**  
COMSATS University Islamabad**

Project Document

for

**Supermarket Sales Predictor**

***By***

**Shahrooz Wali CU/FA17-BCS-080/ISB**

***Submitted To*Dr. Usman Yaseen**

*Bachelor of Science in Computer Science (SP-2021)*

Table of Contents

[1. Introduction 2](#_Toc59825174)

[2. Modules 2](#_Toc59825175)

[2.1 Sales Analyzer 2](#_Toc59825176)

[2.2 Sales Predictor 2](#_Toc59825177)

[3. Dataset: 2](#_Toc59825178)

[4. Work Division 3](#_Toc59825179)

[5. Sale Analysis: 3](#_Toc59825180)

[6. Sale Predication: 4](#_Toc59825181)

**Project Category:**

* Data science
* Analysis and Prediction
* Problem Solving

.

# Introduction

Supermarket sale predictor will be a model that train on the dataset of supermarket sales. As our analysis and prediction is based on the sales that happens on the supermarket. That will help us to understand which things in the supermarket are selling mostly and what things we keep in the supermarket increase how much sales in the market. This model will predicate what steps according to the dataset after analysis of it that increase the client satisfaction and sales of the market

# Modules

Modular division of the whole project is based on the dataset that will be used to train a model according to our requirement.

## Sales Analyzer

In this project we will analyses the sells dataset of a supermarket and try to analyses how it can develop in future and what are the necessary steps needed to be taken for its betterment and customer satisfaction. We will find the co-relation of the sales with customer satisfaction and profit to the supermarket.

## Sales Predictor

In this part of project, we use the report result of the sale analyzer on the bases result we implement different machine learning model to predict which step help us to increase the quality and customer satisfaction for our sales things.

# Dataset:

The growth of supermarkets in most populated cities are 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.

**Attribute information**  
  
Invoice id: Computer generated sales slip invoice identification number  
  
Branch: Branch of supercenter (3 branches are available identified by A, B and C).  
  
City: Location of supercenters  
  
Customer type: Type of customers, recorded by Members for customers using member card and Normal for without member card.  
  
Gender: Gender type of customer  
  
Product line: General item categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel  
  
Unit price: Price of each product in $  
  
Quantity: Number of products purchased by customer  
  
Tax: 5% tax fee for customer buying  
  
Total: Total price including tax  
  
Date: Date of purchase (Record available from January 2019 to March 2019)  
  
Time: Purchase time (10am to 9pm)  
  
Payment: Payment used by customer for purchase (3 methods are available – Cash, Credit card and Wallet)  
  
COGS: Cost of goods sold  
  
Gross margin percentage: Gross margin percentage  
  
Gross income: Gross income  
  
Rating: Customer stratification rating on their overall shopping experience (On a scale of 1 to 10)

# Work Division

Shahrooz Wali FA17-BCS-080

# Sale Analysis:

In the Sale Analysis part of the project, we analyze the sale per hour to take decision which are in demand. In which we can check at any time the demand of sales by point to that time. The packages we use in the analysis of the sale on supermarket dataset are.

1. Numpy
2. Pandas
3. Seaborn
4. Matplotlib
5. Os

These libraries packages are used for the analysis of the sales of the supermarket. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy

**Seaborn** is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics

In the part of the project, we also analyze that the customer can influence the sale of the supermarket. There are two type of customer in our project one is member of the market and the other are normal member of the market.

We also analyze that customer influence effect the customer rating in the cities that we analyze the sale.

# Sale Predication:

In this project part we will analyses the sells dataset of a supermarket and try to analyses how it can develop in future and what are the necessary steps needed to be taken for its betterment and customer satisfaction. This will be followed in a step-by-step analysis report from the historical data and its implementation in future.

1. Data Exploration and insights collection
2. Segregation of important data factors
3. Model building
4. Data visualization
5. Arriving at the conclusion from data
6. Data sources and citation

The packages we use in the analysis of the sale on supermarket dataset are.

1. Numpy
2. Pandas
3. Seaborn
4. Matplotlib
5. Os
6. Train\_Test\_Split

From the analysis of the data across various parameters and processing’s, we have finally alighted upon the following conclusions:

1. The gross percent spreads with the unit price, i.e., for a unit price of higher value, there are a wider range of gross percentages of the commodities in sale
2. Branch C of the supermarket makes the greatest profit
3. City of Naypyidaw draws the maximum percentage of the customers
4. Most people have voted up to a rating of 6.0 which says that the services and facilities are good for all branches
5. From the heatmap we can analyses that tax 5%, total, cogs and gross income are highly correlative
6. Mode of payment does not matter at all
7. The survey was performed on an equal number of men and women