

STRATEGIC DECISION-MAKING USING Power-BI

PRE-REPORT SUBMISSION

Shivam Chandra Shukla

2023JULB01368

Table of Contents

Table of Figures	1
1. PROBLEM STATEMENT	2-4
2. DATA REQUIREMENT	5-6
3. DATA COLLECTION.....	7
4. DATA VALIDATION	8-9
5. DATA CLEANING	10-12
6. TOOLS.....	13
7. DASHBOARD.....	14-16
8. STORYTELLING.....	17-19

1. PROBLEM STATEMENT

Blinkit, a quick-commerce platform, aims to understand and enhance its outlet performance, product sales trends, and customer purchasing behaviours. The dashboard helps in analysing outlet-level performance, item-level sales, and the impact of location and product attributes on overall sales.

Objective

1. To provide actionable insights into sales performance across outlets and product categories.
2. To identify sales trends based on outlet size, location, and product characteristics.
3. To evaluate the distribution of sales among different outlets and locations.
4. To highlight top-performing outlets, categories, and areas for improvement.

Proposed Approach

1. Data Aggregation and Processing:

- Import Blinkit data into Power BI.
- Clean and preprocess the data for consistency.
- Create calculated fields for KPIs like total sales, average sales, and sales contribution percentages.

2. Interactive Visualizations:

- **Use Cards to summarize key KPIs:**

- Total Sales: 18.56M.
- Total Outlets: 10.
- **Apply Slicers for:**
 - Outlet Size (High, Medium).
 - Item Fat Content (Low Fat, Regular).
 - Outlet Location (Tier 1, Tier 2, Tier 3).
- **Visualize sales trends and patterns using:**
 - Pie Chart for sales distribution by outlet location type.
 - Scatter Chart for sales vs. item pricing (MRP).
 - Stacked Column Chart for sales by outlet identifier.
 - Donut Chart for sales by item fat content.
 - Gauge for average sales per outlet.
 - Funnel for sales contribution by location type.
 - Line Chart for sales trends across location types.
 - Matrix Table for the top 5 item types with the highest sales.
 - Tree map for outlet-wise sales distribution.

3. Analysis of Key Drivers:

- Investigate correlations between item MRP and sales.
- Analyse location-based trends in sales.
- Identify high-performing outlets and top-selling product categories.

4. Iterative Improvements:

- Update the dashboard regularly with new data.
- Refine visualizations based on user feedback and emerging business needs.

Expected Outcomes

1. Enhanced Decision-Making:

- Clear identification of top-performing outlets, locations, and product categories.
- Insights into the impact of outlet size, location, and product attributes on sales.

2. Operational Efficiency:

- Data-driven strategies to boost sales and improve outlet performance.

3. Business Growth:

- Targeted marketing and operational strategies tailored to high-performing areas and products.

4. Increased Engagement:

- Intuitive, interactive visualizations that empower stakeholders to explore key metrics.

2. DATA REQUIREMENT

	A	B	C	D	E	F	G	H	I	J	K	L
1	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales
2	FDA15	9.3	Low Fat	0.016047301	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.138
3	DRC01	5.92	Regular	0.019278216	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228
4	FDN15	17.5	Low Fat	0.016760075	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.27
5	FDX07	19.2	Regular	0	Fruits and Vegetables	182.095	OUT010	1998		Tier 3	Grocery Store	732.38

Source: Kaggle Dataset

(<https://www.kaggle.com/datasets/irshad9322/blinkit-sales-project-on-powerbi>)

The data contains the following columns:

1. **Item_Identifier:** Unique product codes.
2. **Item_Weight:** Weight of the item.
3. **Item_Fat_Content:** Fat content category of the item (e.g., Low Fat, Regular).
4. **Item_Visibility:** The percentage of display area allocated to the item in the store.
5. **Item_Type:** General category to which the item belongs (e.g., Dairy, Meat).
6. **Item_MRP:** Maximum retail price of the item.
7. **Outlet_Identifier:** Unique store codes.

8. **Outlet_Establishment_Year:** Year the outlet was established.
9. **Outlet_Size:** Size category of the outlet (e.g., Medium, High).
10. **Outlet_Location_Type:** Store location type (e.g., Tier 1, Tier 3).
11. **Outlet_Type:** Type of outlet (e.g., Supermarket Type1, Grocery Store).
12. **Item_Outlet_Sales:** Sales of the item in the outlet.
13. **Outlet_Age:** Calculated age of the outlet based on the establishment year.

