IST 687, Applied Data Science

Syracuse University (SU)

March 2019

**TRAINING EVALUATION**

**Applied Data Science Project Manuscript**

Team 1 Members:

Cinnamon Johnson, Team Lead

Elia Kostyrka

Eric Rodgers

Christopher Veizaga

Table of Contents

[1. INTRODUCTION 1](#_Toc3480359)

[**1.1.**  **Project Background and Description** 2](#_Toc3480360)

[2. PROJECT SCOPE AND CONTEXT OF ANALYSIS 2](#_Toc3480361)

[**2.1.** **Business Questions** 2](#_Toc3480362)

[**2.2.** **Recommendations** 2](#_Toc3480363)

[**2.3.** **Analysis by Business Question** 3](#_Toc3480364)

[**2.3.1.** **Question 1. What was the compliance rate by Regional Office?** 3](#_Toc3480365)

[**2.3.2.** **Question 2. What was the average pre-test and post-test score by Regional Office?** 5](#_Toc3480366)

[**2.3.3.** **Question 3. How much learning occurred overall?** 8](#_Toc3480367)

[**2.3.4.** **Question 4. Which Regional Offices required the most remediation?** 10](#_Toc3480368)

[**2.3.5.** **Question 5. What job categories scored higher or lower on the pre-test?** 11](#_Toc3480369)

[**2.3.6.** **Question 6: What is the overall sentiment of training?** 14](#_Toc3480370)

[3. DATA COLLECTION 16](#_Toc3480371)

[4. DATA CLEANING AND TRANSFORMATION 16](#_Toc3480372)

[5. DATA MODELING AND ANALYSIS 18](#_Toc3480373)

[**5.1.** **Use of Linear Regression** 18](#_Toc3480374)

[6. APPENDICES 20](#_Toc3480375)

[**6.1.** **Data Dictionary** 20](#_Toc3480376)

# **1. INTRODUCTION**

## **1.1. Project Background and Description**

The final data project for the Applied Data Science course (IST 687, SU) required hands-on application of R code to import, clean, manipulate, analyze, and visualize data. R studio is a free and open-source integrated development environment (IDE) for R, providing a “free software environment for statistical computing and graphics,” and was used to complete the analysis for this project. The project requirements include using a data set with a minimum of 10,000 observations and five variables to answer at least five business questions. This manuscript provides the steps taken to meet these requirements.

# **2. PROJECT SCOPE AND CONTEXT OF ANALYSIS**

The team identified a training evaluation data set containing the results from two tests on the same subject, administered in November and December 2018 to an undisclosed audience. The tests aimed to identify participants who were unfamiliar with a new electronic management system in order to provide remediation in the form of self-paced training, and then evaluate if learning occurred through the administration of a posttest. The team identified key findings that helped answer business questions and potentially assist management to determine areas of special interest for follow-on action.

## **2.1. Business Questions**

Five key business questions that this manuscript addresses are as follows:

* **Question 1:** What was the compliance rate by Regional Office?
* **Question 2:** What was the average pre-test and post-test score by Regional Office?
* **Question 3:** How much learning occurred overall?
* **Question 4:** Which regional offices required the most remediation?
* **Question 5:** What job categories scored higher or lower on the pre-test?
* **Question 6:** What is the overall sentiment of training?

## **2.2. Recommendations**

Based on the analysis of training evaluation for the months of November and December 2018 combined, overall recommendations to management include:

1. **Compliance:** A renewed focus on the regional sites that have a compliance rate below 70%, particularly the sites whose compliance was lower than 25 percent. As the data shows, some of these sites include Indianapolis, San Juan, Cheyene/Denver, Providence, San Diego, Baltimore, Philadelphia, Anchorage, St. Petersburg, AMO, VACO, Wilmington, and Washington D.C. 50% was chosen as the threshold because this is a feasible goal that can be achieved within the next fiscal year.
2. **Test Scores by Regional Office:** Evaluation of test scores by Regional Office shows that Washington, D.C, Indianapolis, Boston, and Wilmington had negative or no improvement after going through remediation. There should be a renewed focus on these areas to determine why the training was not helpful and what can be tweaked or changed to benefit these regions.
3. **Remediation:** Inspecting claims involving EMS procedures processed prior to November 2018 at regional offices that required the most remediation, as there may be a high number of errors in required actions, including those at the San Juan, central office, appeals office, New York, Little Rock, Albuquerque, and Los Angeles offices. Review claims processed prior to November 2018 at offices that required the least remediation to glean possible best practices of required actions, including Fargo, Wilmington, Sioux Falls, and Milwaukee.
4. **Test Scores by Job Category:** Pretest and Post-test scores across job titles moved up once training had been completed ranging between a 22% and 33% difference. In some cases, such as with Post VSR, the lowest scoring job title made the biggest jump in average scores from pretest to post-test scores but still remained the lowest scoring job title. Job Titles with average scores of 80% or above (Pre VSR(81.4%), PreAQRS(85.4%), Post AQRS(82.1%)) should be evaluated further to better understand what could have attributed to overall

better scores from Pretest to Post-test.

1. **Overall Sentiment.** The overall sentiment of the training was overwhelmingly high; however, there was a higher than expected negative sentiment ratio of 22% on whether a participant would recommend the training. Management should work to address frequently used words to why a participant would not recommend training which suggests the relevancy of the tool between post-development positions and medical exams.

## **2.3. Analysis by Business Question**

### **2.3.1. Question 1. What was the compliance rate by Regional Office?**

The compliance rate represents the amount of participants who scored a 100% on the pre-test or (if necessary) went through remediation and scored 100% on the posttest. Only scoring a 100% on the pretest or posttest marks successful completion of the program.

