# Project Phase 1

## Team Rafalowski-Rabil

### Exploratory Data Anlysis

#### Load Libraries

#### Setup\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Load and Clean Data

#### Summary

str(sharkstudent)

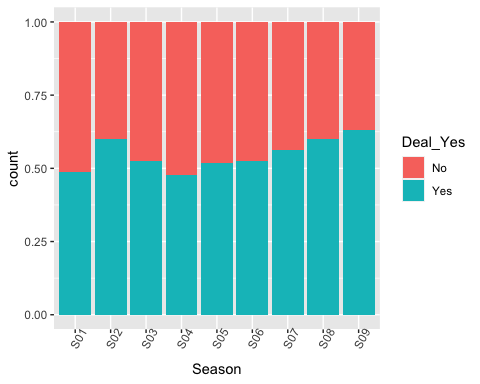
## tibble [551 × 44] (S3: tbl\_df/tbl/data.frame)  
## $ Deal\_Yes : Factor w/ 2 levels "No","Yes": 1 1 2 2 1 1 2 1 2 1 ...  
## $ Number of Presenters : num [1:551] 1 2 1 1 1 1 1 1 1 1 ...  
## $ Eth1 : Factor w/ 5 levels "No presenter 1",..: 3 5 2 3 4 3 3 3 3 5 ...  
## $ Eth2 : Factor w/ 5 levels "No presenter 2",..: 1 3 1 1 1 1 1 1 1 1 ...  
## $ Eth3 : Factor w/ 5 levels "No presenter 3",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Eth4 : Factor w/ 2 levels "No presenter 4",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Eth5 : Factor w/ 2 levels "No presenter 5",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Male1 : Factor w/ 2 levels "No","Yes": 2 2 2 1 1 2 1 2 1 1 ...  
## $ Male2 : Factor w/ 2 levels "No","Yes": 1 2 1 1 1 1 1 1 1 1 ...  
## $ Male3 : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Female1 : Factor w/ 2 levels "No","Yes": 1 1 1 2 2 1 2 1 2 2 ...  
## $ Female2 : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Female3 : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Female4 : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Novelties : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ HealthWellness : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ FoodBeverage : Factor w/ 2 levels "No","Yes": 1 1 2 1 1 1 1 1 1 1 ...  
## $ BusinessServices : Factor w/ 2 levels "No","Yes": 2 2 1 1 1 1 1 1 1 1 ...  
## $ LifestyleHome : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 1 2 2 2 ...  
## $ SoftwareTech : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ ChildrenEducation : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 1 2 1 1 1 ...  
## $ Automotive : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ FashionBeauty : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 2 1 1 1 1 ...  
## $ MediaEntertainment : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ FitnessSportsOutdoors : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ PetProducts : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Travel : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ GreenCleanTech : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ UncertainOther : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ MalePresenter : Factor w/ 2 levels "No","Yes": 2 2 2 1 1 2 1 2 1 1 ...  
## $ FemalePresenter : Factor w/ 2 levels "No","Yes": 1 1 1 2 2 1 2 1 2 2 ...  
## $ MixedGenderPresenters : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ AmountRequested : num [1:551] 1200000 250000 460000 50000 100000 200000 250000 100000 35000 150000 ...  
## $ EquityRequested : num [1:551] 0.1 0.25 0.1 0.15 0.2 0.2 0.2 0.1 0.35 0.25 ...  
## $ ImpliedValuationRequested: num [1:551] 12000000 1000000 4600000 333333 500000 ...  
## $ BarbaraCorcoran : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ MarkCuban : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ LoriGreiner : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ RobertHerjavec : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ DaymondJohn : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ KevinOLeary : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ KevinHarrington : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ Guest : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Season : Factor w/ 9 levels "S01","S02","S03",..: 1 1 1 1 1 1 1 1 1 1 ...

summary(sharkstudent)

