

**ID: 26134**

**NAME: Philemon MUVUNYI**

**COURSE: Introduction to Bigdata**

**Formal Report: Sales Data Analysis**

**1. Introduction**

This report presents an analysis of sales data to understand key trends and patterns in sales performance. The dataset includes information on sales transactions, including salesperson details, country, product, date, amount, and boxes shipped. The analysis aims to provide insights into sales distribution, performance metrics, and potential areas for improvement.

**2. Dataset Overview**

* Basic Information
* Rows: 1,000+ (exact count depends on dataset)

**Columns:**

* Sales Person (Categorical)
* Country (Categorical)
* Product (Categorical)
* Date (Datetime)
* Amount (Numerical, in USD)
* Boxes Shipped (Numerical)

**Key Observations**

* No missing values detected.
* No duplicates found.
* Data types correctly assigned**.**

**Data Summary**

* **Boxes Shipped:**
* Mean: 161.80
* Standard Deviation: 121.54
* Minimum: 1
* Maximum: 709

**3. Analysis and Findings**

**3.1 Sales Distribution by Country**

A box plot was generated to visualize the distribution of boxes shipped across different countries. Key observations include:

* Australia and India show higher variability in the number of boxes shipped.
* The UK and USA have relatively consistent sales volumes.
* New Zealand has fewer transactions but a moderate number of boxes shipped**.**

A graph of blue bars

AI-generated content may be incorrect.

* 1. **Statistical Summary**
* The average number of boxes shipped per transaction is **161.80**, with a wide range (1 to 709).
* The **shipping rate** (mean boxes shipped) is **16,179.80%**, indicating high variability in shipment sizes.
  1. **Data Quality**
* No missing values were detected in the dataset.
* The "Amount" column was successfully cleaned and converted to a numerical format for analysis.

**4. Conclusions**

1. **Sales Performance**: The dataset reveals significant variability in sales performance across countries, with Australia and India showing the highest variability in boxes shipped.
2. **Data Quality**: The dataset is clean and well-structured, with no missing values. The "Amount" column was cleaned for numerical analysis.
3. **Potential Insights**: Further analysis could explore:

* Correlation between "Amount" and "Boxes Shipped."
* Seasonal trends in sales.
* Performance of individual salespersons or products.

**5. Recommendations**

1. **Targeted Strategies**: Focus on countries with high variability (e.g., Australia and India) to understand factors driving sales fluctuations.
2. **Performance Metrics**: Track salesperson performance and product popularity to identify top performers and underperforming areas.
3. **Forecasting**: Use the dataset to build sales forecasting models, leveraging date and amount data for trend analysis.

**Prepared by**: Philemon MUVUNYI  
**Date**: 2nd, July 2025  
**Contact**: muvunyiphilemon22@gmail.com