

DATA SCIENCE PROJECT REPORT
(Project Semester August-December 2020)

***SUPERMARKET SALES
DATA***

Submitted by

MAHIMA UPADHYAYA

Registration No: 11812840

Computer Science and Engineering

Section: KM079

Course Code INT217

Under the Guidance of

Ashu

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Mahima Upadhaya bearing Registration no. 11812840 has completed Data Science project titled, “**Supermarket sales**” under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science

Lovely Professional University

Phagwara, Punjab.

Date:

DECLARATION

I, Mahima Upadhyaya, student of Lovely Professional University under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this assignment is based on my own intensive work and is genuine.

Name of the student: Mahima

Upadhyaya

Date: 05-12-2020

Signature: *Mahima Upadhyaya*

Registration No. 11812840

ACKNOWLEDGEMENT

A project work is a combination of views, ideas, suggestions and contribution of many people. Thus, one of the pleasant parts of writing the report is to thank those who have contributed towards its fulfilment.

I consider it as great privilege to have esteemed Lecturer Ms. Ashu as my project guide. I take this opportunity to express my sincere gratitude to him through constant advice and constructive criticism nourished my interest in the subject and provided a free and pleasant atmosphere to work against all odd situations. I avail this opportunity to extend my heart full thanks and deep respect to faculty member for their able guidance during this project.

My gratitude to all those who responded to my questionnaire in a well-defined manner and helped me acquiring knowledge.

I would like to communicate a deep sense of gratitude to all these people without whom my project would not have been such a great learning experience.

Mahima Upadhyaya

KM079

Reg no: 11812840

Lovely Professional University

Table of Contents

| Sr.No. | Particulars |
|--------|-------------------------------|
| 1. | Introduction |
| 2. | Scope of The Analysis |
| 3. | Existing System |
| 4. | ETL Process |
| 5. | Analysis on Dataset |
| 6. | List of Analysis with Results |
| 7. | Future Scope |
| 8. | References |
| 9. | Bibliography |

Introduction

The dataset is one of the historical sales of supermarket company which has recorded in 3 different branches for 3 months data.

Predictive data analytics methods are easy to apply with this dataset. This dataset can be used for predictive data analytics purpose.

- **Dataset:** <https://www.kaggle.com/aungpyaeap/supermarket-sales>

| Column | Definition |
|-------------------------|------------------------------------|
| Invoice Id | Unique Id for each sale |
| Product line | Name of Commodity sold |
| City | City of sales |
| Branch | Branch of the supermarket |
| Date | Date of selling |
| Payment | Mode of payment |
| Unit Price | Price of a Unit of product |
| Quantity | Total units of products |
| Gender | Gender of the buyer |
| cogs | Cost of goods sold |
| Time | Time of selling |
| Gross Margin Percentage | |
| Gross Income | |
| Rating | Rating given by the customer/buyer |

Scope of The Analysis

The growth of supermarkets in most populated cities is increasing and market competitions are also high.

The dataset is one of the historical sales of supermarket company which has recorded in 3 different branches for 3 months data.

Predictive data analytics methods are easy to apply with this dataset. This dataset can be used for predictive data analytics purpose.

Since such vast field of data present of the supermarket sales is wide range of scope of the analysis of date. For example:

- a) Month wise Gross margin/income.
- b) Product wise sale.
- c) Trends in revenue throughout the year
- d) Branch wise sale
- e) Digital payment wise gross revenue ETC.

Existing System

Before existence of Data Science, analyzing data used to be hectic task and existing system didn't used to analyses the data with perfection.

Without existence of current cutting-edge technology of data science, we can get actionable insights in the dataset of the Indian Trade.

Following are the benefits which weren't present in the existing system of data analyzing:

1. Making Better Decision with The Help of Data
2. Directing actions based on trends- which later defines the goals required for profit.
3. Doing challenging stuffs with the help of prediction which is done by data.
4. Identifying various opportunities to increase the profit,
5. Making decision with Quantifiable, data driven evidence so that loss doesn't happens.
6. Testing the decisions taken by the data and watching and analyzing the trend.

Source of The Dataset

- The dataset is taken from the Kaggle with the name 'Supermarket sales'.

