# Final Project Part 2

### Augi and Watson

shark\_student <- read\_csv("shark\_student.csv")

## Warning: Missing column names filled in: 'X1' [1]

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## .default = col\_double(),  
## Company = col\_character(),  
## SeasonEpisode = col\_character(),  
## CompanyState = col\_character()  
## )  
## i Use `spec()` for the full column specifications.

shark\_student <- shark\_student %>%   
 select(-X1) %>%   
 mutate\_if(is.character,as\_factor) %>%   
 mutate(ReceiveOffer = as\_factor(ReceiveOffer)) %>%   
 mutate(ReceiveOffer = fct\_recode(ReceiveOffer, "No" = "0", "Yes" = "1" )) %>%   
 mutate(RejectOffer = as\_factor(RejectOffer)) %>%   
 mutate(RejectOffer = fct\_recode(RejectOffer, "No" = "0", "Yes" = "1" )) %>%   
 mutate(Deal\_Yes = as\_factor(Deal\_Yes)) %>%   
 mutate(Deal\_Yes = fct\_recode(Deal\_Yes, "No" = "0", "Yes" = "1" )) %>%   
 mutate(Deal\_No = as\_factor(Deal\_No)) %>%   
 mutate(Deal\_No = fct\_recode(Deal\_No, "No" = "0", "Yes" = "1" )) %>%   
 mutate(Eth1 = as\_factor(Eth1)) %>%   
 mutate(Eth1 = fct\_recode(Eth1, "African American" = "1", "White" = "2", "Asian" = "3", "Latino" = "4", "No presenter 1" = "0"))%>%   
 mutate(Eth2 = as\_factor(Eth2)) %>%   
 mutate(Eth2 = fct\_recode(Eth2, "African American" = "1", "White" = "2", "Asian" = "3", "Latino" = "4", "No presenter 2" = "0"))%>%   
 mutate(Eth3 = as\_factor(Eth3)) %>%   
 mutate(Eth3 = fct\_recode(Eth3, "African American" = "1", "White" = "2", "Asian" = "3", "Latino" = "4", "No presenter 3" = "0"))%>%   
 mutate(Eth4 = as\_factor(Eth4)) %>%   
 mutate(Eth4 = fct\_recode(Eth4, "African American" = "1", "White" = "2", "Asian" = "3", "Latino" = "4", "No presenter 4" = "0"))%>%   
 mutate(Eth5 = as\_factor(Eth5)) %>%   
 mutate(Eth5 = fct\_recode(Eth5, "African American" = "1", "White" = "2", "Asian" = "3", "Latino" = "4", "No presenter 5" = "0"))%>%   
 mutate(Male1 = as\_factor(Male1)) %>%   
 mutate(Male1 = fct\_recode(Male1, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Male2 = as\_factor(Male2)) %>%   
 mutate(Male2 = fct\_recode(Male2, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Male3 = as\_factor(Male3)) %>%   
 mutate(Male3 = fct\_recode(Male3, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Male4 = as\_factor(Male4)) %>%   
 mutate(Male4 = fct\_recode(Male4, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Female1 = as\_factor(Female1)) %>%   
 mutate(Female1 = fct\_recode(Female1, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Female2 = as\_factor(Female2)) %>%   
 mutate(Female2 = fct\_recode(Female2, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Female3 = as\_factor(Female3)) %>%   
 mutate(Female3 = fct\_recode(Female3, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Female4 = as\_factor(Female4)) %>%   
 mutate(Female4 = fct\_recode(Female4, "No" = "0", "Yes" = "1" ))

## Warning: Unknown levels in `f`: 1, 3, 4  
  
## Warning: Unknown levels in `f`: 1, 3, 4

## Warning: Unknown levels in `f`: 1

shark\_student <- shark\_student %>%   
 rename("Food\_Beverage"="Food and Beverage") %>%   
 mutate(Food\_Beverage = as\_factor(Food\_Beverage)) %>%   
 mutate(Food\_Beverage = fct\_recode(Food\_Beverage, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Novelties = as\_factor(Novelties)) %>%   
 mutate(Novelties = fct\_recode(Novelties, "No" = "0", "Yes" = "1" )) %>%   
 rename("Health\_Wellness"="Health / Wellness") %>%   
 mutate(Health\_Wellness = as\_factor(Health\_Wellness)) %>%   
 mutate(Health\_Wellness = fct\_recode(Health\_Wellness, "No" = "0", "Yes" = "1" ))%>%   
 rename("Lifestyle\_Home"="Lifestyle / Home") %>%   
 mutate(Lifestyle\_Home = as\_factor(Lifestyle\_Home)) %>%   
 mutate(Lifestyle\_Home = fct\_recode(Lifestyle\_Home, "No" = "0", "Yes" = "1" ))%>%   
 rename("Software\_Tech"="Software / Tech") %>%   
 mutate(Software\_Tech = as\_factor(Software\_Tech)) %>%   
 mutate(Software\_Tech = fct\_recode(Software\_Tech, "No" = "0", "Yes" = "1" ))   
  
shark\_student <- shark\_student %>%   
 rename("Children\_Education"="Children / Education") %>%   
 mutate(Children\_Education = as\_factor(Children\_Education)) %>%   
 mutate(Children\_Education = fct\_recode(Children\_Education, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Automotive = as\_factor(Automotive)) %>%   
 mutate(Automotive = fct\_recode(Automotive, "No" = "0", "Yes" = "1" ))%>%   
 rename("Fashion\_Beauty"="Fashion / Beauty") %>%   
 mutate(Fashion\_Beauty = as\_factor(Fashion\_Beauty)) %>%   
 mutate(Fashion\_Beauty = fct\_recode(Fashion\_Beauty, "No" = "0", "Yes" = "1" ))%>%   
 rename("Media\_Entertainment"="Media / Entertainment") %>%   
 mutate(Media\_Entertainment = as\_factor(Media\_Entertainment)) %>%   
 mutate(Media\_Entertainment = fct\_recode(Media\_Entertainment, "No" = "0", "Yes" = "1" ))   
  
shark\_student <- shark\_student %>%   
 rename("Fitness\_Sports\_Outdoors"="Fitness / Sports / Outdoors") %>%   
 mutate(Fitness\_Sports\_Outdoors = as\_factor(Fitness\_Sports\_Outdoors)) %>%   
 mutate(Fitness\_Sports\_Outdoors = fct\_recode(Fitness\_Sports\_Outdoors, "No" = "0", "Yes" = "1" ))%>%   
 rename("Pet\_Products" = "Pet Products") %>%   
 mutate(Pet\_Products = as\_factor(Pet\_Products)) %>%   
 mutate(Pet\_Products = fct\_recode(Pet\_Products, "No" = "0", "Yes" = "1" ))%>%   
 rename("Business\_Services" = "Business Services") %>%   
 mutate(Business\_Services = as\_factor(Business\_Services)) %>%   
 mutate(Business\_Services = fct\_recode(Business\_Services, "No" = "0", "Yes" = "1" ))%>%   
 mutate(Travel = as\_factor(Travel)) %>%   
 mutate(Travel = fct\_recode(Travel, "No" = "0", "Yes" = "1" ))   
  
shark\_student <- shark\_student %>%   
 rename("Green\_CleanTech"="Green/CleanTech") %>%   
 mutate(Green\_CleanTech = as\_factor(Green\_CleanTech)) %>%   
 mutate(Green\_CleanTech = fct\_recode(Green\_CleanTech, "No" = "0", "Yes" = "1" ))%>%   
 rename("Uncertain\_Other"="Uncertain / Other") %>%   
 mutate(Uncertain\_Other = as\_factor(Uncertain\_Other)) %>%   
 mutate(Uncertain\_Other = fct\_recode(Uncertain\_Other, "No" = "0", "Yes" = "1" ))%>%   
 mutate(MalePresenter = as\_factor(MalePresenter)) %>%   
 mutate(MalePresenter = fct\_recode(MalePresenter, "No" = "0", "Yes" = "1" ))%>%   
 mutate(FemalePresenter = as\_factor(FemalePresenter)) %>%   
 mutate(FemalePresenter = fct\_recode(FemalePresenter, "No" = "0", "Yes" = "1" ))%>%   
 mutate(MixedGenderPresenters = as\_factor(MixedGenderPresenters)) %>%   
 mutate(MixedGenderPresenters = fct\_recode(MixedGenderPresenters, "No" = "0", "Yes" = "1" ))   
  
