Review Questions for Data Analytics Exam II

Excel for Data Analytic I

Formula and Function / Data Preparation / Extraction

1. What is the key difference between a Formula and a Function in Excel?
2. You need to organize a customer list alphabetically by last name. Which feature do you use?
3. Describe a scenario where you would use the Filter tool instead of the Sort tool.
4. What is the primary purpose of Conditional Formatting? Give one example.
5. Which function would you use to count how many cells in a range contain any text?
6. You need to count the number of sales transactions that were over $500. Which function is best?
7. You need to count the number of sales in the "North" region that were over $500. Which function is best?
8. Your data is organized in rows, but you need it in columns. What is this operation called?
9. You have first names in column A and last names in column B. What function can combine them into a full name in column C?
10. Data imported from an external system has inconsistent spacing. Which function is essential for cleaning this text?

Excel for Data Analytic II

Data Analysis Tools
11. You have a financial model and want to find the required sales volume to achieve a specific profit target. Which tool is designed for this?
12. You want to create and compare a Best Case, Worst Case, and Most Likely Case scenario for a budget. Which tool should you use?
13. What is the main difference between a Scenario Manager and a Data Table?
14. The `RANK` function returns a number's _____ within a list.
15. The `PERCENTILE` function returns the value below which a given _____ of observations fall.

Data Warehouse

16. A Data Warehouse is optimized for _____ processing, while a database is optimized for _____ processing. (Fill in the blanks with OLAP/OLTP).
17. List the four key characteristics of a Data Warehouse.
18. What does the acronym ETL stand for? Describe each step briefly.
19. How does the ELT process differ from ETL? Why is ELT becoming more common?
20. A Data _____ is a subset of a Data Warehouse, built for a specific business unit.

Power BI

Data Preparation and Modeling

21. What is the name of the built-in tool in Power BI used for cleaning, transforming, and shaping data?
22. Why is it necessary to create relationships between tables in the Power BI Data Model?
23. What is the cardinality of a typical relationship between a 'Products' table and a 'Sales' table? (e.g., one-to-one, one-to-many).
24. What language is used to create Calculated Columns and Measures in Power BI?
25. What is the critical difference between a Calculated Column and a Measure?
26. Write a simple DAX measure to calculate the sum of a column named 'SalesAmount' in a table called 'Sales'.

Advanced Features

27. Creating a group for "Year > Quarter > Month" is an example of creating a _____.
28. What is the difference between the Drilldown and Drillthrough features?
29. Name three interactive visual elements you can add to a Power BI report to make it dynamic for the user.

Python II: Working with Libraries

30. What is the name of the primary pandas data structure for working with tabular data?
31. A single column of a pandas DataFrame is known as a _____.
32. Which pandas function is most commonly used to read data from a CSV file?
33. You have a DataFrame `df` with missing values. What method would you use to remove any rows containing missing values?

34. What is the name of the fundamental Python library providing multi-dimensional arrays for numerical computation?
35. Which Python library is the foundation for creating static, animated, and interactive visualizations?

Exploratory Data Analysis (EDA) & Feature Engineering

EDA

36. What is the main goal of performing Exploratory Data Analysis (EDA)?
37. Analyzing the distribution of a single variable like 'Age' is an example of _____ analysis.
38. Creating a scatter plot to see the relationship between 'Advertising Spend' and 'Sales' is an example of _____ analysis.
39. What is the primary purpose of using Cross-Validation (like k-Fold) in machine learning?

Feature Engineering

40. Why is Feature Engineering a critical step in the machine learning pipeline?
41. Name two common strategies for handling missing numerical data.
42. Your classification dataset has 990 records of Class 0 and 10 records of Class 1. This is an example of an _____ dataset.
43. Name one technique to handle the imbalance described in question 42.
44. Converting the categorical variable "Size" (with values: Small, Medium, Large) into three binary columns (1/0) is an example of _____ encoding.
45. Converting the same "Size" variable into numerical values like (Small=1, Medium=2, Large=3) is an example of _____ encoding.
46. Why is Feature Scaling (e.g., Standardization) critically important for algorithms like SVM and K-Nearest Neighbors?
47. What is the key difference between Normalization (Min-Max Scaling) and Standardization (Z-Score Scaling)?