

Question 2

Pavitra Sattanapalle

November 28, 2016

```
casino<-read.csv("Casino.csv",header = TRUE,stringsAsFactors = TRUE)
View(casino)
head(casino)
```

```
##      Player Total.Spend Gender Age Source
## 1 Player 2690    13.01958 Female  37    AAA
## 2 Player 3688    16.04045 Female  27   WALK
## 3 Player 1414    17.88949 Female  55    WEB
## 4 Player 4778    24.60281 Female  42    WEB
## 5 Player 2952    27.49134 Female  26    AAA
## 6 Player 3838    29.04269 Female  60    WEB
```

```
summary(casino)
```

```
##      Player      Total.Spend      Gender      Age
## Player 1 : 1 Min. : 13.02 Female:2462 Min. :21.0
## Player 10 : 1 1st Qu.: 184.52 Male :2538 1st Qu.:40.0
## Player 100 : 1 Median : 336.95 Median :48.0
## Player 1000: 1 Mean : 1122.41 Mean :47.1
## Player 1001: 1 3rd Qu.: 756.74 3rd Qu.:57.0
## Player 1002: 1 Max. :15582.18 Max. :89.0
## (Other) :4994
## Source
## AAA :1151
## WALK:1233
## WEB :2616
##
##
##
##
```

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

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

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

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

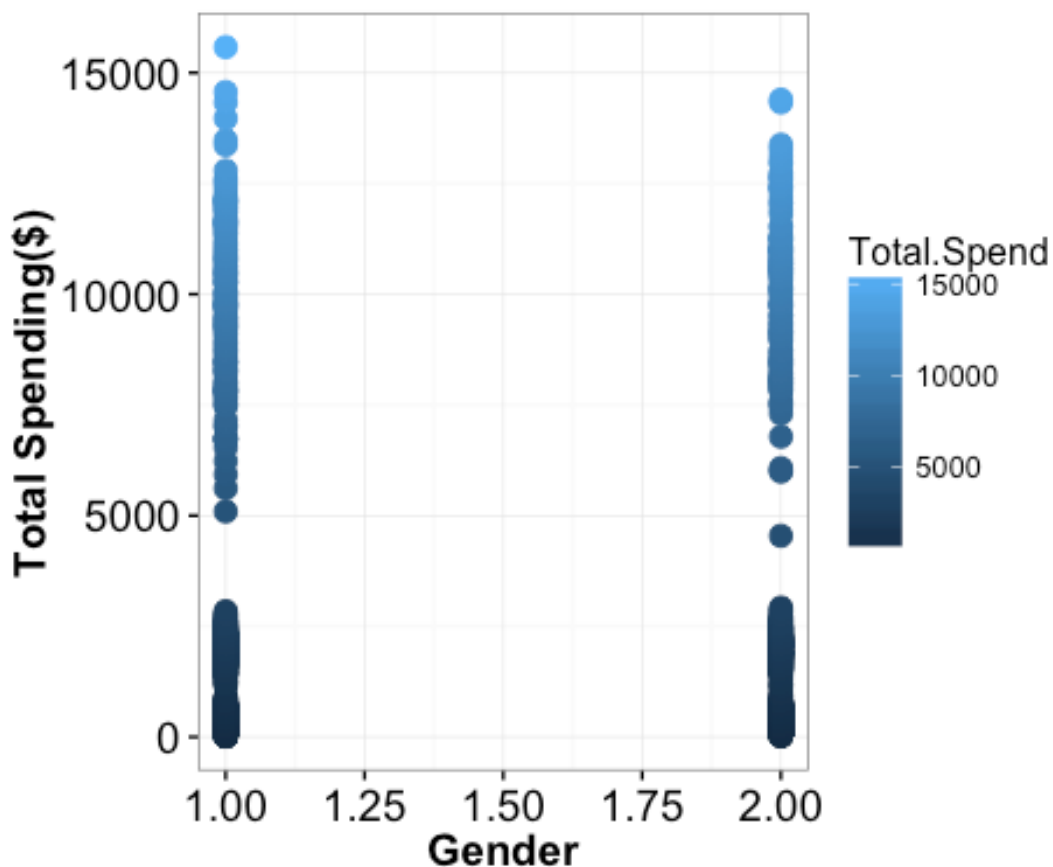
```
cor(casino$Total.Spend,casino$Age)
```

```
## [1] 0.08442126
```