shark\_student <- shark\_student%>%   
 mutate(BarbaraCorcoran = as\_factor(BarbaraCorcoran)) %>%   
 mutate(BarbaraCorcoran = fct\_recode(BarbaraCorcoran, "No" = "0", "Yes" = "1" ))%>%   
 mutate(MarkCuban = as\_factor(MarkCuban)) %>%   
 mutate(MarkCuban = fct\_recode(MarkCuban, "No" = "0", "Yes" = "1" ))%>%   
 mutate(LoriGreiner = as\_factor(LoriGreiner)) %>%   
 mutate(LoriGreiner = fct\_recode(LoriGreiner, "No" = "0", "Yes" = "1" ))%>%   
 mutate(RobertHerjavec = as\_factor(RobertHerjavec)) %>%   
 mutate(RobertHerjavec = fct\_recode(RobertHerjavec, "No" = "0", "Yes" = "1" ))%>%   
 mutate(DaymondJohn = as\_factor(DaymondJohn)) %>%   
 mutate(DaymondJohn = fct\_recode(DaymondJohn, "No" = "0", "Yes" = "1" ))%>%   
 mutate(KevinOLeary = as\_factor(KevinOLeary)) %>%   
 mutate(KevinOLeary = fct\_recode(KevinOLeary, "No" = "0", "Yes" = "1" ))%>%   
 mutate(KevinHarrington = as\_factor(KevinHarrington)) %>%   
 mutate(KevinHarrington = fct\_recode(KevinHarrington, "No" = "0", "Yes" = "1" ))   
  
shark\_student <- shark\_student %>%   
 mutate(Guest = as\_factor(Guest)) %>%   
 mutate(Guest = fct\_recode(Guest, "No" = "0", "Yes" = "1" ))   
  
shark\_student <- shark\_student %>%   
 rename("NumberOfPresenters"="Number of Presenters")  
  
shark\_student <- select(shark\_student, -Male4)   
shark\_student <- shark\_student %>% drop\_na()   
summary(shark\_student)

## Company SeasonEpisode ReceiveOffer  
## Wispots : 1 Shark.Tank.S01E11: 5 No :183   
## CollegeFoxesPackingBoxes: 1 Shark.Tank.S01E01: 4 Yes:295   
## Mr.TodsPieFactory : 1 Shark.Tank.S01E07: 4   
## AvatheElephant : 1 Shark.Tank.S02E04: 4   
## AttachNoted : 1 Shark.Tank.S02E07: 4   
## CrookedJaw : 1 Shark.Tank.S02E09: 4   
## (Other) :472 (Other) :453   
## RejectOffer Deal\_Yes Deal\_No NumberOfPresenters Eth1   
## No :444 No :220 No :258 Min. :1.000 No presenter 1 : 2   
## Yes: 34 Yes:258 Yes:220 1st Qu.:1.000 African American: 40   
## Median :1.000 White :372   
## Mean :1.525 Asian : 16   
## 3rd Qu.:2.000 Latino : 48   
## Max. :4.000   
##   
## Eth2 Eth3 Eth4   
## No presenter 2 :254 No presenter 3 :453 No presenter 4:476   
## African American: 17 African American: 2 White : 2   
## White :150 White : 18   
## Asian : 18 Asian : 2   
## Latino : 39 Latino : 3   
##   
##   
## Eth5 Male1 Male2 Male3 Female1 Female2   
## No presenter 5:477 No :129 No :355 No :462 No :349 No :377   
## White : 1 Yes:349 Yes:123 Yes: 16 Yes:129 Yes:101   
##   
##   
##   
##   
##   
## Female3 Female4 Novelties Health\_Wellness Food\_Beverage Business\_Services  
## No :468 No :476 No :457 No :468 No :399 No :448   
## Yes: 10 Yes: 2 Yes: 21 Yes: 10 Yes: 79 Yes: 30   
##   
##   
##   
##   
##   
## Lifestyle\_Home Software\_Tech Children\_Education Automotive Fashion\_Beauty  
## No :397 No :461 No :430 No :471 No :396   
## Yes: 81 Yes: 17 Yes: 48 Yes: 7 Yes: 82   
##   
##   
##   
##   
##   
## Media\_Entertainment Fitness\_Sports\_Outdoors Pet\_Products Travel   
## No :442 No :432 No :462 No :476   
## Yes: 36 Yes: 46 Yes: 16 Yes: 2   
##   
##   
##   
##   
##   
## Green\_CleanTech Uncertain\_Other MalePresenter FemalePresenter  
## No :477 No :476 No :187 No :363   
## Yes: 1 Yes: 2 Yes:291 Yes:115   
##   
##   
##   
##   
##   
## MixedGenderPresenters CompanyState AmountRequested EquityRequested   
## No :406 CA :123 Min. : 10000 Min. :0.0150   
## Yes: 72 NY : 43 1st Qu.: 90000 1st Qu.:0.1000   
## FL : 38 Median : 150000 Median :0.1500   
## TX : 38 Mean : 260684 Mean :0.1614   
## GA : 19 3rd Qu.: 300000 3rd Qu.:0.2000   
## MA : 18 Max. :5000000 Max. :1.0000   
## (Other):199   
## ImpliedValuationRequested BarbaraCorcoran MarkCuban LoriGreiner RobertHerjavec  
## Min. : 40000 No :176 No : 56 No :152 No : 32   
## 1st Qu.: 500000 Yes:302 Yes:422 Yes:326 Yes:446   
## Median : 1000000   
## Mean : 2637321   
## 3rd Qu.: 2500000   
## Max. :40000000   
##   
## DaymondJohn KevinOLeary KevinHarrington Guest   
## No :154 No : 22 No :430 No :407   
## Yes:324 Yes:456 Yes: 48 Yes: 71   
##   
##   
##   
##   
##

skim(shark\_student)

Data summary

|  |  |
| --- | --- |
| Name | shark\_student |
| Number of rows | 478 |
| Number of columns | 49 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 45 |
| numeric | 4 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

