



Maven Toys Database Management and Analysis Report

Authors:

**Maram Abdallah
Ferdawes Haoualaa
Imen Kenza Chouaieb**

**IT-300 Business Intelligence and Database Management Systems
Pr.Abdelkader and Pr.Azzouz
January 2023**

Table of contents:

1.Introduction

2.Database Management and analysis:

2.1.Data Gathering

2.2.ETL Process

2.3.Modeling and Datawarehouse Creation

2.4.Data Analysis

3.Conclusion

1.Introduction:

The Maven Toys Database project analyzes this business management of inventory and sales based on different factors such as the products' categories and the stores' locations all over Mexico.

These valuable insights will introduce us to the target customers' location and preferences, boosting the company's future expansion and inventory management strategies.

2.Implementation:

2.1.Data Gathering:

The Maven Toys dataset was extracted from [Kaggle](#).

2.2.ETL Process:

We first started by cleaning data by manipulating:

-Columns' names to facilitate the process:

```
#define cleaning 1 : change all columns names to lower case
# code :
sales.columns= sales.columns.str.lower()
inventory.columns= inventory.columns.str.lower()
product.columns= product.columns.str.lower()
stores.columns= stores.columns.str.lower()
```

-Columns' type (eg: the currency and date ones):

```
#define : converting datatype :in stores table: Store_Open_Date to date
#code
stores['store_open_date']=pd.to_datetime(stores['store_open_date'])
#test
stores.info()
```

```
#define : in product table : product_price and product_cost to float
product['product_price'] = product['product_price'].str.replace('$', '').astype(float)
product['product_cost'] = product['product_cost'].str.replace('$', '').astype(float)
```

-New columns needed for analysis (eg: profit):

```
#define cleaning 3 : add a profit column  
# code  
product["profit"]=product["product_price"] - product["product_cost"]  
#test  
product
```

2.3.Modeling and Datawarehouse Creation:

The datasets contains 4 tables: Sales, Inventory, Products and Stores. The core of our datawarehousing infrastructure comprised PostgreSQL as the database management system, SQLAlchemy, Python library, for seamless table mapping within our code, and psycopg2, a Python library, enabling smooth interactions between them.