3. DATA COLLECTION

For the **Blinkit Sales and Performance Dashboard**, data collection is a critical step that ensures all necessary information is gathered accurately and completely. Here's how it applies:

Source of Data

1. Internal Business Systems:

- Outlet-wise sales data collected from Blinkit's internal sales databases.
- Product-related details sourced from inventory management systems.

2. Customer Interaction Data:

- Data on product visibility and pricing from Blinkit's e-commerce platform.

3. Third-party Data (if applicable):

- Market trends and customer preferences from external sources such as market research reports.

4. Data from the Uploaded Excel File:

- Contains:
 - Product attributes (*Item Identifier, Item Type, Item Fat Content, Item MRP, etc.*).
 - Sales figures (*Item Outlet Sales*).
 - Outlet details (*Outlet Identifier, Outlet Location Type, Outlet Size, Outlet Establishment Year*).

4. DATA VALIDATION

To ensure the accuracy and reliability of the collected data, a robust validation process is crucial. Here's how it can be performed:

1. Check for Missing Data

- Identify null or blank fields, particularly in critical columns such as:
 - *Item MRP*.
 - *Item Outlet Sales*.
 - *Outlet Size*.
- Address missing values using:
 - Imputation (e.g., average sales for missing sales data).
 - Discarding rows with excessive missing data if they cannot be reasonably filled.

2. Validate Data Types

- Ensure the consistency of data types:
 - Numeric fields (*Item MRP*, *Item Outlet Sales*) should not contain text or special characters.
 - Categorical fields (*Outlet Size*, *Item Fat Content*) should match predefined categories.

3. Remove Duplicates

- Check for and remove duplicate records, especially for *Item Identifier* and *Outlet Identifier*.

4. Outlier Detection

- Identify outliers in key metrics like:
 - *Item Outlet Sales*: Extremely high or low sales figures.
 - *Item MRP*: Prices significantly deviating from the typical range for a category.
- Investigate and validate outliers with business context.

5. Validate Relationships

- Verify logical consistency:
 - Sales figures (*Item Outlet Sales*) should align with *Item MRP*

and outlet attributes (*Outlet Type*).

- Ensure outlet-specific attributes (*Outlet Size*, *Outlet Location Type*) are consistent with the outlet identifier.

6. Cross-check with Source

- Compare collected data against the original source systems or reports to ensure accuracy.

7. Data Transformation

- Convert raw data into a usable format:
 - Calculate *Outlet Age* from the *Outlet Establishment Year*.
 - Standardize category names (e.g., unify variations like "Low fat" and "low fat").
- Create derived columns/measures if needed (e.g., average sales per outlet).

5. DATA CLEANING

Data Cleaning Process Using Julius AI

1. Importing Data

- Uploaded the raw Blinkit dataset into Julius AI for processing.
- Defined key data columns for the tool to focus on:
 - *Item Identifier*, *Item MRP*, *Item Outlet Sales*, *Outlet Size*, etc.

2. Handling Missing Data

- Julius AI identified missing values in critical fields (e.g., *Item Outlet Sales*, *Outlet Size*).
- Automated Imputation:
 - For numeric fields (*Item Outlet Sales*), Julius applied mean or median imputation.
 - For categorical fields (*Outlet Size*), missing values were filled based on the mode or using machine learning-based predictions.

3. Standardizing Data

- Standardized category labels to ensure consistency:
 - Unified case sensitivity (e.g., "Low Fat" vs. "low fat").
 - Merged similar categories (e.g., "Tier 1" and "Tier-1").
- Converted data types to align with field requirements:
 - Numeric for *Item MRP* and *Item Outlet Sales*.
 - Text for *Item Identifier* and *Outlet Location Type*.

4. Removing Duplicates

- Detected and removed duplicate rows where the same *Item Identifier* and *Outlet Identifier* appeared with identical sales values.

5. Outlier Detection

- Julius AI flagged outliers in:
 - *Item MRP* (e.g., exceptionally high or low prices).
 - *Item Outlet Sales* (e.g., sales figures significantly deviating from the norm).
- Retained valid outliers justified by business logic and corrected erroneous ones.

