Group Discussion

Focusing on the Complexity of the Data Structures:

1. **Data Silos and Redundancy:**
    - "Examine the customer and purchase tables. Do you notice any duplicated information? How might this redundancy affect the accuracy of churn predictions?"
    - "Imagine you need to update a customer's address. How many tables would you need to modify? What are the potential risks of inconsistencies arising if you don't update all the relevant tables?"
2. **Unnecessary Identifiers:**
    - "Why might 'Salesperson ID' or 'Order Number' be included in the purchase table? How might this type of information complicate churn analysis?"
    - "What are the potential consequences of having a 'Product SKU' that isn't standardized across different product tables?"
3. **Data Quality and Inconsistency:**
    - "The 'Full Name' and 'Email Address' were randomly generated. What implications does this have for data quality and reliability in the churn model? How would you assess the accuracy of this data?"
    - "How might inconsistent formats for 'Purchase Date' or 'Service Price' affect the churn prediction model?"
4. **Customer Data Privacy:**
    - "The customer table contains sensitive information like Social Security Numbers and Credit Card numbers. What are the ethical and legal considerations for handling such data? What steps should be taken to ensure data security?"
5. **Churn Driver Identification:**
    - "If you were to build a churn prediction model, what data points from the various tables would you consider most important for predicting churn? Why?"
    - "How might the presence of unnecessary data or inconsistent formats hinder the identification of true churn drivers?"
6. **Model Reliability:**
    - "How would you evaluate the reliability of a churn prediction model built using this fragmented and potentially inaccurate data?"
    - "What are the potential consequences of inaccurate churn predictions for the company's business?"

7. **Data Ownership and Responsibilities:**
    - "Who should be responsible for ensuring data quality and consistency across different systems? How can you establish clear roles and responsibilities for data management?"
8. **Data Standards and Policies:**
    - "What data standards and policies would you recommend for managing data in this environment to improve data quality and consistency?"
9. **Data Security and Compliance:**
    - "What are the key principles for data security and privacy that should be applied to this dataset? How can you ensure the company complies with relevant data protection laws?"
10. Data Culture:
    - "How can the company foster a data-driven culture that values data quality, consistency, and security?"