The compliance rate per region was calculated by finding the percentage of participants who failed either test:

TeamDataCV = select(TeamData, 7, 24)

library("stringr")

# Transforming study status to ones and zeros

TeamDataCV$Compliant <- ifelse(TeamDataCV$'Study Status' ==  
"Complete",1,0)

# Calculate compliance rate by Regional Office

TeamDataCV$Compliant <- as.numeric(TeamDataCV$Compliant)

CompRate <- aggregate(TeamDataCV$Compliant ~ TeamDataCV$`Use  
Regional Office `, FUN = mean)

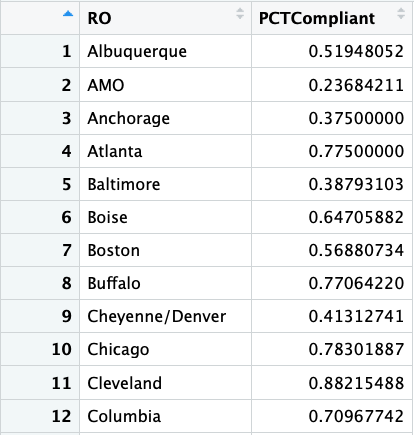
This calculation was then input into a new data frame showing the compliance rate (“PCTCompliant”) per Regional Office (sample of resultant data frame provided):

# Create new df and rename columns

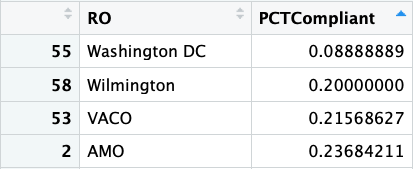
TeamDataCV2 <- data.frame(CompRate)

colnames(TeamDataCV2) <- c("RO ”,"PCTCompliant")

View(TeamDataCV2)



According to the dataset, 42% of the 59 regional offices met the 70% compliance rate for the tests administered in November and December. Only four offices held compliance rates below 25% (Washington DC, Wilmington, the Central Office, and Appeals Office).



The following code produces a graph of the compliance rates by region:

> library("ggplot2")

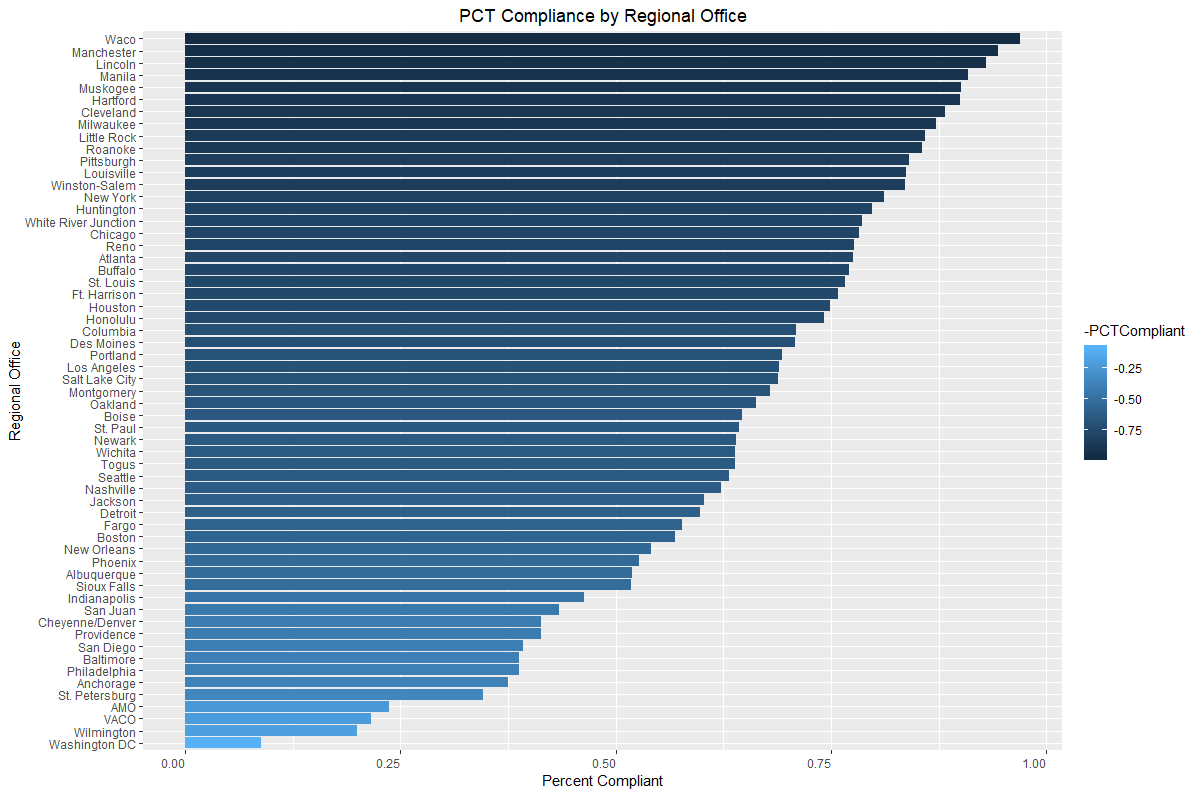
> TeamDataCV2\_bar <- ggplot(TeamDataCV2,aes(x=reorder(RO,PCTCompliant),y=PCTCompliant,fill=-PCTCompliant))

> TeamDataCV2\_bar <- TeamDataCV2\_bar + geom\_col() + coord\_flip()

> TeamDataCV2\_bar <- TeamDataCV2\_bar + ggtitle("PCT Compliance by Regional Office")+ theme (axis.text.x=element\_text(angle=0,hjust=1))

> TeamDataCV2\_bar <- TeamDataCV2\_bar + theme (plot.title=element\_text(hjust=0.5))+xlab("Regional Office")+ylab("Percent Compliant")

> TeamDataCV2\_bar



### **2.3.2. Question 2. What was the average pre-test and post-test score by Regional Office?**

Using the tapply function, average pretest and posttest scores were calculated by region. The following code produced a table that calculated the pretest and posttest scores by region, and the difference between these scores. The “Improvement” column gives an idea of how well different regions responded to remediation:

#get rid of NAs

df.pretest <- df[!is.na(df$`Pretest Score`)]

df.posttest <- df[!is.na(df$`Posttest Score Attempt 1`)]

#average pretest scores by region

pretest.by.region <- data.frame(tapply(df.pretest$`Pretest Score`,df.pretest$`Use Regional Office`, mean))

setDT(pretest.by.region, keep.rownames = TRUE)

colnames(pretest.by.region) <- c("Region","Average Pre-test Score")

#average postest scores by region

posttest.by.region <- data.frame(tapply(df.posttest$`Posttest Score Attempt 1`,df.posttest$`Use Regional Office`, mean))

setDT(posttest.by.region, keep.rownames = TRUE)

