Elia Kostyrka

IST 652, Scripting for Data Analysis, January 2020

Final Project

**TRAINING EVALUTION USING PYTHON**

**OVERVIEW**

Mandatory training on various topics of interest is conducted monthly to help standardize the way in which an unspecified organization conducts business across the six regions in which it operates. Training evaluation is important to help discover training gaps and opportunities for the training program. This analysis was conducted using post-training assessment scores and Kirkpatrick Level 1 reaction data from one training event in an effort to understand how effective mandatory training was for one particular (unspecified) topic.

The training consisted of a slide presentation with example scenarios and links to pertinent regulations and a post-training assessment accompanied by a mock case file that was used to answer test questions. It was conducted online, was self-paced, and could be accessed from within the organization or remotely.

**DATA AND SOURCE**

Data from real-world mandatory training on an unspecified topic was analyzed to determine training effectiveness. The information was sanitized to maintain the anonymity of the organization and its staff.

The data included a combination of structured and unstructured data in a .csv file with over 5,000 observations and over 30 variables. Select critical variables were examined and included the following:

* Score
* Assessment status
* Satisfaction
  + Overall satisfaction
  + Lesson quality
  + Value in improving job
  + Recommend to peers
* Region
* Job title
* Years of experience on the job
* Comments on how to improve training

**Structured data**

**Score.** Scores were captured from a 4-question assessment that was taken after participants completed training. Scores were evenly weighted across all questions and included score averages of 0%, 25%, 50%, 75%, and 100%. Participants that did not attempt the assessment receive a blank.

**Assessment status.** The assessment status presents whether a participant completed the training and is either a “complete” or “incomplete” status. All participants that went through and attempted all questions received a complete status. A small number of participants did not complete all questions and were provided an incomplete status. Those who did not attempt the assessment were given an incomplete status.

**Satisfaction.** Four different questions gauging satisfaction with various aspects of training were analyzed. Satisfaction levels are gauged using a 5-point Likert scale with 5 being very satisfied and 1 being highly dissatisfied. The first question asks for overall satisfaction with training. The second question asks for overall satisfaction with the lesson quality. The third question asks for satisfaction with how valuable the training was to improving job performance. The fourth question asks likelihood to recommend to peers, 5 being extremely likely and 1 being extremely unlikely.

**Region.** Six regions where the organization operates were assessed, are broken down by geographic regions of the United States, and include Main, Pacific, Midwest, Continental, Southeast, and North Atlantic. Each region contains smaller office segments that were not used for this analysis as overarching guidance on regulations is issued at the regional level.

**Job title.** Six job titles were assessed and included staff that belonged to the Quality 1, Quality 2, Review, Service 1, Service 2, or cross-trained jobs roles. Typically, Quality 1 and Quality 2 staff have higher responsibility and require more knowledge in the field. Review staff are a mix of staff selected to conduct spot checks on processes. Service 1 and Service 2 staff typically have less responsibility but a higher workload. Cross-trained staff can perform in either a Quality and Review role but may have more experience in one or the other.

**Years of Experience.** Six different categories of years on the job were assessed and included, Less than 1 year, 1-2 years, 2-5 years, 5-10 years, and 10 or more years.

**Unstructured data**

**Comments.** One open text question on the assessment was analyzed asking participants ways the training conducted could be improved. The comments included dirty data with misspellings, various punctuations, use of numbers, unresponsive comments (e.g., “blah,” n/a, or none), use of uppercase and lower case in the sentences, use of acronyms, and some blank fields.

**BUSINESS QUESTIONS**

Five business questions were used to understand where training gaps and training opportunities might exist for the unspecified topic. They included the following:

1. What subgroup was most familiar with the topic based on the assessment results by region, job title, and years of experience?
2. What subgroup requires remediation based on assessment results by region, job title, and years of experience?
3. What was overall compliance by region?
4. What was overall satisfaction by region, job title, and years of experience?
5. How can training be improved according to participant comments?

Additional sub-questions could be developed to further understand training needs by smaller clusters.

