



SUPERMARKET SALES ANALYSIS

Dataset Exploration Part 1

About The Data Set

In the most crowded cities, the competition among sales market is incessantly increasing. To cope-up and to grow the business, sales analysis becomes crucial for the organisation. This dataset is one among the sales of a supermarket brands, containing data of 3 different branches recorded for 3 months. This dataset got my attention because I can learn to analyse and understand the trends of sales by different variables present here like types of products, date and time, rating and gross income.

This dataset is a population, not a sample. (I have changed the city names for my convenience)

Data Dictionary

This data set consists of 1001 records and 17 columns with a mix of both, qualitative and quantitative data types. Below is the description of each variable present in the data set.

1. **Invoice Id:** It's the unique computer-generated identification number of every printed sales slip.
2. **Branch and City:** 3 different branches of supermarket in 3 different cities categorized as A, B and C.
 - A. Toronto
 - B. Barrie
 - C. Orillia
3. **Customer type:** Type of customer is categorized as member and normal, in which member depicts the customers with membership card and normal depicts the new and those who don't have membership card (dichotomous data).
4. **Gender:** Gender type of customer (dichotomous data).
5. **Product line:** It's a qualitative data of varied item groups such as electronic accessories, fashion accessories, food and beverages, health and beauty, home and lifestyle, sports and travel(nominal).
6. **Unit price:** Price of each product (priced in cad)
7. **Quantity:** Number of products purchased by customer (quantitative data)
8. **Tax:** 5% tax fee for applied for each invoice (continuous data).
9. **Total:** Total price including tax. (priced in cad)
10. **Date:** Date of generated invoice
11. **Time:** Purchase time
12. **Payment:** Way of payment used by customers from 3 options available – Cash/ Credit / E-wallet
13. **COGS:** Cost Of Goods Sold. (priced in cad)
14. **Gross margin percentage** – percentage of gross margin income
15. **Gross margin income** – (gross revenue – COGS)
16. **Rating:** recorded customer satisfaction rating on their overall shopping experience on a scale of 1 to 10 (interval scale).

Research Questions

1. Which city leads in sales? On that note, which location's branch should be chosen for expansion and which category of items it should focus on?
 - As we can see, there is data of 3 cities. So, the sales among them may be compared which could answer above question
2. What is the purchasing preference of men and women? Is there any difference in category they prefer more?
 - In this dataset, there are different categories of items which can be preferred by different genders
3. Is there any relationship between COGS and ratings? Using these, shall gross income be predicted?
 - Generally, ratings not only depend on customer service one gets, but also depend on the price of the goods as customer may compare the price from one store with another. And using the current trend future gross income may be calculated
4. Which is the day, the products are sold maximum? And which hour of the day is busiest?
 - Looking at the date and time the products bought, thought of above question

Demonstrated Tracking

To start analysing sales trends in bigger datasets, time range must be determined. But here there is data of 3 months, so that would be the time scale.

To work on this project, there should be the knowledge of sales analysis. How to compare different variables using the combination of different variables.

Measures should be taken to learn to understand and analyse the questions regarding:

- The most selling products
- Analysing preference of men and women's purchasing pattern
- What can be done to attract the non members to member's list.
- Relation between COGS and ratings
- Relation between unit price and quantity with gross income

References

Historical record of sales data in 3 different supermarkets. 2019. [Supermarket sales | Kaggle](#)