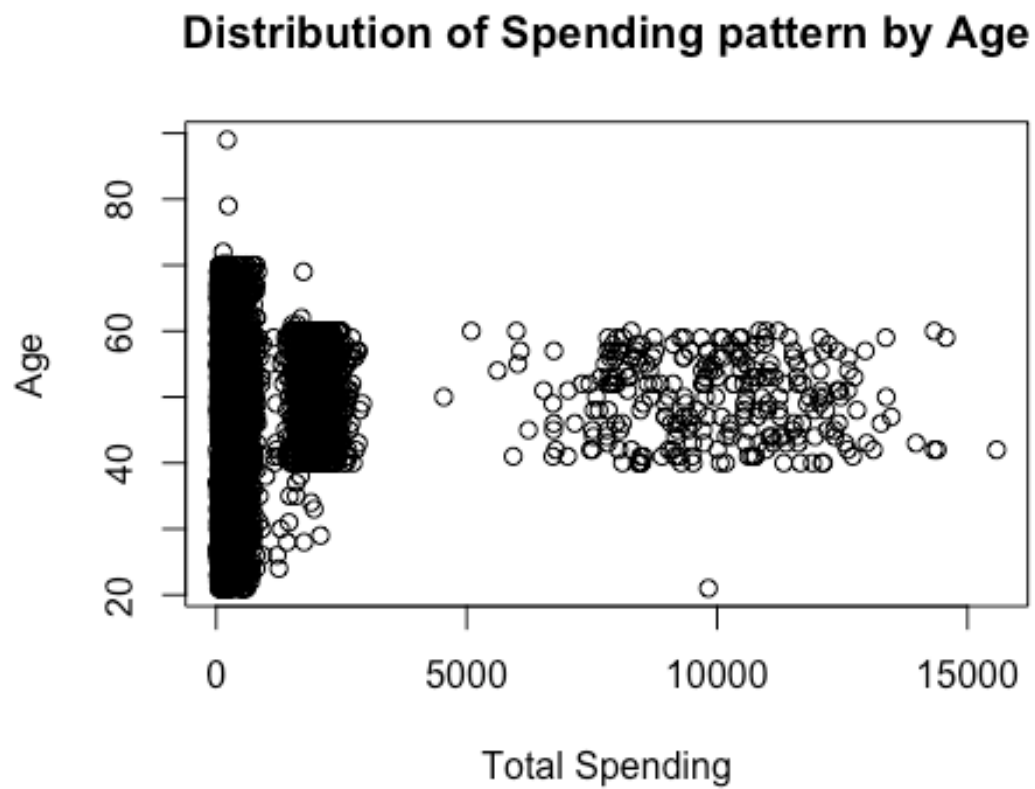
```
casino$Gender<-as.numeric(casino$Gender)  #Specify which value is which gender
casino$Source<-as.numeric(casino$Source)  #Specify which value is which source
cortab<- casino[c(2,3,4,5)]
cor(cortab)

##           Total.Spend      Gender      Age      Source
## Total.Spend  1.00000000 -0.0111156719 0.0844212554 0.124821581
## Gender      -0.01111567  1.0000000000 0.0002173126 -0.005696354
## Age         0.08442126  0.0002173126 1.0000000000 0.018125396
## Source      0.12482158 -0.0056963544 0.0181253958 1.0000000000

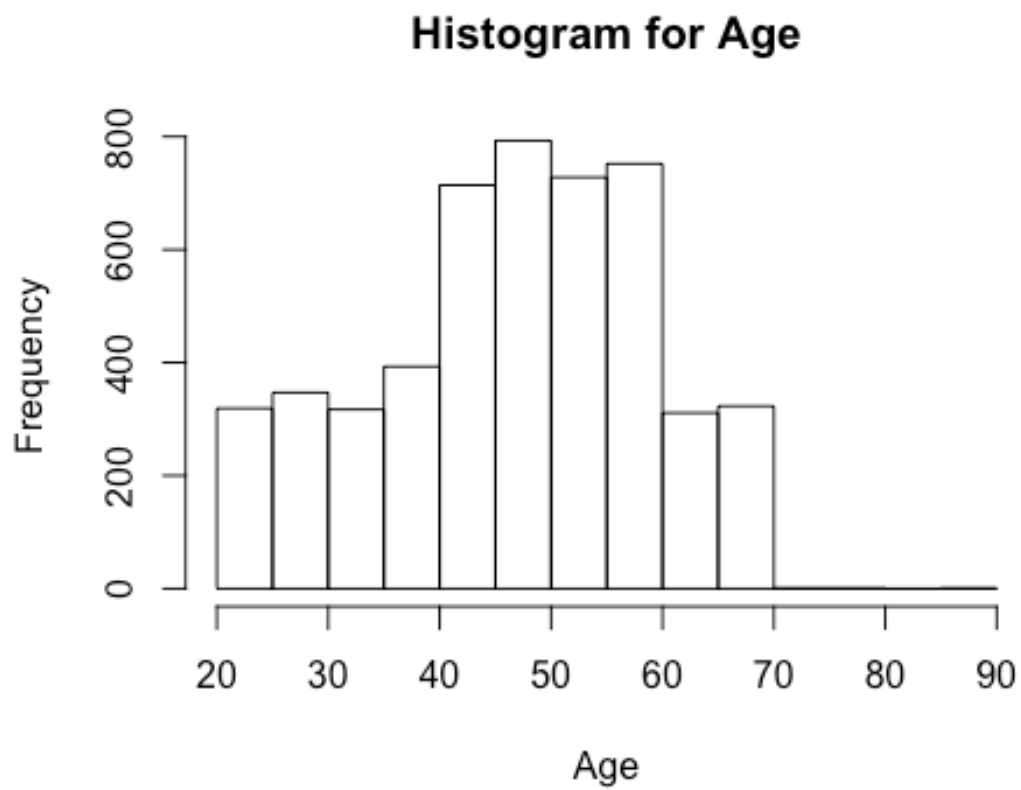
#plot age vs spending
library(ggplot2)
ggplot(data=casino, aes(x=Gender, y=Total.Spend)) +
  geom_point(aes(colour=Total.Spend), size=3) + theme_bw() +
  theme(axis.title.x = element_text(face="bold", vjust=-0.5, size=14),
        axis.text.x = element_text(size=14),
        axis.title.y = element_text(face="bold", vjust=1, size=14),
        axis.text.y = element_text(size=14)) +
  labs(x= "Gender", y = "Total Spending($)")
```



```
plot(casino$Total.Spend,casino$Age, xlab = "Total Spending", ylab = "Age", main = "Distribution of Spending pattern by Age")
```