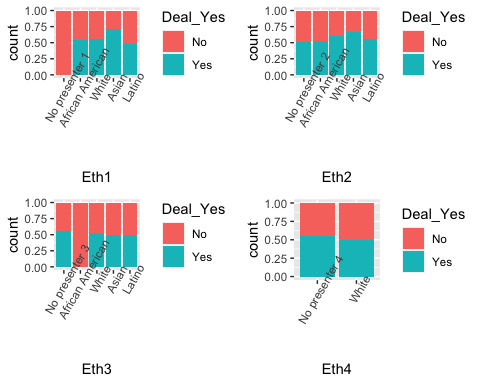
## Deal\_Yes Number of Presenters Eth1 Eth2   
## No :250 Min. :1.000 No presenter 1 : 2 No presenter 2 :285   
## Yes:301 1st Qu.:1.000 African American: 51 African American: 25   
## Median :1.000 White :417 White :173   
## Mean :1.539 Asian : 17 Asian : 18   
## 3rd Qu.:2.000 Latino : 64 Latino : 50   
## Max. :4.000   
##   
## Eth3 Eth4 Eth5 Male1   
## No presenter 3 :523 No presenter 4:549 No presenter 5:550 No :143   
## African American: 3 White : 2 White : 1 Yes:408   
## White : 19   
## Asian : 2   
## Latino : 4   
##   
##   
## Male2 Male3 Female1 Female2 Female3 Female4 Novelties  
## No :407 No :533 No :408 No :429 No :540 No :549 No :530   
## Yes:144 Yes: 18 Yes:143 Yes:122 Yes: 11 Yes: 2 Yes: 21   
##   
##   
##   
##   
##   
## HealthWellness FoodBeverage BusinessServices LifestyleHome SoftwareTech  
## No :539 No :455 No :521 No :458 No :527   
## Yes: 12 Yes: 96 Yes: 30 Yes: 93 Yes: 24   
##   
##   
##   
##   
##   
## ChildrenEducation Automotive FashionBeauty MediaEntertainment  
## No :497 No :544 No :456 No :514   
## Yes: 54 Yes: 7 Yes: 95 Yes: 37   
##   
##   
##   
##   
##   
## FitnessSportsOutdoors PetProducts Travel GreenCleanTech UncertainOther  
## No :496 No :533 No :548 No :550 No :546   
## Yes: 55 Yes: 18 Yes: 3 Yes: 1 Yes: 5   
##   
##   
##   
##   
##   
## MalePresenter FemalePresenter MixedGenderPresenters AmountRequested   
## No :217 No :421 No :464 Min. : 10000   
## Yes:334 Yes:130 Yes: 87 1st Qu.: 100000   
## Median : 150000   
## Mean : 274205   
## 3rd Qu.: 300000   
## Max. :5000000   
##   
## EquityRequested ImpliedValuationRequested BarbaraCorcoran MarkCuban  
## Min. :0.0150 Min. : 40000 No :222 No : 56   
## 1st Qu.:0.1000 1st Qu.: 500000 Yes:329 Yes:495   
## Median :0.1500 Median : 1000000   
## Mean :0.1576 Mean : 2829148   
## 3rd Qu.:0.2000 3rd Qu.: 2500000   
## Max. :1.0000 Max. :40000000   
##   
## LoriGreiner RobertHerjavec DaymondJohn KevinOLeary KevinHarrington Guest   
## No :155 No : 38 No :191 No : 22 No :503 No :461   
## Yes:396 Yes:513 Yes:360 Yes:529 Yes: 48 Yes: 90   
##   
##   
##   
##   
##   
## Season   
## S07 : 87   
## S05 : 83   
## S06 : 78   
## S04 : 71   
## S08 : 65   
## S09 : 65   
## (Other):102

#### Visualizations\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

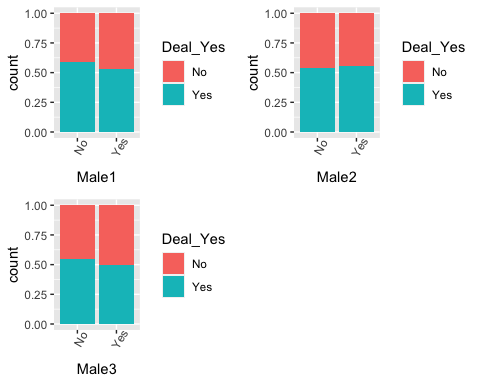
ggplot(sharkstudent, aes(x = Season, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))