6. Validating Relationships

- Verified logical relationships using Julius AI algorithms:
 - Ensured that *Item Outlet Sales* were proportional to *Item MRP* for similar products.
 - Cross-validated *Outlet Size* and *Outlet Location Type* against known data patterns.

7. Exporting the Clean Dataset

- The cleaned data was exported directly from Julius AI into a CSV/Excel format for use in Power BI.

Benefits of Using Julius AI

1. Time Efficiency:

- Automated cleaning reduced manual effort and accelerated the data preparation process.

2. Accuracy:

- AI-driven imputation and error detection ensured a high-quality dataset with minimal discrepancies.

3. Scalability:

- Easily handled large volumes of Blinkit data with complex relationships.

4. Consistency:

- Standardized outputs ensured the data was ready for seamless integration with Power BI.

6. TOOLS

Tools to be Used for Analysis and Visualization:

1. **Power BI:** For creating interactive dashboards and visualizations.
2. **Microsoft Excel:** For data cleaning, validation, and basic exploratory analysis.

7. DASHBOARD

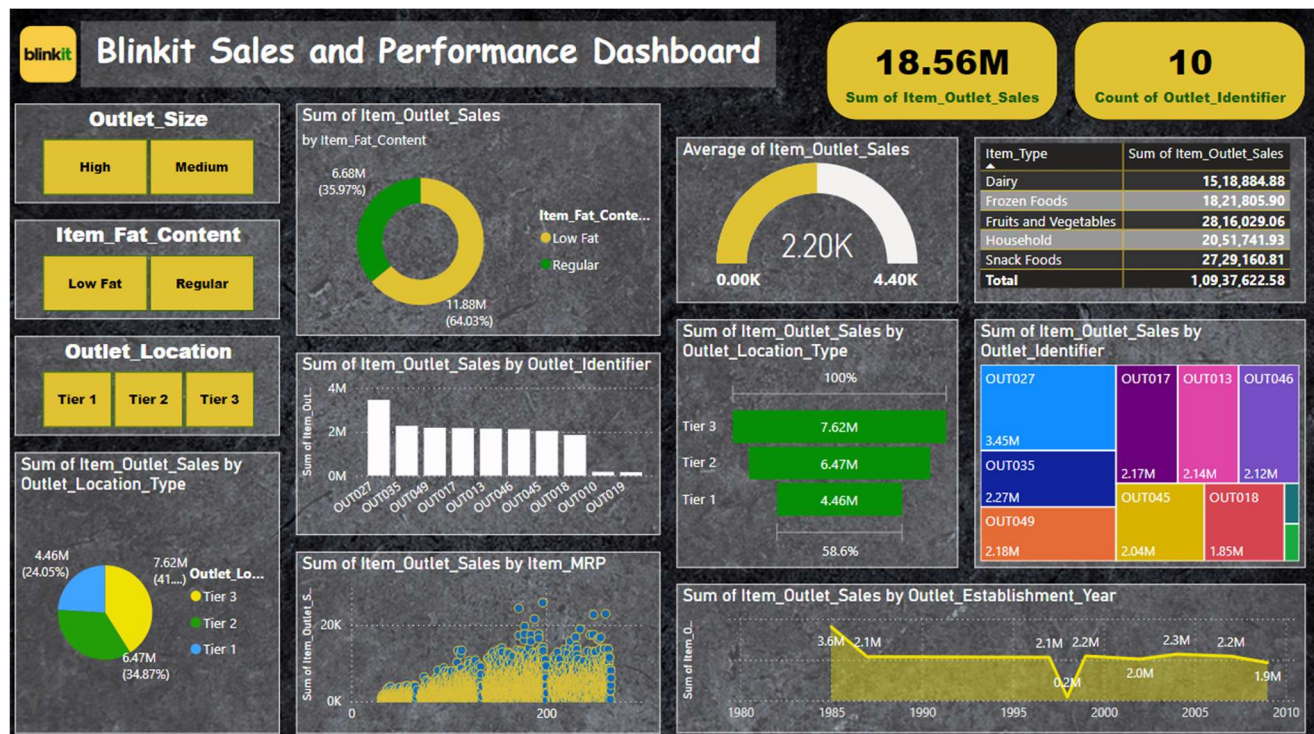


Figure 2, Demo Sales Dashboard

Key Components of the Power BI Dashboard

The Blinkit Sales and Performance Dashboard uses various visualizations to provide insights into sales performance and outlet characteristics. Here's a breakdown of these components, segregated into Univariate, Bivariate, and Multivariate analyses:

1. Univariate Analysis (Single Variable Insights)

These components analyse one variable at a time to understand its distribution or magnitude.