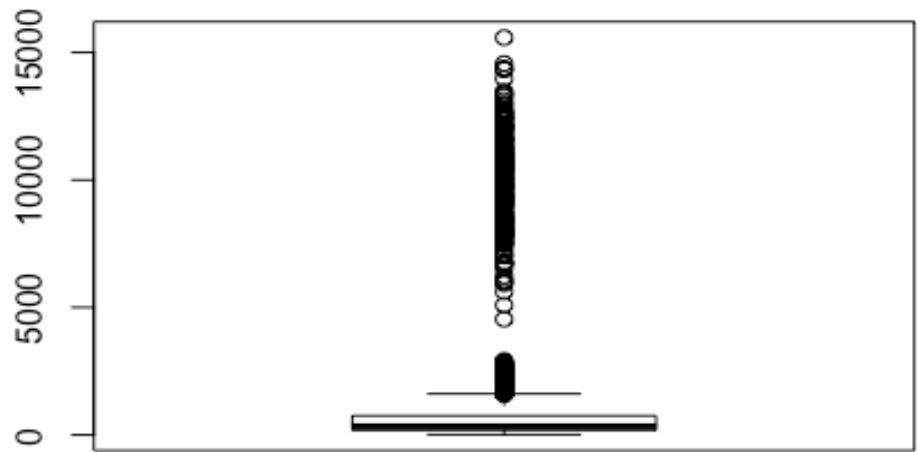


```
hist(casino$Age, xlab="Age", main= "Histogram for Age")
```

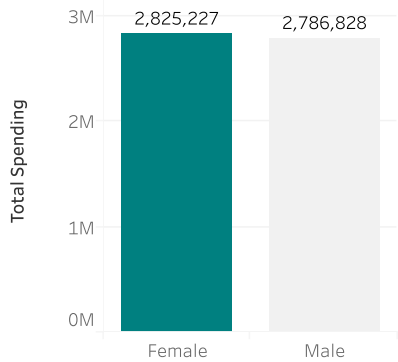


```
boxplot(casino$Total.Spend, main= "Boxplot for Total Spending")
```