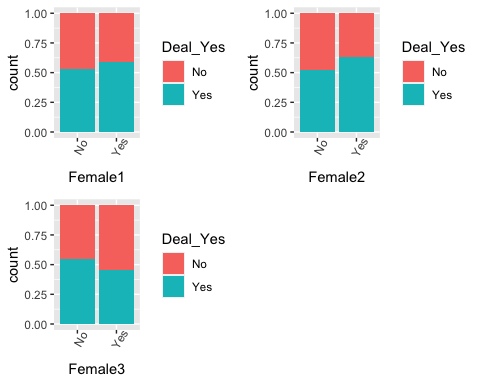
p1 = ggplot(sharkstudent, aes(x = Eth1, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))   
p2 = ggplot(sharkstudent, aes(x = Eth2, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))   
p3 = ggplot(sharkstudent, aes(x = Eth3, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))   
p4 = ggplot(sharkstudent, aes(x = Eth4, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))   
grid.arrange(p1,p2,p3,p4,ncol=2)



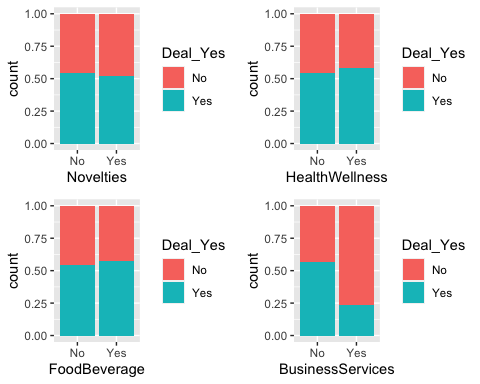
p1 = ggplot(sharkstudent, aes(x = Male1, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))   
p2 = ggplot(sharkstudent, aes(x = Male2, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))   
p3 = ggplot(sharkstudent, aes(x = Male3, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))  
grid.arrange(p1,p2,p3, ncol=2)



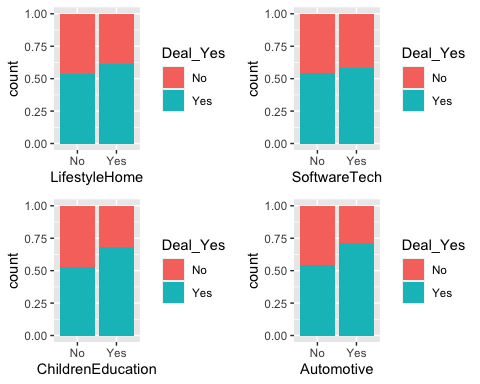
p1 = ggplot(sharkstudent, aes(x = Female1, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))   
p2 = ggplot(sharkstudent, aes(x = Female2, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))   
p3 = ggplot(sharkstudent, aes(x = Female3, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60))  
grid.arrange(p1,p2,p3, ncol=2)



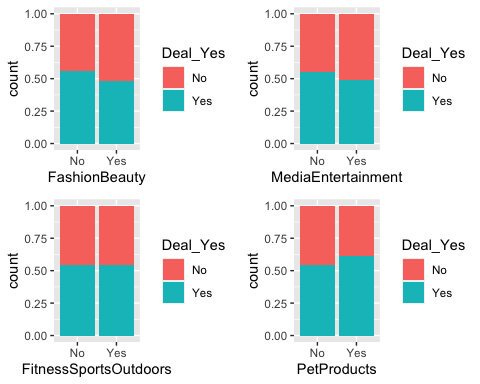
p1 = ggplot(sharkstudent, aes(x = Novelties, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p2 = ggplot(sharkstudent, aes(x = `HealthWellness`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p3 = ggplot(sharkstudent, aes(x = `FoodBeverage`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p4 = ggplot(sharkstudent, aes(x = `BusinessServices`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



