Lab-3-Report

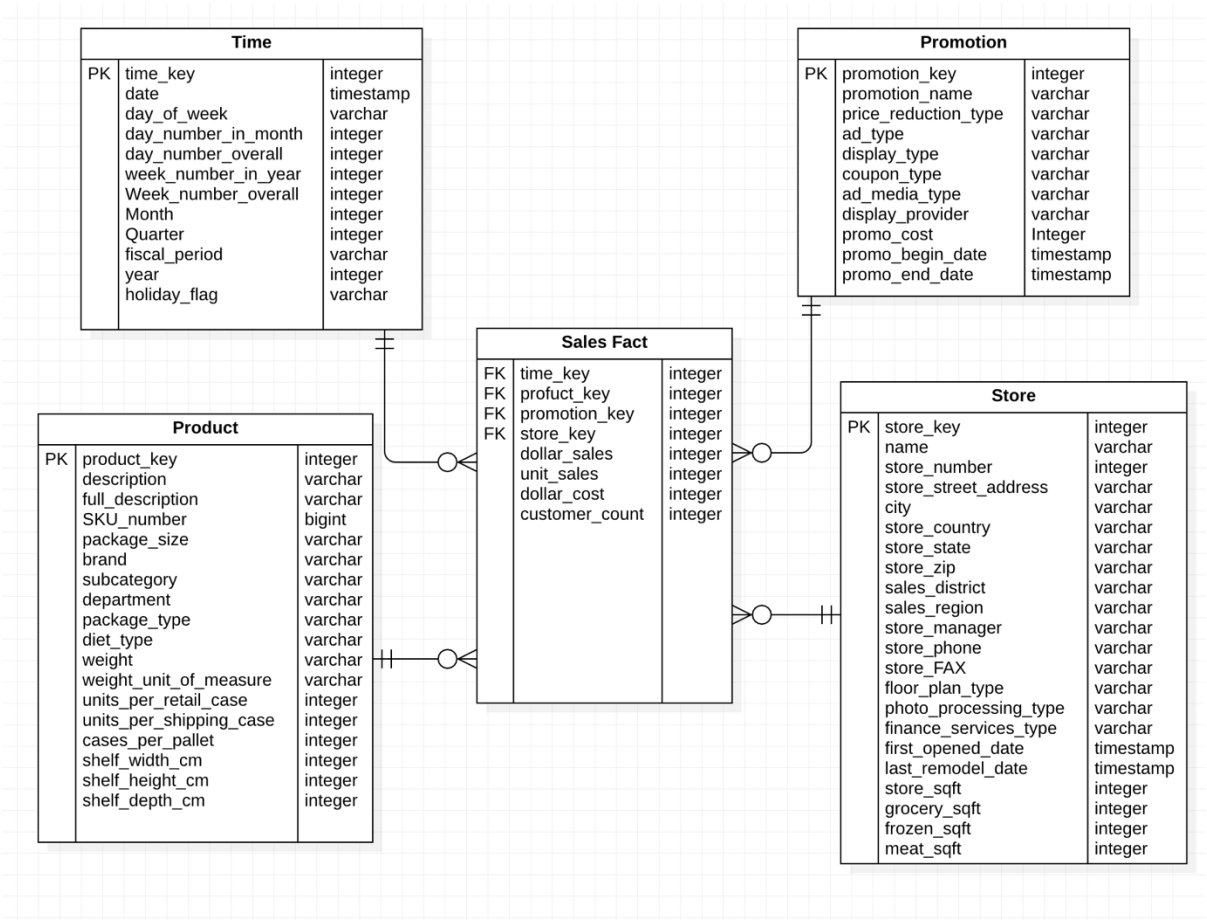
Group members: Sehrish Naqvi, Balaji Vijayraj, Waleed Tariq, Charu Bisht

For G

1.Explain the business you chose to model through a DW.

For this lab, we are working on the same dataset as in Lab 1. Therefore, the business case we chose to model through the data warehouse is of a grocery chain. The aim is to pick a design for the data warehouse that best helps us to achieve meaningful inferences about the grocery chain. We implemented this based on Star Schema

2.Present and describe the ER diagram implemented as solution to assess the business performance of the imaginary company.