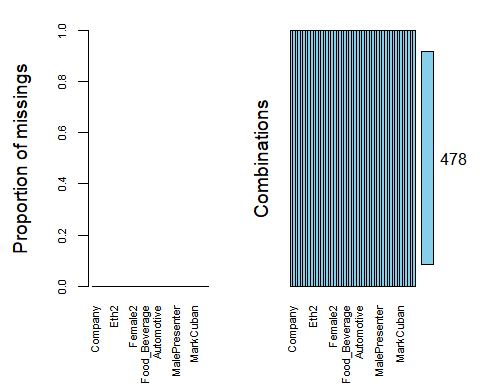
**Variable type: factor**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| Company | 0 | 1 | FALSE | 478 | Wis: 1, Col: 1, Mr.: 1, Ava: 1 |
| SeasonEpisode | 0 | 1 | FALSE | 173 | Sha: 5, Sha: 4, Sha: 4, Sha: 4 |
| ReceiveOffer | 0 | 1 | FALSE | 2 | Yes: 295, No: 183 |
| RejectOffer | 0 | 1 | FALSE | 2 | No: 444, Yes: 34 |
| Deal\_Yes | 0 | 1 | FALSE | 2 | Yes: 258, No: 220 |
| Deal\_No | 0 | 1 | FALSE | 2 | No: 258, Yes: 220 |
| Eth1 | 0 | 1 | FALSE | 5 | Whi: 372, Lat: 48, Afr: 40, Asi: 16 |
| Eth2 | 0 | 1 | FALSE | 5 | No : 254, Whi: 150, Lat: 39, Asi: 18 |
| Eth3 | 0 | 1 | FALSE | 5 | No : 453, Whi: 18, Lat: 3, Afr: 2 |
| Eth4 | 0 | 1 | FALSE | 2 | No : 476, Whi: 2 |
| Eth5 | 0 | 1 | FALSE | 2 | No : 477, Whi: 1 |
| Male1 | 0 | 1 | FALSE | 2 | Yes: 349, No: 129 |
| Male2 | 0 | 1 | FALSE | 2 | No: 355, Yes: 123 |
| Male3 | 0 | 1 | FALSE | 2 | No: 462, Yes: 16 |
| Female1 | 0 | 1 | FALSE | 2 | No: 349, Yes: 129 |
| Female2 | 0 | 1 | FALSE | 2 | No: 377, Yes: 101 |
| Female3 | 0 | 1 | FALSE | 2 | No: 468, Yes: 10 |
| Female4 | 0 | 1 | FALSE | 2 | No: 476, Yes: 2 |
| Novelties | 0 | 1 | FALSE | 2 | No: 457, Yes: 21 |
| Health\_Wellness | 0 | 1 | FALSE | 2 | No: 468, Yes: 10 |
| Food\_Beverage | 0 | 1 | FALSE | 2 | No: 399, Yes: 79 |
| Business\_Services | 0 | 1 | FALSE | 2 | No: 448, Yes: 30 |
| Lifestyle\_Home | 0 | 1 | FALSE | 2 | No: 397, Yes: 81 |
| Software\_Tech | 0 | 1 | FALSE | 2 | No: 461, Yes: 17 |
| Children\_Education | 0 | 1 | FALSE | 2 | No: 430, Yes: 48 |
| Automotive | 0 | 1 | FALSE | 2 | No: 471, Yes: 7 |
| Fashion\_Beauty | 0 | 1 | FALSE | 2 | No: 396, Yes: 82 |
| Media\_Entertainment | 0 | 1 | FALSE | 2 | No: 442, Yes: 36 |
| Fitness\_Sports\_Outdoors | 0 | 1 | FALSE | 2 | No: 432, Yes: 46 |
| Pet\_Products | 0 | 1 | FALSE | 2 | No: 462, Yes: 16 |
| Travel | 0 | 1 | FALSE | 2 | No: 476, Yes: 2 |
| Green\_CleanTech | 0 | 1 | FALSE | 2 | No: 477, Yes: 1 |
| Uncertain\_Other | 0 | 1 | FALSE | 2 | No: 476, Yes: 2 |
| MalePresenter | 0 | 1 | FALSE | 2 | Yes: 291, No: 187 |
| FemalePresenter | 0 | 1 | FALSE | 2 | No: 363, Yes: 115 |
| MixedGenderPresenters | 0 | 1 | FALSE | 2 | No: 406, Yes: 72 |
| CompanyState | 0 | 1 | FALSE | 46 | CA: 123, NY: 43, FL: 38, TX: 38 |
| BarbaraCorcoran | 0 | 1 | FALSE | 2 | Yes: 302, No: 176 |
| MarkCuban | 0 | 1 | FALSE | 2 | Yes: 422, No: 56 |
| LoriGreiner | 0 | 1 | FALSE | 2 | Yes: 326, No: 152 |
| RobertHerjavec | 0 | 1 | FALSE | 2 | Yes: 446, No: 32 |
| DaymondJohn | 0 | 1 | FALSE | 2 | Yes: 324, No: 154 |
| KevinOLeary | 0 | 1 | FALSE | 2 | Yes: 456, No: 22 |
| KevinHarrington | 0 | 1 | FALSE | 2 | No: 430, Yes: 48 |
| Guest | 0 | 1 | FALSE | 2 | No: 407, Yes: 71 |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| NumberOfPresenters | 0 | 1 | 1.53 | 0.61 | 1e+00 | 1e+00 | 1.0e+00 | 2.0e+00 | 4e+00 | ▇▆▁▁▁ |
| AmountRequested | 0 | 1 | 260684.10 | 397883.08 | 1e+04 | 9e+04 | 1.5e+05 | 3.0e+05 | 5e+06 | ▇▁▁▁▁ |
| EquityRequested | 0 | 1 | 0.16 | 0.10 | 1e-02 | 1e-01 | 1.5e-01 | 2.0e-01 | 1e+00 | ▇▂▁▁▁ |
| ImpliedValuationRequested | 0 | 1 | 2637321.18 | 4344427.88 | 4e+04 | 5e+05 | 1.0e+06 | 2.5e+06 | 4e+07 | ▇▁▁▁▁ |

vim\_plot = aggr(shark\_student, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)



set.seed(123)   
shark\_split = initial\_split(shark\_student, prop = 0.7, strata = Deal\_Yes)   
train = training(shark\_split)   
test = testing(shark\_split)   
  
shark\_recipe = recipe(Deal\_Yes ~., train) %>%   
 step\_dummy(all\_nominal(), -all\_outcomes())   
   
rf\_model = rand\_forest() %>%   
 set\_engine("ranger", importance = "permutation") %>% #added importance metric   
 set\_mode("classification")   
   
shark\_wflow =   
 workflow() %>%   
 add\_model(rf\_model) %>%   
 add\_recipe(shark\_recipe)   
   
set.seed(123)   
shark\_fit = fit(shark\_wflow, train)   
   
trainpredrf = predict(shark\_fit, train)   
head(trainpredrf)

## # A tibble: 6 x 1  
## .pred\_class  
## <fct>   
## 1 No   
## 2 No   
## 3 No   
## 4 No   
## 5 No   
## 6 No

confusionMatrix(trainpredrf$.pred\_class, train$Deal\_Yes,   
 positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 154 0  
## Yes 0 180  
##   
## Accuracy : 1   
## 95% CI : (0.989, 1)  
## No Information Rate : 0.5389   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 1   
##   
## Mcnemar's Test P-Value : NA   
##   
## Sensitivity : 1.0000   
## Specificity : 1.0000   
## Pos Pred Value : 1.0000   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5389   
## Detection Rate : 0.5389   
## Detection Prevalence : 0.5389   
## Balanced Accuracy : 1.0000   
##   
## 'Positive' Class : Yes   
##

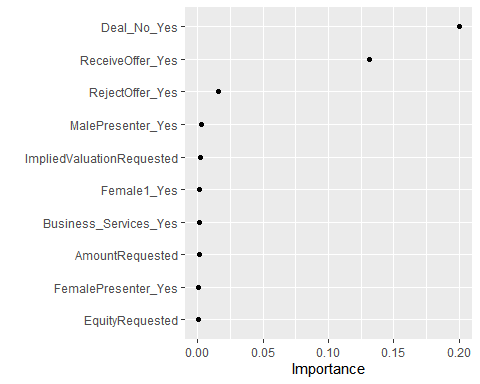
testpredrf = predict(shark\_fit, test)   
head(testpredrf)

## # A tibble: 6 x 1  
## .pred\_class  
## <fct>   
## 1 No   
## 2 Yes   
## 3 No   
## 4 Yes   
## 5 Yes   
## 6 Yes

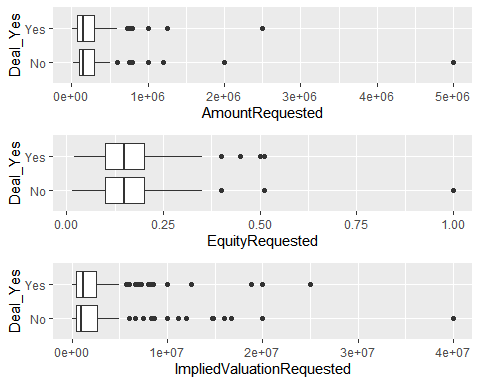
confusionMatrix(testpredrf$.pred\_class, test$Deal\_Yes,   
 positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 65 0  
## Yes 1 78  
##   
## Accuracy : 0.9931   
## 95% CI : (0.9619, 0.9998)  
## No Information Rate : 0.5417   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.986   
##   
## Mcnemar's Test P-Value : 1   
##   
## Sensitivity : 1.0000   
## Specificity : 0.9848   
## Pos Pred Value : 0.9873   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5417   
## Detection Rate : 0.5417   
## Detection Prevalence : 0.5486   
## Balanced Accuracy : 0.9924   
##   
## 'Positive' Class : Yes   
##

shark\_fit %>% pull\_workflow\_fit() %>% vip(geom = "point")