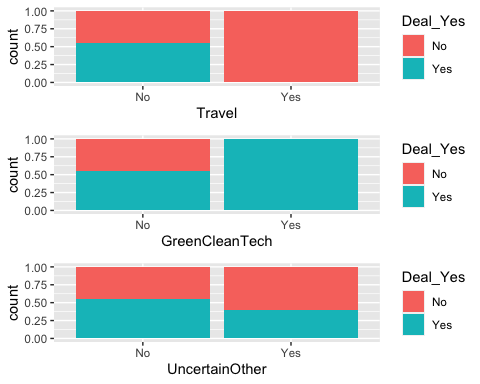
p1 = ggplot(sharkstudent, aes(x = `LifestyleHome`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p2 = ggplot(sharkstudent, aes(x = `SoftwareTech`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p3 = ggplot(sharkstudent, aes(x = `ChildrenEducation`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p4 = ggplot(sharkstudent, aes(x = `Automotive`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



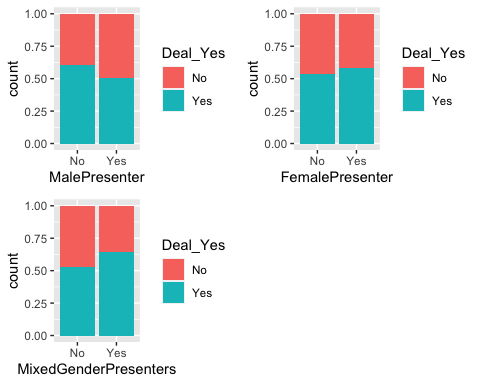
p1 = ggplot(sharkstudent, aes(x = `FashionBeauty`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p2 = ggplot(sharkstudent, aes(x = `MediaEntertainment`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p3 = ggplot(sharkstudent, aes(x = `FitnessSportsOutdoors`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p4 = ggplot(sharkstudent, aes(x = `PetProducts`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



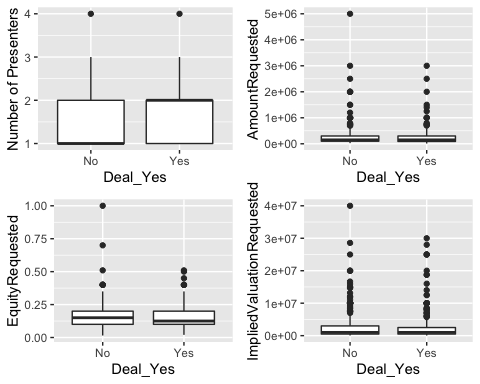
p1 = ggplot(sharkstudent, aes(x = Travel, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p2 = ggplot(sharkstudent, aes(x = `GreenCleanTech`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p3 = ggplot(sharkstudent, aes(x = `UncertainOther`, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3)



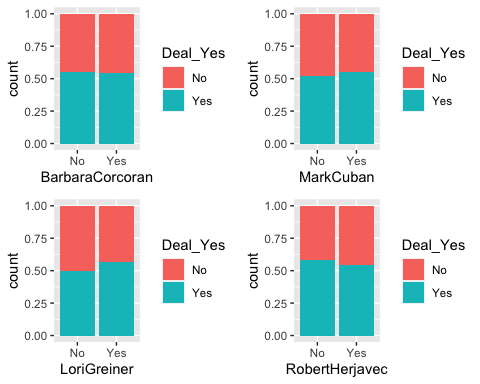
p1 = ggplot(sharkstudent, aes(x = MalePresenter, fill = Deal\_Yes)) + geom\_bar(position="fill")  
p2 = ggplot(sharkstudent, aes(x = FemalePresenter, fill = Deal\_Yes)) + geom\_bar(position="fill")  
p3 = ggplot(sharkstudent, aes(x = MixedGenderPresenters, fill = Deal\_Yes)) + geom\_bar(position="fill")  
grid.arrange(p1,p2,p3, ncol=2)



