**MUNCH BOX SALES BI CASE STUDY**

**Learning Outcomes**

Following are the learning outcomes from this BI project:

1. Data analysis using Excel
   1. Use of vlookup to compare and merge two sheets of a workbook.
   2. Use of if else conditions for analysis.
   3. Use of countif formula to find the number of instances with a particular condition.
   4. To summarize the data in the form of a pivot table with multiple filters on the attributes.
2. Handling and manipulating huge databases using SQL.
   1. Use of join clause to merge tables of a database.
   2. Use of various other clauses like where, group by, order by, having to do calculations and manipulations on the database.
   3. Use of subquery to put a condition on the main data while retrieving the output.
3. Understanding the raw data and fetching useful information from it with the help of data analysis which can be further used by the company to rearrange the stocks of the products in the inventory and keep more of those products which are highly contributing to the revenue of the company and which are highly sold.

**Introduction**

‘Munch Box’ is a manufacturing company of snacks, dairy items, meat and other goods which are produced from fruits and dairy goods. They even have fruits and other non-manufactured items to sell in the store. This company was created in 1995 and has been working as a brand in the market since then to supply goods in Florida.

‘Munch Box’ company has various stores all over the state of Florida selling many different items like:

* Chips
* Beverages
* Dairy products
* Meat
* Fruits
* Frozen foods

And many other snack items and dry goods.

Due to various situations happening in the state like recession, increased interest rate, loss of income of people the overall demand of the goods in the state has decreased considerably which is leading into huge losses and wastage of products like milk, dairy items, meat, fruits and beverages.

**Problem statement (Excel)**

As there has been a lot of loss of the company due to reduced sales in the stores because of the above-mentioned circumstances in the state, the company has decided to reduce the stocks of the products in the warehouse which will prevent all the wastage and losses.

Munch box wants to know which products need to be manufactured more and for which products manufacturing needs to be stopped till the situation gets better economically. For this reason they will need a complete analysis of the products and the sales transactions to see which products are still making money and which are not.

Let us consider you are a Data analyst working externally for this company and the data is provided to you in the form of excel sheets of the transactions, product information and the inventory information.

The excel workbook has 2 sheets as **‘Transactions’** and **‘Inventory\_data’**.

1. The **‘Transactions’** sheet has 594 observations with 3 columns which consists of the data related to the transactions and the products sold in the transactions.
2. The **‘Inventory\_data’** sheet has 85 observations with 7 columns which consists of the data related to product like name, price, wholesale rate, units available in stock/ inventory and type.

The attribute description is as follows:

| **COLUMN NAME** | **DESCRIPTION** |
| --- | --- |
| transaction\_id | ID of the transaction. |
| time | Date and time of the transaction. |
| product\_id | ID of the sold product. |
| product\_name | Name of the product. |
| product\_type | Type of the product. |
| unit | Unit in which the product is measured. |
| price\_unit | Price per unit of the product. |
| wholesale | Price at which the company bought or manufactured. |
| current\_inventory | The number of units available in the inventory or goods in stock. |

Following are your tasks which are needed to be done in excel:

1. **SCENARIO - 1**

The company needs to find the distribution of the transactions with respect to the product type and analyse the best product types which are helping in generating more revenue to the company.

This is not possible to find by just using the first sheet of the excel dataset provided.

We need to merge the sheets and find the best selling product types.

**Task :** Which product type had the most number of transactions?

**Hint:** Use Vlookup.

1. **SCENARIO - 2**

In addition to the first task the company also needs to find the unit price of the products sold in each transaction by merging the unit price information provided in another sheet.

This will help the company to find the average unit price of the top selling product types for all the transactions.

**Task:** Find the average unit price of top 3 product types which have made the most number of transactions.

**Hint:** Use Vlookup.

1. **SCENARIO - 3**

From the analysis of the past transactions data it was found that the threshold of the unit price of a product which comes under the affordable and highly sold category is one dollar.

The current requirement is to categorize the transaction data into 2 types, that is, if the unit price of a product is less than one dollar it will be represented as 0 else 1.

This will help us understand that from the current data how many transactions had those products which are cheap and high in demand as well as the count of the transactions which had products whose unit price is greater than one dollar.

**Task:** Create a new column with values 0 and 1 such that if the unit price in each transaction is less than 1 dollar it will be 0 else otherwise. Find the count of both.