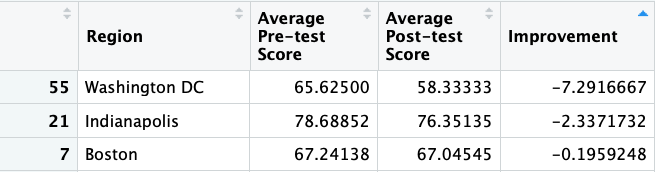
colnames(posttest.by.region) <- c("Region","Average Post-test Score")

tests.by.region <- cbind(pretest.by.region, posttest.by.region$`Average Post-test Score`)

colnames(tests.by.region) <- c("Region","Average Pre-test Score","Average Post-test Score")

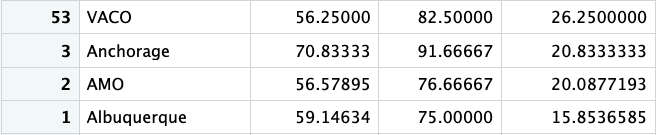
tests.by.region$Improvement <- tests.by.region$`Average Post-test Score` - tests.by.region$`Average Pre-test Score`

Three regions – Washington, D.C; Indianapolis; and Boston – showed negative improvement:

  
  
Wilmington was the only region to show no improvement:



Four regions – VACO; Anchorage; AMO; and Albuquerque – responded very well to remediation, showing an average improvement above 15 points:



The following code produces a bar graph of the improvement in scores by region:

#graph results

scores\_region <- ggplot(tests.by.region, aes(x=Region, y=Improvement))

scores\_region <- scores\_region + geom\_col()

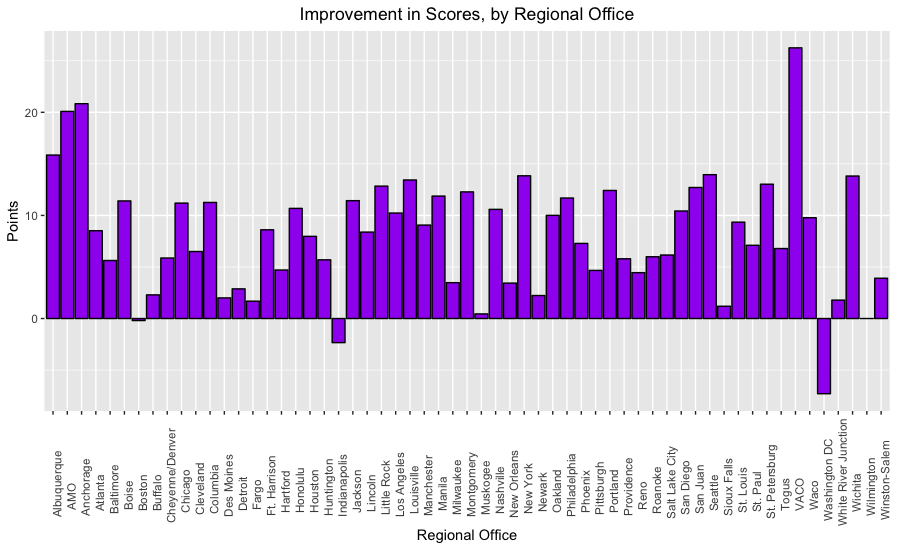
scores\_region <- scores\_region + ggtitle("Improvement in Scores, by Regional Office") + theme (axis.text.x=element\_text(angle=90,hjust=0.2))

scores\_region <- scores\_region + theme (plot.title=element\_text(hjust=0.5))+xlab("Regional Office")+ylab("Points")

scores\_region <- scores\_region + guides(fill=guide\_legend(title="Point Change"))

scores\_region <- scores\_region + geom\_bar(stat="identity", fill="purple", colour="black")

scores\_region



In addition to looking at average pretest and posttest scores by region, the scores were also calculated according to level of experience. Participants were asked to identify their years of experience at the job and to place themselves in one of the following categories: Less than 1 year; 1-2 years; 2-5 years; 5-10 years; or 10 or more years. In addition, they could choose to leave the question unanswered.

The following code produces a data frame showing how well individuals scored on the pretest according to their experience level:

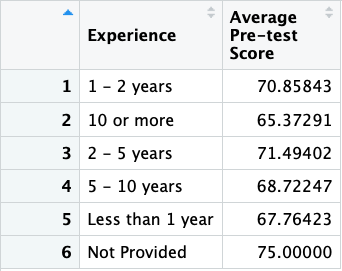
av.exp <- data.frame(tapply(df.pretest$`Pretest Score`, df.pretest$`Demographic Question 3`, mean))

av.exp <- setDT(av.exp, keep.rownames = TRUE)

colnames(av.exp) <- c("Experience","Average Pre-test Score")

av.exp <- av.exp[-1,]

av.exp



The following code produces a histogram of the average pretest scores by experience level:

scores\_experience <- ggplot(av.exp, aes(x=Experience, y=av.exp$`Average Pre-test Score`, fill=av.exp$`Average Pre-test Score`))

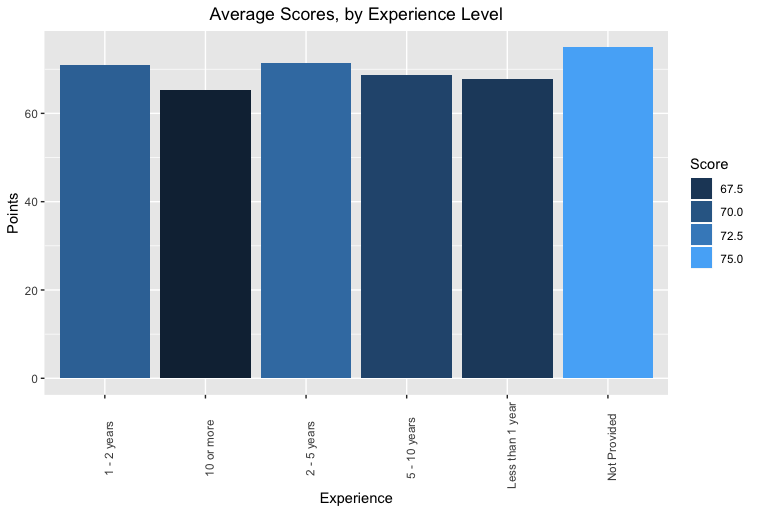
scores\_experience <- scores\_experience + geom\_col()

