36052 Female  66    AAA   0   0      2
## 62 Player 4041    55.95004 Female  66    AAA   0   0      2
```

```
sum(aaafemale3$Total.Spend)
```

```
## [1] 11977.98
```

```
aaafemale4<-subset(casino,subset=casino$gen==0 & casino$Sou==1 & casino$Agegrp==0)
head(aaafemale4,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 2  Player 3688    16.04045 Female  27   WALK   0   1      0
## 26 Player 4123    44.78334 Female  25   WALK   0   1      0
```

```
sum(aaafemale4$Total.Spend)
```

```
## [1] 54690.41
```

```
aaafemale5<-subset(casino,subset=casino$gen==0 & casino$Sou==1 & casino$Agegrp==1)
head(aaafemale5,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 16 Player 3096    39.09869 Female  48   WALK   0   1      1
## 20 Player 2895    41.66454 Female  47   WALK   0   1      1
```

```
sum(aaafemale5$Total.Spend)
```

```
## [1] 343805.3
```

```
aaafemale6<-subset(casino,subset=casino$gen==0 & casino$Sou==1 & casino$Agegrp==2)
head(aaafemale6,2)
```

```
##          Player Total.Spend Gender Age Source gen Sou Agegrp
## 96  Player 4727    63.90525 Female  68   WALK   0   1      2
## 100 Player 4875    64.74071 Female  70   WALK   0   1      2
```

```
sum(aaafemale6$Total.Spend)
```

```
## [1] 8030.14
```

```
aaafemale7<-subset(casino,subset=casino$gen==0 & casino$Sou==2 & casino$Agegrp==0)
head(aaafemale7,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 14 Player 4706    37.54023 Female  37    WEB   0   2      0
## 17 Player 1718    39.26721 Female  35    WEB   0   2      0
```

```
sum(aaafemale7$Total.Spend)
```

```
## [1] 191749.5
```

```
aaafemale8<-subset(casino,subset=casino$gen==0 & casino$Sou==2 & casino$Agegrp==1)
head(aaafemale8,2)
```

```
##        Player Total.Spend Gender Age Source gen Sou Agegrp
## 3 Player 1414    17.88949 Female  55    WEB   0   2      1
## 4 Player 4778    24.60281 Female  42    WEB   0   2      1
```

```
sum(aaafemale8$Total.Spend)
```

```
## [1] 1713894
```

```
aaafemale9<-subset(casino,subset=casino$gen==0 & casino$Sou==2 & casino$Agegrp==2)
head(aaafemale9,2)
```

```
##          Player Total.Spend Gender Age Source gen Sou Agegrp
## 38  Player 3346    47.92477 Female  69    WEB   0   2      2
## 144 Player 3225    73.11537 Female  67    WEB   0   2      2
```

```
sum(aaafemale9$Total.Spend)
```

```
## [1] 21029.48
```

```
aaafemale10<-subset(casino,subset=casino$gen==1 & casino$Sou==0 & casino$Agegrp==0)
head(aaafemale10,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 11 Player 3778    34.05991   Male  31    AAA   1   0      0
## 15  Player 344    38.05616   Male  36    AAA   1   0      0
```

```
sum(aaafemale10$Total.Spend)
```

```
## [1] 70539.31
```

```
aaafemale11<-subset(casino,subset=casino$gen==1 & casino$Sou==0 & casino$Agegrp==1)
head(aaafemale11,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 23 Player 1945     43.4933   Male  50    AAA   1   0      1
## 45 Player 4977     49.5356   Male  65    AAA   1   0      1
```

```
sum(aaafemale11$Total.Spend)
```

```
## [1] 369902.8
```

```
aaafemale12<-subset(casino,subset=casino$gen==1 & casino$Sou==0 & casino$Agegrp==2)
head(aaafemale12,2)
```

```
##          Player Total.Spend Gender Age Source gen Sou Agegrp
## 30  Player 1725    45.60373   Male  66    AAA   1   0      2
## 108  Player 498    67.23070   Male  66    AAA   1   0      2
```

```
sum(aaafemale12$Total.Spend)
```

```
## [1] 7764.355
```

```
aaafemale13<-subset(casino,subset=casino$gen==1 & casino$Sou==1 & casino$Agegrp==0)
head(aaafemale13,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 22 Player 2191    42.25217   Male  30   WALK   1   1      0
## 40 Player 4026    48.53758   Male  37   WALK   1   1      0
```

