# E-Commerce Sales Analysis & Customer Insights Dashboard

**Intern: Anshul Kumawat**
**Domain: Data Analyst**
**Organization: Uniconverge Technology Pvt. Ltd.**
**Duration: 20 Days**

## Introduction

**This report presents the complete 20-day capstone project conducted as part of my Data Analyst internship. The project focused on analyzing a large-scale e-commerce dataset to extract insights using SQL, Python, Machine Learning, Excel, and BI tools like Power BI and Tableau.**

## Data Source

**The dataset consisted of 500,000+ transactional records containing:**

**- Customer information**

**- Product details**

**- Order quantities and pricing**

**- Purchase timestamps**

**- Country and region information**

**The data was provided as CSV files, imported into SQL databases, and further processed using Python and Excel.**

## Phase 1: Data Acquisition & Exploration

**Tools: SQL (MySQL/PostgreSQL), Python (Pandas), Jupyter Notebook**

**- Imported raw dataset into SQL**

**- Executed SQL queries (joins, CTEs, window functions)**

**- Conducted Exploratory Data Analysis (EDA)**

**- Identified missing values, anomalies, and data distributions**

## Phase 2: Data Cleaning & Transformation

**- Removed duplicates and handled missing values**

- Conducted outlier detection using IQR and Z-score

- Performed feature engineering: CLV, RFM, seasonality features

- Cleaned categorical text with regex

- Created aggregated tables for further analysis

### Phase 3: Statistical Analysis

- Generated descriptive statistics

- Built correlation matrix

- Performed hypothesis tests including:

 * Weekend vs Weekday sales (T-test)

 * Country influence on sales (ANOVA / Chi-square)

- Conducted time-series decomposition

### Phase 4: Data Visualization

Tools: Python (Matplotlib, Seaborn), Power BI,

- Created heatmaps, pair plots, time-series graphs

- Developed interactive dashboards with KPIs

- Used Power BI DAX measures for dynamic calculations

### Phase 5: Advanced Analytics & Machine Learning

Algorithms Used:

- K-Means Clustering (Customer Segmentation via RFM)

- Linear Regression (Sales Prediction)

- Logistic Regression (Customer Churn Classification)

- Decision Tree (Product Category Classification)

Model Evaluation Metrics:

- RMSE, $R^2$, Accuracy, Precision, Recall

### Phase 6: Excel Analysis & Final Deliverables

- Created pivot tables and dashboards

- Applied VLOOKUP, INDEX-MATCH, SUMIFS

- Automated tasks using VBA macro

- Conducted What-If analysis

- Compiled 15–20 page report

- Created 10–12 slide presentation

## Conclusion

This project provided hands-on experience across the full data analytics pipeline: data acquisition, cleaning, analysis, modeling, visualization, and reporting. The insights generated can support decision-making in e-commerce businesses and strengthen customer understanding.