Fact Table – Sales Fact

Dimension Table – Time, Product, Promotion Store

The dimension tables have 1 to Many relationship with the Fact Table. The Sales Fact table contains the facts that are needed to access the business performance. It has the Number of products sold, Revenue of the business, Number of customers, and cost of the products. With this, we can find the Revenue, profit, marketing cost and access the performance of the company

Lab-3-Report

3.Answer the question as follows: Which transformation need to be applied to the DW, if the company is interested in knowing the effect of promotions on sales?

To understand better the effects of promotions on sales, we can utilise the following transformations:

- We can aggregate the data by promotion period and see in which period the sales were highest. This can help us understand the seasonal trends that the sales reflect and how to plan and put the promotions in place in the future to reap maximum profit margin. For example, for grocery stores, promotions could indeed push sales, especially around Christmas.

- Another transformation that can be useful is to compare the sales harboured during the promotion period with the baseline period. This will also evaluate when was the highest profit margin reached. We can utilise this data for long periods which can also help to make a prediction model to further enhance the business.

4.Draw conclusions of your work, in terms of benefits and limitations of your model.

Using this model, we were able to identify several important things required to make essential business decisions. One of the important aspects of any business is to understand its operations. It is important to find out how much profit a business is making or how much loss it has endured. It is also important for a business to know how its products/services are being received by its customers. Based on these key points, the business can set new goals or work on their improvement areas to achieve their current goals. For instance, currently, the market is moving towards a green and circular economy and such kind of analysis can help to shape the operations to achieve the same.

Benefits of our model:

As mentioned above, our model was helpful for us to understand the performance of different stores. We talk more about this in the VG part where we explain the different inferences we drew from the model.

Limitation of our model:

Another one of the most important things for any business is its feedback. Unfortunately, our model did not have access to that information so we lacked to draw conclusions on that front. This information can be very vital in terms of understanding customer feedback and customer disengagement. Other than that, we felt that our model contributed to the underestimation of data-loading resources.

Lab-3-Report

For VG

We implemented the data warehouse model in access and loaded the Grocery data. We took the dataset from lab 1 and followed the instructions in the tutorials to load the data. Then we used SQL lite to answer the questions.

Questions 1 - How many sales did the company have?
Query:

```
pd.read_sql("select sum(unit_sales) as Total_Sales from Sales",engine)
pd.read_sql("select * from Sales",engine)
```

Result:

Total_Sales – $550720

Question 2 - How many products did we sale in a store last month?
Query:

```
pd.read_sql("SELECT st.store_number, SUM(unit_sales) AS
Products_sold_in_last_month_Store_wise FROM Sales sa JOIN Store st ON
sa.store_key = st.store_key JOIN Time T ON T.time_key = sa.time_key WHERE
T.date BETWEEN '1995-12-01' AND '1995-12-31' group by st.store_number order
by 1 ",engine)
```

We presented the information store wise.

Result:

	store_number	Products_sold_in_last_month_Store_wise
0	1	5112
1	2	4055
2	3	3881
3	4	5052
4	5	5037
5	6	3641
6	7	4205
7	8	4915
8	9	3947
9	10	5158
10	11	4995
11	12	4361

Lab-3-Report

	store_number	Products_sold_in_last_month_Store_wise
12	13	3860
13	14	4113
14	15	4377
15	16	3935
16	17	5054
17	18	4627
18	19	4052
19	20	3898

Question 3 - Which one of our stores have the highest amount of sells
Query:

```
pd.read_sql("SELECT store_number, SUM(unit_sales) AS Sales_in_a_Store FROM Sales sa JOIN Store st ON sa.store_key = st.store_key group by st.store_number order by 2 desc  ",engine)
```

Result:

Store no – 12

Question 4 - Which products are the most lucrative?
Query:

```
pd.read_sql("SELECT description as Lucrative_Products , sum(dollar_sales) - sum(dollar_cost) as profit FROM Sales sa JOIN Product pr ON pr.product_key = sa.product_key GROUP BY pr.description order by 2 desc;",engine)
```

Result:

Buffalo Jerky is the most lucrative product followed by Chicken Dinner and Turkey Dinner

Question 5 - What was the most lucrative day, month, or year
Query:

```
pd.read_sql("SELECT t.year as Lucrative_Year, sum(sa.dollar_sales) - sum(sa.dollar_cost) as profit FROM Sales sa JOIN Time t ON t.time_key = sa.time_key GROUP BY t.year order by profit desc",engine)
```

Result:

The most lucrative year was 1994 with a profit of $54282