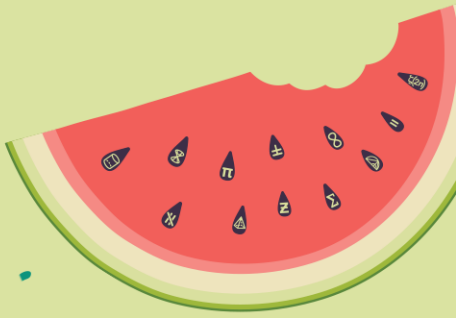
END-OF-SEMESTER
ACADEMIC
PROJECT **2024-25**



Sales Analysis in a Chain of Toy Stores in Mexico

Sinen Fradj
Tunis Business School, Department of Business Analytics

12/21/2024

Business Intelligence ©

TUNIS BUSINESS SCHOOL
UNIVERSITY OF TUNIS

Contents

A. Introduction

1. Data Overview & Key Components

This is a Sales & inventory data for a fictitious chain of toy stores in Mexico called Maven Toys, including information about products, stores, daily sales transactions, and current inventory levels at each location.

- inventory.csv:

Column Name	Data Type	Description	Example Values
store_id	Integer	Store ID	1, 2, 3, 4, ...
Product_ID	Integer	Product ID	1,2,4..
Stock_On_Hand	Integer	stock on hand for stores on a particular product	25, 30, 40

- product.csv:

Column Name	Data Type	Description	Example Values
Product_ID	Integer	Store ID	1, 2, 3, 4, ...
Product_Name	String	name of the product	"maven toy","camion"..
Product_Category	String	category of the product	"electronics","toys"
$Product_Cost$	Float	cost of the product	12$,21$..
Product_Price	FLoat	selling price of the product	12$,54.9$..

- sales.csv:

Column Name	Data Type	Description	Example Values
Sale_ID	Integer	Store ID	1, 2, 3, 4, ...
Date	Date	Product ID	DD-mm-yyyy
Store_ID	Integer	stock on hand for stores on a particular product	25, 30, 40
$Product_ID$	Integer	product ID inherited from products.csv	1,2,3
Units	Integer	number of units sold for that product in that date	12,45,33..

- stores.csv:

Column Name	Data Type	Description	Example Values
Store_ID	Integer	Store ID	1, 2, 3, 4, ...
Store_Name	String	Store name	"maven Junior"..
Store_City	String	City in Mexico where the store is located	25, 30, 40
$Store_Location$	String	Location in the city where the store is located	"new mexico" "bni hassen"

Column Name	Data Type	Description	Example Values
$Store_{O}pen_{D}ate$	Date	Date when the store was opened	DD-mm-yyyy

2. Business Objectives

The objective of this analysis is to assist the manager in answering key business questions regarding the chain of stores. The following questions will be addressed:

1. Which product categories are driving the biggest profits, and does this vary across different store locations?
2. Are there any seasonal trends or patterns in the sales data that could inform business decisions?
3. Are sales being lost due to out-of-stock products at certain store locations?
4. How much money is currently tied up in inventory at the toy stores, and how long will the inventory last given current sales trends?

By answering these questions, we aim to provide actionable insights that can guide the store chain's strategies for profitability, stock management, and overall efficiency.

B. Project Development Phases

1. Data Warehouse schema and tables relations

Star Schema Data Warehouse Description

The star schema illustrated in the diagram represents a data warehouse designed for efficient storage and retrieval of data related to sales and inventory management. The schema consists of a central fact table connected to multiple dimension tables, forming a star-like structure. Each component is described as follows:

Fact Table

The **Fact Table** serves as the core of the schema and captures the metrics related to sales. It contains the following attributes:

- **Product ID**: A foreign key referencing the Products Dimension table.
- **Store ID**: A foreign key referencing the Store Dimension table.
- **Sale ID**: A primary key to uniquely identify each sale transaction.
- **Date ID**: A foreign key referencing the Time Dimension table.