scores\_experience <- scores\_experience + ggtitle("Average Scores, by Experience Level") + theme (axis.text.x=element\_text(angle=90,hjust=0.4))

scores\_experience <- scores\_experience + theme (plot.title=element\_text(hjust=0.5))+xlab("District")+ylab("Points")

scores\_experience <- scores\_experience + guides(fill=guide\_legend(title="Score"))

scores\_experience



As shown by the histogram above, those who did not provide an answer to this question scored best on the pretest (with an average of 75 points) while those who claimed the most experience (10 or more years) fared worst (with an average score of about 65.4 points). This result may seem unintuitive at first. However, when thinking about how employees typically respond to new technology, the picture makes more sense. Employees with more seniority are used to doing things the old way, and are typically much more resistant to learning new systems/processes, whereas newer employees experience much less confusion and emotional upheaval over such a change.

### **2.3.3. Question 3. How much learning occurred overall?**

Training was successful in increasing test scores by 23 points for all participants who required remediation. The average pretest score by Regional Office was 55.61 and the average posttest score was 78.18. The following code shows the average pretest and posttest scores, and the average learning that occurred:

> TeamDataSlim = select(TeamData, 5,6,24)

> TeamDataSlimmer <- na.omit(TeamDataSlim)

> TeamDataSlimmer

> any(is.na(TeamDataSlimmer)) #checks for NA.

[1] FALSE

> AvgPretest <- mean(TeamDataSlimmer$'Pretest Score')

> AvgPretest

[1] 55.61235

> AvgPosttest <- mean(TeamDataSlimmer$`Posttest Score Attempt 1 `)

> AvgPosttest

[1] 78.17586

> AverageLearning <- AvgPosttest - AvgPretest

> AverageLearning

[1] 22.56352

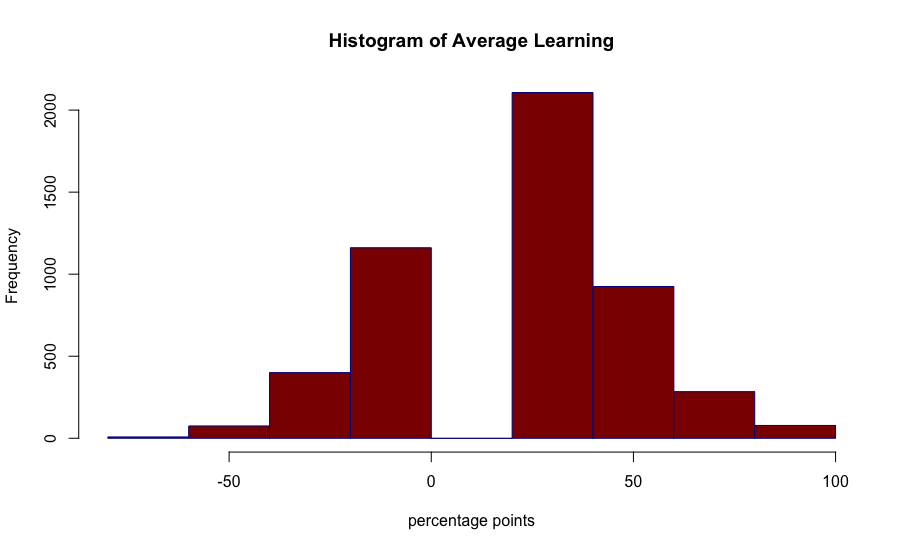
As each question is worth 25 points, most participants increased their score by correctly answering one additional test question.

The following code produces the histogram of average learning.

> Q1 <- hist(TeamDataSlimmer$PCTChange, nclass = 7, plot = FALSE)

> plot(Q1, labels = TRUE)

> plot(Q1, border = "darkblue", col = "darkred", main = "Histogram of Average Learning", xlab = "percentage points")



### **2.3.4. Question 4. Which Regional Offices required the most remediation?**

Remediation represents the amount of participants who failed to score a 100% on the pre-test and thus were required to go through the self-paced training.

The remediation per region was calculated by finding the percentage of participants who failed the pretest:

> library("dplyr")

> TeamDataek = select(TeamData, 4, 5, 24)

> # Clean data, remove NAs and standardize Regional Office names

> library("stringr")

> TeamDataek<-na.omit(TeamDataek)

> # Create categorical "Failed" column

> TeamDataek$Failed <- ifelse(TeamDataek[,2]<100,"1","0")

> View(TeamDataek)

> # Calculate percentage of participants that failed (scored less than 100%) by Regional Office

> TeamDataek$Failed <- as.numeric(TeamDataek$Failed)

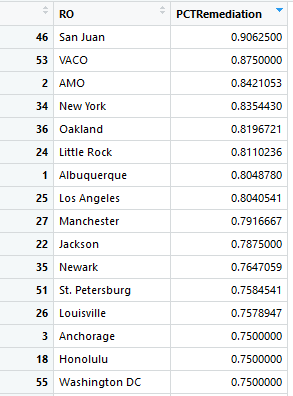
> Remediation<-aggregate(TeamDataek$'Failed' ~ TeamDataek$'Use Regional Office ', FUN = mean)

This calculation was then input into a new data frame showing the remediation percentage (“PCTRemediation”) per Regional Office (sample of resultant data frame provided):

> # Create new df and rename columns

> TeamDataekQ1 <- data.frame(Remediation)

> colnames(TeamDataekQ1) <- c("RO","PCTRemediation")



Eight Regional Offices – San Juan, the Central Office, the Appeals Office, New York, Oakland, Little Rock, Albuquerque, and Los Angeles – appeared to be the least familiar with the study topic prior to taking the pretest and required remediation. One Regional Office stood out as very familiar with the topic (Fargo).