```
sum(aaafemale13$Total.Spend)
```

```
## [1] 58881.38
```

```
aaafemale14<-subset(casino,subset=casino$gen==1 & casino$Sou==1 & casino$Agegrp==1)
head(aaafemale14,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 42 Player 4407    49.23218   Male  56   WALK   1   1      1
## 50 Player 4037    51.32650   Male  65   WALK   1   1      1
```

```
sum(aaafemale14$Total.Spend)
```

```
## [1] 472785.2
```

```
aaafemale15<-subset(casino,subset=casino$gen==1 & casino$Sou==1 & casino$Agegrp==2)
head(aaafemale15,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 10 Player 2209    32.49361   Male  67   WALK   1   1      2
## 31  Player 872    45.62731   Male  69   WALK   1   1      2
```

```
sum(aaafemale15$Total.Spend)
```

```
## [1] 13985.72
```

```
aaafemale16<-subset(casino,subset=casino$gen==1 & casino$Sou==2 & casino$Agegrp==0)
head(aaafemale16,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 8   Player 599    31.42622   Male  33    WEB   1   2      0
## 18 Player 2152    39.88730   Male  26    WEB   1   2      0
```

```
sum(aaafemale16$Total.Spend)
```

```
## [1] 161534.3
```

```
aaafemale17<-subset(casino,subset=casino$gen==1 & casino$Sou==2 & casino$Agegrp==1)
head(aaafemale17,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 12 Player 1993    37.11514   Male  53    WEB   1   2      1
## 13 Player 1783    37.18974   Male  62    WEB   1   2      1
```

```
sum(aaafemale17$Total.Spend)
```

```
## [1] 1611273
```

```
aaafemale18<-subset(casino,subset=casino$gen==1 & casino$Sou==2 & casino$Agegrp==2)
head(aaafemale18,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 41 Player 1628    48.68231   Male  68    WEB   1   2      2
## 98   Player 90    64.35796   Male  68    WEB   1   2      2
```

```
sum(aaafemale18$Total.Spend)
```

```
## [1] 20162.12
```

```
sum<-data.frame(c(sum(aaafemale1$Total.Spend),sum(aaafemale2$Total.Spend),sum(aaafemale3$T
tal.Spend),sum(aaafemale4$Total.Spend),sum(aaafemale5$Total.Spend),sum(aaafemale6$Total.
Spend),sum(aaafemale7$Total.Spend),
                      sum(aaafemale8$Total.Spend),sum(aaafemale9$Total.Spend),sum(aaafe
male10$Total.Spend),sum(aaafemale11$Total.Spend),sum(aaafemale12$Total.Spend),sum(aaafem
ale13$Total.Spend),sum(aaafemale14$Total.Spend),
                      sum(aaafemale15$Total.Spend),sum(aaafemale16$Total.Spend),sum(aaa
female17$Total.Spend),sum(aaafemale18$Total.Spend)))

sum$Sr.No<-1:18
max(sum)
```

```
## [1] 1713894
```

```
min(sum)
```

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

Creates a data frame that returns the total amount spent by Female walk in visitors.

```
walkfemale<-subset(casino,subset=casino$gen==0 & casino$Sou==1)
head(walkfemale,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 2  Player 3688    16.04045 Female  27   WALK   0   1      0
## 16 Player 3096    39.09869 Female  48   WALK   0   1      1
```

```
sum(walkfemale$Total.Spend)
```

```
## [1] 406525.9
```

Creates a data frame that returns the total amount spent by Female visitors who stay overnight.

```
webfemale<-subset(casino,subset=casino$gen==0 & casino$Sou==2)
head(webfemale,2)
```

```
##        Player Total.Spend Gender Age Source gen Sou Agegrp
## 3 Player 1414    17.88949 Female  55    WEB   0   2      1
## 4 Player 4778    24.60281 Female  42    WEB   0   2      1
```

```
sum(webfemale$Total.Spend)
```

```
## [1] 1926673
```

Creates a data frame that returns the total amount spent by Male visitors coming from AAA.

```
aaamale<-subset(casino,subset=casino$gen==1 & casino$Sou==0)
head(aaamale,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 11 Player 3778    34.05991   Male  31    AAA   1   0      0
## 15  Player 344    38.05616   Male  36    AAA   1   0      0
```

```
sum(aaamale$Total.Spend)
```

```
## [1] 448206.4
```

Creates a data frame that returns the total amount spent by Male walk in visitors.

