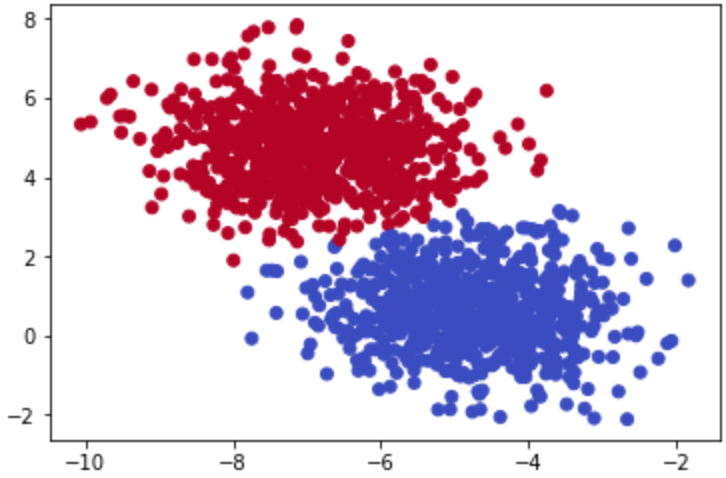
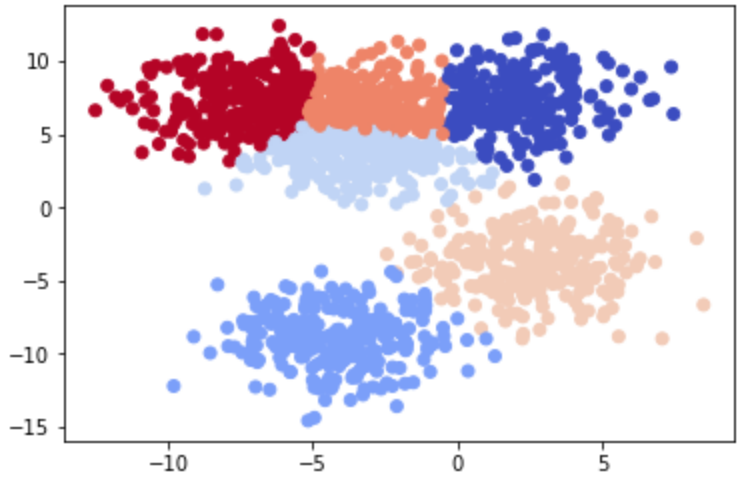
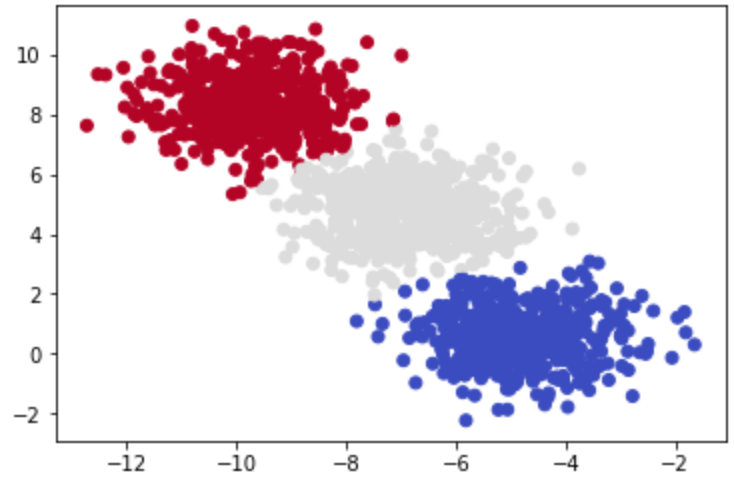
Chapter 1: Introduction to Unsupervised Learning Answer Key

Activity 1 Solutions:







Activity 2 Solutions:

Please refer to the provided Jupyter notebook for solution code.

Practice Question Solutions:

1. How are supervised and unsupervised learning different?
   1. 
2. What is a cluster?
   1. A cluster is a grouping of similar data that exists within your larger dataset.
3. Are clusters always valuable in a dataset?
   1. Not necessarily. Just because you find a pocket of data that is similar to each other does not mean that it is correlated with anything you’re interested in.
4. How are dimensions expressed in a dataset?
   1. Dimensions are expressed as features in a dataset, or simply the number of columns available per row. For example a 2-dimensional dataset will have 2 columns.
5. How many dimensions can be easily visualized by humans?
   1. 3 dimensions
6. What does the “K” represent in K-Means clustering?
   1. K represents the expected number of distinct clusters
7. T/F – The starting points for K-Means are determined by taking the mean of all the points in the space.
   1. False – the starting points for K-Means are randomly initialized
8. What is the formula that underpins K-Means clustering?
   1. Euclidean distance formula
9. When does a K-Means clustering algorithm finish running?
   1. K-Means finishes when it converges, and there is no difference between the newest calculated centroid and old one in the training loop.
10. T/F – K-Means will still find clusters in a dataset even if all of the data is fairly similar.
    1. True. K-Means does it’s best to find the K number of clusters you tell it to. That means even if you have one large cluster of data but tell it to find K clusters within it, it will break it into those K number of groups the best it can.