<https://www.kaggle.com/aungpyaeap/supermarket-sales>

- Author of the Dataset: Aung Pyae
- Data last updated: 2019

Analysis of Dataset

1. **Month wise Gross income trend for each Supermaket branch.**

a) Introduction: The analysis shows the month wise gross income trend for each branch of supermarket. i.e, A,B,C

b) Specific Requirements/Functions and Formulas:

i)Pivot table of Supermarket data

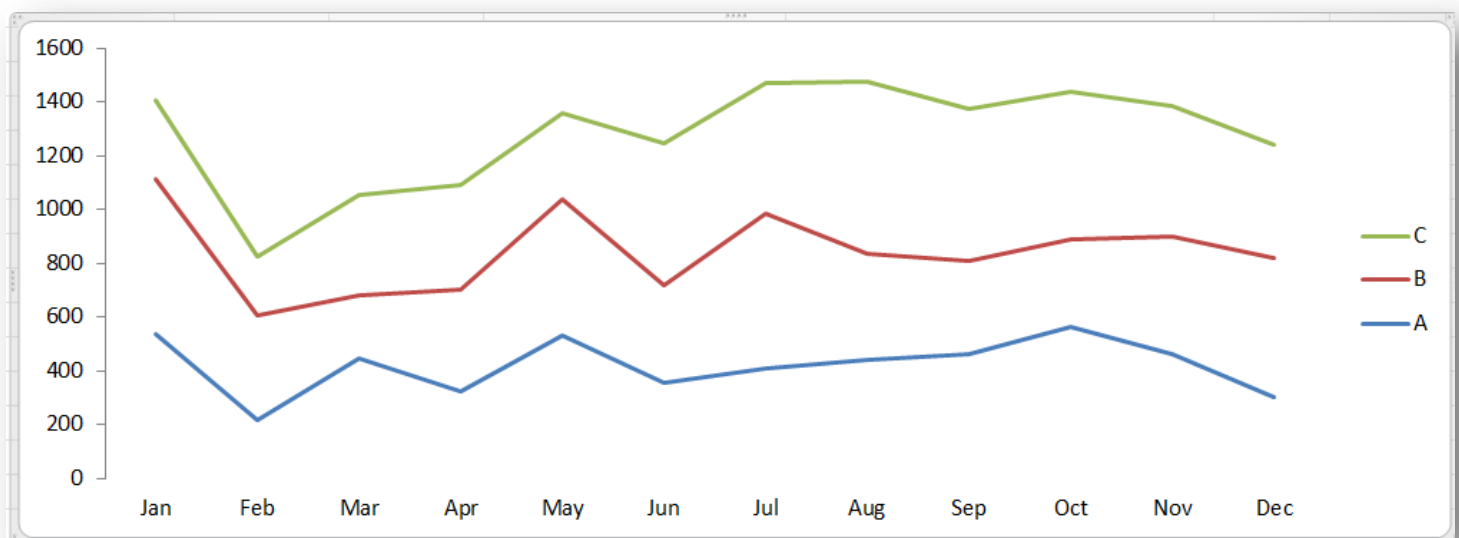
ii) Columns: Gross Margin,
Gross Income, Date,branch

iii) Line Chart

c) Analysis Results:

- Gross income is always high during January and gradually decreases for all supermarket branches
- It is the lowest in February.
- Branch C has the hishest value of gross income for all the months.
- Branch A has the lowest value of gross income for all the months.

d) Visualization:



2. TOP 5 PRODUCT LINES SOLD

a)Introduction: The analysis shows the top 5 product lines sold out of the 7 product lines.

b)Specific Requirements/Functions and Formulas:

i)Pivot table of the data of Import

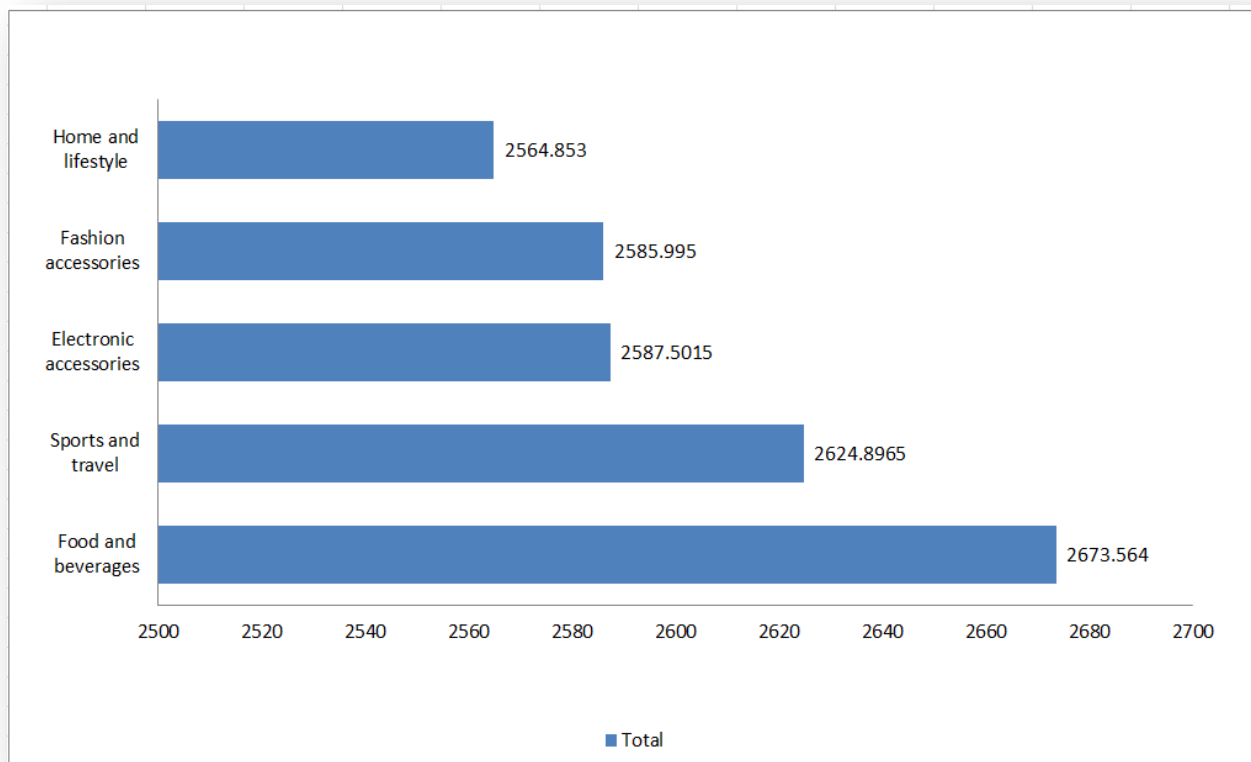
ii)With the help of data Clustered Bar chart is plotted.

iii) Formatted to get top 5 products

c)Analysis Results:

- Food and beverages is the most popular product line followed by sports and travel, electronics and accessories , fashion accessories and lastly home and lifestyle.

d)Visualization:



3. Share of each product line in the income

a) Introduction: The analysis shows the share of each product line in the Final income

b) Specific Requirements/Functions and Formulas:

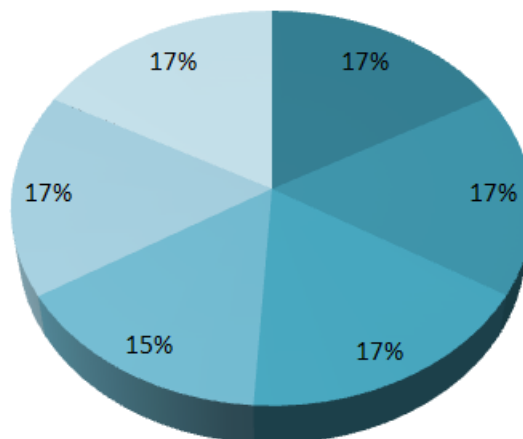
- Pivot table of the data of Supermarket.
- With the help of 3D pie chart is plotted.

c) Analysis Results:

- Home and lifestyle contributes the lowest to the final income. While other product lines contribute almost equally.

d) Visualization:

■ Electronic accessories ■ Fashion accessories ■ Food and beverages
■ Health and beauty ■ Home and lifestyle ■ Sports and travel



4. Ewallet payment by region/city.

a)Introduction: The analysis shows the revenue generated by ewallet in th

b)Specific Requirements/Functions and Formulas:

i)Pivot table of the data of Export

ii)With the help of data doughnut chart is plotted.

