Unit - 4

PART - A

1. Outline the key steps in the data munging process.
2. Explain the key components needed to create a line graph in Python using Matplotlib.
3. List and briefly explain the main steps involved in a machine learning pipeline.
4. Describe how a data pipeline is handled in Python.
5. Define data munging and provide one example of a task involved in the data munging process.
6. Mention one python library, aside from Matplotlib that is commonly used for creating advanced interactive visualizations in data science.
7. State the steps in data munging process
8. What is ML Pipeline?

PART – B

1. Your team has gathered insights from a machine learning model,and now you need to present the results to a non technical audience. Explain how Matplotlib can be used for creating meaningful visualizations to communicate complex findings effectively.
2. A data scientist is working on a project that involves merging data sets from different sources. Discuss the challenges they might encounter during the data munging process and propose strategies to address these challenges
3. Discuss advanced data representation techniques in machine learning using Python, focusing on methods such as deep learning, feature engineering, and dimensionality reduction. Include examples of libraries and frameworks used for implementation.
4. Explain the architecture of a machine learning pipeline, detailing each stage from data collection to model deployment. Justify how ML pipeline tools benefit businesses.
5. What is data munging? Explain the steps involved and list the functionalities of data munging.
6. Discuss in detail the machine learning pipeline architecture and justify how ML pipeline tools benefit businesses.

Part – C

1. Ecolab is a chemical company that wants to go the machine learning way of doing things to improve productivity. Since Ecolab not familiarize with ML they wish to automate the ML process. Suggest a solution architecture so that they can improve productivity and faster deployment of models.
2. A sales organization has huge volume of sales data and they wish to estimate the sales based on weekly, monthly and quarterly basis for different regions. Visualize the above scenario and use appropriate API’s to plot using different types of plots.