SMART INTERNZ - APSCHE **Date:21/04/24**

AI / ML Training

Assignment: Data Wrangling and Regression Analysis

Instructions: Answer the following questions to the best of your ability. Provide concise explanations where necessary.

Section A: Data Wrangling (Questions 1-6)

1. What is the primary objective of data wrangling?
    - a) Data visualization
    - b) Data cleaning and transformation
    - c) Statistical analysis
    - d) Machine learning modeling
2. Explain the technique used to convert categorical data into numerical data. How does it help in data analysis?
3. How does LabelEncoding differ from OneHotEncoding?
4. Describe a commonly used method for detecting outliers in a dataset. Why is it important to identify outliers?
5. Explain how outliers are handled using the Quantile Method.
6. Discuss the significance of a Box Plot in data analysis. How does it aid in identifying potential outliers?

Section B: Regression Analysis (Questions 7-15)

7. What type of regression is employed when predicting a continuous target variable?
8. Identify and explain the two main types of regression.
9. When would you use Simple Linear Regression? Provide an example scenario.
10. In Multi Linear Regression, how many independent variables are typically involved?
11. When should Polynomial Regression be utilized? Provide a scenario where Polynomial Regression would be preferable over Simple Linear Regression.
12. What does a higher degree polynomial represent in Polynomial Regression? How does it affect the model's complexity?
13. Highlight the key difference between Multi Linear Regression and Polynomial Regression.
14. Explain the scenario in which Multi Linear Regression is the most appropriate regression technique.
15. What is the primary goal of regression analysis?

Submission Instructions: Please submit your answers in a neatly organized document, clearly labeling each question with its corresponding number. Ensure your explanations are coherent and demonstrate a solid understanding of the concepts discussed.