[Title]

[Your Name]

[Date]

Table of Contents

[Implementation of Phase 1 3](#_Toc170210879)

[Concept 3](#_Toc170210880)

[Resources 3](#_Toc170210881)

[Results 3](#_Toc170210882)

[Implementation of Phase 2 3](#_Toc170210883)

[Concept 3](#_Toc170210884)

[Resources 4](#_Toc170210885)

[Result 4](#_Toc170210886)

[Reflection 5](#_Toc170210887)

# **Implementation of Phase 1**

## Concept

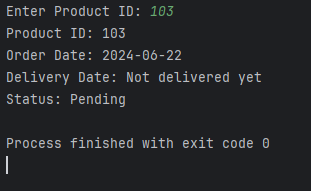
The objective of this program is to create a simple order tracking program in which user can check the statues of a product by entering the product id. There is a function in program named `check\_product\_status` which takes the product ID as argument and find the details of product in the Data Frame include its order details.

## Resources

This program is written in python programming language and using a Sample dummy database for performing operations and executed in whole code by using pandas library. The Data Frame holds a details of Customer order details which include product ID, order date, delivery date and current status of order.

## Results

By Entering the ordered Product ID, user get the details which notify order date, delivery date and current status weather it is dispatched or pending. This is a highly simplified look as to how Python can be utilized for data processing and the creation of dynamic programs. The program utilizes pandas to store and manage datasets and, in response to the user’s input, quickly acquires and outputs the desired data. This kind of functionality is often required in various real-world applications, for instance, in the e-commerce systems where the visibility of Order statuses is a core concern for business functionality and customer satisfaction.



# Implementation of Phase 2

## Concept

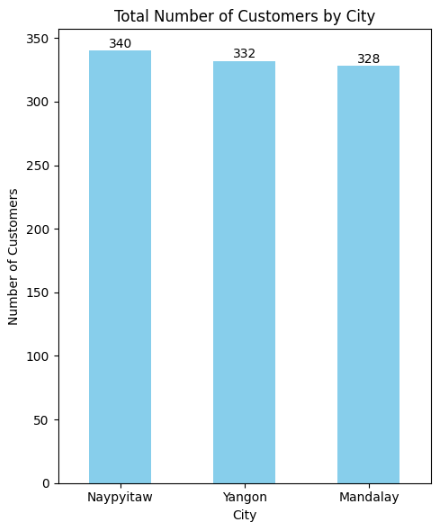
The aim of this project is to apply K-means clustering to a supermarket sales dataset and do the customer segmentation. Customer segmentation is the process through which customers are grouped together according to certain features so that specific marketing strategies can be initiated among them.

## Resources

In this case, the potential target customer segments are based on characteristics like branch, city, customer type, gender, product line, payment method, and rating. I mine this data set from Kaggle and executing on google colab or use any python notebook. For visualizations, two famous libraries are Seaborn and Matplotlib are employed in the present work. We create several plots: the dispersion of clusters over cities and branches, the distribution of customers by type and gender inside clusters, the preferences concerning the type of products by clusters, and the payment methods preferred by customers belonging to the clusters.

## Result

There are miscellaneous attributes of the dataset such as invoice number, branch, city, customer type, gender, products type, unit price, quantity, tax, total, date, time, payment, cost of goods sold, gross margin percentage, gross income and rating. First, the data is preprocessed through feature encoding where categorical variables have to be encoded. To this end, we make use of label encoder belonging to scikit learn dictionaries. Particularly, the customer type, gender, the product line, and the payment method can be turned into numerical traits for clustering.



we calculate the highest number of clusters with the lowest SSE using the elbow method, which presents the plot on the variation of the SSE for various numbers of clusters. Analyzing the elbow curve, we decided to use the value of 4 clusters for K-means algorithm. We then apply K-means model to the selected features andleted to cluster labels of the data points.

To provide the better conceptions regarding the clusters, we establish visualisations. First, we exchange the numerical codes of the city and branch columns by the actual names (the actual names of the cities ‘Yangon,’ ‘Nay Pyi Taw’, and ‘ Mandalay’ for the city and ‘A’, ‘B’, and ‘C’ for the branch). In the same way, the numeric labels in the category of customers and gender are replaced with ‘Member’, ‘Normal’, ‘Male’, ‘Female’. All these replacements make the visualizations easier to understand.

The findings indicate the existence of different customers who have different pattern of purchasing products. For instance, there are tendencies per some clusters to use certain product lines or payment methods. These patterns are well captured by the visualizations thus alternately shedding light on specific targeted marketing strategies. The following is an evaluation of the implementation process of the methodology based on the realisation of the above-stated objectives: The research methodology is useful in segmenting customers with insights gained. The relative ease of use and general approach to data preprocessing followed by clustering and visualization is easy to follow and could be systematically repeated. Some issues were related to working with categorical data and calculating the encoding, as well as achieving logical labels swapping. The findings are in concordance with the objectives of the project, exhibiting precise customer segments that could be used in the company’s marketing strategies.

## Reflection

In sum, the present work demonstrates the applicability of K-means clustering as a descriptor of customers. Defining different customers, on the basis of their characteristics, can be useful to address specific issues in each of them obtaining better results both in satisfying the customers’ needs and in achieving improved business results. In this way penetration of the approach into other industries along with different datasets is possible allowing to investigate the efficiency of data-driven decisions in various domains.

