Business Template

RETAIL SALES

CONTENTS

1 BUSINESS DESCRIPTION

1.1 BUSINESS BACKGROUND

Retail goods are important for everyone. There are huge amount of retail shops in USA and that's why this kind of business is very competitive, so if you want to be successful in this field you should very responsibly approach this case and learn a lot of factors which influence on people's needs. To identify this kind of factors we need to collect and analyze sales records.

1.2 PROBLEMS BECAUSE OF POOR DATA MANAGEMENT

Poor data management practices in retail can severely impact business success by leading to inefficiencies and missed opportunities. Without utilizing effective data analysis tools and strategic planning, retail operations may struggle to meet customer demands and adapt to market changes. Understanding and responding to consumer preferences are essential for staying competitive in the retail industry.

1.3 BENEFITS FROM IMPLEMENTING A DATA WAREHOUSE

Using of data warehouse can help you with the problems described above. Implementing a data warehouse can answer you the following questions:

- Which products have the highest prices?
- Which ones have the widest distribution of prices?
- Identify highest required goods
- Analyze profits

Further processing data would also let you:

- Correlate specific product features with changes in price.
- If there are any sales differences between locations.
- And many other.

1.4 DATASETS DESCRIPTION

The first dataset contains the following information about sales on the USA market.

Product Information:

product_id: The unique identifier of the product

product_name: The name of product

product_length : The length of product

product_depth : The depth of product

product_width : The width of product

hierarchy1_id : Hierarchy first level of product

hierarchy2_id : Hierarchy second level of product

Sales Information:

invoice_number : The unique identifier of sales record
date : Date of sales
quantity : Quantity sold
daily_rev: Daily total sales revenue
stock: Stock quantity
price: Product sales price
cost: The cost of the product.
promo_type_1 : First type of promotion applied
promo_bin_1 : Binned promotion rate
promo_type_2: Second type of promotion applied
sales_channel: Channel of sales (online or offline)

Store Information:

store_id: Unique identifier of store
store_name : Name of the store
store_type_id : Store type name
store_size : Number of employees

Location Information:

city_id: Unique identifier of city
city_name: Name of the city
store_address : Address of the store
store_state : State
country_id : Country id
country_name : Name of country

Customer Information:

customer_id : Unique identifier of customer
f_name: First name of the customer
l_name: Last name of the customer
email: Email of the customer

Employee Information:

employee_id: Unique identifier of employee

The second dataset contains the following information ….

Product Information:

- product_id: The unique identifier of the product
- product_name: The name of product
- product_length : The length of product
- product_depth : The depth of product
- product_width : The width of product
- hierarchy1_id : Hierarchy first level of product
- hierarchy2_id : Hierarchy second level of product

Sales Information:

- invoice_number : The unique identifier of sales record
- date : Date of sales
- quantity : Quantity sold
- stock: Stock quantity
- price: Product sales price
- cost: The cost of the product.
- promo_type_1 : First type of promotion applied
- promo_bin_1 : Binned promotion rate
- sales_channel: Channel of sales (online or offline)

Store Information:

- store_id: Unique identifier of store
- shop_website : Website of online shop

Customer Information:

- customer_id : Unique identifier of customer
- f_name: First name of the customer
- l_name: Last name of the customer
- email: Email of the customer
- cust_phone: Phone number of the customer

Employee Information:

employee_id : Unique identifier of employee

employee_name : Name of the employee

employee_last_name : Last name of the employee

employee_email : Email of employee

The datasets provide a comprehensive overview of offline and online sales, allowing for analysis and exploration of trends, sales performance, customer preferences. In second data set we don't have location information because of online shop, also store information is different , we have only web site of the shop. In second dataset we have cluster information such as name of the cluster , but in first dataset we do have only identifier not whole information.

1.5 GRAIN / DIM / FACT

In our data warehouse model, the grain for the sales transaction fact table is defined at the level of each individual sales transaction. This means that every row in the fact table corresponds to a distinct transaction made by a customer, capturing detailed information such as the specific products purchased, quantities, prices, cost, stock and promotion types applied.

