**TOY SALES ANALYTICS**

**BUSINESS PROBLEM**

Angani company wants to better understand its toy sales performance and gain actionable insights that can help in improving sales. The manager is looking for a data-driven way to make strategic decisions. This project aims to create a dashboard that provides clear and intuitive insights.

Key questions:

1. What are the sales trends over time?
2. Which products perform best and which are underperforming?
3. Who are the most valuable customers?
4. What actionable insights can help the company improve sales?

**Objective**:

Main objective:

* Create a dynamic, intuitive power BI dashboard that consolidates these insights.

This dashboard will help the manager:

* Track overall sales trends over time by region and branch
* Identify top-selling and underperforming products by overall category performance
* Identify high-value performance and geographical sales distribution.
* Use actionable insights to make informed decisions.

**DATA UNDERSTANDING**

This dataset consists of three main components:

1. Main Data - contains individual sales transaction records.
2. Product Master - contains information on the different products available for sale linking to the Main Data via *productCode.*
3. Data Dictionary - provides description for each field across *Main Data* and *Product Master.*

The main data contains columns that provide:

* Order details
* Temporal information
* Product information
* Customer information
* Branch information

The data has 22 columns and 2 columns contain missing values:

* postalCode - contains 76 missing values
* Territory - contains 1074 missing values

**DATA PREPARATION**

1. **Handling missing values in the Territory column**

To address the missing values, we decided to impute them based on the *country* column. For entries where the *country* is USA or Canada and *Territory* is missing, we set *Territory* to NA (North America).

1. **Handling missing values in the postalCode column**

We decided to fill these missing values with 0 as a place holder, representing unknown postal codes.

1. **Combining contact first and last names**

We combined *contactFirstName* and *contactLastName* into a new column *contactName.* After merging, we removed the original *contactFirstName* and *contactLastName* columns to keep the data streamlined.

1. **Creating a Single Date Column**

We combined the *YEAR\_ID*, *MONTH\_ID* and *DAY\_ID* into a new column *Date*. After merging we removed the original columns to reduce redundancy.

1. **Merging datasets**

To enrich the *Main Data* with product-specific details, we performed a merge with the *Product Master* dataset using the *productCode* column as the key.

* Merge Type: we used an inner join to include only records with a matching *productcode* in both datasets.
* The merged dataset now combines the sales transaction data with additional product details. This merged dataset, *merged\_df*, will be the foundation for further exploration and analysis.

**Expected Visuals for Analysis**

To gain insights into our data and assist in problem solving, we plan to create several key visualisations. These visuals will be used to identify patterns and guide decision-making for the dashboard. Some of the visuals include:

1. Distribution of Sales - Understand the overall distribution of sales values, identify outliers and see how sales vary across different orders.
2. Sales by Territory - Compare the total sales across territories to see which regions are leading in revenue.
3. Product line distribution - Understand the popularity and contribution of each product line to total sales.
4. Sales Over Time - Track sales trends over time to spot seasonality and growth patterns.
5. Sales by Branch - Evaluate how each branch contributes to the company total sales