Prediction Challenge 1

By: Raashi Maheshwari

Overall Summary/Breakdown of DataSet

```
> table(HireTrainApr10[HireTrainApr10$Coding == "Excellent" & HireTrainApr10$Hired=='Yes',])
, , Major = CS, College = BestCollege, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = DataScience, College = BestCollege, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = IT, College = BestCollege, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = Stats, College = BestCollege, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = CS, College = BYU, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = DataScience, College = BYU, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = IT, College = BYU, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = Stats, College = BYU, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = CS, College = Peters, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0
```

Overall Summary/Breakdown of DataSet Conti.

, , Major = DataScience, College = Peters, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

, , Major = IT, College = Peters, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

, , Major = Stats, College = Peters, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

, , Major = CS, College = PJIT, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

, , Major = DataScience, College = PJIT, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

, , Major = IT, College = PJIT, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

, , Major = Stats, College = PJIT, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

, , Major = CS, College = Redbrick, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

, , Major = DataScience, College = Redbrick, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

, , Major = IT, College = Redbrick, Hired = No

Coding	Impression			
	Confident	Nerdy	Outgoing	Shy
Excellent	0	0	0	0
OK	0	0	0	0
Weak	0	0	0	0

Overall Summary/Breakdown of DataSet Conti.

```
, , Major = Stats, College = Redbrick, Hired = No

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         0     0        0   0
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = CS, College = BestCollege, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent        10     5        8   3
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = DataScience, College = BestCollege, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent        11     8       12   9
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = IT, College = BestCollege, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         8     0       11   1
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = Stats, College = BestCollege, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         8    10        7   5
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = CS, College = BYU, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent        12    13        9  10
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = DataScience, College = BYU, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         5     8        7   4
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = IT, College = BYU, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         7     9        7   7
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = Stats, College = BYU, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent        11     5        7   5
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = CS, College = Peters, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         9     5       11   8
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = DataScience, College = Peters, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         8    12       14   5
  OK                0     0        0   0
  Weak              0     0        0   0
```

Overall Summary/Breakdown of DataSet Conti.

```
, , Major = IT, College = Peters, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         5    12        9  11
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = Stats, College = Peters, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent        13     9        6   7
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = CS, College = PJIT, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         5     9        6  10
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = DataScience, College = PJIT, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent        11     1        7   8
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = IT, College = PJIT, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         8     9        4   6
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = Stats, College = PJIT, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent        10     8        9   7
  OK                0     0        0   0
  Weak              0     0        0   0
```

```
, , Major = CS, College = Redbrick, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         7     6        7   8
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = DataScience, College = Redbrick, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         4     9       11  13
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = IT, College = Redbrick, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent         6     3        9   7
  OK                0     0        0   0
  Weak              0     0        0   0

, , Major = Stats, College = Redbrick, Hired = Yes

           Impression
Coding      Confident Nerdy Outgoing Shy
  Excellent        12     6       13   6
  OK                0     0        0   0
  Weak              0     0        0   0
```

Analysis

- The previous four slides contain 40 tables that give us extremely detailed breakdown of the dataset
- Broken down by majors and universities as well as impression and coding, the tables are very helpful in understanding what factors play an important role in getting a person hired.
- They are very precise in giving us exact numbers that help us understand what factors play an important role in getting a person hired.
- We only need the information on what gets a person hired, hence the code I wrote was so that we only get the tables that give us the information on what factors play a role in getting people hired. The tables do not give us any statistical information about the people who were not hired. We do not know the coding skills, impression, major or university for the people that were not hired as the tables for those who are not hired is just all zeros. However, not having this information doesn't impact us as much because we want to figure out what GETS people HIRED.

