**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, *Sales*, will be the foundation for further exploration and analysis.
* Also we confirmed on the shape, columns, full information, whole description and the null values of the *sales* data.

**DATA PREPROCESSING**

In this step we decided to add calculated columns for more business insights.

**1. Checking the numeric columns**

First, we began by checking the numeric columns which will help in identifying the fields to use for calculations, visualizations and key performance indicators(KPIs)

**2. Calculating the Total costs**We calculated the *TotalCosts* which is the value of *QuantityOrdered* by the *CostPrice.* This column represents generated revenue from purchasing a specific quantity of a product before you sell it to the customers.

This will be helpful for calculating profitability.

**3. Calculating the Total Sales Amount**

We decided to calculate the *TotalSalesAmount* which is the value of *QuantityOrdered* by the *PriceEach* of each product. This column represent the revenue generated from selling a specific quantity of a product at a given price.

This is essential for understanding overall sales performance and will help in calculation of the profit.

**4. Calculating profit**

To get the *profit* we took the difference between *TotalsalesAmount* and *TotalCosts.* This will be crucial for assessing profitability and performance.

**5. Calculating profit margin**

We changed the profit into percentage through dividing *Profit* by *TotalCost* and then multiplying by 100. This allowed us to assess the profitability of our sales relative to our costs.

**6. Calculating total discounted**

It showed us how much revenue was lost due to discounts. This was calculated by *MSRP* multiply by *QuantityOrdered* minus the *TotalSalesAmount.*

**EXPLORATORY DATA ANALYSIS (EDA)**

The following visual analysis were made to be used to identify patterns and guide decision-making for the dashboard:

**1. Sales by Product Line Analysis**

We created a bar chart which visualize the *total sales* amount by *productline.*

This visualization helped in identifying the top performing product lines. We identified that *classic cars* have the most sales.

**2. Consumption of the products by each country**

We identified the top 5 countries by product line sales. We used a grouped bar plot where by *countries* are on the x-axis and *product lines* on the y-axis being represented by different colors.

This visualization enabled a clear comparison of sales performance for each product line within the top 5 countries, showing the strongest product line in each country.

**3. Product Line Performance Across Territories**

We used a bar plot to visualize sales data, with each *productline* being displayed on the x-axis and the *total sales* on the y-axis. The territory is represented by different colors for allowing the direct comparison of each product line performed in various territories.

**4. Distribution of Sales by Branch**

We used a donut chart to visualize the distribution of sales across branches. The chart displays each branch's proportion as a percentage, with different colors representing different branches. It helped in giving a clear view of the sales contribution of each branch. **Beergenville** branch had the most sales(49.0%),**Geiselweg** had the least sales of(5.6%)

**5. Top 10 Contact Persons with the Most Sales** We created a bar plot to display the *top* *10* *contact* *persons* on the x-axis and their *total* *sales* on the y-axis. This analysis was valuable for identifying high-performing contacts. The contact person with the most sales is **Diego Freyre.**

**6. Top 5 Customers with the Highest Sales**

We created a bar chart to visualize the top 5 customers, with *CustomerName* on the x-axis and *Total* *Sales* on the y-axis. This chart highlighted the top customers based on sales, offered insights into which customers generated the most revenue. The customer with the highest sales is **Euro Shopping Channel**, as indicated by the tallest bar in the chart.

**7. Top 10 Customers with the Most Quantity Ordered**

We plotted a bar chart to visualize the top 5 customers, with *CustomerName* on the x-axis and *Total Quantity Ordered* on the y-axis. This chart highlighted customers who have ordered the most products by quantity, providing insights into high-demand customers. The customer with the most quantity ordered is **Euro Shopping Channel**, as indicated by the tallest bar in the chart.

**8. Distribution of order status**

We used a bar plot to display the *Number of orders* by *each Order status*. This plot enabled the identification of the number of orders in each Order status which shows that **shipped** had the most number of orders.

**9. Total Profit by Branch**

We created a bar plot where x-axis represented the *branch* and y-axis represented the *total profit.* This bar plot gave us a clear comparison of profit in each branch. **Beergenville** has highest profit overall.

**10. Top 5 products by revenue**

We used a bar plot to visualize the top 5 products by revenue. Where we used the x-label as the *product code* and y-label as the *totalsales*. By calculating total sales revenue per product and identifying the top 5 products, this helped us pinpoint which products generated the most revenue.