**CLASSIC WHEELS 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 do workforce management and cost-cutting.

**Introduction**

‘Classic Wheels’ is a product based manufacturing company of motorcycle spare parts and modification parts in India. This company was founded and established in 1995. Classic Wheels provides online as well as offline shopping of the motorcycle parts and their only offline store mall is in Bangalore with approximately 20 employees working in the offline store from the Sales department.

Their online store is available in the form of website and an application which has various shopping segments like:

* Seats
* Tyres
* Batteries
* Carburettor
* Clutch plates

And many other modifications and spare parts.

Various other brands and companies in this same industry are expanding in many parts of India which is creating a tough competition for Classic Wheels and as a result there was a drop in the Sales and revenue. Especially, in the offline store there was a considerable drop in sales of the products.

**Problem statement (Excel)**

The offline store of the company is facing a huge loss because of the drop in the sales and if it gets continued the competitor companies will take over the market and there will be revenue loss of the company.

As a result the company has decided to study the sales data of various products available in the offline store to know which products are making more money to the company and which are not.

Along with the study of sales and products the company wants to analyse the performance of the employees working in the offline store and on that basis workforce management actions will be taken for the employees which will result in cost cutting and higher revenue of the company.

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 sales, product pricing information and the employees responsible for the sale in the offline store.

The Excel workbook contains 3 sheets as **‘Sales Data’, Product Master’ and ‘Emp Master’.**

1. The sheet **‘Sales Data’** contains 6490 observations with 4 columns regarding the sales of the products and the employee responsible for the sale.
2. The sheet **‘Product Master’** contains observations with 3 columns regarding the product information like product name and price.
3. The sheet **‘Emp Master’** contains 20 observations with 3 columns regarding the employee information like employee name and the supervisor of the employee.

The attribute description is as follows:

| **COLUMN NAME** | **DESCRIPTION** |
| --- | --- |
| Date | Date when the product was sold |
| EMP ID | ID of the employee who sold the product to the customer. |
| Product ID | ID of the sold product |
| Unit Sold | How many units of the product were sold on that day by the particular employee. |
| Product Name | Name of the product |
| Price per unit | Price of a product per unit |
| EMP Name | Name of the employee |
| Supervisor | Name of the supervisor under which an employee is working. |

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

1. **SCENARIO - 1**

The company needs to find the total amount of all the products combined which were sold under any particular employee. From the ‘Sales Data’ it is not possible to directly find the total cost of each sale as there is no column of unit price or total cost in the ‘Sales Data’ sheet.

What can be done to solve this problem? Let us look at the given task and we will see if we can answer the question after completing the task.

**Task:** Find the total amount (units sold \* price per unit) of each instance of the sales data.

**Hint:** Derive the unit price column from another sheet using Vlookup.

1. **SCENARIO - 2**

As the total amount (price per unit \* number of units sold) of each sale is not directly given in the ‘Sales Data’ sheet, finding the amount generated under each supervisor directly from the raw ‘Sales Data’ is not possible. However, from the first task we can find the total amount of each sale. Similarly, the requirement of the company is to find the supervisor responsible for each sale, that is, to find the supervisor of the employee who is responsible for a sale.

By doing this the performance of the supervisors can also be evaluated.

**Task:** The company wants to know under which supervisor the least total amount was generated from the sales data.

**Hint:** Derive supervisor column from another sheet using vlookup then find the sum of sales amount under each supervisor.

1. **SCENARIO - 3**

From the past sales data analysis of the company it has been observed that the minimum requirement of the total cost of a product sale in a day is 15,000 which can be considered as a good sale helping the revenue of the company. Each instance from the ‘Sales Data’ needs a label whether it fits the criteria of a good sale or not.

This process will help the company to know which are the common products generating a good amount of money and under which employee and supervisor.

**Task:** Create a new column with values 0 and 1 where total cost less than 15,000 will represent 0 otherwise 1 and find the top 2 supervisors having the most number of sales below 15,000.

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

1. **SCENARIO - 4**

Data analysis on the units sold of each product from the past data suggests that for each product if a minimum 100 units are sold then that product is considered as a good revenue generating product.

Such products need to be in the list even in the future for the betterment of the company.

From the ‘Sales Data’ we need to categorize each transaction with respect to the number of units sold.

**Task:** Create a new column with values 0 and 1 where units sold less than 100 will represent 0 otherwise 1 for each instance. Find the top 3 products having the most number of days where less than 100 units are sold.

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

1. **SCENARIO - 5**

The company needs an understanding of the analysis on the revenue generated by each employee for each date which will help them decide which products to keep in the store and which needs to be eliminated and similar action on the employees based on their performance.