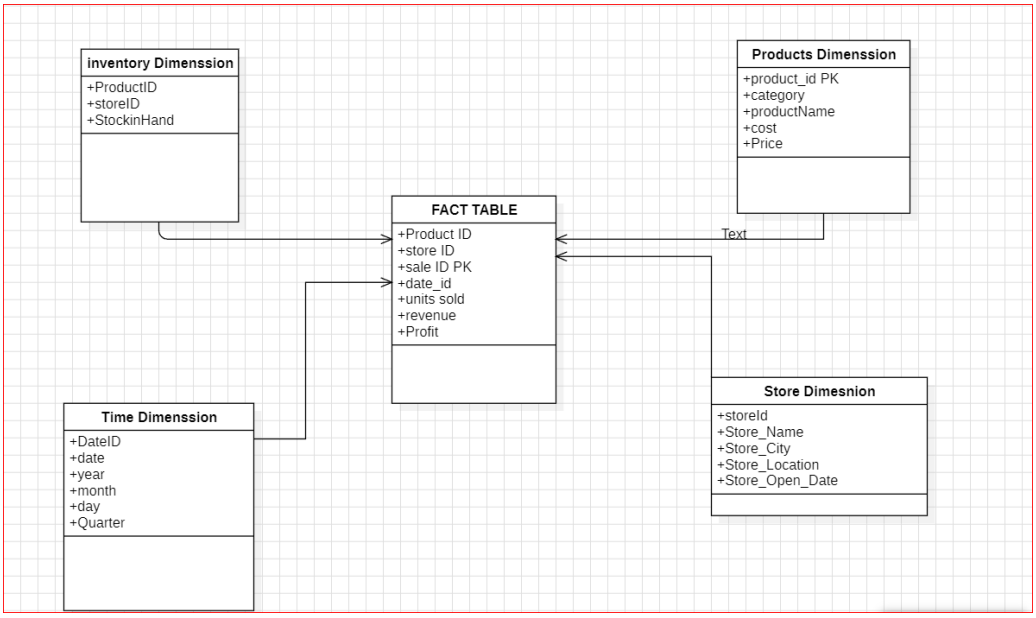


Figure 1: Star Schema Data Warehouse

- **Units Sold**: The quantity of products sold.
- **Revenue**: The total revenue generated from the sale.
- **Profit**: The net profit from the sale.

Dimension Tables

The dimension tables provide descriptive information for different aspects of the sales process:

- **Products Dimension**:
  - Contains details about the products such as Product ID (primary key), category, product name, cost, and price.
- **Store Dimension**:
  - Stores information about the stores such as Store ID (primary key), store name, city, location, and open date.
- **Time Dimension**:
  - Captures time-related attributes such as Date ID (primary key), date, year, month, day, and quarter.
- **Inventory Dimension**:
  - Tracks product inventory by Product ID (foreign key), Store ID (foreign key), and stock in hand.

Rationale for Using Star Schema

The star schema is chosen because it suits the data requirements and offers the following advantages:

- **Simplicity**: The design is straightforward and easy to understand, facilitating data retrieval and analysis.
- **Query Performance**: The denormalized structure reduces the number of joins required during queries, enhancing performance.
- **Scalability**: The schema allows for easy addition of new dimensions or metrics without affecting existing data.
- **Analytical Efficiency**: The design supports OLAP operations like slicing, dicing, and aggregation, making it ideal for business analytics.