```
walkmale<-subset(casino,subset=casino$gen==1 & casino$Sou==1)
head(walkmale,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 10 Player 2209    32.49361   Male  67   WALK   1   1      2
## 22 Player 2191    42.25217   Male  30   WALK   1   1      0
```

```
sum(walkmale$Total.Spend)
```

```
## [1] 545652.3
```

Creates a data frame that returns the total amount spent by Male visitors who stay overnight.

```
webmale<-subset(casino,subset=casino$gen==1 & casino$Sou==2)
head(webmale,2)
```

```
##         Player Total.Spend Gender Age Source gen Sou Agegrp
## 8   Player 599    31.42622   Male  33    WEB   1   2      0
## 12 Player 1993    37.11514   Male  53    WEB   1   2      1
```

```
sum(webmale$Total.Spend)
```

```
## [1] 1792969
```

Gives a regression model where the dependent variable is Total Spent while the independent variables are Gender, Age and Source segmentations respectively.

```
model<-lm(casino$Total.Spend~casino$gen + casino$Sou + casino$Agegrp)
summary(model)
```

```
##
## Call:
## lm(formula = casino$Total.Spend ~ casino$gen + casino$Sou + casino$Agegrp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1876.0  -960.6  -612.1    74.7 14105.6
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     355.51      79.97   4.446 8.95e-06 ***
## casino$gen      -43.49      62.12  -0.700    0.484
## casino$Sou      336.88      38.01   8.863  < 2e-16 ***
## casino$Agegrp   447.32      57.05   7.840 5.45e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2196 on 4996 degrees of freedom
## Multiple R-squared:  0.02765,    Adjusted R-squared:  0.02707
## F-statistic: 47.36 on 3 and 4996 DF,  p-value: < 2.2e-16
```

The following data set comprises of the Casino data using which we need to segment our customers based on certain parameters in order to engage them to spending more at our Casino. 1. Model Output: - In our Model, we have considered “total spend” as our dependent variable and gender, source, age segmentations as the independent variables.
- The Sources and Age Groups are statistically significant and have a positive correlation on Spending. With every unit increase in the Sources(independent variable), the amount spent(dependent variable) increases by a factor of 336.88 + 355.51 (Coefficient of Independent variables for sources and the constant respectively).Similarly with every unit increase in Age group (independent variable), the amount spent(dependent variable) increases by a factor of 447.32 + 355.51. - Since Gender has an inverse relationship with Spending, it supports the fact that it is statistically insignificant.

2. Model equation:

```
X = 355.51 + A (-43.49) + B (336.88) + C (447.32)
```

Where, X = Casino$TotalSpendA = Casino$Gen B = Casino$SouC = Casino$Agegrp

3. Approach: Step 1: The data had factor variables which had to be converted into numerical value. Step 2: As a part of Segmentation, the Source, Age and Gender variables were converted into numerical variables Gender: Male = 1, Female = 0 Source: AAA = 0, WALK = 1, WEB = 2 Age: 20 - 40, 41 - 65, 66 and above. Step 3: We found the Correlation among the different variables which were a part of Segmentation. Step 4: We applied the Linear Regression Model method to determine the R-square value and understand the strength of the model.

4. We have a direct relationship of our revenues with the Sources and age groups of Customers. Having further segmented each of these two groups, we can infer that

- Males who walk-in tend to spend more (correlation value: 0.0131).
- Females who are a part of the Association and stay overnight tend to spend more (correlation value: 0.008, 0.01031 respectively. Hence it would be likely that we target this segment of customers to increase the spending and therby the revenue of the casino.

5. The R square value of the model is only 2.7% which is very weak and makes our model less reliable. The gender is least significant variable. However, the Source and Age Group is statistically significant and hence we are 99% confident that the data these variables are significant and have a strong impact on the revenue model.

6. We see that the maximum revenue ($17,13,893) is generated by females who are within the age group of 41-65 and stayed overnight. This is followed by females who belong to the age group 65 and above, and aprt of Association generated a revenue of $4,26,655.012. Also, females who walked in (age group: 41-65), generated about $3,43,805.34. Hence we would target the females who are within the age group of 41-65 and stay overnight.

We see that the maximum revenue ($16,11,272.551) is generated by Males who are within the age group of 41-65 and stayed overnight. This is followed by Males who belong to the age group 41-65, walked in of $4,72,,785.211. Also, Males who are part of association (age group: 41-65), generated about $3,69,902.786. We would target males, who are within the age group of 41-65 and stayed overnight as they generate maximum income.