**PYTHON PROGRAM**

The analysis plan included using Python to describe the data (descriptive statistics) and aggregate the data by subgroups using Pandas and other libraries. Output in the form of descriptions and data visualizations using Matplotlib and seaborn were employed to help interpret the results. Logistic regression using statsmodels was employed to identify variables of importance. Text mining techniques was employed using nltk and Textblob to determine the most frequent themes participants noted in their comments on how to improve training. Other libraries were incorporated as needed to improve the analysis.

**Cleaning and Transforming the Data.**

Score data was categorized into discrete data for modeling, ScorePerfect (100 = Yes, <100 = No). Years of experience data was categorized into two bins, Experience = Not Much (< 5 years), Much (>5 years), for initial exploratory data analysis. Blank scores were deleted for some calculations using drop.na, including for tabulating distribution and score averages. Text mining including transforming the dirty data by making it all lowercase, removing punctuation, digits, and stop words, and removing some additional common words (e.g., “improve”, “training”).

**Python Libraries**

Pandas was used to read in the .csv file.

Seaborn distribution plot was used to plot the distribution of scores overall.

Seaborn count plot was used to develop plots of categorical features, including number of completions and number of perfect scores.

Seaborn bar plot was used to develop plots of average scores by different categories.

Seaborn box plot was used to develop plots of average overall satisfaction by different categories.

Pandas cross tab and matplotlib were used to compute a cross tabulation of two variables at a time to determine the frequency of one of the variables, either perfect score or number of completions.

Additional plots were developed but were not used in the analysis and included a Seaborn facet grid, regplot (linear regression plot), and heat map.

Patsy and statsmodels were used to develop a logistical regression model. Patsy was used to transform the data into a form that statsmodels could use in the algorithm.

Nltk, Texblob, punkt, and matplotlib were used for text mining. The comment field was selected and transformed into a data frame. Blank comment fields were removed, punctuation was removed, digits were removed, odd characters were removed, text was transformed to lowercase and printed to a string. The string was written to a text file that was then read in for continued text mining. The length of the text included 204,399 words before nltk was used to remove stopwords which reduced the number of words to 24,418 useful words. Punkt was used to tokenize the words. Nltk ngrams was used to identify top bigrams showing the highest frequency of paired words. Matplotlib was used to plot the most frequent themes.

Various functions were used throughout the program to describe or show the data as it was being transformed, including head(), describe(), and print().

**OUTPUT FILE**

The output of the Python program included a series of descriptions and plots that helped to interpret the results of the assessment.

**General Description of the Data**

A distribution plot (Figure 1) showed that more participants got 3 or 4 questions correct on the assessment.

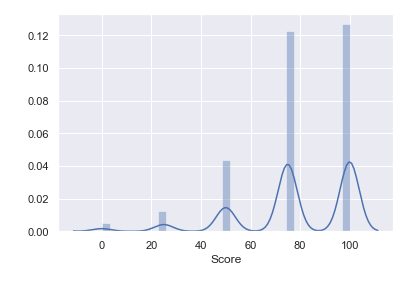


Figure 1. Distribution of Scores

Descriptive statistics (Figure 2) showed that there were 5,337 learners assigned the assessment. The average score was 78.56%, the minimum score was 0% and maximum score was 100%. Participant average satisfaction scores for training using a 5-point Likert scale were between 3.76 and 3.82 (between neutral and satisfied). The average rating for likely to recommend to peers was 3.82, followed by 3.81 for value to improve their job, 3.78 for lesson quality, and 3.76 for Overall satisfaction.

