**Work Document**

**Objective:**

To analyse sales performance and customer behaviour patterns across four leading toy retailers, identify market trends, and uncover actionable insights to improve revenue, customer loyalty, and engagement.

**Data Collection:**

For this project, four structured csv datasets — Customers, Products, Transactions, and Interactions — were used to analyze sales and customer behavior across leading toy brands. The data, limited to around 500 rows for efficiency, was synthetically generated with support from ChatGPT, ensuring realistic links between tables and alignment with the project’s goals.

**Data Preparation:**

* The datasets contained duplicate values, so within excel used “remove duplicates” feature to clear the duplicates.
* Most of the columns were containing white spaces so used trim function to clear the white spaces.
* The date was not in the proper format within the datasets, I got the format as “January 1st ,2025”, using the date format feature in excel standardized the date to “01-01-2025” format.
* Customer Dataset was not containing proper naming format, So corrected the name column using PROPER().
* Observed some missing values in some of the columns, so if the column was number column replaced missing value with AVERAGE(), if it was text column used MODE().

**PowerBI**

* Loaded the data into PowerBI transformed the data in powerquery by making first row as header.
* Changed the datatypes according to the column.
* Introduced Required columns using conditional column and custom column, for example using conditional column introduced LoyaltyStatus column, derived year from date.
* Used Measures and Calculated columns as per requirement for example calculated measures like, total sales, total customers, average price,total quantity sold and some more.
* For calculated columns made use of Dax.

**Measures:**

Total Sales = SUM(transaction\_data\_updated[Total\_Spend])

Total Quantity Sold = SUM(transaction\_data\_updated[Quantity])

Total Customers = DISTINCTCOUNT(transaction\_data\_updated[Customer\_ID])

**Caculated Columns:**

* LoyaltyStatus = SWITCH(TRUE(),transaction\_data\_updated[Total\_Spend] < 100, "Bronze",transaction\_data\_updated[Total\_Spend] < 180, "Silver",transaction\_data\_updated[Total\_Spend] < 250, "Gold","Platinum")
* transactionYear = YEAR(transaction\_data\_updated[Date])
* PurchaseCategory =

VAR CustomerID = transaction\_data\_updated[Customer\_ID]

VAR PurchaseCount =

CALCULATE(

COUNT(transaction\_data\_updated[Transaction\_ID]),

FILTER(transaction\_data\_updated,transaction\_data\_updated[Customer\_ID] = CustomerID

)

)

RETURN

IF(PurchaseCount > 1, "Repeat", "One-TIme")

**Data Model:**

