**BAN 545 Term Project**

PART-2: Creation of the Data Warehouse

Access DB:

A screenshot of a computer

Description automatically generated

Query1:

SELECT SalesFact.Order\_Quantity, SalesFact.Unit\_Price, SalesFact.Total\_Sales, CustomerDimension.[Customer Country], CustomerDimension.[Customer State], CustomerDimension.Customer\_Gender, ProductDimension.Product

FROM DateDimension INNER JOIN (CustomerDimension INNER JOIN (ProductDimension INNER JOIN SalesFact ON ProductDimension.ProductKey = SalesFact.ProductKey) ON CustomerDimension.CustomerKey = SalesFact.CustomerKey) ON DateDimension.DateKey = SalesFact.DateKey

WHERE (((SalesFact.Order\_Quantity)>=4) AND ((SalesFact.Total\_Sales)>10000) AND ((CustomerDimension.Customer\_Gender)="F"));

**Analysis:**

* **Description:** This query selects the sales data for all the orders to be purchased by females, quantity of orders to be 4 or above, the total sale price to be more than $10000. Other essential information also includes the customer’s country, state, gender, and the product type.
* **Purpose:** The goal of this query is to look for patterns of high order values placed by females and their buying pattern analysis.
* **High-Value Products:** The Unit\_Price of products that are ordered is between $3,375 and $3,578 which means that most of the orders contain high-margin goods including Mountain-100 Black/Silver and Road-150 Red bikes. These products assist in achieving total sales of more than $10,000 in one day.
* **Product Categories:** The products fall into two subcategories: Road Bikes and Mountain Bikes, with the Road Bike category receiving slightly more attention. Which points towards the levels demand for road biking as well as off road biking products.

**Insights:**

* **Geographic and Demographic Opportunities**: These results show that the United States’ California and Oregon are standout markets for high-value feminine consumer sales. Possible extension of these marketing activities may help to increase profits.
* The participation of Foreign buyers from Germany and France, illustrates the need to or open up or expand into these markets with premium products targeted at women.
* **Balanced Product Preference:** The strength shown by both road bikes and mountain bikes is nearly balanced, which shows a chance to promote both categories almost equally when developing succeeding marketing strategies.

**Output:**

A screenshot of a computer

Description automatically generated

Query-2:

SELECT SalesFact.OrderID, SalesFact.Order\_Quantity, SalesFact.Unit\_Price, SalesFact.Total\_Sales, CustomerDimension.[Customer State], CustomerDimension.[Customer Country], ProductDimension.Product, ProductDimension.Sub\_Category

FROM CustomerDimension INNER JOIN (DateDimension INNER JOIN (ProductDimension INNER JOIN SalesFact ON ProductDimension.ProductKey = SalesFact.ProductKey) ON DateDimension.DateKey = SalesFact.DateKey) ON CustomerDimension.CustomerKey = SalesFact.CustomerKey

WHERE (((SalesFact.Total\_Sales)<1000) AND ((CustomerDimension.[Customer Country])="United States") AND ((ProductDimension.Sub\_Category)="Road Bikes"));

**Analysis:**

* **Description:** This query points to orders that have total sales value below $ 1000, for customers who are from the United States and buyers who bought road bikes. It captures the order ID, quantity, unit price, total sales, customer location, product and product sub-type.
* **Purpose:** The rationale for this query is to identify which low volume product (road bikes) and in which geographical market (United States) may offer areas for improved sales or promotions.
* **Products Involved:** The products range mainly road bikes with such brands as Road-750 Black, Road-650Red, Road-650 Black. These include 44, 48, 52, 58 and it shows that the customers’ preferences regarding bike sizes and models are diverse.
* **Geographic Trends:** The orders are divided between California, Washington, Oregon and then occasionally other states. California is most active in the number of orders placed, which supports the theory of its dominance in road bike sales, even in the economy segment.
* **Price Points:** The Unit\_Price is grouped mainly around $540, $ 783 and this makes the prices of these models fall on the mid-range category. These price points may also not be encouraging volume buys or higher ticket consumer transactions.

**Insights:**

* **Focus on Upselling:** Bill consumer 1 is more of individual units hence different forms of actual consumer 1 persuasion techniques such as giving out bundle deals of road bikes and related accessories including helmets and maintenance kits.
* **Exploring Demand for Premium Options:** There is potential that in mid-range price tier level will discourage buying in large volumes. The creation of the new model of road bike within the higher price segment with more emotionally appealing characteristics might help customers to make a higher LTV purchase.
* **Potential Product Modifications:** As for the Road-750 Black and Road-650 Red models, which are in high demand, manufacturers should increase the model range or add some new functions for existing models.

