**CAPSTONE PROJECT**

**USING SQL**

**Submitted to – Odin School**

**Submitted by – Sushree Deepa Jena**

**Roll Number – S7615**

**Sales Analysis for Amazon**

**Introduction**

This analysis will analyse some previous sales data of Amazon to see sales performance in various branches of Amazon by considering different factors and will get some insights so that sales will improve by focusing on weaker factors.

**Dataset**

Amazon dataset which has 17 columns and 1000 rows

**Analysis**

**Product Analysis**

Analysis done on the data to understand the different product lines, the product lines performing best and the product lines that need to be improved.

**Sales Analysis**

This analysis gives us sales trend . It will help us measure the effectiveness of each sales strategy the business applies and what modifications are needed to gain more sales.

**Customer Analysis**

This analysis aims to uncover the different customer segments, purchase trends and the profitability of each customer segment.

**Data Wrangling**

This is the first step where inspection of data is done to make sure NULL values and missing values are detected and data replacement methods are used to replace missing or NULL values.

Following steps have done in this phase:

* A database named “Amazon” has created in MySQL WorkBench.
* In that database “Amazon\_Data” table has created in which 17 columns has included.
* Then csv file is imported to that table which has 17 columns and 1000 rows.
* Then for finding any NULL value present or not “IS NULL” condition has checked in Where clause.
* There were no NULL value in our imported data.

**Feature Engineering**

This will help us generate some new columns from existing ones.

Following steps have used in this phase:

* A new column "timeofday" added by using “ALTER Table” command which will give insight of sales in the Morning, Afternoon and Evening. This will help to analyse on which part of the day most sales are made.
* A new column named "dayname" added which contains the extracted days of the week on which the given transaction took place (Mon, Tue, Wed, Thur, Fri). This will to analyse on which week of the day each branch is busiest.
* A new column named "monthname" has been added that contains the extracted months of the year on which the given transaction took place (Jan, Feb, Mar). Help determine which month of the year has the most sales and profit.

**Product Analysis**

* The three branches offer a variety of products across six different Product Lines.
* Food and Beverages have the highest contribution in revenue generation and it outranks other product lines in sales.
* Food and Beverages have incurred the highest Value Added Tax.
* It has been observed that Health and Beauty line of Products have fared poorly in sales as compared to others.
* The product line that gathered most favourable ratings is Food and Beverage whereas Home and Lifestyle line of products receives paned reviews from customers.

**Sales Analysis**

* There are three payment methods available to customers – Cash, E-wallet and credit card, out of which customers mostly prefer using E-Wallet.
* The month of January has seen the highest footfalls in the store which can be attributed to the New Year Shopping spree which also explains why Cost of Goods Sold was highest in January.
* Branch C which lies in Naypytaw has generated highest revenue among the three branches whereas Mandalay’s Branch B fared lowest.
* An important observation regarding shopping time is that it starts at a good scale in the morning, peaks by the afternoon and slows down by the evening.
* Yangon has incurred highest Value Added Tax amongst the three cities.

.

**Customer Analysis**

* Customers are classified into two categories – Members and Normal(non-members).
* It has been identified that female customers often associates themselves with Fashion Accessories whereas male customers mostly prefer Health and Beauty.
* Among the two customers types, Members have higher contribution in the overall revenue as well as in Value Added Tax.
* Another observation in context of gender is that female customers have more occurrences than males.
* Most of the Female customers have preferred Branch C as their favourite shopping destination whereas Branch A has recorded highest male customers.
* Afternoon is the peak shopping time which is also reflected in the ratings which were mostly received during that time.

**Recommendations**

* Better promotional strategies are needed for the Health and Beauty product line in all three branches.
* Branch managers must ensure the Food and Beverage product line has enough inventory so that there is never a shortage during periods of strong demand.
* To raise ratings, consideration must be given to the quality of goods sold within the Home and Lifestyle product line.
* Branch located at Mandalay need a better marketing campaign and launch exclusive offers to improve their sales track record.
* To offer end-of-day deals, branches must be aware of which product lines clients like when they shop in the evening. In this context, a consumer survey or feedback can be helpful.
* Branches need to focus upon how they can bring more male customers by advertising their most preferred product line.
* Stores also need to focus on the normal customer types as their revenue share is quite less as compared to the members.