2. ETL procss and Data preparation

After converting all the files to their corresponding formats, I began by exploring the data. This included checking for null values, identifying missing foreign keys between the dataframes, and ensuring data consistency. Detailed steps for these processes can be found in the EDA notebook. (see Reference 2)

- **Creation of Time Dimension**:

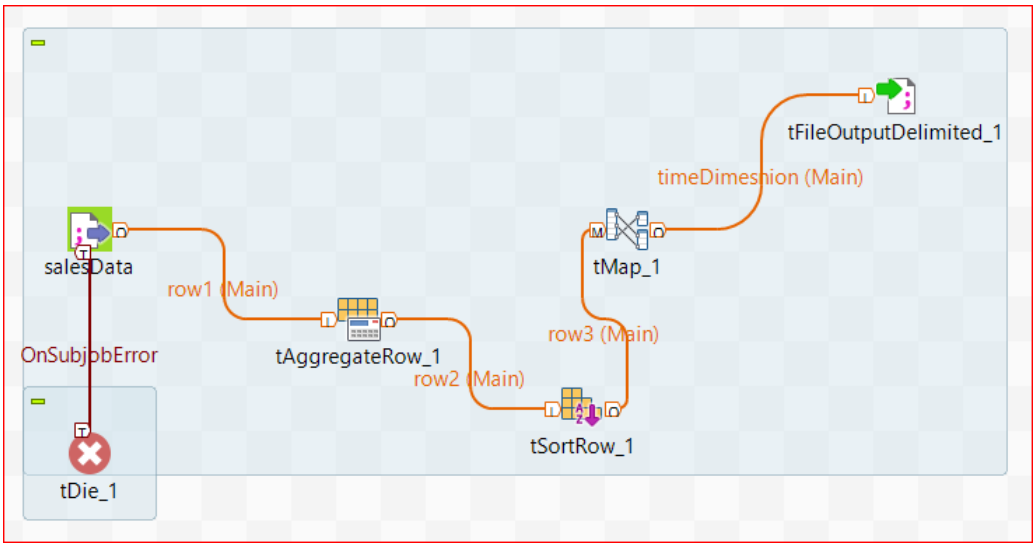


Figure 2: Time Dimesnion Job in Talend

  - Since the time dimension is a crucial part of any data warehouse, as it allows us to visualize data along the time axis, I created this dimension using the Date column from the sales.csv file. By extracting the day, month, and year from the full date, I structured the data accordingly. Additionally, I sorted the column to ensure it follows a chronological order.(using tSortRow_1)

- **Creating my Data Warehouse using MySQL**:
  - Following the class Diagram (Figure 1) I created my full schema using mysqlqueries since i found it more convenient than using Talend,you will find the sql file here (see Reference 2)
- **Loading the Dimesnions on our WDB**:

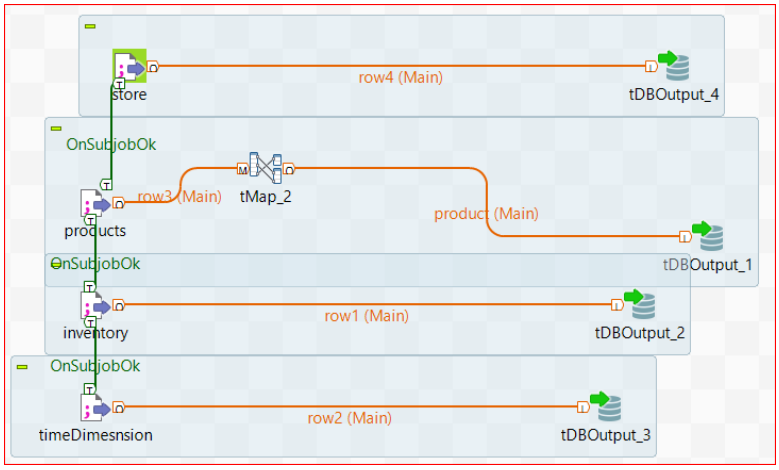


Figure 3: Job For Loading Dimesnions

  - since there are links between Dimensions (Foreign key references),the best strategy here is to make a squential job where there is an order of execution between the tasks therefore we don't fall in a foreign key constraint violation in Figure 3 we can see that the loading of a table is subject to the one above it .
- **Creating Fact Table and loading it in the WDB**:

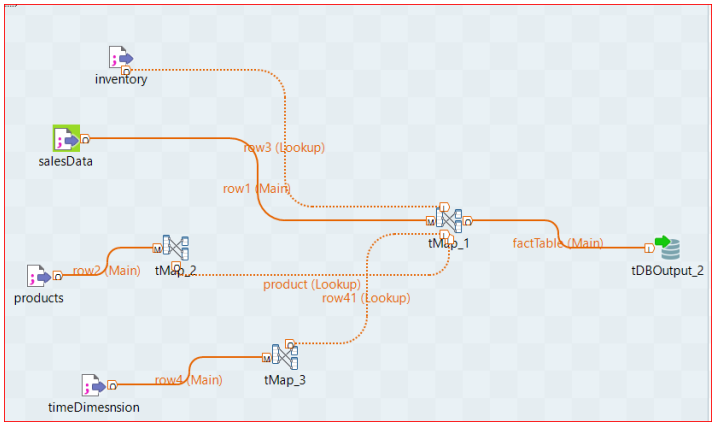


Figure 4: Job for Fact Table

  - as shown on Figure 4 I used tmap_1 to merge the necessary columns and converting some columns to the right format (tMap_2 I converted the cost and price of the product to an integer by removinig the $

symbol) and change the date format from DD-mm-yyyy to DD/mm/yyyy in the tMap_3.subsequently ,we load our new fact table to the WDB

3. Data Analysis using PowerBI

After connecting to the sql server using power BI and laoding our data we are performing ROLAP to generate graphs and reports the next 2 pages are reports that we generate it using powerBI.

Interpretations from the generated Grapsh:

- Maven Toys Ciudad de Mexico 2 has the highest number of units sold throughout the period of analysis with 42752 unit.
- Despite not having the highest number of units sold Maven Toys Cuernavaca 1 returns the highest profit
- in general we can extrapolate that the most sold product category in all the stores is Toys
- Maven Toys Campeche 1 hase 2.25% of the marekt share
- the best selling Product is Colorbuds
- we can notice a major chute in sales between 2019 and 2020 ,one possible explanation is the COVID-19 pandemic in the same period
- the first 2 Quarters of the year always yield the highest Revenue than the last 2 Quarters. The early months (especially Q1) include holidays like Christmas and New Year, which often lead to high demand for toys as gifts. Additionally, in Mexico, the Easter holidays (Semana Santa) in Q2 may also contribute to higher sales for toys, arts, and crafts.