**Output:**

A screenshot of a computer

Description automatically generated

Query-3:

SELECT SalesFact.OrderID, CustomerDimension.Customer\_Age, ProductDimension.Product, DateDimension.Date, CustomerDimension.[Customer State], SalesFact.Total\_Sales

FROM ProductDimension INNER JOIN (DateDimension INNER JOIN (CustomerDimension INNER JOIN SalesFact ON CustomerDimension.CustomerKey = SalesFact.CustomerKey) ON DateDimension.DateKey = SalesFact.DateKey) ON ProductDimension.ProductKey = SalesFact.ProductKey

WHERE (((CustomerDimension.Customer\_Age)>25) AND ((CustomerDimension.[Customer State])="California") AND ((SalesFact.Total\_Sales)>10000));

**Analysis:**

* **Description:** This query selects the orders from customers, which are older than 25 years in California, and have total sales more than 10k$. Information, it contains the order identification, the customer’s age, the product, order date, customer state, and total sales.
* **Purpose:** This can be done by adapting the following query: \* Show me high value orders originating from California Recently, this type of inquiry assisted in identifying crucial demographic information that impacts sales.
* **Products Involved:** Two products dominate the list: Road 150 red and mountain 100 black. Some of the Road-150 Redes are available in several models (48, 56, 62), thus proving its popularity with these customers.

**Insights:**

* **Focus on Road-150 Red Series:** The Road-150 Red series is present continually across each of the shared configurations. Beta’s current appeal to this demographic might be improved by offering more options (sizes, colors) or even by building better, more expensive options.
* **Product Positioning:** The demographic values both road and mountain biking and hence the presence of the Mountain-100 Black along side road bikes. This could capitalize on their interest in things such as helmets and other accessories or new features in bikes.

**Output:**

A screenshot of a computer

Description automatically generated

Query-4:

SELECT SalesFact.OrderID, SalesFact.Order\_Quantity, SalesFact.Total\_Sales, DateDimension.MonthName, ProductDimension.Product, CustomerDimension.[Customer Country]

FROM ProductDimension INNER JOIN (DateDimension INNER JOIN (CustomerDimension INNER JOIN SalesFact ON CustomerDimension.CustomerKey = SalesFact.CustomerKey) ON DateDimension.DateKey = SalesFact.DateKey) ON ProductDimension.ProductKey = SalesFact.ProductKey

WHERE (((DateDimension.MonthName)="January") AND ((CustomerDimension.[Customer Country])="Australia"));

**Analysis:**

* **Description:** This query helps to get the orders that were placed during January by customers from Australia. They are order ID, quantity ordered, total sales, month of the order, the product, and customer country.
* **Purpose:** This is a QA to find out the activities of sales and customers in Australia in January to enable them to form a marketing strategy depending on the season.
* **Product Diversity:** The data includes a mix of Mountain Bikes and Road Bikes, with products like: With Mountain-200 Black, Mountain-200 Silver, Road-550-W Yellow and Road-150 Red. This means that both product categories have a lot of demand in January so to speak.
* **Frequent Products:** Some codes such as Mountain-200 Silver and Road-550-W Yellow may repeat themselves in the current dataset. These products can be the most popular in Australia in January.

**Insights:**

* **Focus on Top-Selling Products:** Some of the products that might be recommended for the fluctuation analysis are Mountain-200 Silver and Road-550-W Yellow as a result of their sales trends.
* **Targeting High-Value Customers:** Large-value purchases (i.e., purchases in excess of $9000) imply a group of clients willing to pay a premium price for products. It might be interesting to target this consumers with more personal offers or to launch a loyalty programs that can enhance its revenue from this group.

**Output:**

A screenshot of a computer

Description automatically generated

Query-5:

SELECT SalesFact.OrderID, SalesFact.Order\_Quantity, SalesFact.Total\_Sales, DateDimension.WeekdayorWeekend, ProductDimension.Product, ProductDimension.Sub\_Category, CustomerDimension.Customer\_Age, CustomerDimension.[Customer Country], CustomerDimension.[Customer State]

FROM ProductDimension INNER JOIN (DateDimension INNER JOIN (CustomerDimension INNER JOIN SalesFact ON CustomerDimension.CustomerKey = SalesFact.CustomerKey) ON DateDimension.DateKey = SalesFact.DateKey) ON ProductDimension.ProductKey = SalesFact.ProductKey

WHERE (((DateDimension.WeekdayorWeekend)="Weekend") AND ((ProductDimension.Sub\_Category)="Mountain Bikes") AND ((CustomerDimension.Customer\_Age)>30));