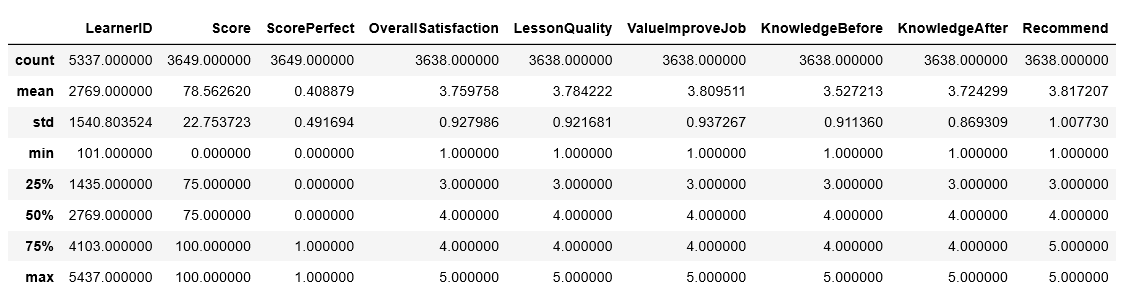


Figure 2. Descriptive Statistics of the Numerical Variables

Count plots (Figures 3 and 4) showed that 3,649 of the 5,337 (68%) learners completed the assessment and of those that completed the assessment 41% (1,492 of the 3,649) received a perfect score of 100%.

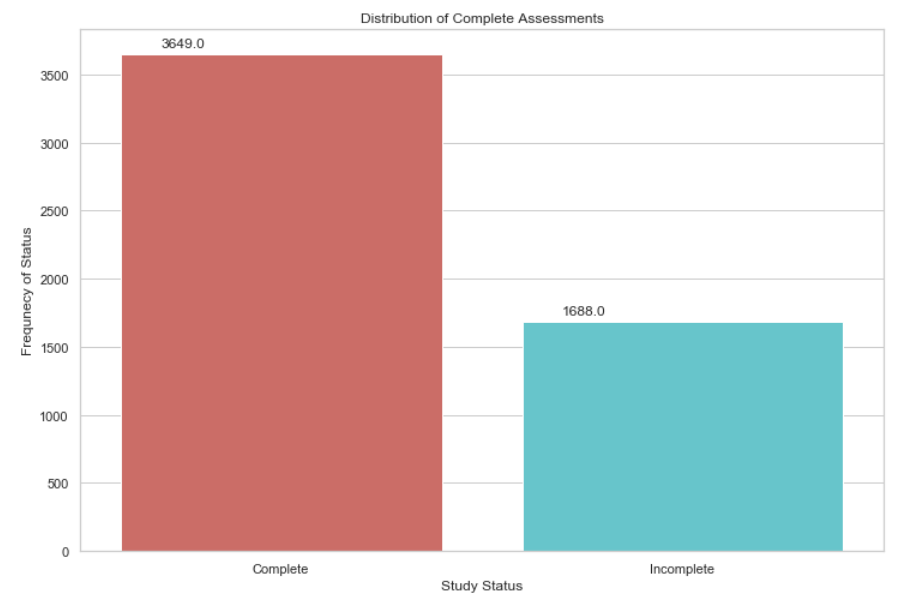


Figure 3. Distribution of Study Status

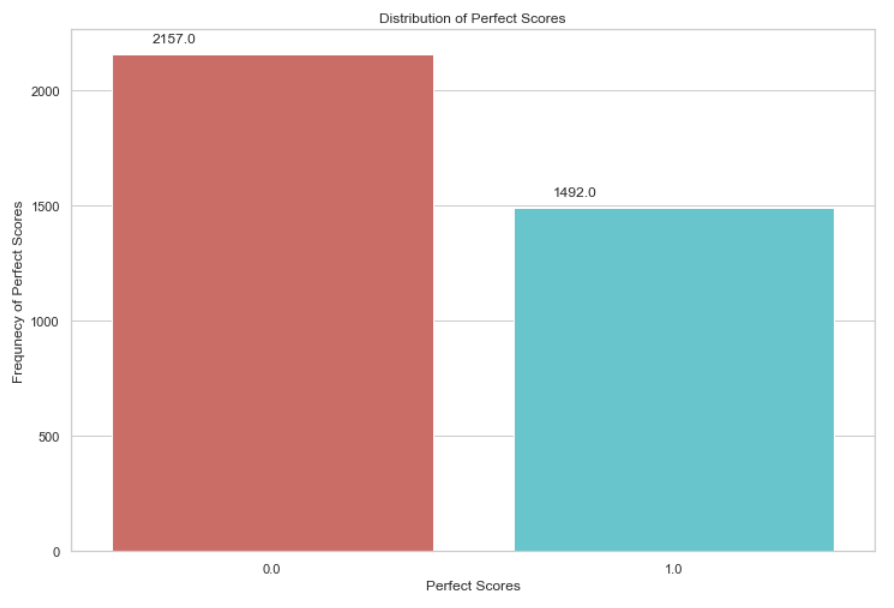


Figure 4. Distribution of Perfect Scores.

