IT in Business Lab CL1001

Project Phase II: Shopify Store Data Analytics & Dashboard

Shopify Store Name: Studio Nest

Home Decore Products

Group Members:

Syeda Alishba Raza: i25-5093
Adeena Waqas: i25-5145
Muhammad Abdullah: i25-5009
Maaz Shah: i25-5071
Syeda Ratba: i25-5029

Table of Contents

1.Introduction **3**
1.1 About Studio Nest 3
1.2 Purpose of This Project 4
2. Dataset Description **4**
2.1 Variables Included 4
3. Intentional Noise Added to the Raw Dataset **5**
3.1 Missing Values 5
3.2 Duplicate Records 5
3.3 Inconsistent Formatting 5
3.4 Outliers 6
3.5 Mixed Data Types 6
4. Data Cleaning and Preprocessing **7**
4.1 Removal of Duplicates 7
4.2 Fixing Missing Values 7
4.3 Correcting Formatting 7
4.4 Outlier Treatment 7
4.5 Derived Columns Created 8
5. Business Analytics (Based on Our BI Dashboard) **8**
5.1 Summary of KPIs 8
5.2 Sales by Category 8
5.3 Sales by City 9
5.4 Monthly Sales Trend 9
5.5 Sales by Region 9
5.6 Payment Method Analysis 9
5.7 Top Products 10
6. Dashboard Overview **10**
6.1 KPIs (4) 10
6.2 Visualizations 10
6.3 Slicers (3) 11
7. Conclusion **11**
7.1 Key Learnings 11
7.2 Recommendations 11
8. Member Contributions **12**

1. Introduction

1.1 About Studio Nest

Studio Nest, launched in 2023, is a home décor and lifestyle brand offering a curated collection of wall art, mirrors, rugs, cushions, lamps, small furniture, and decorative accessories. In addition to physical products, Studio Nest provides interior styling consultations and custom décor design services.

With growing digital demand, the brand expanded to Shopify in Phase I of this course. Phase II builds on that foundation by analyzing store-level data and creating a **professional analytical dashboard**.

1.2 Purpose of This Project

This project focuses on transforming raw Shopify-style transaction data into meaningful business insights. The goals include:

- Creating or collecting realistic store data
- Introducing controlled imperfections
- Cleaning and preprocessing the dataset
- Conducting descriptive business analytics
- Developing a Power BI dashboard for decision-making

2. Dataset Description

This report documents the full analytical workflow and final insights.

A dataset of **97 orders (197 units sold)** between **January–October 2025** was created based on Studio Nest's product catalog.

2.1 Variables Included

The raw dataset contains the following fields:

Variable	Description
Order_ID	Unique order reference number
Order_Date	Date of Transaction
Product_Name	Names of Purchased Product
Category	Mirrors, Rugs, Wall Art, Cushions, Lamps etc
Material	Wood, Metal, Cotton, Glass
Brand	Studio Nest
Price	Unit Selling Price
Quantity	Units Sold
Customer_City	City of the buyer
Region	Province/territory
Payment_Method	COD, JazzCash, Easypaisa, Credit Card, Bank Transfer
Total_Sales	Price x Quantity

Data was created to closely resemble real Shopify order exports.

3. Intentional Noise Added to the Raw Dataset

As required by the guidelines, realistic imperfections were inserted into the dataset.
Below is evidence from your raw data (copied from your table):

3.1 Missing Values

- Some rows contained missing or blank **City**, **Category**, and **Price** fields in the original file before formatting.

- Example: (In raw Excel) Category blank for some products before mapping.

3.2 Duplicate Records

At least 10 entries were intentionally duplicated.
Examples:

- Duplicate rugs and mirror transactions (e.g., Order 1013 repeated in variations).
- Multiple rows with identical Product + Price + Quantity patterns.

3.3 Inconsistent Formatting

Observed in your raw dataset:

- **City formatting inconsistencies**:
    - “Faislabad” (misspelled)
    - “Bahawalpur”, “BahawalPur”, “bahawalpur” (in earlier versions)
    - “Karachi”, “karachi”, “KARACHI”
- **Price written with spaces and currency symbols**:
    - " Rs5,500.00 "
    - "Rs 5,500"
- **Irregular Date formats**:
    - “01/09/2025”
    - “9/1/2025”
    - “09-06-2025”

3.4 Outliers

Several artificial outliers were embedded:

- Extremely high quantity on certain rows (e.g., Quantity > 4 for small items)
- Price inconsistencies for the same product (e.g., Rug prices fluctuating between 6,600 and 7,700)

3.5 Mixed Data Types

- Price and Total Sales were stored as **text**, not numeric values.
- Dates stored as text strings.

This noise ensured real-world data cleaning challenges.

4. Data Cleaning and Preprocessing

The raw dataset was cleaned using Excel and Power BI to generate the final “Clean Data” file.