Coding Skills vs Hired

```
> table(HireTrainApr10$Coding == "Excellent" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1369   631
> table(HireTrainApr10$Coding == "Weak" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1931    69
> table(HireTrainApr10$Coding == "OK" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1445   555
>
```

- The tables to the left show us that although we may have thought that coding skills play a huge role in deciding who gets hired, coding skills may not be the ONLY huge factor in decision making
- The tables do show, however, that not a lot of people who have “weak” coding skills get hired. Candidates with “excellent” or “OK” coding skills are given preference to be hired

Impression vs Hired

```
> table(HireTrainApr10$Impression == "Confident" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1679   321
> table(HireTrainApr10$Impression == "Nerdy" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1677   323
> table(HireTrainApr10$Impression == "Outgoing" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1675   325
> table(HireTrainApr10$Impression == "Shy" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1714   286
>
```

- The tables to the left show us that although we may have thought that a first impression matters to the recruiting/hiring manager, it doesn't necessarily mean that you will or will not be hired.
- The tables show that regardless of the initial impression the candidate made, the hiring decision was not based solely on whether the applicant was confident, nerdy, outgoing or shy
- However, we can notice that candidates who were "Confident", "Nerdy" and "Outgoing" were more likely to be hired than candidates who were "shy"

College vs Hired

```
> table(HireTrainApr10$College == "BestCollege" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1816   184
> table(HireTrainApr10$College == "BYU" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1737   263
> table(HireTrainApr10$College == "Peters" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1732   268
> table(HireTrainApr10$College == "PJIT" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1767   233
> table(HireTrainApr10$College == "Redbrick" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1693   307
>
```

- We can see, similarly to the other attributes, that college is not the sole thing that gets a candidate hired.
- However, it can be seen that there is a lower number of candidates that are hired from BestCollege versus having a slightly higher number of candidates that get hired from the remaining colleges (BYU, Peters, PJIT, and Redbrick)
- However, the numbers are not significantly too apart hence there is a chance that college does not impact the hiring decision

Major vs Hired

```
> table(HireTrainApr10$Major == "CS" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1668   332
> table(HireTrainApr10$Major == "DataScience" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1687   313
> table(HireTrainApr10$Major == "IT" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1702   298
> table(HireTrainApr10$Major == "Stats" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1688   312
```

- Major usually plays a big role in the hiring decision however, the tables to the left show us that the number of candidates that got hired from all four majors is relatively the same across all four majors (CS, DataScience, IT, and Stats)
- This shows us that major doesn't necessarily have a very big impact on the hiring decision.

Analysis

- As seen previously, none of the individual factors (coding, impression, major and college) seem to have a huge impact individually on the hiring decision
- This leads me to believe that a combination of these factors is what will play a role in deciding who gets hired versus who doesn't
- Using this knowledge, I made more tables regarding a combination of these factors, which allowed me to understand what combination of factors actually make an impact in who gets hired.

BestCollege vs Major vs Hired

```
> table(HireTrainApr10[HireTrainApr10$College == "BestCollege" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 51  44
> table(HireTrainApr10[HireTrainApr10$College == "BestCollege" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 66  35
> table(HireTrainApr10[HireTrainApr10$College == "BestCollege" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 51  54
> table(HireTrainApr10[HireTrainApr10$College == "BestCollege" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 50  51
>
```

BYU vs Major vs Hired

```
> table(HireTrainApr10[HireTrainApr10$College == "BYU" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 33  84
> table(HireTrainApr10[HireTrainApr10$College == "BYU" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 35  67
> table(HireTrainApr10[HireTrainApr10$College == "BYU" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 41  58
> table(HireTrainApr10[HireTrainApr10$College == "BYU" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 42  54
>
```

Peters vs Major vs Hired

```
> table(HireTrainApr10[HireTrainApr10$College == "Peters" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 32  70
> table(HireTrainApr10[HireTrainApr10$College == "Peters" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 39  62
> table(HireTrainApr10[HireTrainApr10$College == "Peters" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 30  65
> table(HireTrainApr10[HireTrainApr10$College == "Peters" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 28  71
>
```