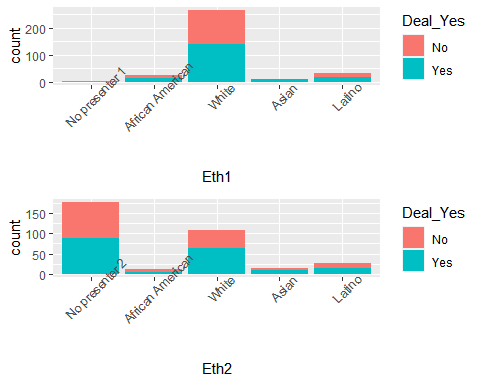
p1 = ggplot(train, aes(x = AmountRequested,y = Deal\_Yes)) + geom\_boxplot()   
  
p2 = ggplot(train, aes(x = EquityRequested, y = Deal\_Yes)) + geom\_boxplot()   
  
p3 = ggplot(train, aes(x = ImpliedValuationRequested, y = Deal\_Yes)) + geom\_boxplot()   
  
  
grid.arrange(p1,p2,p3)



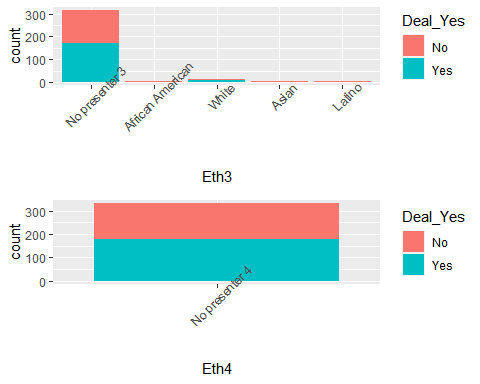
p1 = ggplot(train, aes(x = Company, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = SeasonEpisode, fill = Deal\_Yes)) + geom\_bar()   
  
p3 = ggplot(train, aes(x = ReceiveOffer, fill = Deal\_Yes)) + geom\_bar()   
  
p4 = ggplot(train, aes(x = RejectOffer, fill = Deal\_Yes))+ geom\_bar()  
  
grid.arrange(p1,p2,p3,p4)



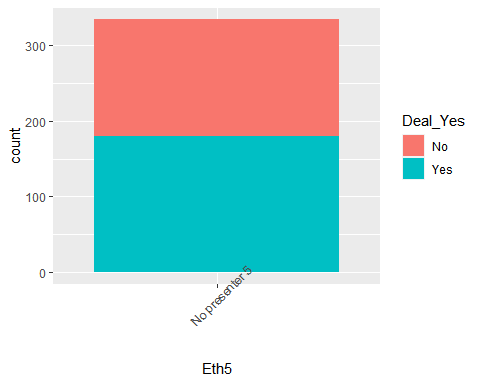
p1 = ggplot(train, aes(x = Eth1, fill = Deal\_Yes)) + geom\_bar() + theme(axis.text.x = element\_text(angle = 45))  
  
p2 = ggplot(train, aes(x = Eth2, fill = Deal\_Yes)) + geom\_bar() + theme(axis.text.x = element\_text(angle = 45))  
  
  
grid.arrange(p1,p2)



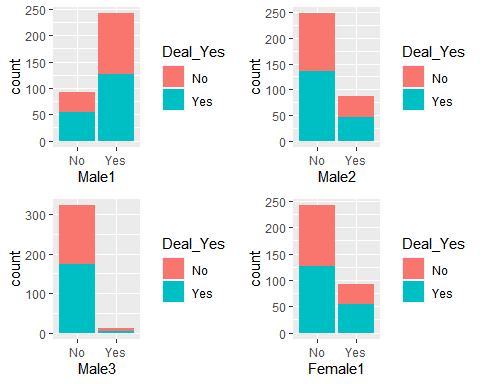
p3 = ggplot(train, aes(x = Eth3, fill = Deal\_Yes)) + geom\_bar()+theme(axis.text.x = element\_text(angle = 45))  
  
  
p4 = ggplot(train, aes(x = Eth4, fill = Deal\_Yes))+ geom\_bar() +theme(axis.text.x = element\_text(angle = 45))  
  
  
grid.arrange(p3,p4)



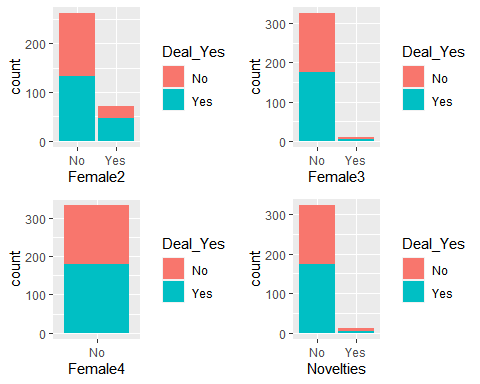
p5 = ggplot(train, aes(x = Eth5, fill = Deal\_Yes))+ geom\_bar() +theme(axis.text.x = element\_text(angle = 45))  
  
grid.arrange(p5)



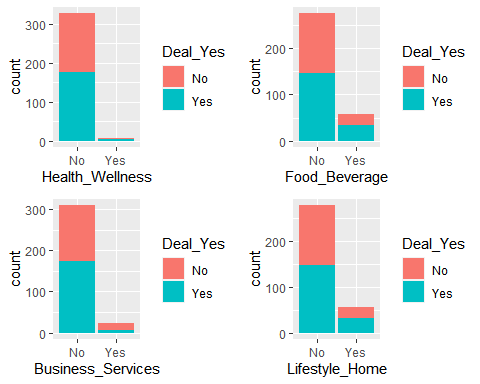
p1 = ggplot(train, aes(x = Male1, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = Male2, fill = Deal\_Yes)) + geom\_bar()   
  
p3 = ggplot(train, aes(x = Male3, fill = Deal\_Yes)) + geom\_bar()   
  
p4 = ggplot(train, aes(x = Female1, fill = Deal\_Yes))+ geom\_bar()  
  
grid.arrange(p1,p2,p3,p4)



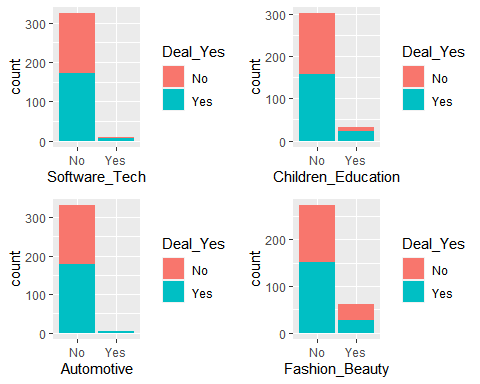
p1 = ggplot(train, aes(x = Female2, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = Female3, fill = Deal\_Yes)) + geom\_bar()   
  
p3 = ggplot(train, aes(x = Female4, fill = Deal\_Yes)) + geom\_bar()   
  
p4 = ggplot(train, aes(x = Novelties, fill = Deal\_Yes))+ geom\_bar()  
  
grid.arrange(p1,p2,p3,p4)