The following code produces a graph of the compliance rates by region:

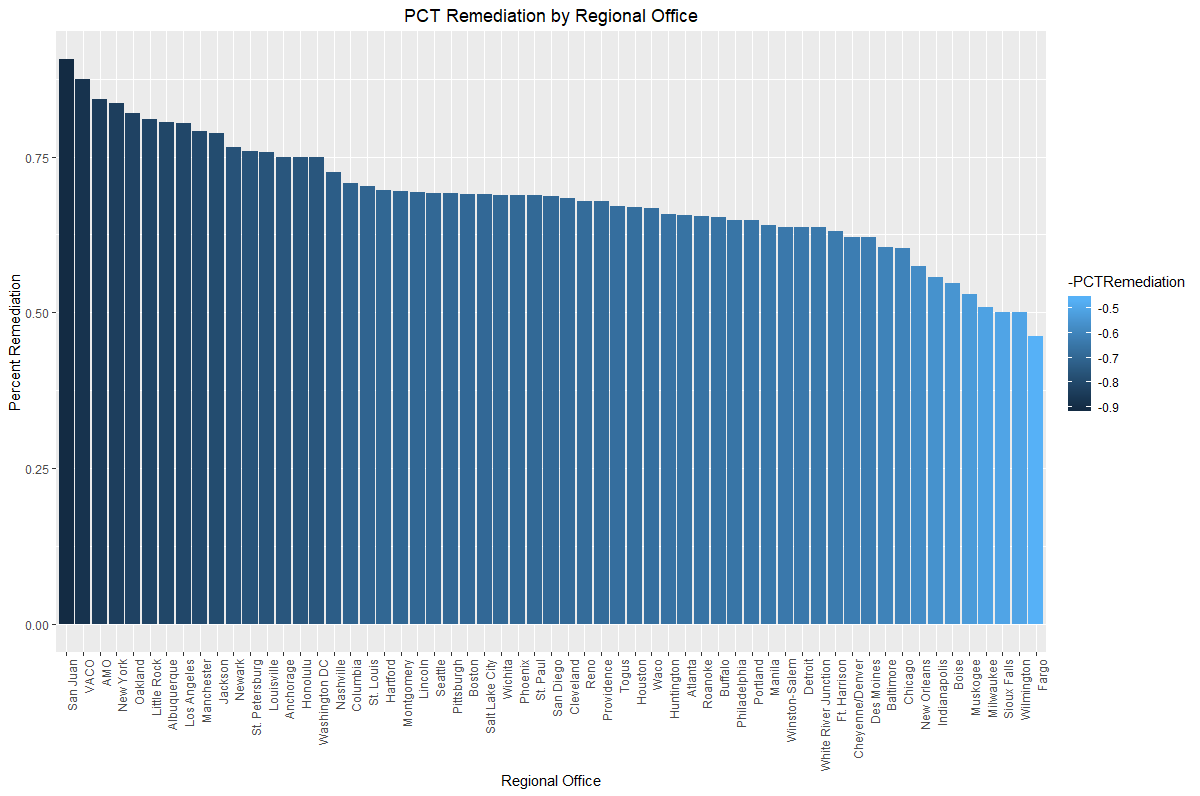
> TeamDataekQ1\_bar <- ggplot(TeamDataekQ1,aes(x=reorder(RO,-PCTRemediation),y=PCTRemediation,fill=-PCTRemediation))

> TeamDataekQ1\_bar <- TeamDataekQ1\_bar + geom\_col()

> TeamDataekQ1\_bar <- TeamDataekQ1\_bar + ggtitle("PCT Remediation by Regional Office")+ theme (axis.text.x=element\_text(angle=90,hjust=1))

> TeamDataekQ1\_bar <- TeamDataekQ1\_bar + theme (plot.title=element\_text(hjust=0.5))+xlab("Regional Office")+ylab("Percent Remediation")

> TeamDataekQ1\_bar



### **2.3.5. Question 5. What job categories scored higher or lower on the pre-test?**

There are seven job titles within the data set, two of which are categorized as “Not Provided” and “Other.” These two ambiguous categories were left in the dataset as test scores for both the pretest and posttest were averaged out per job title, and not as a complete aggregate for the data set. If scores were combined and weighted, it would have made more sense to clean the data of the Job Title categories “Not Provided” and “Other.” Two individual tables were created to clearly show the differences in scores between job titles for both pretest and posttest and then combined. For both pretest and posttest scores, the job title Pre AQRS had the highest average score and Post VSR had the lowest scores for both pretest and post-test.

> # Filter Data by selecting particular columns for business question

> TeamDataSlim = select(TeamData, 26, 24, 25, 5,6)

> # Omit NA’s, assigning new variable to TeamDataSlimmer

> TeamDataSlimmer <- na.omit(TeamDataSlim)

> View(TeamDataSlimmer)

> # New dataframe (TeamDataSlimmer), changed column names

> newdataframe <- data.frame(TeamDataSlimmer)

> colnames(newdataframe) <- c("Learner ID", "Region", "Job Title", "Pretest Score","Posttest score")

> View(newdataframe)

> # New variable assignment, filtering out only pretest scores as a tapply summary

> PretestScores <- data.frame(tapply(newdataframe$`Pretest Score`, list(newdataframe$`Job Title`) ,mean))

> # Changed data frame into data table to maintain rownames

> PretestScores = as.data.table(PretestScores, keep.rownames = TRUE)

> colnames(PretestScores) <- c("Job Title", "Pre-test Scores Averages")

> View(PretestScores)

> # New variable assignment, filtering out only posttest scores as a tapply summary

> PosttestScores <- data.frame(tapply(newdataframe$`Posttest score`, list(newdataframe$`Job Title`) ,mean))

> # Changed data frame into data table to maintain rownames

> PosttestScores = as.data.table(PosttestScores, keep.rownames = TRUE)

> colnames(PosttestScores) <- c("Job Title", "Post-test Scores Averages")

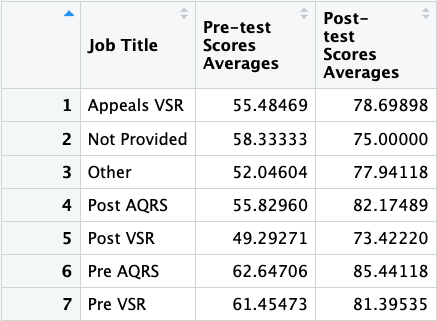
> View(PosttestScores)

> # New variable assignment, binding pretest and posttest scores, deleting duplicate column

> tests.by.jobtitle <- cbind(PretestScores, PosttestScores)

> tests.by.jobtitle <- tests.by.jobtitle[,-3]

> View(tests.by.jobtitle)



