**PHASE 3**

To start the market basket insights project by loading and preprocessing the transaction data, you can follow these steps:

* Load the transaction dataset. This can be done using a variety of programming languages and libraries. For example, in Python, you can use the pandas library to read a CSV file containing the transaction data.
* Load the transaction dataset

Import pandas as pd

**# Load the transaction dataset**

**Df = pd.read\_csv(‘transaction\_data.csv’)**

* Preprocess the data. This may involve the following steps:

Remove any unwanted columns or rows. For example, you may want to remove any rows that contain missing values.

**# Remove any duplicate transactions**

**Df = df.drop\_duplicates()**

* Convert categorical data to numerical data. This is necessary for many association analysis algorithms. For example, you could use one-hot encoding to convert each categorical variable into a binary variable.

**From sklearn.preprocessing import OneHotEncoder**

**# One-hot encode the categorical variables**

**Encoder = OneHotEncoder(handle\_unknown=’ignore’)**

**Df\_encoded = pd.DataFrame(encoder.fit\_transform(df[[‘item1’, ‘item2’]]))**

* Handle duplicate transactions. You may want to remove duplicate transactions, or alternatively, you could count each transaction multiple times depending on how often it occurs in the dataset.

**CODE & OUTPUT:**

Import pandas as pd

# Load the transaction dataset

Transaction\_data = pd.read\_xlsx(‘Assignment-1\_Data.xlsx’)



# Preprocess the data

# Remove any unwanted columns or rows

Transaction\_data = Assignment-1\_Data.dropna()

transaxtionData ddply(itemslist, c(“BillNo”, “Date”), <- function(df1)paste(df1$Itemname, collapse = “,”))

# to store this transaction data into .csv

transaxtionData$BillNo <-NULL

transaxtionData$Date <- NULL

#will gave the name to column “item”

Colnames(transaxtionData) <-C(“items”)

Write.csv(transaxtionData, “assigmentl\_itemslist.csv”, quote= FALSE, row.names = FALSE)

