1. What are some of the issues that you may encounter when performing data mining?

You can run into three types of issues when performing data mining: mining methodology and user interaction, performance issues, and diverse data types issues. Mining methodology and user interaction refers to the incorporation of knowledge into data mining methods. Performance issues refers to the efficiency of data mining methods. And diverse data types issues refer to the types of data you are working with.

1. What is the difference between a Quantile plot and a Quantile-Quantile plot?

A quantile plot displays the quantiles of one set of values while a quantile-quantile plot displays the quantiles of two sets of values.

1. When dealing with similarity or dissimilarity why is it important to treat ordinal categorical data differently from other categorical data?

It’s important to treat ordinal categorical data differently from other categorical data because ordinal data is ordered. There is a distinct order of categories that observations fall under, while other categorical data does not have this. This is important for measuring similarity and dissimilarity so we can normalize and measure accurately.

1. When should you not use Cosine similarity?

You should not use cosine similarity when you are trying to measure the Euclidean distance between observations.

1. Write down any questions you may have.