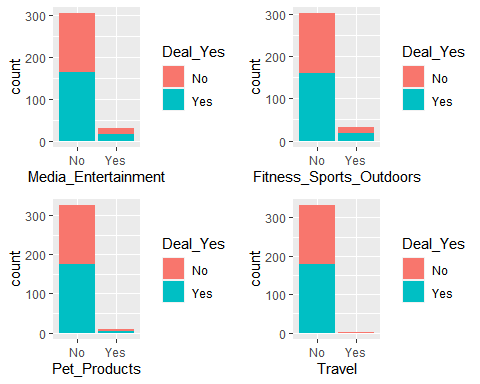
p1 = ggplot(train, aes(x = Health\_Wellness, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = Food\_Beverage, fill = Deal\_Yes)) + geom\_bar()   
  
p3 = ggplot(train, aes(x = Business\_Services, fill = Deal\_Yes)) + geom\_bar()   
  
p4 = ggplot(train, aes(x = Lifestyle\_Home, fill = Deal\_Yes))+ geom\_bar()  
  
grid.arrange(p1,p2,p3,p4)



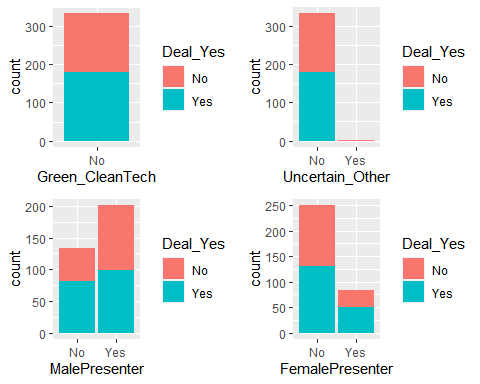
p1 = ggplot(train, aes(x = Software\_Tech, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = Children\_Education, fill = Deal\_Yes)) + geom\_bar()   
  
p3 = ggplot(train, aes(x = Automotive, fill = Deal\_Yes)) + geom\_bar()   
  
p4 = ggplot(train, aes(x = Fashion\_Beauty, fill = Deal\_Yes))+ geom\_bar()  
  
grid.arrange(p1,p2,p3,p4)



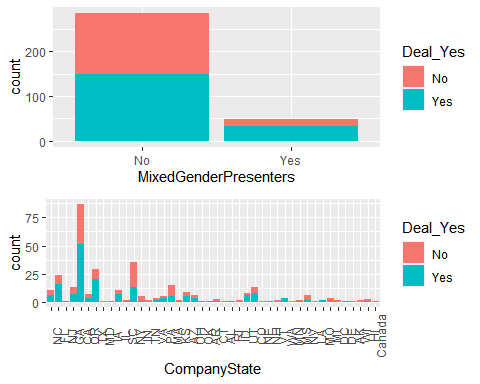
p1 = ggplot(train, aes(x = Media\_Entertainment, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = Fitness\_Sports\_Outdoors, fill = Deal\_Yes)) + geom\_bar()   
  
p3 = ggplot(train, aes(x = Pet\_Products, fill = Deal\_Yes)) + geom\_bar()   
  
p4 = ggplot(train, aes(x = Travel, fill = Deal\_Yes))+ geom\_bar()  
  
grid.arrange(p1,p2,p3,p4)



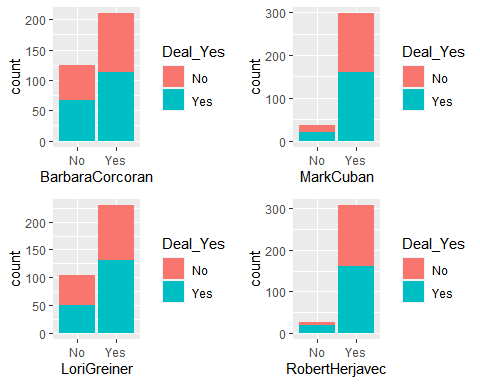
p1 = ggplot(train, aes(x = Green\_CleanTech, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = Uncertain\_Other, fill = Deal\_Yes)) + geom\_bar()   
  
p3 = ggplot(train, aes(x = MalePresenter, fill = Deal\_Yes)) + geom\_bar()   
  
p4 = ggplot(train, aes(x = FemalePresenter, fill = Deal\_Yes))+ geom\_bar()  
  
grid.arrange(p1,p2,p3,p4)



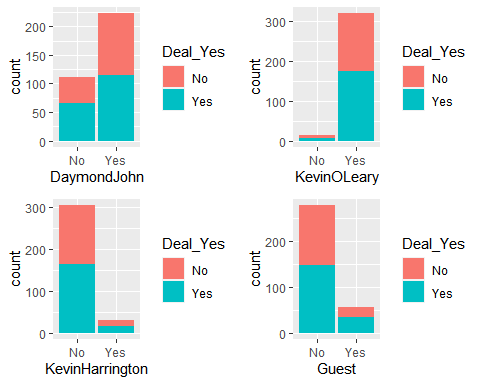
p1 = ggplot(train, aes(x = MixedGenderPresenters, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = CompanyState, fill = Deal\_Yes)) + geom\_bar() +theme(axis.text.x = element\_text(angle = 90))  
  
  
  
grid.arrange(p1,p2)



p1 = ggplot(train, aes(x = BarbaraCorcoran, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = MarkCuban, fill = Deal\_Yes)) + geom\_bar()   
  
p3 = ggplot(train, aes(x = LoriGreiner, fill = Deal\_Yes)) + geom\_bar()   
  
p4 = ggplot(train, aes(x = RobertHerjavec, fill = Deal\_Yes))+ geom\_bar()  
  
grid.arrange(p1,p2,p3,p4)



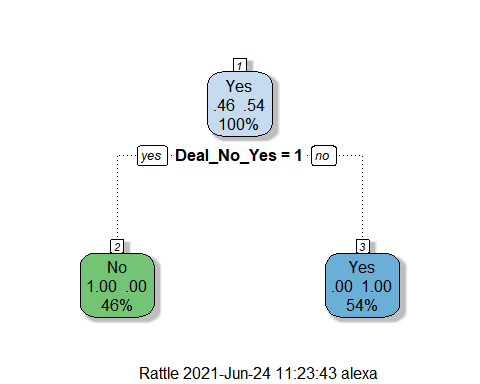
p1 = ggplot(train, aes(x = DaymondJohn, fill = Deal\_Yes)) + geom\_bar()   
  
p2 = ggplot(train, aes(x = KevinOLeary, fill = Deal\_Yes)) + geom\_bar()   
  
p3 = ggplot(train, aes(x = KevinHarrington, fill = Deal\_Yes)) + geom\_bar()   
  
p4 = ggplot(train, aes(x = Guest, fill = Deal\_Yes))+ geom\_bar()  
  
grid.arrange(p1,p2,p3,p4)



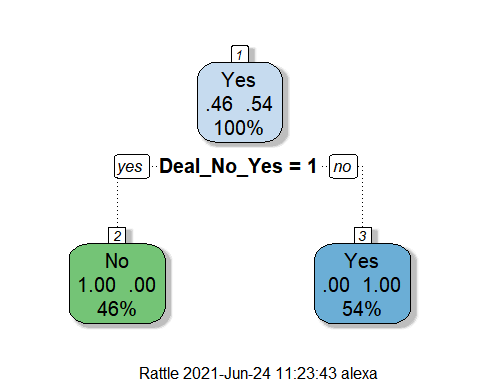
#Phase Two Starts Here

deal\_yes\_recipe = recipe(Deal\_Yes ~., train) %>%  
 step\_dummy(all\_nominal(),-all\_outcomes())  
  
deal\_yes\_model = decision\_tree() %>%   
 set\_engine("rpart", model = TRUE) %>%   
 set\_mode("classification")  
  
deal\_yes\_wflow =   
 workflow() %>%   
 add\_model(deal\_yes\_model) %>%   
 add\_recipe(deal\_yes\_recipe)  
  
deal\_yes\_fit = fit(deal\_yes\_wflow, train)

tree = deal\_yes\_fit %>%   
 pull\_workflow\_fit() %>%   
 pluck("fit")  
  
fancyRpartPlot(tree)