PJIT vs Major vs Hired

```
> table(HireTrainApr10[HireTrainApr10$College == "PJIT" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 32  61
> table(HireTrainApr10[HireTrainApr10$College == "PJIT" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 34  55
> table(HireTrainApr10[HireTrainApr10$College == "PJIT" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 28  61
> table(HireTrainApr10[HireTrainApr10$College == "PJIT" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 32  56
>
```

Redbrick vs Major vs Hired

```
> table(HireTrainApr10[HireTrainApr10$College == "Redbrick" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 33  78
> table(HireTrainApr10[HireTrainApr10$College == "Redbrick" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 27  79
> table(HireTrainApr10[HireTrainApr10$College == "Redbrick" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 31  77
> table(HireTrainApr10[HireTrainApr10$College == "Redbrick" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 30  73
>
```

Analysis

- As seen in the previous four slides, I made tables of each major in each school to see how many people got hired for each section.
- We can see that the number of people that got hired from BestCollege is lower as compared to the other four colleges, which leads me to believe that there are not a lot of candidates that get hired from BestCollege.
- For the remaining colleges, the number of people hired remain relatively close, showing me that out of the remaining four colleges, where the candidate went does not make that big of an impact.
- However, one key thing that is crucial to be noticed is the fact that for each college, all the majors hired seem to have relatively close numbers, showing me that the major of the candidate didn't really make a difference in whether the candidate was hired or not
- After focusing on these two aspects, I could notice that these didn't seem to make too much of an impact on the hiring decision, which lead me to believe that the coding skills and the impression is what makes the most impact.
- As seen from the forty tables in the beginning, I noticed the trends between coding skills and impressions, as that made the most impact. I saw that weak coding skills didn't usually get people hired while "Excellent" and "OK" skills did. I also noticed that "Shy" people tended to not get hired, while "Outgoing", "Nerdy" and "Confident" got hired. Using this information (from the forty tables in the beginning) and this information (major and college doesn't have a great impact on hiring decision), I created the decision vector that can be seen on the next slide.

Cross Validation

```
> myprediction <- HireTrainApr10
>
> decision <- rep('No', nrow(myprediction))
> decision[myprediction$Impression == "Confident" & myprediction$Coding == "Excellent"] <- 'Yes'
>
> decision[myprediction$Impression == "Outgoing" & myprediction$Coding == "Excellent"] <- 'Yes'
>
>
> decision[myprediction$Impression == "Nerdy" & myprediction$Coding == "Excellent"] <- 'Yes'
>
> decision[myprediction$Impression == "Shy" & myprediction$Coding == "Excellent"] <- 'Yes'
>
> decision[myprediction$Impression == "Outgoing" & myprediction$Coding == "OK"] <- 'Yes'
> decision[myprediction$Impression == "Nerdy" & myprediction$Coding == "OK"] <- 'Yes'
>
> decision[myprediction$Impression == "Confident" & myprediction$Coding == "OK"] <- 'Yes'
>
> decision[myprediction$Impression == "Shy" & myprediction$Coding == "OK"] <- 'Yes'
> decision[myprediction$Impression == "Confident" & myprediction$Coding == "Weak"] <- 'No'
>
> decision[myprediction$Impression == "Shy" & myprediction$Coding == "Weak"] <- 'No'
>
> decision[myprediction$Impression == "Outgoing" & myprediction$Coding == "Weak"] <- 'No'
>
> decision[myprediction$Impression == "Nerdy" & myprediction$Coding == "Weak"] <- 'No'
>
> myprediction$Hired <- decision
>
> error <- mean(HireTrainApr10$Hired != myprediction$Hired)
>
> error
[1] 0.103
>
```

- Using the idea that impression and coding make the most impact on getting a candidate hired, I came up with the decision vector seen on the left.
- Testing these conditions/checks on the testing dataset provided to us gave me an error value of 10.3% which is not the best, however, it gives us a general idea of what gets a person hired which is what we were looking for in the first place