Fact Description

Column name	Description	Data Type
invoice_number	The unique identifier of sales record	VARCHAR(100) <pk>
date	Date of sales,date identifier	DATE <fk>
product_id	The unique identifier of product	VARCHAR(10) <fk>
employee_id	The unique identifier of employee	INT <fk.
store_id	The unique identifier of store	VARCHAR(10) <fk>
customer_id	The unique identifier of customer	BIGINT <fk>
quantity	Quantity product sold	INT
daily_rev	Daily revenue	DECIMAL(10,2)
stock	Stock count	INT
price	Price of sale per product	DECIMAL(10,2)
cost	Price of cost per product	DECIMAL(10,2)
promo_type_1	Promotion 1 applied	VARCHAR(50)
promo_bin_1	Binned promotion rate	VARCHAR(50)
promo_type_2	Promotion 2 applied	VARCHAR(50)
sales_channel	Channel of sale	VARCHAR(50)

Example with filled data

invoice_number	date	product_id	cluster_id	store_id	customer_id	quantity	daily_rev	stock	price	cost	promo_type_1	promo_bin_1	promo_type_2	sales_channel
INV-100001	2023-08-17	P0184	4	S0085	136	1	0	7	64.9	53.2	PR06	low	PR03	offline

Store Dim Description

Column name	Description	Data Type
store_id	The unique identifier of store	VARCHAR(20) <pk>
store_name	Name of the store	VARCHAR(100)
storetype_id	The unique identifier of type of store	VARCHAR(20) <fk>
storetype_name	Name of the type	VARCHAR(50)
store_size	Count employees	INT
city_id	The unique identifier of city	VARCHAR(20) <fk>
city_name	Name of the city	VARCHAR(100)
store_address	Address	VARCHAR(200)
store_state	State	VARCHAR(30)
country_id	The unique identifier of country	INT <fk>
country_name	Name of the country	VARCHAR(100)

Example with filled data

store_id	store_name	soretype_id	storetype_name	store_size	city_id	city_name	store_address	store_state	country_id	country_name
S0085	Local Legends	ST01	Electronics Store	60	C006	Atlanta	890 Spruce Lane	GA	1	United Stated

Store Dim Description in second dataset

Column name	Description	Data Type
store_id	The unique identifier of store	VARCHAR(20) <pk>
shop_website	Web site of the store	VARCHAR(100)

Example with filled data

store_id	shop_website
S0085	electronicsgalore.com

Customer Dim Description

Column name	Description	Data Type
customer_id	The unique identifier of customer	INT <pk>
f_name	First name of the cutomer	VARCHAR(50)
l_name	Last name of the cutomer	VARCHAR(50)
email	Email	VARCHAR(100)
cust_phone	Phone number	VARCHAR(30)

Example with filled data

customer_id	f_name	l_name	email	cust_phone
136	ANITA	MORALES	ANITA.MORALES@sakilacustomer.org	+9406440

Product Dim Description

Column name	Description	Data Type
product_id	The unique identifier of product	INT <pk>
product_name	Name of the product	VARCHAR(100)
product_length	Length of the product	DECIMAL(5,2)
product_depth	Depth of the product	DECIMAL(5,2)
product_width	Width of the product	DECIMAL(5,2)
hierarchy1_id	Hierarchy id 1	VARCHAR(30)
hierarchy2_id	Hierarchy id 2	VARCHAR(30)

Example with filled data

customer_id	f_name	l_name	email	cust_phone
136	ANITA	MORALES	ANITA.MORALES@sakilacustomer.org	+9406440

Employee Dim Description

Column name	Description	Data Type
employee_id	The unique identifier of employee	INT <pk>
employee_name	First name of the employee	VARCHAR(50)
employee_last_name	Last name of the employee	VARCHAR(50)
employee_email	Email of the employee	VARCHAR(100)

Example with filled data

customer_id	f_name	l_name	email	cust_phone
50	ANITA	MORALES	ANITA.MORALES@gmail.com	+9406440548

Date Dim Description

Column name	Description	Data Type
date	Date of sale	DATE <pk>

Example with filled data

date
2023-06-26

2 BUSINESS LAYER 3NF

In BL_3NF layer we store data in third normal form and also specify SCD dimensions.

For all tables in this layer will have SOURCE TRIPLET.

Tables CE_SALES_SCD , CE_CUSTOMERS_SCD, CE_PRODUCTS_SCD , CE_EMPLOYEES_SCD are SCD type 2 so I added start_dt , end_dt and is_active columns.

All the rest tables are SCD type 1 , insert_dt, update_dt added for storing change dates only.

3 BUSINESS LAYER DIMENSIONAL MODEL

4 LOGICAL SCHEME

5 DATA FLOW

6 FACT TABLE PARTITIONING STRATEGY