Store with the highest units sold

Maven Toys Ciudad de Mexico 2

42757

Sum of UnitsSold

Market share of each store

Maven Toys Aguascalientes 1

1.64%

%GT Sum of UnitsSold

Maven Toys Campeche 1

2.25%

%GT Sum of UnitsSold

Maven Toys Campeche 2

1.49%

%GT Sum of UnitsSold

Maven Toys Chetumal 1

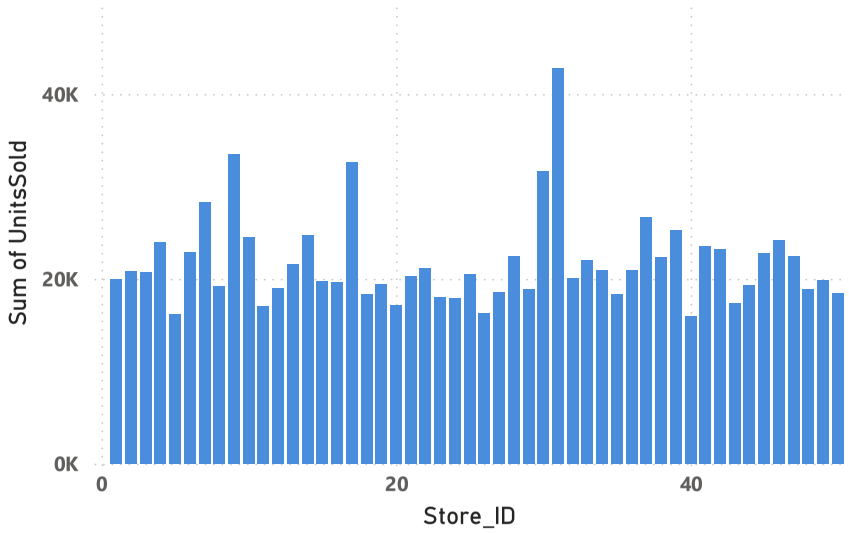
1.74%

%GT Sum of UnitsSold

Maven Toys Chihuahua 1

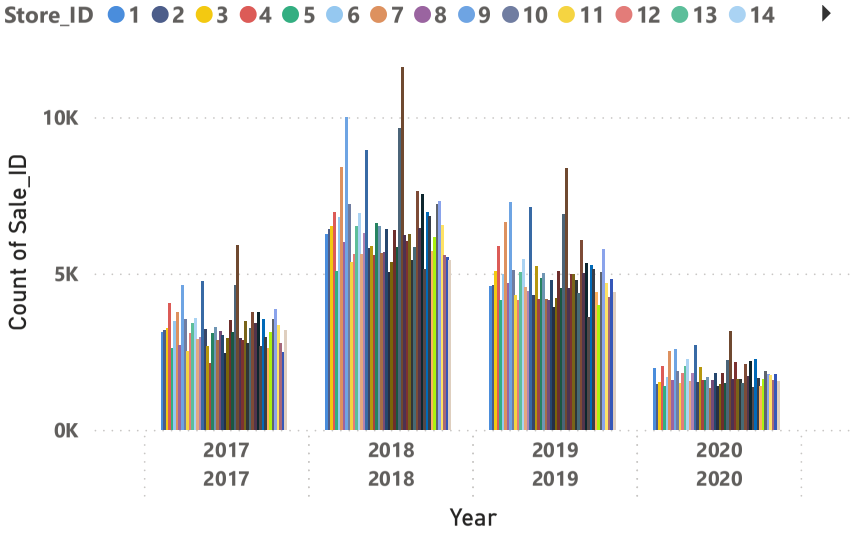
Store_Name	Sum of Profit
Maven Toys Cuernavaca 1	56,811.00
Maven Toys La	57,407.00
Total	4,014,029.00