Kaggle Submission Code

```
> test_challenge1 <- read.csv("C:/Users/msraa/Downloads/test_challenge1.csv", stringsAsFactors=TRUE)
>   View(test_challenge1)
> sample_submission_challenge1 <- read.csv("C:/Users/msraa/Downloads/sample_submission_challenge1.csv", stringsAsFactors=TRUE)
>   View(sample_submission_challenge1)
> test <- test_challenge1
> submission <- sample_submission_challenge1
> myprediction <- test
> decision <- rep('No', nrow(myprediction))
>
> decision[myprediction$Impression == "Confident" & myprediction$Coding == "Excellent"] <- 'Yes'
>
> decision[myprediction$Impression == "Outgoing" & myprediction$Coding == "Excellent"] <- 'Yes'
>
> decision[myprediction$Impression == "Nerdy" & myprediction$Coding == "Excellent"] <- 'Yes'
>
> decision[myprediction$Impression == "Shy" & myprediction$Coding == "Excellent"] <- 'Yes'
> decision[myprediction$Impression == "Outgoing" & myprediction$Coding == "OK"] <- 'Yes'
>
> decision[myprediction$Impression == "Nerdy" & myprediction$Coding == "OK"] <- 'Yes'
>
> decision[myprediction$Impression == "Confident" & myprediction$Coding == "OK"] <- 'Yes'
>
> decision[myprediction$Impression == "Shy" & myprediction$Coding == "OK"] <- 'Yes'
>
> decision[myprediction$Impression == "Confident" & myprediction$Coding == "Weak"] <- 'No'
>
> decision[myprediction$Impression == "Shy" & myprediction$Coding == "Weak"] <- 'No'
>
> decision[myprediction$Impression == "Outgoing" & myprediction$Coding == "Weak"] <- 'No'
>
> decision[myprediction$Impression == "Nerdy" & myprediction$Coding == "Weak"] <- 'No'
>
```

Kaggle Submission Continued

```
> submission$Prediction <- decision
> submission
     Id Prediction
1     1        Yes
2     2         No
3     3         No
4     4        Yes
5     5         No
6     6        Yes
7     7        Yes
8     8        Yes
9     9         No
10   10         No
11   11         No
12   12        Yes
13   13         No
14   14        Yes
15   15        Yes
16   16         No
17   17        Yes
18   18        Yes
19   19         No
20   20        Yes
21   21         No
22   22         No
23   23         No
24   24        Yes
25   25        Yes
26   26         No
```

To create the submission file, I used the following command:

```
> write.csv(submission, "submission.csv", row.names = FALSE)
```

Lines continue in a similar fashion

(listing all the predictions for 500 rows)

Analysis: All Code

```
> table(HireTrainApr10[HireTrainApr10$Coding == "Excellent" & HireTrainApr10$Hired=='Yes',])
```

```
> table(HireTrainApr10$Coding == "Excellent" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1369   631
> table(HireTrainApr10$Coding == "Weak" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1931    69
> table(HireTrainApr10$Coding == "OK" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1445   555
>
```

Analysis: All Code Continued

```
> table(HireTrainApr10$Impression == "Confident" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1679   321
> table(HireTrainApr10$Impression == "Nerdy" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1677   323
> table(HireTrainApr10$Impression == "Outgoing" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1675   325
> table(HireTrainApr10$Impression == "Shy" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1714   286
>
```

```
> table(HireTrainApr10$College == "BestCollege" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1816   184
> table(HireTrainApr10$College == "BYU" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1737   263
> table(HireTrainApr10$College == "Peters" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1732   268
> table(HireTrainApr10$College == "PJIT" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1767   233
> table(HireTrainApr10$College == "Redbrick" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1693   307
>
```

