Sample Q&A; – EdTech Course (Data Science Fundamentals)

Q1: What is Data Science?

A: Data Science is the field that combines statistics, programming, and domain knowledge to extract insights from data.

Q2: What is the role of Python in Data Science?

A: Python is widely used for data analysis, visualization, and machine learning due to its rich libraries.

Q3: What is Machine Learning?

A: Machine Learning is the process of teaching computers to learn from data and make predictions.

Q4: What is a dataset?

A: A dataset is a structured collection of data, often in tabular form.

Q5: What is exploratory data analysis (EDA)?

A: EDA involves summarizing and visualizing data to understand patterns and relationships.

Q6: What is feature engineering?

A: Feature engineering is the process of transforming raw data into meaningful features for machine learning models.

Q7: What is a supervised learning model?

A: A supervised model learns using labeled data to make future predictions.

Q8: What is an unsupervised learning model?

A: Unsupervised models identify patterns in unlabeled data, such as clusters.

Q9: What is the difference between AI and ML?

A: AI is the broader concept of creating intelligent systems, while ML is a subset that uses data to learn.

Q10: What libraries are used for data visualization?

A: Common libraries include Matplotlib, Seaborn, and Plotly.

Q11: What is a neural network?

A: A neural network is a machine learning model inspired by the structure of the human brain.

Q12: What is overfitting?

A: Overfitting happens when a model performs well on training data but poorly on new data.

Q13: What is underfitting?

A: Underfitting occurs when a model is too simple and cannot capture data patterns.

Q14: What is a confusion matrix?

A: A confusion matrix is a table used to evaluate classification model performance.

Q15: What is cross■validation?

A: Cross■validation is a technique used to evaluate how well a model generalizes to unseen data.

Q16: What is Big Data?

A: Big Data refers to extremely large datasets that require advanced tools for storage and analysis.

Q17: What are SQL queries used for?

A: SQL queries are used to store, retrieve, and manipulate data in databases.

Q18: What is a data pipeline?

A: A data pipeline is an automated process for moving and transforming data between systems.

Q19: What is cloud computing in Data Science?

A: Cloud computing provides scalable resources for storing and processing data.

Q20: Do I need math for Data Science?

A: Basic knowledge of statistics, probability, and linear algebra is helpful but not mandatory at the start.