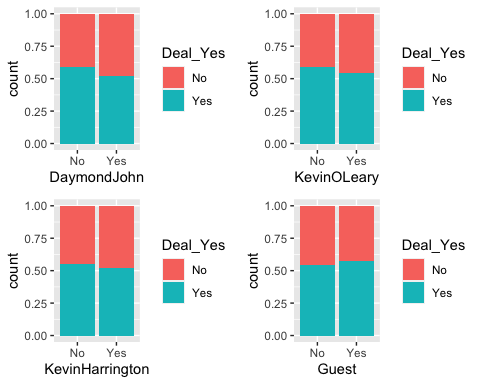
p1 = ggplot(sharkstudent, aes(x = Deal\_Yes, y = `Number of Presenters`)) + geom\_boxplot()  
p2 = ggplot(sharkstudent, aes(x = Deal\_Yes, y = AmountRequested)) + geom\_boxplot()  
p3 = ggplot(sharkstudent, aes(x = Deal\_Yes, y = EquityRequested)) + geom\_boxplot()  
p4 = ggplot(sharkstudent, aes(x = Deal\_Yes, y = ImpliedValuationRequested)) + geom\_boxplot()  
grid.arrange(p1,p2,p3,p4, ncol = 2)



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

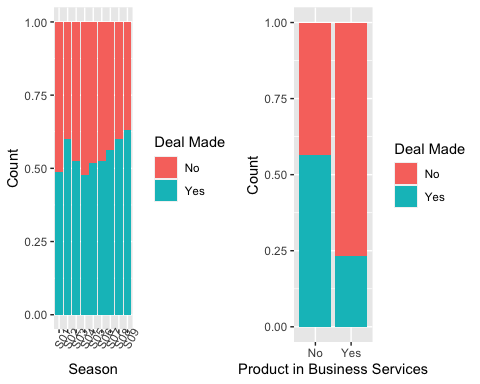


p1 = ggplot(sharkstudent, aes(x = DaymondJohn, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p2 = ggplot(sharkstudent, aes(x = KevinOLeary, fill = Deal\_Yes)) + geom\_bar(position = "fill")  
p3 = ggplot(sharkstudent, aes(x = KevinHarrington, fill = Deal\_Yes)) + geom\_bar(position = "fill")  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

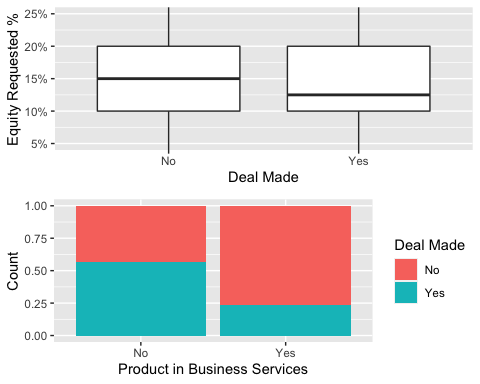


#### Final Visualizations\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

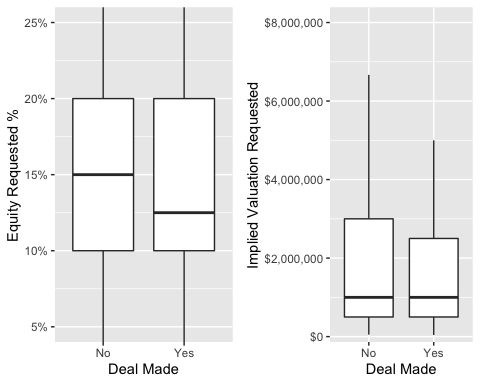
p1=ggplot(sharkstudent, aes(x = Season, fill = Deal\_Yes)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle =60)) + labs( y= "Count") + theme(plot.title = element\_text(hjust = 0.5,face="bold"))+ scale\_fill\_discrete(name="Deal Made")   
  
p2=ggplot(sharkstudent, aes(x = `BusinessServices`, fill = Deal\_Yes)) +   
 geom\_bar(position = "fill") +  
 labs(y = "Count", x = "Product in Business Services", fill = "Deal Made") +  
 theme(plot.title = element\_text(hjust = 0.5,face="bold"))  
  
grid.arrange(p1,p2,ncol=2)