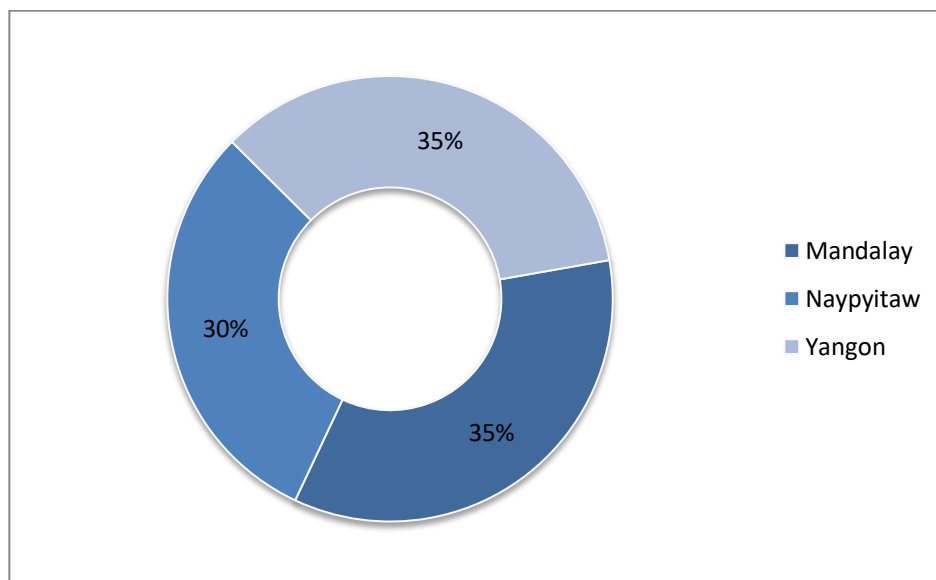
c)Analysis Results:

i) Yangon is the city with highest digital payments.

ii) Mandalay has the least ewallet payments of all

others.

d) Visualization:



5. Monthly Revenue by gender.

c)Introduction: The analysis shows the montly revenue by gender.

d)Specific Requirements/Functions and Formulas:

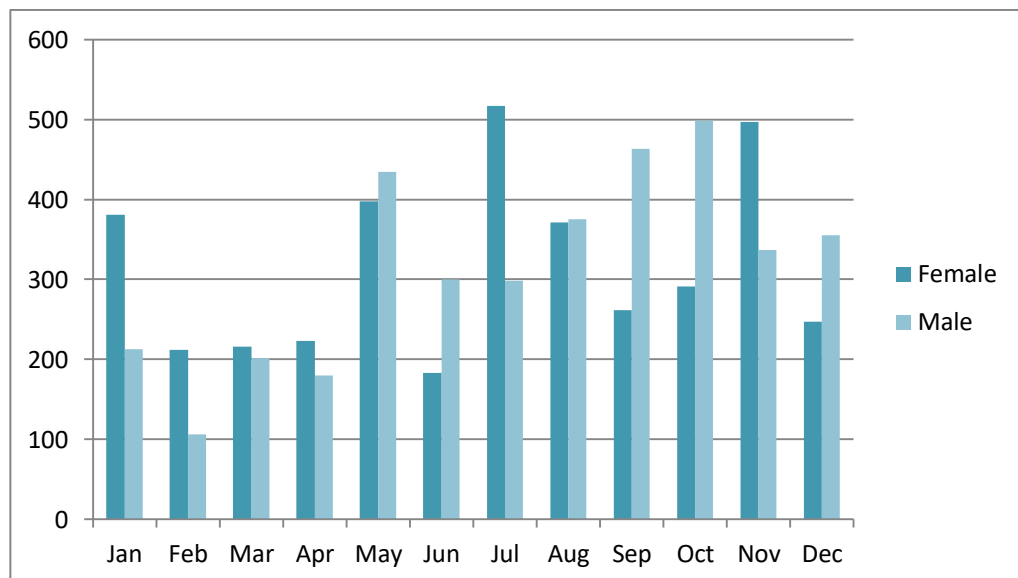
i) Pivot table of the data of Import.

ii) With the help of data Clustered Bar chart is plotted.

c)Analysis Results:

- Females contributed to most purchases.
 - In the month of july females purchased the most.
 - In the month of October the purchase made by males was highest out of all the other months.

d) Visualization:



List of Analysis with Results

For this dataset:

- In the month of July females purchased the most.
- In the month of October the purchase made by males was highest out of all the other months.
- Females contributed to most purchases
- Home and lifestyle contributed the lowest to the final income. While other product lines contributed almost equally.
- Yangon is the city with highest digital payments.
- Mandalay has less ewallet payments of all others.
- Food and beverages is the most popular product line followed by sports and travel, electronics and accessories , fashion accessories and lastly home and lifestyle.
- Gross income is always high during January and gradually decreases for all supermarket branches
- Gross Income is the lowest in February.
- Branch C has the hishest value of gross income for all the months.
- Branch A has the lowest value of gross income for all the months.

References

1. www.kaggle.com
2. www.youtube.com
3. www.google.com
4. www.stackoverflow.com
5. www.github.com

Bibliography

1. Microsoft Excel 2016 Bible: The Comprehensive Tutorial Resource by John Walkenbach, Wiley
2. Fundamentals of Business Analytics by R.N. Prasad, Seema Acharya, Wiley