Project goal:
The goal of this project is to design a dashboard that helps a digital marketing manager understand sales performance and discount effectiveness on a furniture e-commerce website. The insights aim to support business decisions.

Dashboards overview:
Dashboard 1 “Sales overview”:
Goal- track overall sales and discount per month
Dashboard 2 “Product overview”
Goal - track sales per month on product category level.

The data has been generated syntactically to showcase data cleaning and data profiling skills.

Generate the data set. The original dataset can be found here
Steps:

1) Explore the data
2) Add primary key to each table
3) Fix date and time formatting issue
    a) Find out all the possible combinations of data formats
    b) Remove extra characters
    c) Convert the formal to one standard format yyyy-mm-dd hh:mm:ss
4) Convert the text names into one standard format - lower case
5) Fix product category names
    a) Standardize product names by renaming similar categories to one
    b) Fill in missing product categories based on product name
        i) Later I have decided to use only product category column as product name column is mostly repetition of product category.
6) Fix the amounts
    a) Removing any extra spaces and characters
    b) Standardize thousands separator by removing ‘dot’
7) Convert the data to the relevant data types
8) Handle NULLs:
    a) Find NULLs and calculate % of total to understand the completeness of data.
    b) Handle NULLs by analysing the data type and filling in values where possible. The decision overview whether to keep NULLs
        i) order_date_ts and user_id were kept as NULL since they do not affect overall calculations
        ii) product_name, product_category, and event_name were replaced with *“unknown”* to maintain data consistency
        iii) order_total was recalculated based on product price and quantity
        iv) Discount_applied_currency NULLs were replaced with *0* to indicate no discount
        v) Session_id NULLs were kept as they represent incomplete event tracking
9) Calculate additional columns for more complete data analysis. Additional columns include
    a) session_duration_min= session_start_time - session_end_time

    b) order_total = (actual_product_price * quantity) - discount
    c) Discount percentage = discount_applied_currency/ (actual_product_price*quantity)* 100
10) Create the final table to export to the visualisation tool excluding columns with NULLs such as order_date_ts, user_id, session_id, session_duration_min
11) Check accuracy and completeness of data
    a) Check duplicates
    b) Check extreme values

Insights:
Dashboard 1:

1) Sales increased from April to June and then slightly decreased from July to September.
2) Orders followed the same pattern, however, product quantity sharply increased in August but sales did not see such increase, suggesting that in August AOV is lower and quantity is the main driver for sales.
3) Average discount has been increasing over the past months (Aug-Oct). Moreover the average discount is around 18% suggesting that discount is a strong driver for sales.

Dashboard 2:

1) Products from the bedroom category and desk categories account for around 60% of total sales.
2) Product mix is concentrated on 2 categories which might be risky if one category starts underperforming. To mitigate, the company can broaden the product portfolio by adding new variants.
3) In October bedroom and couch categories had the most positive change in sales in October vs September. The most negative change has chair and desk categories.