Analysis: All Code Continued

```
> table(HireTrainApr10$Major == "CS" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1668   332
> table(HireTrainApr10$Major == "DataScience" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1687   313
> table(HireTrainApr10$Major == "IT" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1702   298
> table(HireTrainApr10$Major == "Stats" & HireTrainApr10$Hired == "Yes")

FALSE  TRUE
 1688   312
```

```
> table(HireTrainApr10[HireTrainApr10$College == "BestCollege" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 51  44
> table(HireTrainApr10[HireTrainApr10$College == "BestCollege" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 66  35
> table(HireTrainApr10[HireTrainApr10$College == "BestCollege" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 51  54
> table(HireTrainApr10[HireTrainApr10$College == "BestCollege" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 50  51
>
```

Analysis: All Code Continued

```
> table(HireTrainApr10[HireTrainApr10$College == "BYU" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 33  84
> table(HireTrainApr10[HireTrainApr10$College == "BYU" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 35  67
> table(HireTrainApr10[HireTrainApr10$College == "BYU" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 41  58
> table(HireTrainApr10[HireTrainApr10$College == "BYU" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 42  54
>
```

```
> table(HireTrainApr10[HireTrainApr10$College == "Peters" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 32  70
> table(HireTrainApr10[HireTrainApr10$College == "Peters" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 39  62
> table(HireTrainApr10[HireTrainApr10$College == "Peters" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 30  65
> table(HireTrainApr10[HireTrainApr10$College == "Peters" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 28  71
>
```

Analysis: All Code Continued

```
> table(HireTrainApr10[HireTrainApr10$College == "PJIT" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 32  61
> table(HireTrainApr10[HireTrainApr10$College == "PJIT" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 34  55
> table(HireTrainApr10[HireTrainApr10$College == "PJIT" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 28  61
> table(HireTrainApr10[HireTrainApr10$College == "PJIT" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 32  56
>
```

```
> table(HireTrainApr10[HireTrainApr10$College == "Redbrick" & HireTrainApr10$Major =='CS',]$Hired)

 No Yes
 33  78
> table(HireTrainApr10[HireTrainApr10$College == "Redbrick" & HireTrainApr10$Major =='IT',]$Hired)

 No Yes
 27  79
> table(HireTrainApr10[HireTrainApr10$College == "Redbrick" & HireTrainApr10$Major =='Stats',]$Hired)

 No Yes
 31  77
> table(HireTrainApr10[HireTrainApr10$College == "Redbrick" & HireTrainApr10$Major =='DataScience',]$Hired)

 No Yes
 30  73
>
```

Analysis: All Code Continued

```
> myprediction <- HireTrainApr10
>
> decision <- rep('No', nrow(myprediction))
> decision[myprediction$Impression == "Confident" & myprediction$Coding == "Excellent"] <- 'Yes'
>
> decision[myprediction$Impression == "Outgoing" & myprediction$Coding == "Excellent"] <- 'Yes'
>
>
> decision[myprediction$Impression == "Nerdy" & myprediction$Coding == "Excellent"] <- 'Yes'
>
> decision[myprediction$Impression == "Shy" & myprediction$Coding == "Excellent"] <- 'Yes'
>
> decision[myprediction$Impression == "Outgoing" & myprediction$Coding == "OK"] <- 'Yes'
> decision[myprediction$Impression == "Nerdy" & myprediction$Coding == "OK"] <- 'Yes'
>
> decision[myprediction$Impression == "Confident" & myprediction$Coding == "OK"] <- 'Yes'
>
> decision[myprediction$Impression == "Shy" & myprediction$Coding == "OK"] <- 'Yes'
> decision[myprediction$Impression == "Confident" & myprediction$Coding == "Weak"] <- 'No'
>
> decision[myprediction$Impression == "Shy" & myprediction$Coding == "Weak"] <- 'No'
>
> decision[myprediction$Impression == "Outgoing" & myprediction$Coding == "Weak"] <- 'No'
>
> decision[myprediction$Impression == "Nerdy" & myprediction$Coding == "Weak"] <- 'No'
>
> myprediction$Hired <- decision
>
> error <- mean(HireTrainApr10$Hired != myprediction$Hired)
>
> error
[1] 0.103
>
```