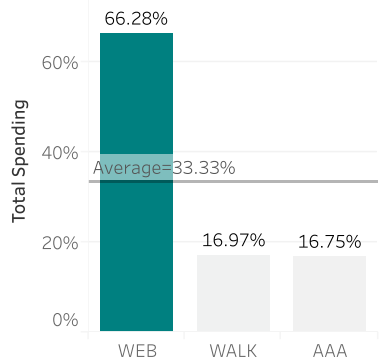
Boxplot for Total Spending



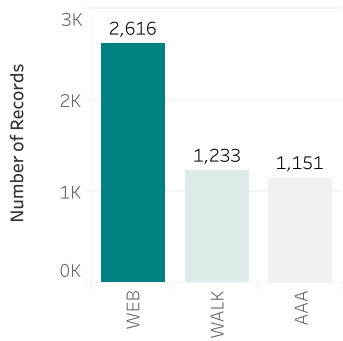
Total Spending per Gender



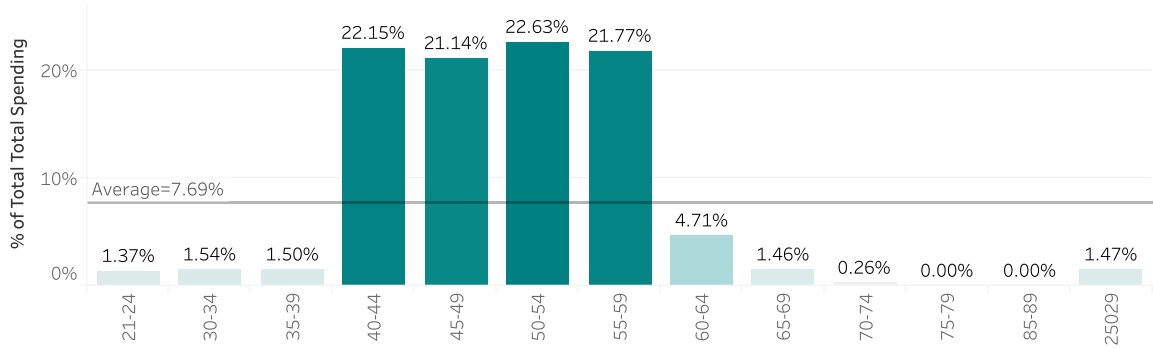
Total Spending per Source



Count of Spending Sources



Total Spending per Age



```

#linear regression model
lmcasino<-lm(data = casino, formula = casino$Total.Spend~casino$Gender+casino
$Age+casino$Source)
summary(lmcasino)

##
## Call:
## lm(formula = casino$Total.Spend ~ casino$Gender + casino$Age +
##     casino$Source, data = casino)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1783.1  -989.1  -662.9   18.1 14274.1
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -273.943    176.589  -1.551   0.121
## casino$Gender  -46.445     62.283  -0.746   0.456
## casino$Age     14.783      2.516   5.875 4.51e-09 ***
## casino$Source 335.851     38.117   8.811 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2202 on 4996 degrees of freedom
## Multiple R-squared:  0.02244,    Adjusted R-squared:  0.02185
## F-statistic: 38.23 on 3 and 4996 DF,  p-value: < 2.2e-16

```

Business Implications

1. What is your model's output?

The Linear Regression Model has an R-squared value of 0.02244 which implies a model strength of 2.2%. The model is an extremely weak model and will not produce statistically significant results.

2. What is your model's equation?

```
lm(formula = casinoTotal.Spend ~ casinoGender + casinoAge + casinoSource, data =
casino)
```

3. What is the approach for developing this model?

Name the dataset. Talk about the variables present. Mention initial analysis included boxplots/histograms summary statistics etc. Since only 3 independent variables are there, all 3 were selected to create this model and find then significance on the dependent variable.

4. Strengths, Weaknesses, Reliability of the Model

From our Linear Model we can see that the Age of customers and the Source are highly significant when it comes to Total Spending.

The Gender however is not at all significant when it comes to measuring Total Spending.

The significant factors can be further utilised to drive business decisions for the casino. Since Age and Source are highly significant, casino management can target their efforts towards maximising income from these sources.

The model in itself is not reliable at all as it has an extremely low R-squared value. The low R-squared value indicates that changes in these predictors are not related to changes in the response variable and that this model does not explain a lot of the response variability.

5. Business Implications and Recommendations

- The first business implication would be to create a more comprehensive dataset. This could be done by either recording more data, or by recording data over a wider spread of variables so that dependencies can be better observed.
- The current dataset only captures the Age, Source, Total Spending and Gender of customers. If The Casino is able to capture data such as spending on individual games, average spending across hours of operation, marital status of customers etc., it will prove useful in generating stronger models that better predict patterns.
- From the plots generated, we can see that the income from Web customers is the highest. The casino should focus their efforts towards targeting this segment of customers. They could launch attractive marketing campaigns and special pricing/discount schemes for web traffic to maximise their income from this segment of the customer population.
- From the plots generated, we can see that the maximum income comes from customers between the ages of 40 - 60. The Casino should further try and appeal to this demographic of the customer population. They could do this by creating an ambience/environment that people between the ages of 40-60 generally like. They could create a separate family dining section as well as a separate kids play area, so that customers from this demographic could come to The Casino with their families and still play.