**Business Question 2. What subgroup was most familiar with the topic based on the assessment results by region, job title, and years of experience?**

Grouped bar charts (Figures 5, 6, and 7) showed that 46% of learners in the Midwest Region got a perfect score, followed by 44% of Continental Region. On the contrary, the Main, Pacific, and Southeast regions had the fewest number of staff that scored a perfect score (26%, 37%, and 38% respectively). A higher percentage of participants had more perfect scores by job title. 70% of Quality 2 staff got a perfect score, followed by 65% of the Quality1 staff. And participants with more years of experience were more familiar with the topic. 48% of personnel with 5-10 years of experience got a perfect score, followed by 46% with 10 or more years of experience

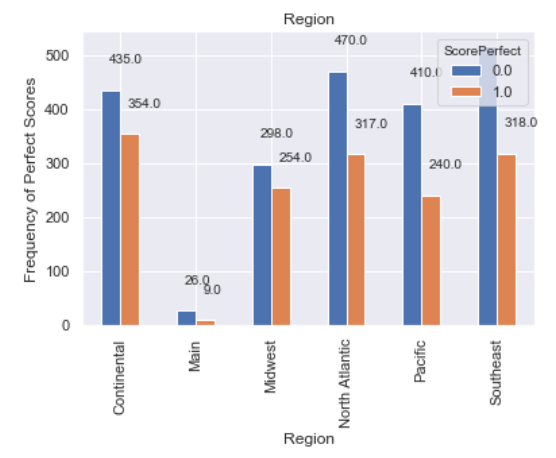


Figure 5. Perfect Scores by Region

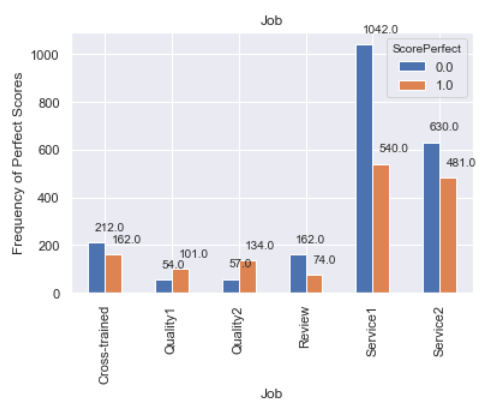


Figure 6. Perfect Scores by Job Title

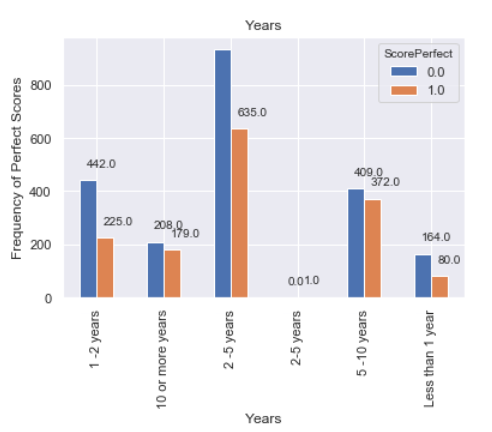


Figure 7. Perfect Scores by Year of Experience

**Business Question 2: What subgroup requires remediation based on assessment results by region, job title, and years of experience?**

Bar plots (Figures 8, 9, and 10) showed that participants in the Main Region had the lowest score average, 69%. In addition, participants from the Review job title had the lowest score average, 73%. And lastly, participants with 1-2 years of experience had the lowest score average, 76%.

