Walmart Sales Analysis

SQL Project

BY : ABHINAV AKARSH

CONTENTS

ABOUT

This project aims to explore the Walmart Sales data to understand top performing branches and products, sales trend of different products, customer behaviour. The aims is to study how sales strategies can be improved and optimized. The dataset was obtained from the Kaggle Walmart Sales Forecasting Competition.

"In this recruiting competition, job-seekers are provided with historical sales data for 45 Walmart stores located in different regions. Each store contains many departments, and participants must project the sales for each department in each store. To add to the challenge, selected holiday markdown events are included in the dataset. These markdowns are known to affect sales, but it is challenging to predict which departments are affected and the extent of the impact." source

PURPOSE

The major aim of this project is to gain insights into the sales data of Walmart to understand the different factors that affect sales of the different branches.

It also helps to understand customer's taste and preference and their purchasing patterns.

With the help of insights, Management will take data driven decisions which will help to flourish the business.

ABOUT DATA

Column	Description	Data Type
invoice_id	Invoice of the sales made	VARCHAR(30)
branch	Branch at which sales were made	VARCHAR(5)
city	The location of the branch	VARCHAR(30)
customer_type	The type of the customer	VARCHAR(30)
gender	Gender of the customer making purchase	VARCHAR(10)
product_line	Product line of the product solf	VARCHAR(100)
unit_price	The price of each product	DECIMAL(10, 2)
quantity	The amount of the product sold	INT
VAT	The amount of tax on the purchase	FLOAT(6, 4)
total	The total cost of the purchase	DECIMAL(10, 2)
date	The date on which the purchase was made	DATE
time	The time at which the purchase was made	TIMESTAMP
payment_method	The total amount paid	DECIMAL(10, 2)
cogs	Cost Of Goods sold	DECIMAL(10, 2)
gross_margin_percentage	Gross margin percentage	FLOAT(11, 9)
gross_income	Gross Income	DECIMAL(10, 2)
rating	Rating	FLOAT(2, 1)

ANALYSIS LIST

1.PRODUCT ANALYSIS - Conducted analysis on the data to understand the different product lines, which products lines performing best and the product lines that need to be improved.

2.SALES ANALYSIS - This analysis aims to answer the question of the sales trends of product. The result of this can help use measure the effectiveness of each sales strategy the business applies and what modifications are needed to gain more sales.

3.CUSTOMER ANALYSIS - This analysis aims to uncover the different customers segments, purchase trends and the profitability of each customer segment.

APPROACH USED

1.<u>Data Wrangling:</u> 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.

i. Build a database

ii.Create table and insert the data.

iii.Select columns with null values in them. There are no null values in our database as in creating the tables, we set **NOT NULL** for each field, hence null values are filtered out.

Creation of database and Insertion of data

```sql
CREATE DATABASE IF NOT EXISTS SalesDataWalmart;

CREATE TABLE SALES  (
invoice_id VARCHAR(30)  NOT NULL PRIMARY KEY ,
branch VARCHAR(5) NOT NULL,
city VARCHAR(30) NOT NULL,
customer_type VARCHAR(30) NOT NULL,
gender VARCHAR(10) NOT NULL,
product_line VARCHAR(100) NOT NULL,
unit_price DECIMAL(10,2) NOT NULL,
quantity INT NOT NULL,
VAT FLOAT (6,4) NOT NULL,
total DECIMAL (10,4) NOT NULL,
date DATETIME NOT NULL,
time TIME NOT NULL,
payment_method VARCHAR(10) NOT NULL,
COGS DECIMAL(10,2) NOT NULL,
gross_margin_pct FLOAT (11,9),
gross_income DECIMAL (12,4) NOT NULL,
ratings FLOAT (2,1) NOT NULL
)
```

2.<u>Feature Engineering :</u> This is the second step and it will help to generate some new columns from existing ones.

i.Added a new column named *time_of_day* to give insights of sales in the morning,afternoon and evening.This will help answer the question on which part of the day sales are made.

ii.Added a new column named *day_name* that contains the extracted days of the week on which the given transaction took place.This will help answer the questions on which week of the day each branch is busiest.

iii.Added a new column named *month_name* that contains the extracted month of the year on which the given transaction took place.It will help determine which month of the year has most sales and profit.