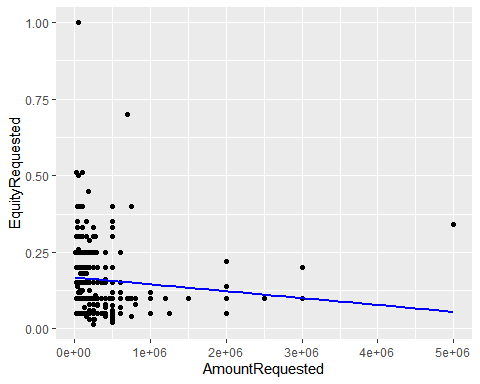


fancyRpartPlot(tree, tweak=1.25)



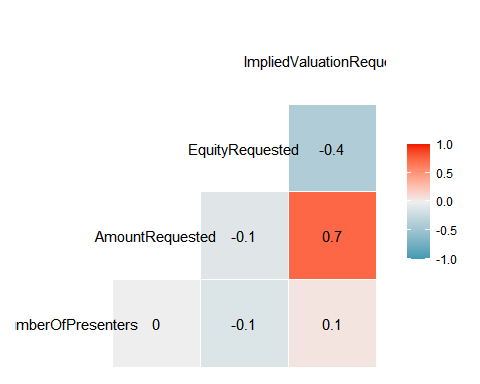
ggplot(shark\_student, aes(AmountRequested, EquityRequested)) +geom\_point() +geom\_smooth(method="lm",se=FALSE, color="blue")

## `geom\_smooth()` using formula 'y ~ x'



ggcorr(shark\_student, label=TRUE)

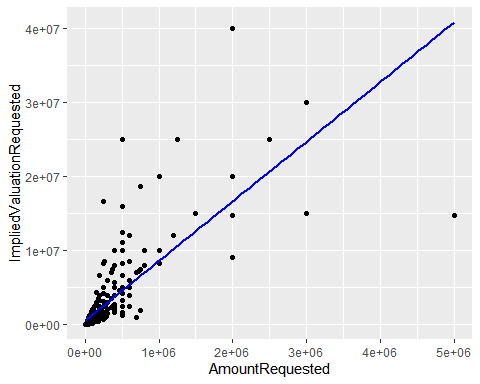
## Warning in ggcorr(shark\_student, label = TRUE): data in column(s) 'Company',  
## 'SeasonEpisode', 'ReceiveOffer', 'RejectOffer', 'Deal\_Yes', 'Deal\_No',  
## 'Eth1', 'Eth2', 'Eth3', 'Eth4', 'Eth5', 'Male1', 'Male2', 'Male3',  
## 'Female1', 'Female2', 'Female3', 'Female4', 'Novelties', 'Health\_Wellness',  
## 'Food\_Beverage', 'Business\_Services', 'Lifestyle\_Home', 'Software\_Tech',  
## 'Children\_Education', 'Automotive', 'Fashion\_Beauty', 'Media\_Entertainment',  
## 'Fitness\_Sports\_Outdoors', 'Pet\_Products', 'Travel', 'Green\_CleanTech',  
## 'Uncertain\_Other', 'MalePresenter', 'FemalePresenter', 'MixedGenderPresenters',  
## 'CompanyState', 'BarbaraCorcoran', 'MarkCuban', 'LoriGreiner', 'RobertHerjavec',  
## 'DaymondJohn', 'KevinOLeary', 'KevinHarrington', 'Guest' are not numeric and  
## were ignored



This looks like a correlation! Positive relationship between the Amount Requested and Implied Valuation, and a Negative relationship between Equity Requested and Implied Valuation.

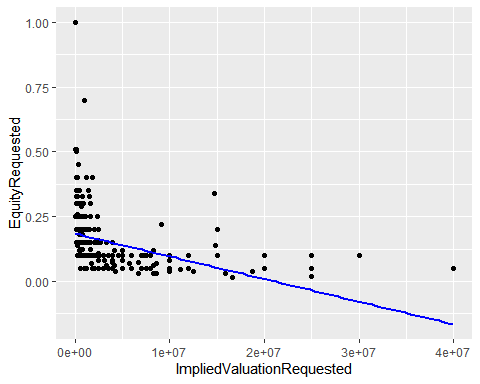
ggplot(shark\_student, aes(AmountRequested, ImpliedValuationRequested)) +geom\_point() +geom\_smooth(method="lm",se=FALSE, color="blue")

## `geom\_smooth()` using formula 'y ~ x'



ggplot(shark\_student, aes(ImpliedValuationRequested, EquityRequested)) +geom\_point() +geom\_smooth(method="lm",se=FALSE, color="blue")

## `geom\_smooth()` using formula 'y ~ x'



#Logistic Regressions

shark\_reg <- recipe(ImpliedValuationRequested ~ AmountRequested, shark\_student)  
  
lm\_model =   
 linear\_reg() %>%   
 set\_engine("lm")  
  
lm\_wflow =   
 workflow() %>%   
 add\_model(lm\_model) %>%   
 add\_recipe(shark\_reg)  
  
lm\_fit = fit(lm\_wflow, shark\_student)  
  
summary(lm\_fit$fit$fit$fit)

##   
## Call:  
## stats::lm(formula = ..y ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -26185384 -917131 -639618 -51083 23323586   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.332e+05 1.602e+05 3.328 0.000942 \*\*\*  
## AmountRequested 8.072e+00 3.370e-01 23.949 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2929000 on 476 degrees of freedom  
## Multiple R-squared: 0.5465, Adjusted R-squared: 0.5455   
## F-statistic: 573.5 on 1 and 476 DF, p-value: < 2.2e-16

#p-value is significant, r-squared is 0.5455

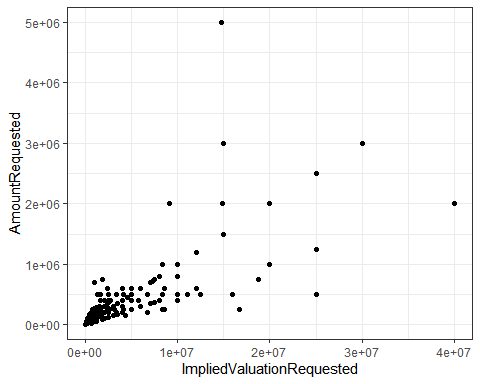
shark\_reg2 <- recipe(ImpliedValuationRequested ~ EquityRequested, shark\_student)  
  
lm\_model2 =   
 linear\_reg() %>%   
 set\_engine("lm")  
  
lm\_wflow2 =   
 workflow() %>%   
 add\_model(lm\_model2) %>%   
 add\_recipe(shark\_reg2)  
  
lm\_fit2 = fit(lm\_wflow2, shark\_student)  
  
summary(lm\_fit2$fit$fit$fit)

##   
## Call:  
## stats::lm(formula = ..y ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3940397 -1831486 -1029374 231879 35459603   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5394852 353007 15.283 <2e-16 \*\*\*  
## EquityRequested -17089104 1869032 -9.143 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4011000 on 476 degrees of freedom  
## Multiple R-squared: 0.1494, Adjusted R-squared: 0.1476   
## F-statistic: 83.6 on 1 and 476 DF, p-value: < 2.2e-16

P-value is significant, r-squared is 0.1476. This is a weak correlation and I wouldn’t use this.