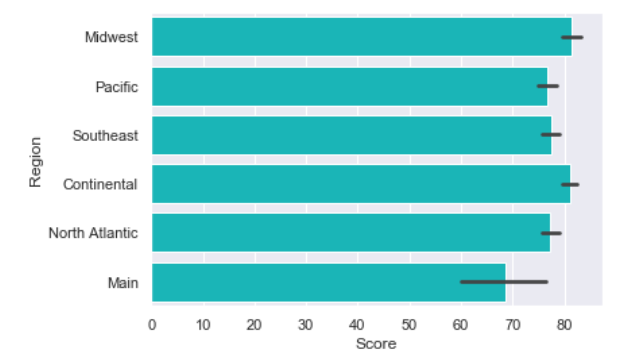


Figure 8. Average score by Region

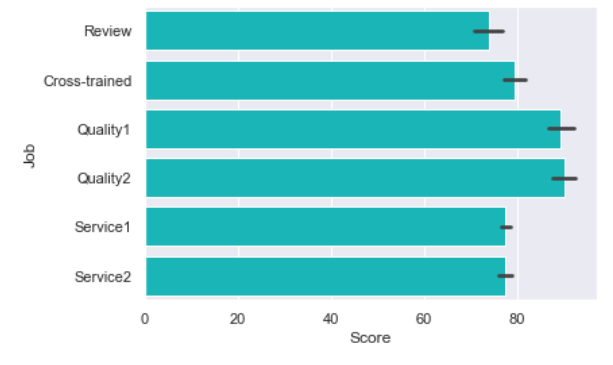


Figure 9. Average score by Job Title

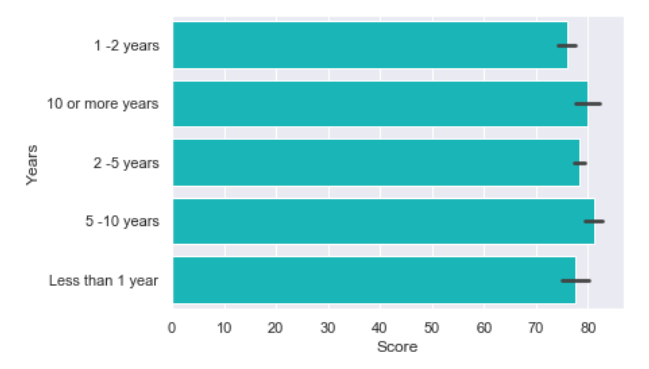


Figure 10. Average score by Years of Experience

Logistical regression (Figure 11) was used to determine how well the job title, years of experience, and region variables of the model explained the dependent variable (perfect score). Groups less likely to produce a perfect score included participants in the Service 1 and Review job roles, and participants in the Main, Pacific, and Southeast regions.