Sum of UnitsSold by Store_ID

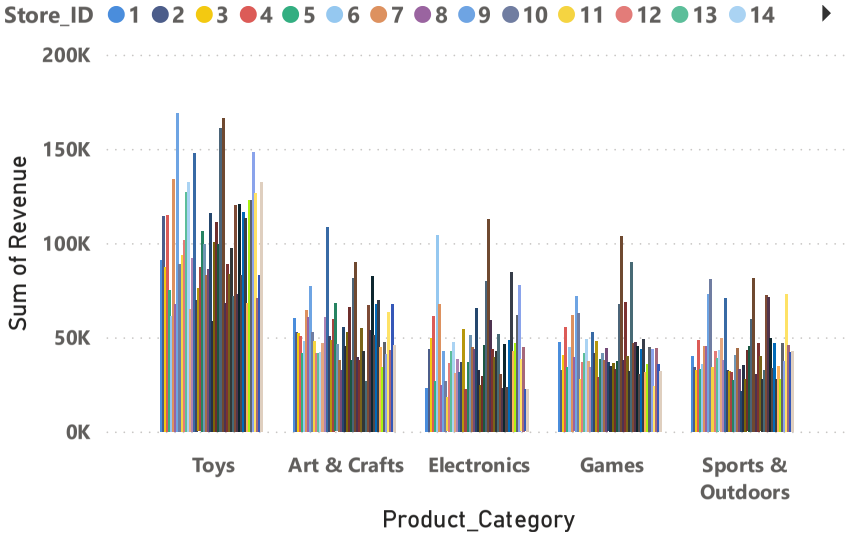


Sum of Revenue by Store_City and Store_ID

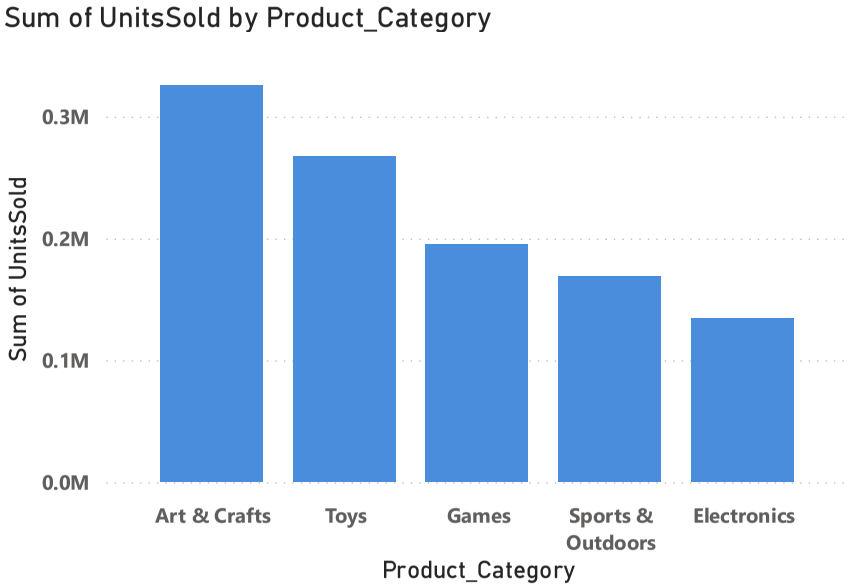
Count of Sale_ID by Year, Year and Store_ID

