

EDA_VERIFY - Questions

Q1. What is the primary goal of VERIFY in the SEETREATVERIFY pipeline?

- A. To clean remaining missing values
- B. To confirm the dataset behaves like stable operational data
- C. To remove outliers
- D. To merge reference and treated data

Correct answer: B. To confirm the dataset behaves like stable operational data

Q2. Which result indicates a FAILED structural verification?

- A. OTIF = 92%
- B. Weight_kg median varies by product category
- C. Delivery_Time_hrs still contains negative values
- D. EUR/kg matches reference

Correct answer: C. Delivery_Time_hrs still contains negative values

Q3. Why compare df_treat with df_clean_reference?

- A. To copy values directly
- B. To ensure TREAT produces realistic operational behaviour
- C. To identify duplicate rows
- D. To merge two datasets

Correct answer: B. To ensure TREAT produces realistic operational behaviour

Q4. Which KPI failure suggests the dataset is NOT ready?

- A. OTIF cannot compute due to NaN in On_Time_Flag
- B. Delivery time varies by region
- C. Sales per kg differs slightly from reference
- D. Boxplot shows normal spread

Correct answer: A. OTIF cannot compute due to NaN in On_Time_Flag

Q5. Which VERIFY check ensures category-level realism?

- A. Counting rows per Carrier
- B. Grouping Weight_kg by Product_Category and comparing medians
- C. Replacing missing Notes
- D. Sorting Delivery_Time_hrs

Correct answer: B. Grouping Weight_kg by Product_Category and comparing medians