The following code produces a histogram showing the pretest scores by Job Title:

> JobTitlePreTestScore\_bar <- ggplot(tests.by.jobtitle,aes(x=reorder(tests.by.jobtitle$`Job Title`,tests.by.jobtitle$`Pre-test Scores`),y=tests.by.jobtitle$`Pre-test Scores`,fill=-tests.by.jobtitle$`Pre-test Scores`))

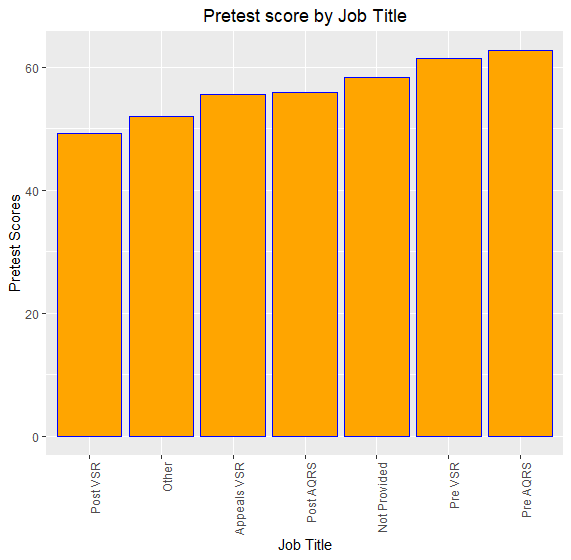
> JobTitlePreTestScore\_bar <- JobTitlePreTestScore\_bar + geom\_col(fill = "orange", color = "blue")

> JobTitlePreTestScore\_bar <- JobTitlePreTestScore\_bar + ggtitle("Pretest score by Job Title")+ theme (axis.text.x=element\_text(angle=90,hjust=1))

> JobTitlePreTestScore\_bar <- JobTitlePreTestScore\_bar + theme (plot.title=element\_text(hjust=0.5))+xlab("Job Title")+ylab("Pretest Scores")

> JobTitlePreTestScore\_bar <- JobTitlePreTestScore\_bar + labs(fill = "Pretest Scores")

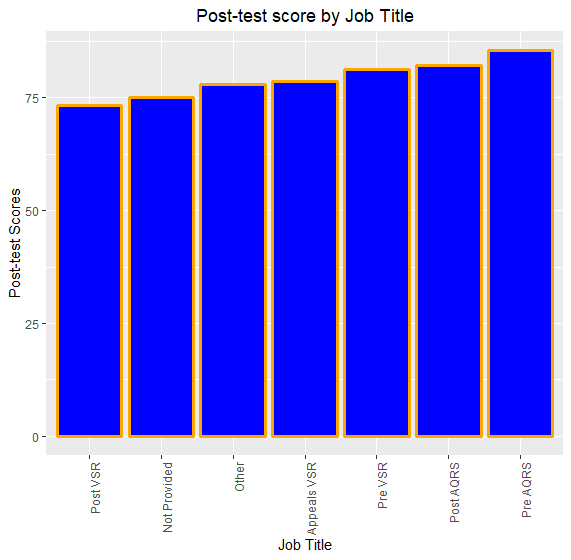
> JobTitlePreTestScore\_bar



The following code produces a histogram showing the posttest scores by Job Title:

> # Jobtitle by PosttestScore ggplot to show pretest scores by job title

> JobTitlePostTestScore\_bar <- ggplot(tests.by.jobtitle,aes(x=reorder(tests.by.jobtitle$`Job Title`,tests.by.jobtitle$`Post-test Scores`),y=tests.by.jobtitle$`Post-test Scores`,fill=-tests.by.jobtitle$`Post-test Scores`))  
> JobTitlePostTestScore\_bar <- JobTitlePostTestScore\_bar + geom\_col(fill = "blue", color = "orange", size = 1.25)  
> JobTitlePostTestScore\_bar <- JobTitlePostTestScore\_bar + ggtitle("Post-test score by Job Title")+ theme (axis.text.x=element\_text(angle=90,hjust=1))  
> JobTitlePostTestScore\_bar <- JobTitlePostTestScore\_bar + theme (plot.title=element\_text(hjust=0.5))+xlab("Job Title")+ylab("Post-test Scores")  
> JobTitlePostTestScore\_bar<- JobTitlePostTestScore\_bar + labs(fill = "Posttest Scores")  
> JobTitlePostTestScore\_bar



### 

### **2.3.6. Question 6: What is the overall sentiment of training?**

The total sentiment ratio of whether a participant would recommend training to their peers was overwhelmingly positive (120%); however the negative sentiment was slightly higher than expected, 22%. Frequently used words to why a participant would NOT recommend training suggests the relevancy of the tool between post-development positions and medical exams.

> # Positive words matched

> matched <- match(words,pos,nomatch=0)

> mCounts <- wordCounts[which(matched!=0)]

> length(mCounts)

[1] 142

> mwords <- names(mCounts)

> nPos <- sum(mCounts)

> nPos

[1] 2245

> # Negative words matched

> matched <- match(words,neg,nomatch=0)

> mCounts2 <- wordCounts[which(matched!=0)]

> length(mCounts2)

[1] 119

> mwords <- names(mCounts2)

> nNeg <- sum(mCounts2)

> nNeg

[1] 412

> totalWords <- length(words)

> ratioPos <- nPos/totalWords

> ratioPos

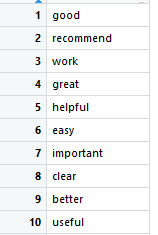
[1] 1.202464

> ratioNeg <- nNeg/totalWords

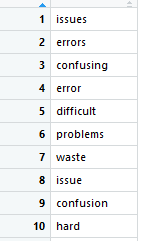
> ratioNeg

[1] 0.2206749

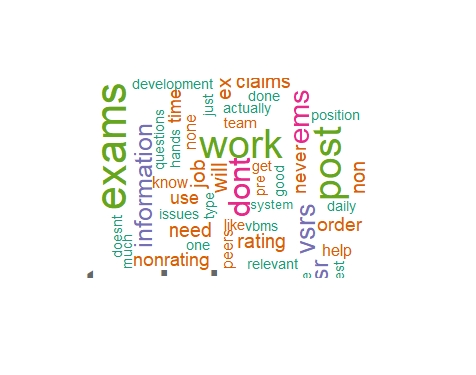
Example Positive Words:



Example Negative Words:



The following word cloud shows that the highest reason for **NOT** recommending the training included words associated with post-development positions and work or work-relevancy of the topic with medical exams.

