1. What is the primary objective of data wrangling?
a) Data visualization

Data visualization is a crucial aspect of machine learning that enables analysts to understand and make sense of data patterns, relationships, and trends. Through data visualization, insights and patterns in data can be easily interpreted and communicated to a wider audience, making it a critical component of machine learning.

- b) Data cleaning and transformation

Data cleaning is the process of making sure that the data is accurate and consistent, while data wrangling is the process of manipulating the data to make it usable for analysis. Both steps are essential for the process of working with data, and they need to be performed before any analysis takes place.

- c) Statistical analysis

Data analysts typically spend the majority of their time in the process of data wrangling compared to the actual analysis of the data. The process of data wrangling may include further munging, data visualization, data aggregation, training a statistical model, as will as many other potential uses.

- d) Machine learning modeling

Data wrangling is the process of transforming and structuring data from one raw form into a desired format with the intent of improving data quality and making it more consumable and useful for analytics or machine learning. It's also sometimes called data munging.

2. Explain the technique used to convert categorical data into numerical data.

One common method is to assign labels based on the alphabetical order of categories, though the labels could also be assigned randomly or based on the order of appearance in the data. Once these assignments are determined, the categorical values in the dataset are replaced with their corresponding numerical labels.

3. How does LabelEncoding differ from OneHotEncoding?

To prevent biases from being introduced, One-Hot Encoding is preferable for nominal data (where there is no inherent order among categories). Label encoding, however, might be more appropriate for ordinal data (where categories naturally have an order). The effect of dimensionality should also be taken into account.

4.Describe a commonly used method for detecting outliers in a dataset. Why is it important to identify outliers?
If the data, or feature of interest is normally distributed, you may use standard deviation and z-score to label points that are farther than three standard deviations away from the mean as outliers. If the data is not normally distributed, you can use the interquartile range or percentage methods to detect outliers.

PART-B

7. What type of regression is employed when predicting a continuous target variable?
It is one of the most widely known modeling technique. Linear regression is usually among the first few topics which people pick while learning predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear

8. Identify and explain the two main types of regression.
The two basic types of regression are simple linear regression and multiple linear regression, although there are non-linear regression methods for more complicated data and analysis.

9. What is the primary goal of regression analysis?
Objective of Regression analysis is to explain variability in dependent variable by means of one or more of independent or control variables.