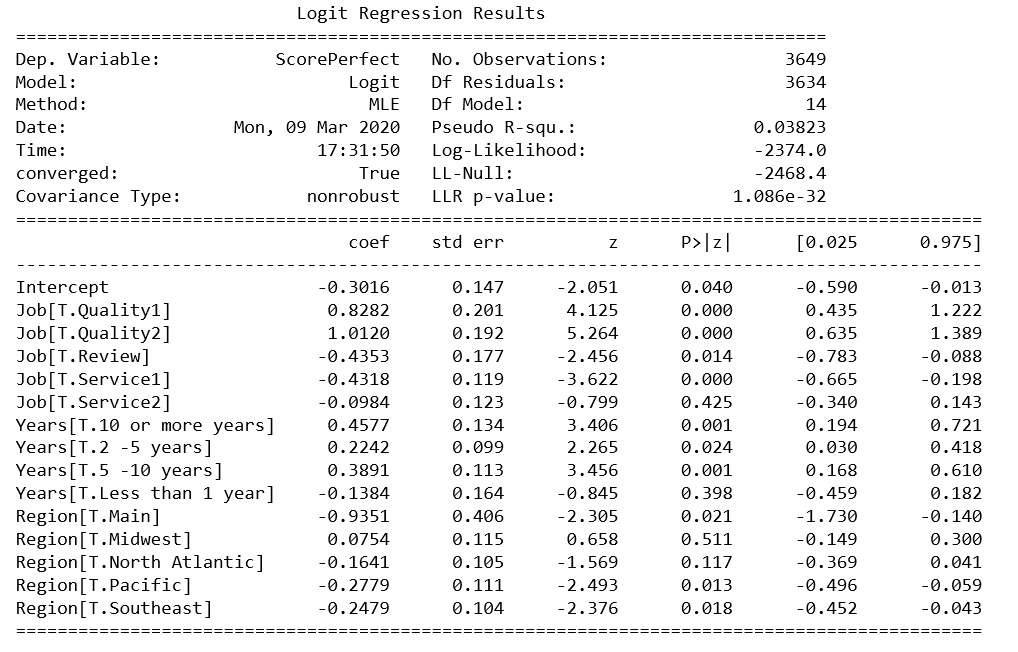
****

Figure 11. Logistical Regression

**Business Question 3: What was overall compliance by region?**

A grouped bar chart (Figure 12) showed that compliance was less than 50% for a majority of the regions, except for the Main region with 68% compliance. Compliance ranged from 23% to 68% for all regions. The regions with the lowest compliance rate included the Continental (23%) and Southeast regions (28%). The remaining regions had less than 40% participation and included the North Atlantic region (30%), Pacific region (37%), and the Midwest region (37%). It could be assessed that because the Pacific and Southeast regions had the fewest number of staff that scored a perfect score, these two regions possess a larger percentage of participants unfamiliar with the topic.

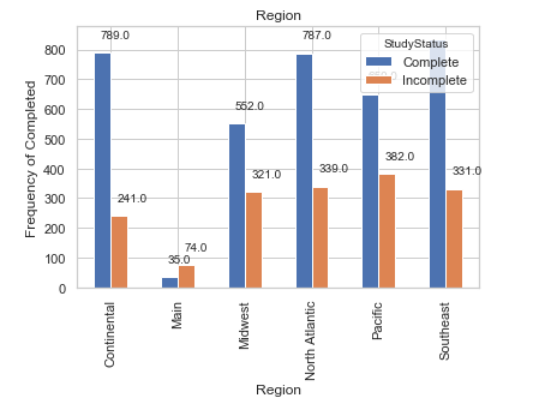
****

Figure 12. Compliance by Region

**Business Question 4. What was overall satisfaction by region, job title, and years of experience?**

Box plots (Figures 13, 14, and 15) showed that participants in the Southeast Region had higher overall satisfaction ratings than all other groups. Additionally, Quality1 participants had higher overall satisfaction ratings than all other jobs. Notably, participants in the job had the highest percentage of individuals who got a perfect score on the assessment indicating that the it was well-constructed training. And lastly participants with less than 1-year experience had higher overall satisfaction ratings than all other years of experience groups.