- **Cards:**
 - Sum of Outlet Sales: 18.56M.
 - Count of Outlet Identifiers: 10.
- **Gauge:**
 - Average of Item Outlet Sales: 2.20K.

- **Donut Chart:**
 - Distribution of Item Fat Content (*Low Fat: 64.03%, High Fat: 35.97%*).
- **Funnel Chart:**
 - Sum of Item Outlet Sales by Outlet Location Type.
- **Pie Chart:**
 - Sum of Item Outlet Sales by Outlet Location Type (Tier 1, Tier 2, Tier 3).

2. Bivariate Analysis (Two Variables)

These components explore the relationship between two variables to identify patterns or correlations.

- **Scatter Chart:**
 - *Sum of Item Outlet Sales by Item MRP*: Analysing the relationship between pricing and sales.
- **Stacked Column Chart:**
 - *Sum of Item Outlet Sales by Outlet Identifier*: Evaluating sales performance across different outlets.
- **Line Chart:**
 - *Sum of Item Outlet Sales by Outlet Location Type*: Understanding sales trends across locations over time.

3. Multivariate Analysis (Multiple Variables)

These components examine the interplay of multiple variables simultaneously to uncover deeper insights.

- **Table/Matrix:**
 - Displays the *Top 5 Item Types* with their *Sum of Item Outlet Sales*, enabling analysis of categories with the highest contributions.
- **Tree map:**
 - Visualizes the *Sum of Item Outlet Sales by Outlet Identifier* to

compare sales contributions across outlets.

- **Slicers (Interactive Filters):**

- **Outlet Size (*High, Medium*):** Filters based on outlet size.
- **Item Fat Content (*Low Fat, Regular*):** Filters based on product fat content.
- **Outlet Location (*Tier 1, Tier 2, Tier 3*):** Filters based on location.

Summary Table

Analysis

Type

Dashboard Component Purpose

Univariate	Cards, Gauge, Donut	Provides an overview of key metrics and single-variable distributions.
	Chart, Funnel Chart, Pie Chart	
Bivariate	Scatter Chart, Stacked	Analyses relationships between two variables (e.g., sales by price, sales by location).
	Column Chart, Line Chart	
Multivariate	Table/Matrix, Tree map, Slicers	Explores the interaction of multiple dimensions (e.g., item type, outlet identifier, location).

8. STORYTELLING

Introduction

Blinkit, a leading player in the quick commerce industry, aims to optimize sales performance and enhance customer satisfaction through data-driven insights. This Power BI dashboard leverages sales and outlet data to uncover trends, identify challenges, and provide actionable recommendations for improved business operations.

Challenges Identified

1. Uneven Sales Distribution:

- Significant variations in *Item Outlet Sales* across different *Outlet Identifiers* and *Outlet Location Types*.

2. Impact of Pricing:

- A potential lack of correlation between *Item MRP* and *Item Outlet Sales* indicates suboptimal pricing strategies for certain products.

3. Outlet Performance Disparities:

- Outlets of varying sizes (*High, Medium*) demonstrate inconsistent sales performance despite similar resources.

4. Consumer Preference Variability:

- Sales are influenced by *Item Fat Content* preferences (e.g., Low Fat dominates but High Fat holds a significant share).

Key Insights

1. Sales Trends by Location:

- Tier 1 locations contribute the largest share of sales, followed by Tier 2 and Tier 3.

2. Product Attributes:

- *Low Fat* items generate 64% of total sales, showing a clear consumer preference.

3. Outlet-Specific Insights:

- Certain outlets underperform, requiring targeted interventions to boost their contribution to overall sales.

4. Pricing Sensitivity:

- Scatter chart analysis reveals that moderately priced items achieve better sales volume compared to extremely high-priced products.

Recommendations

1. Optimize Outlet Strategies:

- Focus on underperforming outlets through tailored marketing and operational support.

2. Dynamic Pricing:

- Implement tiered pricing models based on consumer sensitivity and location-specific trends.

3. Product Assortment:

- Expand the range of *Low Fat* products to cater to prevailing consumer preferences.

4. Location Prioritization:

- Enhance infrastructure and marketing in Tier 2 and Tier 3 cities to unlock untapped potential.