**Hint:**  Use if condition to create a new column.

1. **SCENARIO - 4**

The company has a requirement of finding the current status of the inventory and what is the total amount combined of all the products available in the inventory.

In the inventory dataset we have information regarding the unit price and the number of units of a product available in the inventory from which the total amount of a product contributing to the inventory price can be found out.

This will help us understand which products and product types are contributing the highest in the current inventory price.

**Task:** In the current inventory there are some units of each product. Find the top 10 products contributing the highest in terms of total price in the inventory.

**Hint:** Total cost = Unit price \* total units in inventory.

1. **SCENARIO - 5**

The company needs an understanding of the analysis on the revenue generated by each product type for each transaction which will help them decide which products to keep in the store and which needs to be eliminated.

It will be a bit difficult if this analysis is represented in the form of rows and columns where unit prices of the products are categorized according to the grouping of product types for each transaction hence a summarized format is required.

**Task:** Find the sum of unit price for all the transaction ids combined by each product type.

**Hint:** Create a pivot table.

**Problem statement (SQL)**

There are other tasks and questions which are needed for the analysis on the same data which are easier to do in SQL.

The database is provided to you from various tables containing inventory information and transaction history.

The database has 2 tables as **‘transaction\_data’** and **‘inventory\_data’**.

1. The **‘transaction\_data’** table has 594 observations with 3 columns which consists of the data related to the transactions and the products sold in the transactions.
2. The **‘inventory\_data’** table has 85 observations with 7 columns which consists of the data related to product like name, price, wholesale rate, units available in stock/ inventory and type.

The attribute description is the same as the excel data description mentioned above.

Following are your tasks which are needed to be done in SQL:

1. **SCENARIO - 1**

The company needs to find the distribution of the transactions with respect to the products and analyse the best products which are helping in generating more revenue to the company.

For this purpose we need to find those products which are highly sold and in most of the transactions.

These products will be kept in bulk stocks in the inventory which will keep generating a good amount of money to the company.

**Task:** Find the top 3 products having the most number of transactions.

**Hint:** Use group by and order by clauses.

1. **SCENARIO - 2**

According to the past study, snack type products are generally sold on a daily basis and are one of the important product types in the revenue generating products. For the current task we will specifically focus on the snack type products.

Even from the snack type products we will filter out top 10 snack items which need to be kept in bulk in the inventory.

**Task:** Find the top 10 products having the most number of transactions from snacks product type.

**Hint:** Join the required tables and use group by, order by and where clauses.

1. **SCENARIO - 3**

How can we find the days of the week or the time of the day when the highest transactions took place which had the highest number of products in it?

The solution of this task will help us answer this question.

By doing this the company will know on which days and at what time of the day the store needs to be fully stocked with all the products and their varieties which will make the customers shop more.

**Task:** Find the top 5 transactions which had the most number of products in it.

**Hint:** Use group by and order by clauses

1. **SCENARIO - 4**

From the tasks done in excel it is clear that ‘produce’ product type has a maximum number of transactions and hence the company wants to focus on that product type.

This task will help the company to know which transactions had the expensive products from the ‘produce’ type and accordingly study the time and days at which such expensive transactions take place so that they can keep the store fully stocked at that particular time and days.

**Task:** Find the list of transactions having products from ‘produce’ type and unit price more than the average unit price of ‘produce’ product type.

**Hint:** Join the required tables and use where clause and subquery.

1. **SCENARIO - 5**

From the past analysis of the transactions data it was found that the minimum amount for a product type to be considered as highly sold or popular it has to cross the threshold of 150 dollars from the monthly transactions.

We have a similar requirement of analysis from the current transaction data.

**Task:** From the transactions table find the sum of unit price for all the unique product types and filter the list of those product types having a sum of unit price greater than 150 dollars from the transactions.

**Hint:** Use group by and having clauses and join.

**Process for the submission**

Follow the steps mentioned below to submit the BI project.

1. Understand the data provided in the excel folder and create another excel workbook to do the tasks related to excel.
2. Understand the data provided in the form of tables for SQL database and upload those files in your Mysql workbench.
3. To do the tasks related to SQL write your queries in one workspace and save it as a text file.
4. Save the excel and SQL data files as well as solution files in one zip folder.