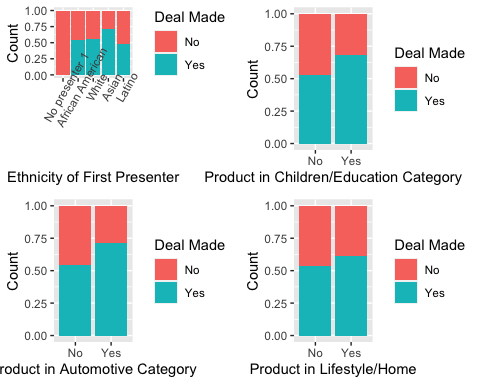
p1=ggplot(sharkstudent, aes(x = Deal\_Yes, y = EquityRequested)) + #   
 geom\_boxplot(outlier.shape = NA) +  
 coord\_cartesian(ylim = quantile(sharkstudent$EquityRequested, c(0.1, 0.9))) +  
 scale\_y\_continuous(labels = scales::percent\_format(accuracy = 1)) +  
 labs(y = "Equity Requested %", x = "Deal Made") +  
 theme(plot.title = element\_text(hjust = 0.5,face="bold"))  
grid.arrange(p1,p2)



p2=ggplot(sharkstudent, aes(x = Deal\_Yes, y = ImpliedValuationRequested)) + geom\_boxplot(outlier.shape = NA)+   
 coord\_cartesian(ylim = quantile(sharkstudent$ImpliedValuationRequested, c(0.1, 0.9))) + scale\_y\_continuous(labels=scales::dollar\_format()) + labs( x="Deal Made", y= "Implied Valuation Requested") + theme(plot.title = element\_text(hjust = 0.5,face="bold"))  
  
grid.arrange(p1,p2,ncol=2)



p1 = ggplot(sharkstudent, aes(x = Eth1, fill = Deal\_Yes)) + geom\_bar(position = "fill") +   
 geom\_bar(position = "fill") +  
 labs(y = "Count", x = "Ethnicity of First Presenter", fill = "Deal Made") +  
 theme(plot.title = element\_text(hjust = 0.5,face="bold")) + theme(axis.text.x = element\_text(angle =60))   
  
p2 = ggplot(sharkstudent, aes(x = `ChildrenEducation`, fill = Deal\_Yes)) + geom\_bar(position = "fill")+   
 geom\_bar(position = "fill") +  
 labs(y = "Count", x = "Product in Children/Education Category", fill = "Deal Made") +  
 theme(plot.title = element\_text(hjust = 0.5,face="bold"))  
  
p3 =ggplot(sharkstudent, aes(x = Automotive, fill = Deal\_Yes)) + geom\_bar(position = "fill") +   
 geom\_bar(position = "fill") +  
 labs(y = "Count", x = "Product in Automotive Category", fill = "Deal Made") +  
 theme(plot.title = element\_text(hjust = 0.5,face="bold"))  
  
p4 =ggplot(sharkstudent, aes(x = LifestyleHome, fill = Deal\_Yes)) + geom\_bar(position = "fill") +   
 geom\_bar(position = "fill") +  
 labs(y = "Count", x = "Product in Lifestyle/Home", fill = "Deal Made") +  
 theme(plot.title = element\_text(hjust = 0.5,face="bold"))  
  
grid.arrange(p1,p2,p3,p4,ncol=2)



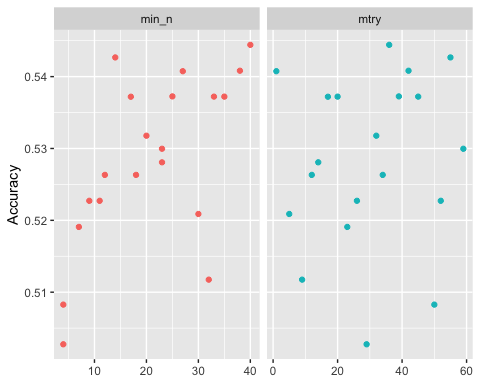
#### Random Forest\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

set.seed(123)  
rf\_folds = vfold\_cv(sharkstudent, v = 5)

