## Week 1a – Improving Data Through EDA

1. Which of the following are categories of data quality tools?
   1. Both ‘Cleaning tools’ and ‘Monitoring tools’
2. What are the features of low data quality?
   1. Unreliable info
   2. Incomplete data
   3. Duplicated data
3. What are the objectives of exploratory data analysis?
   1. Check for missing data and other mistakes.
   2. Gain maximum insight into the data set and its underlying structure.
   3. Uncover a parsimonious model, one which explains the data with a minimum number of predictor variables.
4. Exploratory Data Analysis is majorly performed using the following methods:
   1. Both Univariate and Bivariate
5. Which of the following are components of Exploratory Data Analysis?
   1. Accounting and Summarizing
   2. Anomaly Detection
   3. Statistical Analysis and Clustering

## Week 1b – Machine Learning in Practice

1. Which of the following machine learning models have labels, or in other words, the correct answers to whatever it is that we want to learn to predict?
   1. Supervised Model
2. Which model would you use if your problem required a discrete number of values or classes?
   1. Classification Model
3. To predict the continuous value of our label, which of the following algorithms is used?
   1. Regression
4. What is the most essential metric a regression model uses?
   1. Mean squared error as their loss function
5. Why is regularisation important in logistic regression?
   1. Avoids overfitting

## Week 2a – Training AutoML Models Using Vertex AI

1. What is the main benefit of using an automated Machine Learning workflow?
   1. It reduces the time it takes to develop trained models and assess their performance.
2. For a user who can use SQL, but has little Machine Learning experience and wants a ‘Low-Code’ solution, which Machine Learning framework should they use?
   1. BigQuery ML
3. If a dataset is presented in a Comma Separated Values (CSV) file, which is the correct data type to choose in Vertex AI?
   1. Tabular
4. If the business case is to predict fraud detection, which is the correct Objective to choose in Vertex AI?
   1. Regression/Classification
5. Which of the following are stages of the Machine Learning workflow that can be managed with Vertex AI?
   1. Create a dataset and upload data.
   2. Train an ML model on your data.
   3. Deploy your trained model to an endpoint for serving predictions.
6. What is the default setting in AutoML Tables for the data split in model evaluation?
   1. 80% Training 10% Validation, 10% Testing
7. MAE, MAPE, RMSE, RMSLE and R2 are all available as test examples in the Evaluate section of Vertex AI and are common examples of what type of metric?
   1. Linear Regression Metrics
8. What does the Feature Importance attribution in Vertex AI display?
   1. How much each feature impacts the model, expressed as a percentage
9. Which of the following metrics can be used to find a suitable balance between precision and recall in a model?
   1. F1 Score

## Week 2b – BigQuery Machine Learning

1. Which of the following are advantages of BigQuery ML when compared to Python based ML frameworks?
   1. BigQuery ML custom models can be created without the use of multiple tools
   2. BigQuery ML automates multiple steps in the ML workflow
   3. Moving and formatting large amounts of data takes longer with Python based models compared to model training in BigQuery
2. Which of these BigQuery supported classification models is most relevant for predicting binary results, such as True/False?
   1. Logistic Regression
3. For Classification or Regression problems with decision trees, which of the following models is most relevant?
   1. XGBoost
4. Where labels are not available, for example where customer segmentation is required, which of the following BigQuery supported models is useful?
   1. K-Means Clustering
5. What are the 3 key steps for creating a Recommendation System with BigQuery ML?
   1. Prepare training data in BigQuery, train a recommendation system with BigQuery ML, use the predicted recommendations in production
   2. Import training data to BigQuery, train a recommendation system with BigQuery ML, tune the hyperparameters
   3. Prepare training data in BigQuery, select a recommendation system from BigQuery ML, deploy and test the model
   4. Prepare training data in BigQuery, specify the model options in BigQuery ML, export the predictions to Google Analytics

## Week 3a – Optimisation

1. For the formula used to model the relationship i.e. y = mx + b, what does ‘m’ stand for?
   1. It captures the amount of change we've observed in our label in response to a small change in our feature.
2. What are the basic steps in an ML workflow (or process)?
   1. Collect data
   2. Check for anomalies, missing data and clean the data
   3. Perform statistical analysis and initial visualization
3. Which of the following loss functions is used for classification problems?
   1. Cross entropy
4. Which of the following gradient descent methods is used to compute the entire dataset?
   1. Batch gradient descent.
5. Which of the following are benefits of Performance metrics over loss functions?
   1. Performance metrics are easier to understand and are directly connected to business goals.

## Week 3b – Generalisation and Sampling

1. Which is the best way to assess the quality of a model?
   1. Observing how well a model performs against a new dataset that it hasn't seen before.
2. How do you decide when to stop training a model?
   1. When your loss metrics start to increase
3. Which of the following actions can you perform on your model when it is trained and validated?
   1. You can write it once, and only once against the independent test dataset.
4. Which of the following allows you to create repeatable samples of your data?
   1. Use the last few digits of a hash function on the field that you're using to split or bucketize your data.
5. Which of the following allows you to split the dataset based upon a field in your data?
   1. FARM\_FINGERPRINT, an open-source hashing algorithm that is implemented in BigQuery SQL.