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

**Task:** Find the sum of total cost of all the products combined by each customer for each date and list down the top 8 employees having least total which indicates that the employee was not able to do more sales.

**Hint:** Create a pivot table.

**Problem statement (SQL)**

Along with the study of products and sales from the offline store and the study of employees working in the offline store the company finds a need of analysing the overall data of the employees working for the company in both offline and online mode as well as working in the offline store and in the office.

The company finds a need for workforce management and retrenchment of the employees for cost cutting and revenue increase.

This can be done by the analysis of employees and their salaries and departments which is stored as a database.

Let us consider you are a Data analyst working externally for this company and the database is provided to you of various tables containing employee information, department information, salary information and various other tables.

The database consists of 6 tables as **‘departments’, ‘employees’, ‘new\_dept\_emp’, ‘new\_dept\_manager’, ‘newsalaries’ and ‘newtitle’.**

1. The **‘departments’** table has 9 observations with 2 columns which consist of the data regarding the department names.
2. The **‘employees’** table has 300024 observations with 6 columns which consists of the employee data like name, hiring date, birth date and gener.
3. The **‘new\_dept\_emp’** table has 300053 observations with 3 columns which consists of the data regarding the department to which the employee belongs.
4. The **‘new\_dept\_manager’** table has 9 observations with 3 columns which consists of the department manager data.
5. The **‘newsalaries’** table has 300024 observations with 3 columns which consists of the data regarding the salaries of the employees.
6. The **‘newtitle’** table has 170896 observations with 3 columns which consists of the data regarding the designation of the employees.

The attribute description is as follows:

| **COLUMN NAME** | **DESCRIPTION** |
| --- | --- |
| dept\_no | ID of the department |
| dept\_name | Name of the department in which the employee or the manager works. |
| emp\_no | ID of the employee |
| from\_date | The start date of an instance. It can be the start date for the employee/ manager to join a department, or the start date of a particular amount of salary for an employee/ manager. |
| to\_date | The end date of an instance. It can be leaving the company, end date for the employee/ manager to work in a department, end date of a particular amount of salary for an employee/ manager. |
| birth\_date | Date of birth of an employee |
| first\_name | First name of the employee |
| last\_name | Last name of the employee |
| gender | Gender of the employee |
| hire\_date | Date of hiring of the employee |
| salary | Salary of the employee. For one employee there can be multiple salaries depending upon the time period. As years pass by salary changes of an employee |
| title | Designation of an employee. |

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

1. **SCENARIO - 1**

The company is in urgent need for workforce management which means they need to reduce the number of employees for cost cutting due to the recent crises they have been facing.

One way to do this is to do analysis on the highly paid employees such that if their work is not satisfactory they can be terminated which will help largely in the cost cutting process.

Terminating low salaried employees will not help the company much.

**Task:** Find the top 10 employees (names) having the highest salaries.

**Hint:** Join the tables of employees and salary.

1. **SCENARIO - 2**

Analysing the departments of high salaried employees and the salaries from such departments is as important as the first task of salary analysis.

The company has a requirement of finding the departments in which high salaried positions exist.

This analysis will help the company to reduce the workforce specifically of such departments which will tremendously help in the cost cutting process of the company.

**Task:** Find the department names of the top 10 employees having the highest salaries.

**Hint:** Join the required tables and sort the values.

1. **SCENARIO - 3**

Similar to the previous scenario, not just the high salaried departments need to be focused on but also high employed departments analysis can help the company to do workforce management.

The company has decided a threshold of the number of employees for each department and to take an action on the departments with the number of employees more than the threshold they first need to find the highly employed departments.

**Task:** Find the top 3 departments having the highest number of employees.

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

1. **SCENARIO - 4**

From the observations and analysis of the excel dataset we can say that the major focus of the company is to reduce the workforce of the sales department. As the number of sales are affected majorly the company finds the need to urgently reduce the workforce of sales departments.

This issue can be solved by focusing on those sales employees whose salary is greater than the average salary of the sales department.

**Task:** Find the list of employees from the sales department having a salary more than the average sales department salary.

**Hint:** Use where clause and subquery

1. **SCENARIO - 5**

According to the current situation of the company, they have decided the threshold/ margin of 65,000 monthly average salary of the departments which will help them in cost cutting as well as in workforce management.

How can we find the designations or job titles which are having high salaries?

If this question is resolved the company can focus on such designations or job titles and will stop their new hiring as well as can focus on the retrenchment of the employees having these designations for cost cutting purposes.

**Task:** Find the titles having an average salary greater than 65,000.

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

**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. A video explaining the project.
5. Save the excel and SQL data files as well as solution files in one zip folder.