4.1 Removal of Duplicates

- All duplicate rows (at least 10) removed using Excel’s “Remove Duplicates”
- Unique Order_IDs validated

4.2 Fixing Missing Values

- Missing **Price** values imputed using **median category price**
- Missing cities corrected using **customer region patterns**
- Blank categories identified through product names and assigned correctly

4.3 Correcting Formatting

- Standardized all city names using PROPER():
  "lahore" → "Lahore"

- Corrected misspellings:
  "Faislabad" → "Faisalabad"

- Dates converted from text to proper date format (DD/MM/YYYY)

- Removed spaces from price fields and converted to numeric values

4.4 Outlier Treatment

- Outliers detected using boxplots in Power BI

- Unrealistic prices and quantities capped to category norms

- Standard category price bands created

4.5 Derived Columns Created

- Total Sales recalculated:
  Total_Sales = Price × Quantity

- Month column added for time-series analysis

- Region-based segmentation created

Result:

A **fully cleaned dataset of 197 transaction units** ready for dashboard analysis.

5. Business Analytics (Based on Our BI Dashboard)

5.1 Summary of KPIs

KPI	Value
Total Orders	97

Total Sales	Rs 873,200
Units Sold	197
Unique Cities	10

5.2 Sales by Category

- **Wall Art** and **Rugs** are the top revenue-driving categories.
- **Small Furniture** also performs strongly.
- **Vases and Decor Sets** contribute minimally.

5.3 Sales by City

Top 3 performing cities:

1. **Karachi**
2. **Lahore**
3. **Peshawar**

Small but steady demand from: Multan, Rawalpindi, Islamabad, Faisalabad, Hyderabad.

5.4 Monthly Sales Trend

- Peak in **September**, indicating seasonal festival/home décor demand.
- Sharp decline in October (only early-month transactions).
- January and June have low volume.

5.5 Sales by Region

- **Sindh** contributes the highest share (50%)
- **Punjab** follows (34%)
- **Islamabad Capital** is 3rd
- **KPK** contributes the least

5.6 Payment Method Analysis

- **Credit Card (24%)** most used
- **Bank Transfer & Easypaisa** close behind
- **Cash on Delivery** is least preferred (shows digital trust)

5.7 Top Products

1. Wooden Side Table
2. Round Wall Mirror
3. Minimalist Table Lamp
4. Scented Candle Set

These are high-margin products, ideal for promotions.

6. Dashboard Overview

Your Power BI dashboard includes:

6.1 KPIs (4)

- Total Orders
- Total Sales
- Units Sold
- Unique Cities

6.2 Visualizations

- Sales by Category (bar chart)
- Sales by City
- Sales by Region
- Monthly Sales Trend
- Payment Method Distribution
- Top Products
- Category Slicer
- Additional slicers for filtering

6.3 Slicers (3)

- Category
- City
- Payment Method

All charts that we have produced respond interactively to filter selections.

7. Conclusion

This Phase II project successfully transformed Studio Nest's operational data into meaningful business insights. By applying data cleaning, preprocessing, and dashboard development techniques, the project demonstrates the value of analytics in modern e-commerce.

7.1 Key Learnings

- Real-world data requires significant cleaning before analysis.
- Power BI enables powerful trend, customer, and product insights.
- Digital payments are preferred over COD.
- Karachi and Lahore are the strongest markets for Studio Nest.

7.2 Recommendations

- Expand marketing in Sindh and Punjab
- Focus on high-performing categories (Rugs, Wall Art, Small Furniture)
- Strengthen October and early-year sales through promotions
- Enhance SEO and integrate Instagram Shopping
- Build customer loyalty programs using Shopify plugins

8. Member Contributions

Muhammad Abdullah (i25-5009)

- Remove duplicate records
- Handle missing values (imputation, removal, or estimation)
- Normalize numerical values
- Introduce small random text variations into categorical fields
- Simulate missing values for stress-testing
- Create bar charts for category comparisons

Syeda Ratba (i25-5029)

- Calculate basic statistics (average, minimum, maximum, etc.)
- Identify trends or patterns in the data
- Find correlations between important variables
- Display key visuals such as sales trends or user activity
- Connect the dashboard to the cleaned dataset

Maaz Shah (i25-5071)

- Present key insights and visuals
- Speak about challenges and solutions
- Add small random errors to test how the system handles bad data
- Create fake missing values to check the cleaning process
- Slightly change some numbers or text to test data quality tools

Adeena Waqas (i25-5145)

- Write a summary of what the data shows
- Describe the steps taken during cleaning and analysis
- Add tables, tools for summary etc
- Provide conclusions and recommendations
- Add filters (like date or category filters)

S. Alishba Raza (i25-5093)

- Design slides explaining each phase of the project
- Show before/after examples of cleaned data
- Make bar charts, line charts, and pie charts to visualize results
- Format charts for readability (titles, labels, colors)
- Data Analysis for presentation