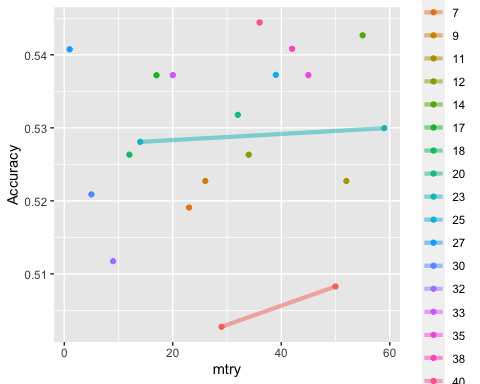
set.seed(123)  
shark\_recipe = recipe(Deal\_Yes ~., sharkstudent) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes())  
 # step\_other()  
 # step\_novel()  
  
rf\_model = rand\_forest(mtry = tune(), min\_n = tune(), trees = 100) %>% #add tuning of mtry and min\_n parameters  
 #setting trees to 100 here should also speed things up a bit, but more trees might be better  
 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)  
rf\_res = tune\_grid(  
 shark\_wflow,  
 resamples = rf\_folds,  
 grid = 20 #try 20 different combinations of the random forest tuning parameters  
)

## i Creating pre-processing data to finalize unknown parameter: mtry

#Plot for Tuning  
rf\_res %>%  
 collect\_metrics() %>%  
 filter(.metric == "accuracy") %>%  
 select(mean, min\_n, mtry) %>%  
 pivot\_longer(min\_n:mtry,  
 values\_to = "value",  
 names\_to = "parameter"  
 ) %>%  
 ggplot(aes(value, mean, color = parameter)) +  
 geom\_point(show.legend = FALSE) +  
 facet\_wrap(~parameter, scales = "free\_x") +  
 labs(x = NULL, y = "Accuracy")



rf\_res %>%  
 collect\_metrics() %>%  
 filter(.metric == "accuracy") %>%  
 mutate(min\_n = factor(min\_n)) %>%  
 ggplot(aes(mtry, mean, color = min\_n)) +  
 geom\_line(alpha = 0.5, size = 1.5) +  
 geom\_point() +  
 labs(y = "Accuracy")



#### Random Forest Tuning\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#RANDOM FOREST  
set.seed(123)  
rf\_folds = vfold\_cv(sharkstudent, v = 5)  
  
set.seed(123)  
shark\_recipe = recipe(Deal\_Yes ~., sharkstudent) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes())  
 # step\_other()  
 # step\_novel()  
  
rf\_model = rand\_forest(mtry = tune(), min\_n = tune(), trees = 100) %>% #add tuning of mtry and min\_n parameters  
 #setting trees to 100 here should also speed things up a bit, but more trees might be better  
 set\_engine("ranger", importance = "permutation") %>% #added importance metric  
 set\_mode("classification")  
  
shark\_wflow =   
 workflow() %>%   
 add\_model(rf\_model) %>%   
 add\_recipe(shark\_recipe)  
  
  
rf\_grid = grid\_regular(  
 mtry(range = c(2, 30)), #these values determined through significant trial and error  
 min\_n(range = c(15, 25)), #these values determined through significant trial and error  
 levels = 5  
)  
  
set.seed(123)  
rf\_res\_tuned = tune\_grid(  
 shark\_wflow,  
 resamples = rf\_folds,  
 grid = rf\_grid #use the tuning grid  
)

#### Random Forest Var of Importance\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

best\_rf = select\_best(rf\_res\_tuned, "accuracy")  
  
final\_rf = finalize\_workflow(  
 shark\_wflow,  
 best\_rf  
)  
  
final\_rf

## ══ Workflow ════════════════════════════════════════════════════════════════════  
## Preprocessor: Recipe  
## Model: rand\_forest()  
##   
## ── Preprocessor ────────────────────────────────────────────────────────────────  
## 1 Recipe Step  
##   
## • step\_dummy()  
##   
## ── Model ───────────────────────────────────────────────────────────────────────  
## Random Forest Model Specification (classification)  
##   
## Main Arguments:  
## mtry = 9  
## trees = 100  
## min\_n = 22  
##   
## Engine-Specific Arguments:  
## importance = permutation  
##   
## Computational engine: ranger

set.seed(123)  
final\_rf\_fit = fit(final\_rf, sharkstudent)  
  
set.seed(123)  
final\_rf\_fit %>% pull\_workflow\_fit() %>% vip(geom = "point", mapping=aes\_string(fill="Variable")) + labs(title="Random Forest Variables of Importance", y="Importance", x="Variable") + theme(plot.title = element\_text(hjust = 0.5,face="bold"))

