# Task One – Online Retail Dataset Analysis (Nexus AI Internship)

### Introduction

This project is part of my Nexus AI Internship (Task One).  
The objective was to analyze the Online Retail II dataset to clean and prepare the data, perform descriptive statistics, and generate insights using Python (Pandas, Matplotlib, Seaborn).

### Steps Performed

**1. Data Loading & Inspection**

* Loaded dataset (online\_retail\_II.xlsx).
* Checked dataset shape (525,461 rows × 8 columns).
* Inspected column names and data types using df.info().

**2. Data Cleaning**

* Checked and reported missing values.
* Dropped rows with missing values.
* Removed duplicate records.
* Corrected column data types (InvoiceDate → datetime, Customer ID → Int64).

**3. Feature Engineering**

* Created a new column Sales = Quantity × Price.

**4. Descriptive Statistics**

* Calculated Mean, Median, Mode, and Standard Deviation for Quantity and Sales.
* Found that most transactions involved small quantities and low sales values, with a few high-value outliers.

**5. Visualizations**

* Histogram → Distribution of Sales (most sales small, some refunds, few large outliers).
* Bar Chart → Top 10 products by revenue.
* Line Chart → Daily sales trend over time.
* Bar Chart → Sales distribution by country.

### Key Insights

1. **Sales Distribution** → Most transactions are of small value; a few large outliers exist; negative values indicate refunds.
2. **Top Products by Sales** → A few products (e.g., Regency Cakestand 3 Tier) contribute a large share of revenue.
3. **Sales Trend Over Time** → Sales were volatile in 2010 but peaked strongly in Q4 (holiday season).
4. **Sales by Country** → The United Kingdom dominates sales; other countries contribute smaller portions.
5. **Top Customers by Sales** → A small number of customers generate most revenue, showing customer concentration.

### Tools Used

* **Python**: Pandas, NumPy, Matplotlib, Seaborn
* **Kaggle Notebook**: Data cleaning, analysis, visualization

### Learning Outcome

Through this task, I improved my skills in:

* Handling missing values, duplicates, and data type corrections.
* Creating new features like Sales.
* Using descriptive statistics to understand data distribution.
* Building effective visualizations for business insights.