Sales Performance Analysis Project Report

This project analyzes the Superstore Sales dataset to uncover trends, anomalies, and insights that align with the Google Data Analytics Apprenticeship JD responsibilities. It demonstrates SQL querying, Python analysis, anomaly detection, data visualization, and dashboarding with Tableau. The objective was to assess data accurately, identify anomalies, and quantify business impact across regions, categories, and customers.

Dataset: Kaggle - Superstore Sales Dataset **Size:** ~10,000+ records **Fields:** Customer, Sales, Profit, Region, Category, Sub-Category, Order Date

Key SQL Insights:

Query	Result / Insight
Total Revenue	≈ $2.29M across dataset
Top Customer by Sales	Sean Miller ($25,043)
Regional Sales & Profit	West ($725k revenue, $108k profit) leads all regions
Yearly Revenue Trend	2014–2017 steady growth, peak in 2017 ($733k)

Visual Insights:

- Profitability varies significantly by region: West outperforms South and Central. - Top 5 customers account for a significant share of sales. - Discount strategies negatively impacted profit margins in some sub-categories. - Sales show strong upward trend over 2014–2017, proving business growth trajectory.

Business Impact:

- **Quantified Growth:** Revenue grew from $484k in 2014 to $733k in 2017 (51% growth). - **Customer Segmentation:** High-value customers like Sean Miller drive $25k+ sales individually. - **Regional Strategy:** West region contributes ~32% of total revenue, should remain a strategic focus. - **Profit Margins:** Some sub-categories show profit margins below 5%, requiring pricing review. - **Anomalies:** Detected sudden spikes in discounts that caused negative profit margins in Office Supplies.

This project successfully fulfills the responsibilities outlined in the Google JD: ✔ Data accuracy & anomaly detection using SQL & Python ✔ Knowledge sharing via documentation ✔ Custom analysis to uncover industry insights ✔ Quantified business impact across regions, categories, and customers ✔ Practice of the full data lifecycle with SQL, Python, Excel, and Tableau