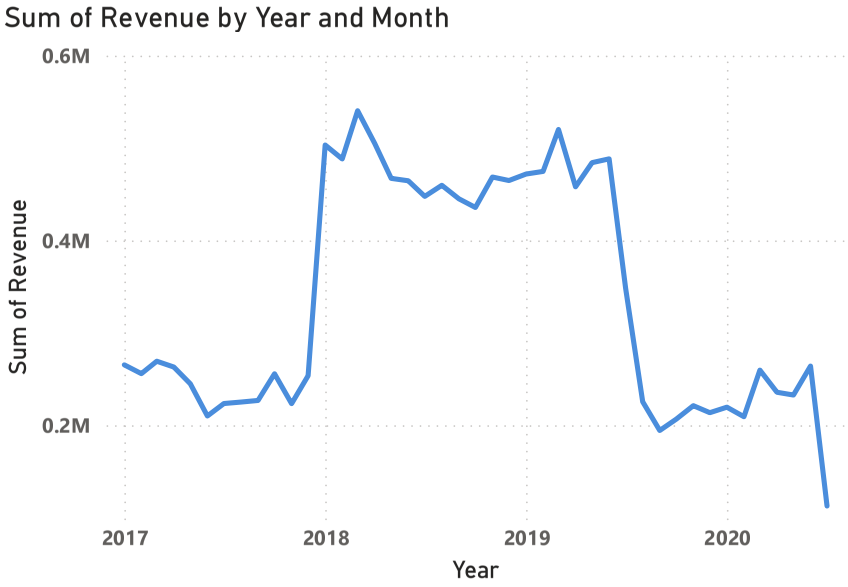


Sum of Revenue by Product_Category and Store_ID

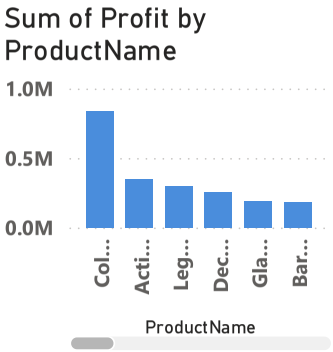


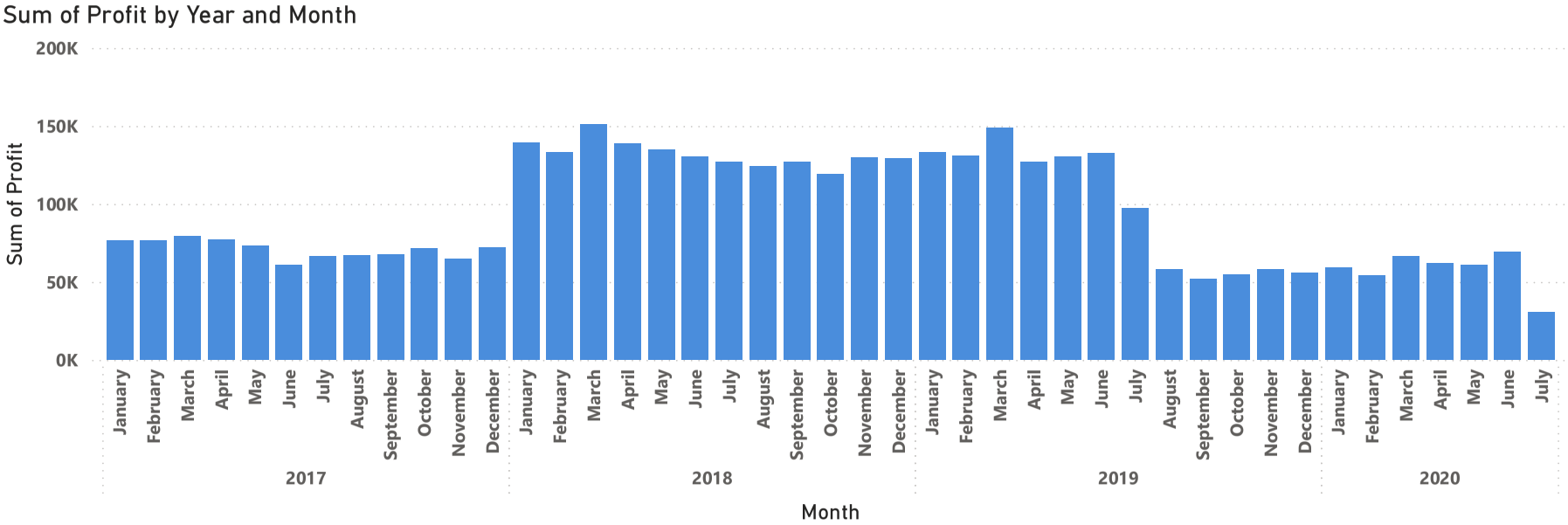
ProductName	Sum of UnitsSold
Colorbuds	104368
PlayDoh Can	103128
Barrel O' Slime	91663
Deck Of Cards	84034
Magic Sand	60598
Splash Balls	60248
Lego Bricks	59737
Action Figure	57958
Rubik's Cube	45672
Animal Figures	39089
Dino Egg	38236
Glass Marbles	37518
Mini Ping Pong Set	33604
Total	1090565





Quarter	Sum of Revenue
Qtr 1	4,476,829.64
Qtr 2	4,316,972.33
Qtr 3	2,907,958.58
Qtr 4	2,742,811.77
Total	14,444,572.32





C. Conclusion

1. Project Overview

This project provides a comprehensive solution for data-driven decision-making, offering insights into sales trends, market share, and the impact of seasonal factors, all facilitated by an efficient Star Schema Data Warehouse and Power BI reporting.

2. Possible Enhancements

- the project needs a better error handling strategies in the ETL phase .
- At the outset of the project, I initially intended to use MOLAP (Multidimensional OLAP) with multi-dimensional cubes through SQL Server Analysis Services. However, I encountered a fatal error during installation that I was unable to resolve despite multiple attempts. As a result, I have decided to proceed with ROLAP (Relational OLAP) for the time being, until I can fix the issue and return to the MOLAP setup.

References

- GitHub Repository: EDA notebook
- GitHub Repository: SQL query