****

# **3. DATA COLLECTION**

A total of 5,804 participants were assigned to take the exam in November 2018 and 5,975 participants were assigned to the exam in December 2018, totaling 11,779 observations for the dataset. Tests were administered through a learning management system. Both centered on one topic – the new electronic management system. A final learner transcript containing learner activity that included score per module (pre-test module, training module, and post-test module) was pulled one day after the studies were closed. The data was summarized in a roll-up that included one row per learner, their user profile information, associated test scores, and demographic information (job title, years of experience, and assigned regional office). In addition, participants were offered an opportunity to complete a post-evaluation survey after completing the study through a link an external survey link. Post-evaluation survey data was pulled three days after the study was closed. The data was merged with learner information from the roll-up. The merged roll-up file was used for this analysis.

# **4. DATA CLEANING AND TRANSFORMATION**

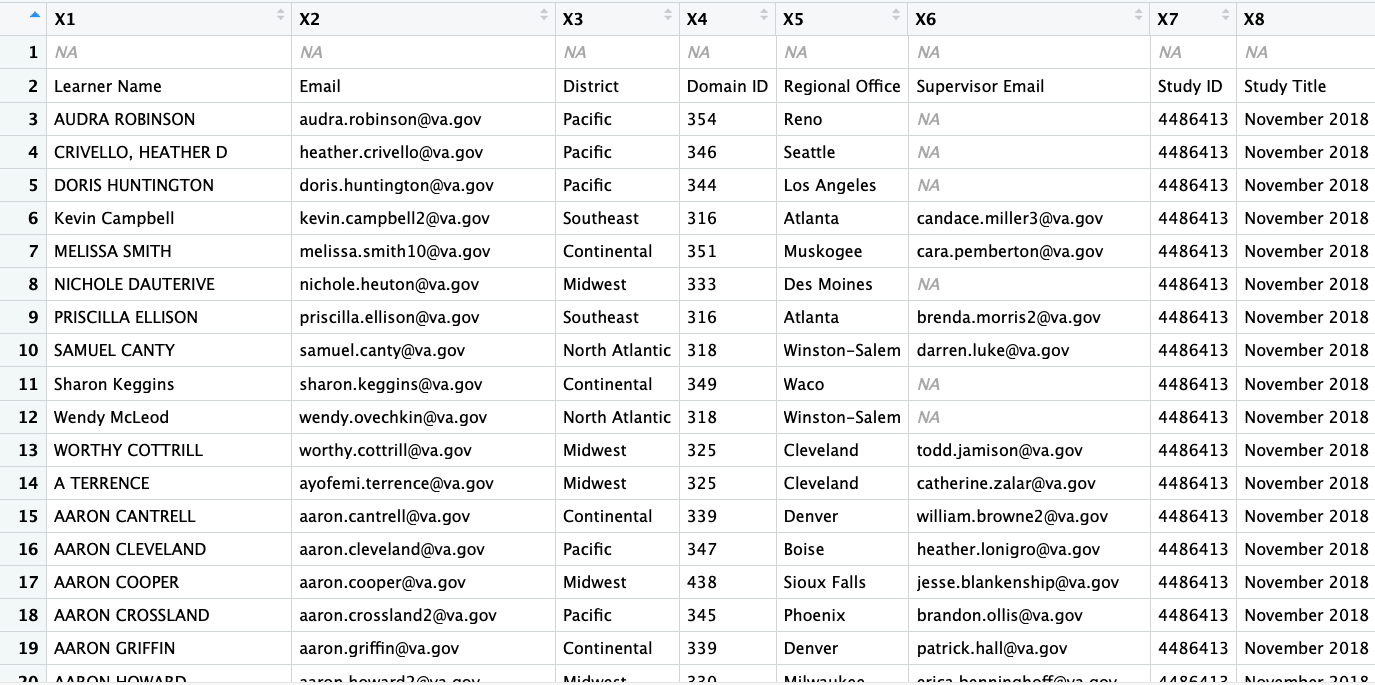
The following code was used to import the data:

library(openxlsx)

# save worksheet as "Project Data"

df <- read.xlsx("Project Data.xlsx")

The following screenshot shows a part of the resulting data that was imported into R (some columns have been hidden for privacy concerns):



The following code was used to clean and sanitize the data:

# rename columns by row 2

colnames(df) <- df[2,]

# remove unnecessary columns

df2 <- df[-1:-3,-c(1, 2, 5, 6, 10, 12:14, 24, 25 )]

#remove columns for open-ended questions

TeamData <- df2[,-c(14,15,24:26)]

#changing variable classes

TeamData$District <- as.factor(TeamData$District)

TeamData$`Study Status` <- as.factor(TeamData$`Study Status`)

TeamData$`Post Eval Question 1` <- as.factor(TeamData$`Post Eval Question 1`)

TeamData$`Pretest Score` <- as.numeric(TeamData$`Pretest Score`)

TeamData$`Posttest Score Attempt 1 ` <- as.numeric(TeamData$`Posttest Score Attempt 1 `)

# Rename and standardize Regional Office names

TeamData$'Use Regional Office '[TeamData$'Use Regional Office' == 'Denver/Cheyenne'] = 'Cheyenne/Denver'

TeamData$'Use Regional Office '[TeamData$'Use Regional Office' == 'Other (AMO)'] = 'AMO'

TeamData$'Use Regional Office '[TeamData$'Use Regional Office' == 'Other (VACO)'] = 'VACO'

TeamData$'Use Regional Office '[TeamData$'Use Regional Office' == 'Washington'] = 'Washington DC'

TeamData$`Use Regional Office ` <- as.factor(TeamData$`Use Regional Office `)

The following screenshot shows a part of the resulting data frame (“TeamData”) after the above code was run in R:

# 

# 

# **5. DATA MODELING AND ANALYSIS**

## **5.1. Use of Linear Regression**

Using the lm() function, a regression was run to reveal potential relationships between pretest score and 1) experience, and 2) district.

