Blinkit Sales Data Analysis Project

Excel Data Analysis & Dashboarding Project
Prepared by: Ashish Jangir
Role: Data Analyst

1. Executive Summary

This project presents a comprehensive Excel-based data analysis performed on the Blinkit grocery sales dataset sourced from Kaggle. The primary objective of this project was to clean raw transactional data, identify data quality issues, perform exploratory data analysis, and develop an interactive dashboard to generate actionable business insights. The analysis enables stakeholders to understand sales performance, customer behavior, payment preferences, and store-level efficiency.

2. Introduction

In the modern retail and e-commerce industry, data-driven decision-making plays a crucial role in improving operational efficiency and customer satisfaction. Blinkit, as a last-minute grocery delivery platform, generates large volumes of transactional data across multiple cities, store types, and customer segments. However, raw data often contains inconsistencies, formatting issues, and errors that reduce analytical accuracy. This project demonstrates how Microsoft Excel can be effectively used to transform raw data into meaningful business insights.

3. Project Objectives

The key objectives of this project are:
- To identify and resolve data quality issues in raw datasets
- To perform structured data cleaning and transformation in Excel
- To analyze sales trends, customer demographics, and store performance
- To design an interactive Excel dashboard for business reporting
- To generate insights that support strategic decision-making

4. Dataset Description

The dataset used in this project was obtained from Kaggle and represents simulated Blinkit grocery sales data. It consists of multiple interrelated tables designed to mimic a real-world retail data model.

Tables Included:
1. Products_Details – Product name, category, brand, unit size, and MRP
2. Customers_Details – Customer demographics such as city, gender, and age group
3. Stores_Details – Store name, location, and store type
4. Orders_Details – Order date, quantity, selling price, discount, final price, and payment method

5. Data Collection Source (Kaggle)

The dataset was downloaded from Kaggle, a widely used platform for data science and analytics practice. To simulate real-world data challenges, one row of unclean and messy data was intentionally introduced using AI-assisted modifications. This approach ensures that the project reflects realistic data quality issues encountered in corporate environments.

6. Data Quality Challenges

During the initial data exploration phase, several data quality issues were identified:
- Presence of one intentionally messy and unclean data row
- Inconsistent date formats across records
- Currency symbols embedded within numeric fields
- Inconsistent text formatting and spacing
- Potential duplicate or mismatched IDs across tables

If left unresolved, these issues could significantly impact KPI calculations and dashboard accuracy.

7. Data Cleaning Methodology

A systematic data cleaning process was implemented using Microsoft Excel. Various Excel features and functions were applied to ensure data accuracy and consistency.

Key Cleaning Techniques:
- Text to Columns for separating combined values
- TRIM and CLEAN functions to remove extra spaces and non-printable characters
- Standardization of date formats
- Removal of currency symbols from numeric fields
- Data validation and filtering to identify anomalies
- Referential integrity checks between tables using IDs

8. Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted using Pivot Tables and Pivot Charts in Excel. The purpose of EDA was to summarize the dataset, identify trends, and uncover patterns that influence business performance.

EDA Focus Areas:
- Sales performance by product category
- Sales distribution across cities
- Customer demographics and age groups
- Payment method preferences
- Quarterly sales trends
- Store type-wise performance

9. Dashboard Development

An interactive Excel dashboard was designed to present insights in a visually intuitive manner. The dashboard integrates multiple KPIs, charts, and slicers, allowing users to filter data dynamically.

Dashboard Components:
- KPI cards for Total Sales, Average Sales, Total Orders, and Average Discount
- Donut charts for payment type and gender distribution
- Bar charts for category-wise and city-wise sales
- Line chart for quarterly sales trend
- Slicers for Category, City, Gender, and Order Date

10. Key Business Insights

The following key insights were derived from the analysis:

1. Product Category Insights:
- Fruits and Vegetables generate the highest total sales, indicating strong customer demand.
- Bakery and Beverages contribute comparatively lower revenue and may require targeted promotions.

2. Customer Demographics Insights:
- The 25–34 age group places the highest number of orders, making it the most valuable customer segment.
- Both male and female customers contribute almost equally to total sales, indicating balanced engagement.

3. Payment Method Insights:
- Digital payment methods such as Wallet and UPI dominate transactions.
- Cash payments are comparatively lower, reflecting a shift towards digital adoption.

4. City-wise Insights:
- Tier-1 cities such as Bangalore, Chennai, and Delhi generate the highest sales volume.
- Sales distribution across cities is relatively balanced, suggesting widespread market penetration.

5. Store Type Insights:
- Warehouse stores contribute the highest revenue among all store types.
- Dark Stores and Partner Stores show similar order volumes but slightly lower revenue.

6. Time-based Insights:
- Sales peak during the third quarter (Q3), indicating seasonal or promotional impact.

11. Business Recommendations

Based on the insights obtained, the following recommendations are proposed:
- Increase inventory and marketing focus on high-performing categories such as Fruits and Vegetables
- Introduce targeted campaigns for the 25–34 age group
- Promote digital payment methods through exclusive offers
- Optimize warehouse operations to support high-demand periods
- Launch city-specific promotional strategies

12. Limitations of the Project

Despite providing valuable insights, this project has certain limitations:
- The dataset is simulated and may not reflect real-time business conditions
- External factors such as supply chain disruptions were not considered
- Advanced predictive modeling was not included

13. Future Scope

The scope of this project can be extended in the future by:
- Integrating Power BI for advanced visualization
- Implementing automated data cleaning using Power Query
- Performing predictive sales forecasting
- Developing customer segmentation models

14. Conclusion

This project demonstrates a complete end-to-end Excel data analysis workflow. It highlights strong data cleaning, analytical, and dashboarding skills while addressing real-world data

challenges. The insights generated through this analysis can help Blinkit improve sales strategies, customer targeting, and operational efficiency.

15. Skills & Tools Used

- Microsoft Excel (Advanced)
- Data Cleaning & Transformation
- Pivot Tables & Charts
- Dashboard Design
- Business Analysis
- Analytical Thinking
- Data Storytelling