3.Exploratory Data Analysis : EDA is done to find the answer of the listed questions and aim of the project

1. Find out the most common payment method used by customers during shopping.

REPORT

SQL QUERY

```
select payment_method,
       count(payment_method) as total_count
from sales
group by payment_method
order by total_count desc
```

OUTPUT SHEET

payment_method	total_count
Cash	344
Ewallet	342
Credit card	309

2.Find out the most selling product line in the department

SQL QUERY

```
select product_line,
       sum(quantity) as total
from sales
group by product_line
order by total desc
```

OUTPUT SHEET

product_line	total
Electronic accessories	961
Food and beverages	952
Home and lifestyle	911
Sports and travel	902
Fashion accessories	902
Health and beauty	844

3.FIND OUT REVENUE OF ALL THE PRODUCT LINE IN THE DEPARTMENT.

SQL QUERY

```
select product_line,
       round(sum(total),1) as total_revenue
from sales
group by product_line
order by total_revenue desc
```

OUTPUT SHEET

product_line	total_revenue
Food and beverages	56144.8
Fashion accessories	54305.9
Sports and travel	53936.1
Home and lifestyle	53861.9
Electron	.2
Health and beauty	48854.4

Home and lifestyle

4.FIND OUT WHICH CITY GENERATES THE LARGEST REVENUE.IT SHOULD BE FOLLOWED BY SECOND AND THIRD LARGEST ALSO.

SQL QUERY

```
select city,branch,
       round(sum(total),0) as total_revenue
from sales
group by city,branch
order by total_revenue desc
```

OUTPUT SHEET

city	branch	total_revenue
Naypyitaw	C	110491
Yangon	A	105861
Mandalay	B	104535

5.FIND OUT THE LIST OF PRODUCT LINES HAVING HIGHEST VAT (VALUE ADDED TAX)

SQL QUERY

```
select product_line,
       round(avg(VAT),2) as avg_vat
from sales
group by product_line
order by avg_VAT desc
```

OUTPUT SHEET

product_line	avg_vat
Home and lifestyle	16.03
Sports and travel	15.76
Health and beauty	15.41
Food and beverages	15.37
Electronic accessories	15.15
Fashion accessories	14.53

6.FIND OUT THE COMMON PRODUCT LINE BY GENDER IN THE DEPARTMENT

SQL QUERY

```
select product_line,
       gender,
       count(gender) as total_count
from sales
group by product_line,gender
order by total_count desc
```

OUTPUT SHEET

product_line	gender	total_count
Fashion accessories	Female	96
Food and beverages	Female	90
Health and beauty	Male	88
Sports and travel	Female	86
Electronic accessories	Male	86
Food and beverages	Male	84
Electronic accessories	Female	83
Fashion accessories	Male	82
Home and lifestyle	Male	81
Home and lifestyle	Female	79
Sports and travel	Male	77
Health and beauty	Female	63

7.FIND OUT THE AVERAGE RATINGS OF ALL THE PRODUCT LINES IN THE STORE.

SQL QUERY

```
select product_line,
       round(avg(ratings),2) as avg
from sales
group by product_line
order by avg desc
```

OUTPUT SHEET

product_line	avg
Food and beverages	7.11
Fashion accessories	7.03
Health and beauty	6.98
Electronic accessories	6.91
Sports and travel	6.86
Home and lifestyle	6.84

8.FIND OUT THE NUMBER OF SALES MADE IN EACH TIME OF DAY IN THE WEEKDAY

SQL QUERY

```
select time_of_day,
       count(invoice_id) as total_sales
from sales
group by time_of_day
```

OUTPUT SHEET

time_of_day	total_sales
Evening	429
Afternoon	376
Morning	190

9.FIND OUT WHICH CITY HAS THE HIGHEST VAT(VALUE ADDED TAX)

SQL QUERY

```
select city,
       avg(vat) as vat
from sales
group by city
order by vat desc
```

OUTPUT SHEET

city	vat
Naypyitaw	16.09010850
Mandalay	15.13020824
Yangon	14.87020798

10.FIND OUT DAYS OF WEEK WHICH HAVE BEST AVERAGE RATINGS

SQL QUERY

```
select day_name,
       round(avg(ratings),2) as avg
from sales
group by (day_name)
order by avg desc
```

OUTPUT SHEET

day_name	avg
Monday	7.13
Friday	7.06
Tuesday	7
Sunday	6.99
Saturday	6.9
Thursday	6.89
Wednesday	6.76

CONCLUSION

- Cash is the most common payment method used with total count of 344.
- Electronic accessories was the top selling product line in the store followed by food & beverages.
- Among all the product lines, Food & beverages contributes highest to the revenue. (RS.56144)
- “Home & lifestyle” has the highest VAT. It is then followed by “Sports and travel” & “Health & beauty”
- “Evening” was the best time for people to do their shopping with the total count of 429 transactions.