**Analysis:**

* **Description:** This query searches for orders of mountain bikes made on weekends, for customers who are over 30 years of age. This comprise of order identification number, order quantity, total sale value, order type: weekday or weekend, product type, sub-type, customer age and geographical region.
* **Purpose:** The idea of this query is to find the buying behaviour of older customers on the weekend and specifically regarding mountain bikes. It gives an idea about pattern in the leisurely buying and can be useful while planning weekend oriented campaigns.
* **Customer Demographics:** Most of the customers are more than 30years and the youngest is 72 years of age. This location factors comprise countries such as Canada, Australia, France, Germany, United Kingdom, and United States, hence meaning that we have got a global customer base.
* **Order Quantity and Total Sales:** The order quantity is from 1 up to 4 while total sales stand at $2,295 up to $13,500. Other big orders include Mountain-100 Black for $13,500, 48, show that there can be pretty big-ticket items in this category.
* **Frequent Products:** In this segment, such stocks as Mountain-200 Black, Mountain-200 Silver and Mountain-100 Black are in demand, fixing a preference’ certain bike models.

**Insights:**

* **Geographical Focus:** The fact that the above regions make very frequent orders suggest that localized marketing campaigns or improving the stock availability for the regions could be of some gain.
* **Premium Product Focus:** High ticket items found in bikes such as Mountain-100 Black suggest that demand for premium brands exist. Financing this product or giving bundled accessories could make more people buy these products.
* **Weekend Promotions:** Such orders were placed on the weekends only, which means that businesses could come up with special promotions such as weekend promotions or promotions exclusive to mountain bikes only.
* **Leisure-Oriented Purchases:** It is fare to conclude that weekend purchase of mountain bikes could be motivated by recreational activities or exercise. This could possibly be made alongside that regularity with events such as outdoor or recreational events so that they are promoted.

**Output:**

A screenshot of a computer

Description automatically generated

PART-3: Lessons Learned Report

**Lessons Learned:**

**Problems and Solutions:**

1. **Initial File Format Issue**  
   One of the challenges we faced at the inception of the project was to add more sheets to the data extracted from the Canva. The sheets were not saving. After doing research on the internet, we came to realize that the problem was as a result of file format.  
   **Solution:**  
   All the data transformations have been made from CSV format into Excel format that is represented by the extension XLSX. This helped to clear the problem and we were able to insert and save the sheets appropriately.
2. **Building Fact and Dimension Tables without a Primary Key**  
   First, we tried to create the dimension and fact tables not understanding the significance of a primary key. For creating proper interconnectivity between the tables, it is essential to have a primary key.  
   **Solution:**  
   We have designed a new and modified sales sheet where an OrderID column has been added at the top as key.
3. **Confusion Regarding Unique Customer Attributes**  
   When constructing the CustomerKey, we experienced the confusion since none of the customer characteristics (age, gender, country, state) are unique. We were unsure how to establish a reference to link the CustomerKey with other customer attributes.  
   **Solution:**  
   To this, we addressed this in the transformed sales sheet by creating a new column known as CustomerKey which solves the issue of having similar customers with similar characteristics.
4. **Handling Duplicate Customers in the Customer Dimension Table**  
   During the creation of dimension tables, we initially removed duplicate values for dates, products, and customers. Later, I realized that customers could have identical attributes (age, gender, country, state), which might lead to data loss.  
   **Solution:**  
   Instead of removing duplicate values in the customer dimension table, we have assigned a unique CustomerKey to each customer, even if they had identical attributes.

**Improvements for Future Projects:**

From this project, I learned on data transformation, how to build queries and data analysis throughout the data processing period. It added the importance to consider carefully of data preparation like creating primary and customer keys and also gave insights in terms of fact and dimensions how relational databases work.

If given the opportunity to repeat this project, I would:

1. Choose a dataset with better-defined primary keys in each dimension to the creation of linkage schemes.
2. Spend more time in data pre-processed and reprocessed to check all the dimension and facts are correct, thereby reducing the confusion level and mistakes during analysis.
3. To cover more detailed aspects of customers’ behavior the given range of the queries should be expanded with added specification of choice frequency and time of the day.

**Conclusion:**

Within this project, we have analyzed sales data through various queries based on customer behavior, products choice, regions and more. Through the creation and utilization of various SQL queries, we identified several key patterns in the data that can guide strategic decision-making in sales and marketing:

1. From this list, customers placing high value orders include female customers and older people, and in certain geographical areas can be seen as areas of high potential for promotional campaigns for premium products.

2. Low value orders suggest that there might be better practices for sales to enhance such as offering offers or increasing product presence.

3. Mountain bikes sales on the weekends, and peak month trends like January for Australia exposed the best time for coupon promotions and marketing strategies.

4. Some demographic findings point to locality in its marketing, for instance California US and Australian customers differ.

Due to fact and dimension tables the analysis was structured and the facts provided in tables showed how customers use the product at different time periods. Analyzing shopping behavior data may help retailers get better control of inventories, refine their customer categorization, as well as increase the promotional campaign effectiveness.