A different version of the graph above without the blue line.

ggplot(shark\_student, aes(x=ImpliedValuationRequested,y=AmountRequested)) + geom\_point() + theme\_bw()



recipe3 = recipe(ImpliedValuationRequested ~ EquityRequested + NumberOfPresenters + AmountRequested, shark\_student)  
  
lm\_model3 =   
 linear\_reg() %>%   
 set\_engine("lm")  
  
lm\_wflow3 =   
 workflow() %>%   
 add\_model(lm\_model3) %>%   
 add\_recipe(recipe3)  
  
lm\_fit3 = fit(lm\_wflow3, shark\_student)  
  
  
summary(lm\_fit3$fit$fit$fit)

##   
## Call:  
## stats::lm(formula = ..y ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22237670 -1038530 -538667 209473 22226305   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.656e+06 4.049e+05 6.560 1.41e-10 \*\*\*  
## EquityRequested -1.412e+07 1.215e+06 -11.625 < 2e-16 \*\*\*  
## NumberOfPresenters 1.567e+05 1.951e+05 0.803 0.422   
## AmountRequested 7.755e+00 2.980e-01 26.021 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2579000 on 474 degrees of freedom  
## Multiple R-squared: 0.6498, Adjusted R-squared: 0.6476   
## F-statistic: 293.2 on 3 and 474 DF, p-value: < 2.2e-16

When all numeric variables are run together, we have no significance with Number of Presenters, but our R-squared goes up to 0.6476.

recipe4 = recipe(ImpliedValuationRequested ~ EquityRequested + AmountRequested, shark\_student)  
  
lm\_model4 =   
 linear\_reg() %>%   
 set\_engine("lm")  
  
lm\_wflow4 =   
 workflow() %>%   
 add\_model(lm\_model4) %>%   
 add\_recipe(recipe4)  
  
lm\_fit4 = fit(lm\_wflow4, shark\_student)  
  
  
summary(lm\_fit4$fit$fit$fit)

##   
## Call:  
## stats::lm(formula = ..y ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22128641 -1040604 -615675 251533 22293229   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.914e+06 2.461e+05 11.84 <2e-16 \*\*\*  
## EquityRequested -1.424e+07 1.206e+06 -11.80 <2e-16 \*\*\*  
## AmountRequested 7.752e+00 2.979e-01 26.02 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2578000 on 475 degrees of freedom  
## Multiple R-squared: 0.6493, Adjusted R-squared: 0.6479   
## F-statistic: 439.8 on 2 and 475 DF, p-value: < 2.2e-16

The p-value is significant and the r-sqaured stayed high at 0.6479 when taking out the Number of Presenters variable.

### Kaggle Submission Work

# Final Project Part 2

### Augi and Watson

shark\_student1 <- shark\_student1 %>%  
 mutate(BarbaraCorcoran = as\_factor(BarbaraCorcoran)) %>%  
 mutate(BarbaraCorcoran = fct\_recode(BarbaraCorcoran, "No" = "0", "Yes" = "1" ))%>%  
 mutate(MarkCuban = as\_factor(MarkCuban)) %>%  
 mutate(MarkCuban = fct\_recode(MarkCuban, "No" = "0", "Yes" = "1" ))%>%  
 mutate(LoriGreiner = as\_factor(LoriGreiner)) %>%  
 mutate(LoriGreiner = fct\_recode(LoriGreiner, "No" = "0", "Yes" = "1" ))%>%  
 mutate(RobertHerjavec = as\_factor(RobertHerjavec)) %>%  
 mutate(RobertHerjavec = fct\_recode(RobertHerjavec, "No" = "0", "Yes" = "1" ))%>%  
 mutate(DaymondJohn = as\_factor(DaymondJohn)) %>%  
 mutate(DaymondJohn = fct\_recode(DaymondJohn, "No" = "0", "Yes" = "1" ))%>%  
 mutate(KevinOLeary = as\_factor(KevinOLeary)) %>%  
 mutate(KevinOLeary = fct\_recode(KevinOLeary, "No" = "0", "Yes" = "1" ))%>%  
 mutate(KevinHarrington = as\_factor(KevinHarrington)) %>%  
 mutate(KevinHarrington = fct\_recode(KevinHarrington, "No" = "0", "Yes" = "1" ))

### SUbmission 1

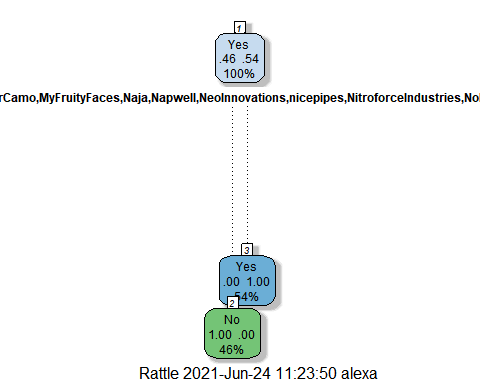
shark\_student1 <- shark\_student1 %>%  
 mutate(Guest = as\_factor(Guest)) %>%  
 mutate(Guest = fct\_recode(Guest, "No" = "0", "Yes" = "1" ))%>%   
 rename("Number\_of\_Presenters"="Number of Presenters")  
  
# View(shark\_student)  
  
shark\_student <- shark\_student %>%  
 dplyr::select(-ReceiveOffer) %>%  
 dplyr::select(-RejectOffer) %>%  
 dplyr::select(-Deal\_No)  
  
imp\_deal = mice(shark\_student, m=5, method='pmm', printFlag=FALSE)

## Warning: Number of logged events: 42

# summary(imp\_deal)  
  
deal\_complete = complete(imp\_deal)   
# summary(deal\_complete)  
  
set.seed(12345)   
shark\_split = initial\_split(deal\_complete, prop = 0.7, strata = Deal\_Yes)   
train = training(shark\_split)  
test = testing(shark\_split)  
  
shark\_recipe = recipe(Deal\_Yes ~ ., train)  
  
tree\_model = decision\_tree() %>%   
 set\_engine("rpart", model = TRUE) %>% #don't forget the model = TRUE flag  
 set\_mode("classification")  
  
shark\_wflow =   
 workflow() %>%   
 add\_model(tree\_model) %>%   
 add\_recipe(shark\_recipe)  
  
shark\_fit = fit(shark\_wflow, train)  
  
shark\_fit %>%  
 pull\_workflow\_fit() %>%  
 pluck("fit")