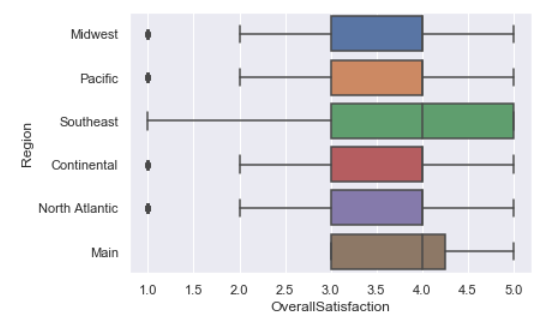


Figure 13. Overall satisfaction by Region

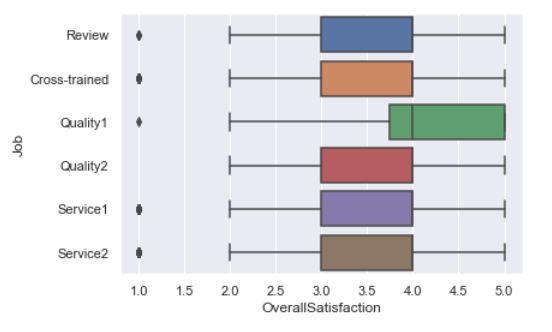


Figure 14. Overall Satisfaction by Job Title

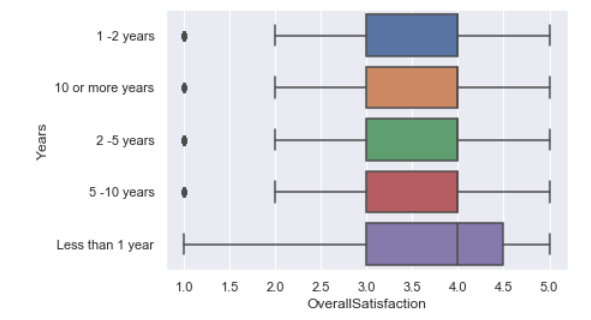


Figure 15. Overall Satisfaction by Years of Experience

**Business Question 5: How can training be improved according to participant comments?**

A bar plot (Figure 16) showed that participants noted that overall training was good and that the delivery was fine; however, participants mentioned that three things could be done to improve training, including 1) adding more questions (and likely the associated scenarios), using screen shots for visual display of the concept being covered, and 3) adding audio. Both the request for screen shots and audio indicate a need to incorporate more multi-media.

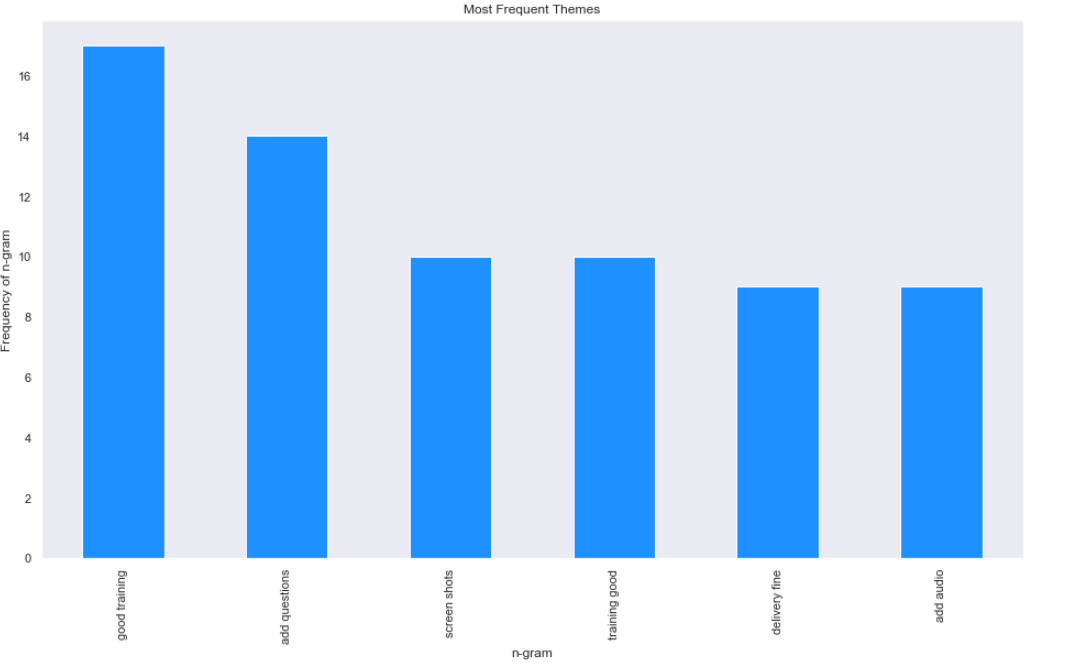
****

Figure 16. Most Frequent Themes on Ways to Improve Training

**FINDINGS AND RECOMMENDATIONS**

Overall, compliance and score average were below the benchmarks with compliance below 70% (68%) and the assessment score average below 80% (78.56%), indicating that more messaging needs to be conducted on the importance of the topic in daily operations.

Less than half (41%) of the participants that completed the training scored 100% on the assessment indicating that more training is needed on the topic.

Overall satisfaction ratings for lesson quality, value to improve participant’s job, and whether they would recommend the training ranged from 3.78 to 3.82 with overall satisfaction at 3.76 (between neutral and satisfied), indicating that the training needs to be improved before being re-administered.

Participants most familiar with the training topic included staff from the Midwest region, Quality 1 and Quality 2 job titles, and more experienced staff (5+ years), indicating a training opportunity to use these participants to conduct follow-on training on the topic.

The organization should prioritize training participants that either did not complete the assessment or scored poorly on the assessment and belong to the Main and Pacific regions, are in the Service1 and Review job roles across all regions, and have 1-2 years of experience or less across all regions.