The following code was used to run the analysis of the relationship between pretest scores and experience level:

**mod.exp <- lm(df.pretest$`Pretest Score` ~ df.pretest$`Demographic Question 3`)**

**summary(mod.exp)**

The results of the regression are as follows:

Call:

lm(formula = df.pretest$`Pretest Score` ~ df.pretest$`Demographic Question 3`)

Residuals:

   Min     1Q Median      3Q Max

-71.494 -20.858   4.142 28.506 34.627

Coefficients:

                                                   Estimate Std. Error t value Pr(>|t|)

(Intercept)                                          70.8584 0.6158 115.058 < 2e-16 \*\*\*

df.pretest$`Demographic Question 3`10 or more        -5.4855 1.2366 -4.436 9.3e-06 \*\*\*

df.pretest$`Demographic Question 3`2 - 5 years        0.6356 0.8135 0.781 0.4347

df.pretest$`Demographic Question 3`5 - 10 years      -2.1360 0.9245 -2.310 0.0209 \*

df.pretest$`Demographic Question 3`Less than 1 year  -3.0942 1.2680 -2.440 0.0147 \*

df.pretest$`Demographic Question 3`Not Provided       4.1416 27.4933 0.151 0.8803

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 27.49 on 7522 degrees of freedom

Multiple R-squared:  0.00475, Adjusted R-squared:  0.004089

F-statistic:  7.18 on 5 and 7522 DF,  p-value: 1.03e-06

Results show a very low R-Squared value, meaning that experience level does not accurately predict how well a participant may score on the pretest.

The following code was used to run the analysis of the relationship between pretest scores and district:

**mod.district <- lm(df.pretest$`Pretest Score` ~ df.pretest$District)**

**summary(mod.district)**

The results of the regression are as follows:

Call:

lm(formula = df.pretest$`Pretest Score` ~ df.pretest$District)

Residuals:

   Min     1Q Median      3Q Max

-72.164 -20.409   4.591 28.910 43.056

Coefficients:

                                 Estimate Std. Error t value Pr(>|t|)

(Intercept)                         56.944 6.479 8.789 <2e-16 \*\*\*

df.pretest$DistrictContinental      14.145 6.510 2.173 0.0298 \*

df.pretest$DistrictMidwest          15.220 6.524 2.333 0.0197 \*

df.pretest$DistrictNorth Atlantic   13.465 6.515 2.067 0.0388 \*

df.pretest$DistrictPacific          10.524 6.525 1.613 0.1068

df.pretest$DistrictSoutheast        10.980 6.522 1.684 0.0923 .

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 27.49 on 7522 degrees of freedom

Multiple R-squared:  0.004498, Adjusted R-squared:  0.003836

F-statistic: 6.797 on 5 and 7522 DF,  p-value: 2.474e-06

The results of this regression are even worse, showing an R-squared value of less than 0.004. This means that district also does not predict a participant’s pretest score.

# **6. APPENDICES**

## **6.1. Data Dictionary**

|  |  |  |
| --- | --- | --- |
| Variable | Meaning | Type |
| Learner Name | Learner information | System selected |
| Learner Email | Learner information (key identifier) | System selected |
| District | District (five nationally) | System selected |
| Domain ID | Regional Office numerical identifier | System selected |
| Regional Office | Regional Office (57 regional offices plus a central office and appeals office) | System selected |
| Supervisor Email | Learner supervisor information | System selected |
| Study ID | Official study ID | System selected |
| Study Title | Official study title | System selected |
| Pretest Score\* | Average score for first test before training (score distribution 25 points per question) | Integer |
| Pretest Status | Pretest status based on score, 100% required to pass | Binomial (Pass or Fail) |
| Posttest Score Attempt 1\* | First posttest attempt average score after training (score distribution 25 points per question) | Integer |
| Posttest Score Attempt 1 Status | First posttest status based on score, 100% required to pass | Binomial (Pass or Fail) |
| Final Posttest Score | Final posttest (of multiple if taken beyond first attempt) average score after training (score distribution 25 points per question) | Integer |
| Final Posttest Status | First posttest status based on score, 100% required to pass | Binomial (Pass or Fail) |
| Study Status | Final compliance status on test | Binomial (Complete or Incomplete) |
| Pretest Question 1 | First pretest question | Multiple choice |
| Pretest Question 2 | Second pretest question | Multiple choice |
| Pretest Question 3 | Third pretest question | Multiple choice |
| Pretest Question 4 | Fourth pretest question | Multiple choice |
| Posttest Question 1 | First posttest question | Multiple choice |
| Posttest Question 2 | Second posttest question | Multiple choice |
| Posttest Question 3 | Third posttest question | Multiple choice |
| Posttest Question 4 | Fourth posttest question | Multiple choice |
| Demographic Question 1 | Job title | Drop-down |
| Demographic Question 2 | If selected “Other” write-in option | Write-in |
| Demographic Question 3 | Years of experience in current job | Drop-down |
| Demographic Question 4 | Additional technical training needed | Open ended |
| Post Eval Question 1 | Office Location | Drop-down |
| Post Eval Question 2 | Overall Satisfaction | Likert 5-point scale |
| Post Eval Question 3 | Overall lesson quality | Likert 5-point scale |
| Post Eval Question 4 | Ease of navigation through training | Likert 5-point scale |
| Post Eval Question 5 | Value of training to improve job performance | Likert 5-point scale |
| Post Eval Question 6 | Perceived knowledge before training | Likert 5-point scale |
| Post Eval Question 7 | Perceived knowledge after training | Likert 5-point scale |
| Post Eval Question 8 | Likely to recommend training to peers | Likert 5-point scale |
| Post Eval Question 9 | Reason for recommending or not recommending training | Open ended |
| Post Eval Question 10 | Recommendation to improve training | Open ended |
| Post Eval Question 11 | Additional thoughts on training | Open ended |
| Use Regional Office | Used to override system selected data with post-evaluation survey response | Corrected Regional Office to use |
| Use Job Title | Used to override system selection with demographic survey response | Corrected Job Title to use |
| LearnerID | Fictionally assigned Learner ID (key identifier once learner email is removed) | Learner information |
| \*Learning is measured using the percent change between pretest and posttest scores for those that required remediation (scored less than 100% on the pretest) and who had both a pretest and first attempt post test score. | | |