**Key Takeaways from Week 1 - Module 1:**

Data exploration is a critical step.

Use a data quality report to help you quickly explore the data. Pro tip: save the python code and reuse it.

Be cautious when removing data quality issues.

There are multiple options for visualizations. Use whichever you are most comfortable with.

**Key Takeaways from Week 2 - Module 1:**

* There are differences between Data Exploration and Data Presentation.
  + Data exploration means the deep-dive analysis of data in search of new insights.
  + Data presentation means the delivery of data insights to an audience in a form that makes clear the implications.
* Common ways to view relationships between features in a dataset.
  + Scatter plots (continuous features)
  + Small multiples bar plots (categorical features)
  + Small multiples Density Histograms (categorical vs continuous)
  + Multiple Box plots (categorical vs continuous)
  + Covariance: measures the direction of the variables.
  + Correlation: measures the strength and direction of the variables.
* Basic Machine Learning Process:
  + Preprocess
  + Train/Test Split
  + Model -> Fit -> Predict
  + Evaluate
* Feature engineering.
  + Normalization or continuous variables
  + Encoding categorical variables
  + Sampling
  + Under/Over Sampling
  + Handling Missing values
* KNN: similarity-based theory approach that makes predictions even with noisy data based on the proximity of neighboring data points.
* Evaluating machine learning models.
  + Confusion Matrix
  + Accuracy, Precision, Recall, and F1

**Key Takeaways from Week 3 - Module 2:**

Decision Tree Algos: Splits the data based on values until only one label (prediction class) is reached. In its simplest form a decision tree can be viewed as a series of if-else statements.

* Entropy: a measure of the randomness in the labels of a set of data. Generally entropy falls between 0 and 1. Zero means it has ZERO randomness. The closer we get to one the higher the randomness of the data. Very simple example using Spam or Ham emails:
* *Entropy value of 0:* Zero Entropy means your data only contains Spam categorized emails. Alternatively, if you had all Ham categories in your data it would also mean zero entropy. Its not random.
* *Entropy value between 0 and 1:* Data contains spam and ham categorized emails but is unbalanced. For example, 3 Ham and 7 Spam in a data set of 10 emails. Alternately, 7 Ham and 3 Spam is the same.
* *Entropy value of 1:* Means the data has a high level of randomness. Data would be perfectly balanced between categories. For example: 5 Ham and 5 Spam labeled emails out of a total of 10.
* Information Gain (IG): Also measured between 0-1 and helps the algo decide which column (feature) that is used to split. The column (feature) with the highest value of IG that is closest to one is chosen. Zero means no information and one means it has a high amount of information. The highest value closest to 1 is the column that has the most information for making a decision!
* Information Gain Calculation:  IG = (Parent Node Total Entropy) - (Weighted Sum of Child Nodes Entropy)
* ID3 Algo: Uses IG for making splits/decisions. Starts at the top or root node (the entire data set) and splits the data samples on the column (feature) with the highest information gain.  As new branches are created, each child node also splits the data on the column (feature) with the highest information gain. This is done until a value of 0 entropy is found in the remaining data.

**Key Takeaways from Week 5 - Module 3:**

* Ensemble methods improve results. Ensembles use multiple models that by themselves are weak learners but combined can predict very well. You are basically building multiple experts at classifying specific subsets of the data.
  + Bagging Method: Combining multiple learners that learn independently and in parallel.
  + Boosting Method: Combining multiple learners that learn sequentially and adapt to improve performance.
* Model evaluation options for Continuous Targets:
  + ***Mean squared error (MSE):*** captures the average differences between the expected values and predicted values. The smaller the value the better the performance. Be careful because this is relative and domain specific.
  + ***Root mean squared error (RMSE):* is** the square root of MSE. It provides the same information but in the same units as the target values. Smaller values indicate better performance. Be careful because this is relative and domain specific. RMSE overemphasizes large differences. In other words, if the data you are measuring against has a lot of outliers then RMSE will be larger. RMSE is most effective when the data is not expected to have a lot of outliers.
  + ***Mean absolute error (MAE):***  also captures average differences but in a linear way. Not influenced by large or small differences. Represented in the target value units. MAE is most effective when outliers are expected.  Also relative and domain specific.
  + ***R-squared coefficient (R2):*** Ranges from 0 to 1. The closer the value is to 1 the better the performance of the model. R2shows how well your model explains the variance in the target variable. For example if R2 is .70 that means your model explained 70% of the variance in the target values. This metric is not relative or domain specific.