## n= 334   
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 334 154 Yes (0.4610778 0.5389222)   
## 2) Company=AlaskaGlacialMudCo.,AlphaM,Amber,AtlanticCandies,AttachNoted,BakersEdge,BalloonDistractions,BarbellApparel,Beardbrand,BedJet,BeneaththeInk,Bermies,besomebody,Bibbitec,BodyWalking,BoogieBoxFitness,BrewersCowIceCream,BuckMason,BufferBit,BusinessGhost,CAB-20RockBand,CampNoCounselors,CaptainIceCream,Cheekd,ChristmasTreeHugger,CocoTaps,CoffeeMeetsBagel,CollegeFoxesPackingBoxes,CopadiVino,CorksAway,CougarEnergy,CrioBru,cropsticks,DDPYoga,DoorBot,DudeRobe,Ecomowers,EcoNuts,eCreamery,ElephantChat,EllipticalStroller,EmergenSee,energybits,EssoWatches,EveDrop,FairytaleWishes,FashionTap,FatherFigure,FitDeck,FitnessStride,Five-MinuteFurniture,FortMagic,Frameri,FunCakes,GeoOrbital,GranolaGourmet,HammerNails,HargittMarineServices,HipChix,HoopMaps,InchBug,InkFlip,Intelli-Stopper,iSlide,JungleJumparoo,KwytzaChopstickArt,LaidBrand,LedgePillow,Lite-Netics,LocTote,Lumi,MagicalMoments,Moberi,MosBows,MuddyWaterCamo,MyFruityFaces,Naja,Napwell,NeoInnovations,nicepipes,NitroforceIndustries,NoFlyCone,Nootrobox,nPowerPegKineticEnergyGenerator,OfftheCob,OraOrganics,ParkerMapleFarm,PartiePoche,Pavlok,PCClassesOnline,PetPaint,petplate,PillarsofSlippers,PMSBites,PrettyPaddedRoom,ProofEyewear,PSIBands,PuppyCake,QuickstopFireSprinklerTools,RentAGoat,ReVestor,RoloDoc,RompNRoll,Salespreneur,SamsonMartin,Scan.Me,ScotteVest,ShowerPill,SimpleHabit,Slawsa,SnagaStool,SoapSox,Solemender,SoulsCalling,SpoonfulofComfort,StormStoppers,Sunnova,TaalumaTotes,Tablejacks,TheBearTheRat,TheBobblePlace,TheBouqsCompany,TheCaddyGirls,TheCookieKahuna,TheFizz,TheFloatingMugCo.,TheFunHouseProject,TheGreenGarmento,TheHomeT,TheOilerie,ThePaleoDietFoods,TheShemie,TheTwinZPillow,TheWingman,THINgloss,ThreeDayRule,Throx,TieTry,TraditionalFisheries,TrimiTanks,trippie,TwisterCleaner,UncleZipsBeefJerky,Validated,VictoriasKitchen,VictoryCoffees,Vinamor,VirtuixOmni,VirtuSphere,Wispots,WypAviation,XeroShoes,YourselfExpression,Zuvaa 154 0 No (1.0000000 0.0000000) \*  
## 3) Company=180Cup,AldoOrtaJewelry,Angellift,AquaVault,AvatheElephant,Avocaderia,Back9Dips,BantamBagels,BeatBoxBeverages,BeeSweetLemonade,biaggi,biem,BillyBlanksJr.sDanceWithMe,BodyJac,BonAffair,Boobypack,BootIllusions,BootyQueenApparel,BottleBreacher,BrazynLife,Breathometer,brightwheel,BubbasQBBQ,Caffeindicator,CateApp,ChangEd,Chilantro,ChordBuddy,Cinnaholic,CitiKitty,CocoJack,CoffeeJoulies,CoolWazoo,Coolwraps,Cordaroys,Coverplay,Dbestproducts,Detrapel,DNASimple,dollopgourmet,DriveSuits,ElementBars,EnsoRings,evREwares,FiberFix,FirefighterTurnoutBags,FirstDefenseNasalScreen,Fizzics,FreshPatch,FrillClothing,FunBites,GarageDoorLock,GobieH2O,GoodHangups,GraceLace,GreaseBags,GreaseMonkeyWipes,GreenBox,GrillCharms,Grypmat,H3oSports,HangEase,HappyFeet,HeartPup,HeidiHo,HellsBellsHelmets,HenrysHumdingers,HillbillyBrand,HoodiePillow,HotTot,HowDoYouRoll,iLumi,Inboard,Invisiplug,iReTron,IWanttoDrawaCatforYou,JacksonsHonest,JacksStandsMarketplaces,JoesGourmetFishFry,KronosGolf,Liddup,Lightfilm,LockerBones,Lollacup,luminAID,Mee-MasLouisianaGumboBrick,MenschonaBench,MilkSnob,MissionBelt,MistoBox,ModernChristmasTrees,ModMomFurniture,MonkeyMat,Mr.TodsPieFactory,NaturallyPerfectDolls,NeatCheeks,NightRunner270,nomiku,Nubrella,OAFoods,OneSole,Origaudio,PaparazziProposals,PaperBoxPilots,PetePedro,PhoneSoap,Pinblock,PinkShutter,Pipsnacks,PittMoss,Plated,PopUpPlay,PotatoParcel,PRO-NRG,ProntoBev,PupBox,Q-Flex,ReelyHookedFishCo.,Remyxx,Reviver,Ride-OnCarry-On,RockBands,RokBlok,Ruckpack,RuffleButtsandRuggedButts,RuggedManiacObstacleRace,RyansBarkery,ScratchGrainBakingCo.,ScreenMend,Seventy2,ShellBobber,ShowNo,SiliDogs,SleepingBaby,Smartwheel,SnarkyTea,SouthernCultureArtisanFoods,SpyEscapeandEvasion,SquattyPotty,Stasher,StellaValle,Sun-Staches,sunscreeenr,SurfSetFitness,SynDaverLabs,Tactibite,TeddyNeedsaBath,TekDry,TheChefinBlack,TheCutBuddy,TheDoughBar,TheDripDrop,TheLastLid,TheOriginalComfy,ThePaintedPretzel,TheSBUUnicycle,TheSleepStyler,TheSpooner,ThirdWaveWater,TippiToes,TipsyElves,TITIN,Toor,Tranquilo,Turbobaster,U-Lace,UniKeyTechnologies,VeggieMamaGardenPops,VerbalizeIt,WaiveCar,WickedGoodCupcakes,WildSquirrelNutButter,WineBalloon,WineDesign,WorldRecordStriperCompany,YouKickAss,YouSmellSoap,Yubo,ZinePak,Zipz 180 0 Yes (0.0000000 1.0000000) \*

tree = shark\_fit %>%   
 pull\_workflow\_fit() %>%   
 pluck("fit")  
fancyRpartPlot(tree)



# newdata = data.frame(shark\_student1)  
# x1 = predict(shark\_fit, newdata, type="prob")  
# x1 <- select(x1, .pred\_Yes)  
#   
# kaggle = shark\_student1 %>% rowid\_to\_column("ID") %>% select(ID)  
# kaggle = bind\_cols(kaggle, x1)  
# kaggle <- kaggle %>% mutate(.pred\_Yes = as\_factor(.pred\_Yes)) %>%  
# mutate(.pred\_Yes = fct\_recode(.pred\_Yes, "No" = "0", "Yes" = "1" )) %>%  
# rename("Deal\_Yes"=".pred\_Yes")  
#   
# write.csv(kaggle, "kaggle\_submit.csv", row.names=FALSE)

### Submission 2/3

shark\_student <- shark\_student %>% mutate\_if(is.character,as\_factor)   
shark\_student1 <- shark\_student1 %>% mutate\_if(is.character,as\_factor)  
shark\_student1 <- shark\_student1 %>%  
 dplyr::select(-SeasonEpisode)  
shark\_student1 <- shark\_student1 %>%  
 dplyr::select(-Company)  
shark\_student1 <- shark\_student1 %>%  
 dplyr::select(-X1)  
shark\_student <- shark\_student %>%  
 dplyr::select(-SeasonEpisode)  
shark\_student <- shark\_student %>%  
 dplyr::select(-Company)  
shark\_student1 <- shark\_student1 %>%  
 dplyr::select(-CompanyState)  
shark\_student <- shark\_student %>%  
 dplyr::select(-CompanyState)  
  
set.seed(1234)  
imp\_age = mice(shark\_student, m=5, method='pmm', printFlag=FALSE)

## Warning: Number of logged events: 2

shark\_complete = complete(imp\_age)   
# summary(shark\_complete)  
  
  
shark\_recipe = recipe(Deal\_Yes ~., shark\_complete) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes())  
  
rf\_model = rand\_forest() %>%   
 set\_engine("ranger") %>%   
 set\_mode("classification")  
  
shark\_wflow =   
 workflow() %>%   
 add\_model(rf\_model) %>%   
 add\_recipe(shark\_recipe)  
  
set.seed(123)  
shark\_fit = fit(shark\_wflow, shark\_complete)  
  
predRF = predict(shark\_fit, shark\_complete)  
# head(predRF)  
  
# newdata = data.frame(shark\_student1)  
# x1 = predict(shark\_fit, newdata)  
# x1 <- dplyr::select(x1, .pred\_class)  
#   
# kaggle = shark\_student1 %>% rowid\_to\_column("ID") %>% dplyr::select(ID)  
# kaggle = bind\_cols(kaggle, x1)  
# kaggle <- kaggle %>% mutate(.pred\_class = as\_factor(.pred\_class)) %>% rename("Deal\_Yes"=".pred\_class")  
#   
# write.csv(kaggle, "kaggle\_submit.csv", row.names=FALSE)

confusionMatrix(predRF$.pred\_class, shark\_complete$Deal\_Yes, positive="Yes")

## Confusion Matrix and Statistics

## Reference

## Prediction No Yes

## No 195 7

## Yes 25 251

## Accuracy : 0.9331

## 95% CI : (0.9068, 0.9538)

## No Information Rate : 0.5397

## P-Value [Acc > NIR] : < 2.2e-16

## Kappa : 0.8644

## Mcnemar's Test P-Value : 0.002654

## Sensitivity : 0.9729

## Specificity : 0.8864

## Pos Pred Value : 0.9094

## Neg Pred Value : 0.9653

## Prevalence : 0.5397

## Detection Rate : 0.5251

## Detection Prevalence : 0.5774

## Balanced Accuracy : 0.9296

## 'Positive' Class : Yes