Assignment – 1

1. What is data wrangling, and why is it important in data analysis?
2. List the main challenges encountered in data wrangling.
3. Describe a scenario where data wrangling plays a critical role in improving the accuracy of a machine learning model.
4. What is web scraping, and when is it appropriate to use it in data acquisition
5. Compare the advantages and limitations of file I/O and database access for data acquisition
6. What is Exploratory Data Analysis (EDA), and how does it help in identifying data quality issues
7. Explain how EDA can be used to identify missing values and inconsistencies in a dataset.
8. How do missing data, outliers, and duplicates affect the results of a machine learning model

9. Describe a scenario where data wrangling plays a critical role in improving the accuracy of a machine learning model.

Assignment – 2

1. Explain the concept of normalization in data transformation. Why is it important?
2. What is one-hot encoding? Provide an example
3. Discuss the potential effects of not transforming your data before analysis. Use examples to illustrate your points.
4. Design a step-by-step approach for transforming a raw dataset with mixed types of data (numerical and categorical) into a structured format suitable for analysis
5. Define the operations of stack and unstack in pandas. Provide an example.
6. Create a pivot table from a sample dataset to summarize sales data by region and product. Explain your process.
7. Illustrate the steps to convert a wide-format dataset into long format using pandas.
8. What are the key techniques for feature extraction from text data?
9. Describe how you would convert a categorical variable into numerical form for modelling.
10. Analyze how aggregation can simplify complex datasets for clearer insights.
11. Compare different summarization techniques in terms of their utility and limitations.
12. Create a case study where data aggregation significantly improved decision-making in a business context.