1. What are the key tasks involved in getting ready to work with machine learning modeling?

2. What are the different forms of data used in machine learning? Give a specific example for each of them.

3. Distinguish:

1. Numeric vs. categorical attributes

2. Feature selection vs. dimensionality reduction

4. Make quick notes on any two of the following:

1. The histogram

2. Use a scatter plot

3.PCA (Personal Computer Aid)

5. Why is it necessary to investigate data? Is there a discrepancy in how qualitative and quantitative data are explored?

6. What are the various histogram shapes? What exactly are ‘bins'?

7. How do we deal with data outliers?

8. What are the various central inclination measures? Why does mean vary too much from median in certain data sets?

9. Describe how a scatter plot can be used to investigate bivariate relationships. Is it possible to find outliers using a scatter plot?

10. Describe how cross-tabs can be used to figure out how two variables are related.