Business Data Management
(Helping Offline Retail Stores through E-Commerce)

Final Submission Report

Submitted by

Varun Ram S
21F1001490

January – 2023

(September 2022 Term – Cycle 4)

DECLARATION

I hereby declare that the Final Submission for "*Business Data Management"* Project titled *"Helping Offline Retail Stores through E-Commerce*" is my own work and the idea proposed is completely a result of my thought process and ideas.

I further declare that to the best of my knowledge this Final Submission report does not contain any part of work that has been submitted for the award of any degree either in this university or in other university / Deemed University without proper citation.

Varun Ram S

21F1001490

Table of Contents

Contents	Page No.

Title of Project: Helping Offline Retail Stores through E-Commerce

Executive Summary

Consumer electronics companies and other industries with offline storefronts or outlets have been severely impacted by the explosive growth of e-commerce websites and delivery services. With the push of a button on an app and a quick transaction through various digital payment platforms that are crucial to the development of Digital India, many individuals were able to stay at home during the pandemic that hit back in 2019. Services and goods were then delivered to their doorsteps. The idea of a digital India looks fantastic, but a substantial portion of the population still prefers to physically visit a store in their neighbourhood and select a product of their choice. The same sizeable population is, however, declining as a result of the increased competition provided by online goliaths. As a result, several stores have been forced to close, hurting all of the staff members who worked there. This project offers strategies for combating competition from these internet behemoths while also outlining precautions that can be taken to prevent losses through various ad hoc techniques. This report also contains the Data Collected, the detailed analysis and process used for analysing the data and giving a number of results and recommendations that the store is advised to incorporate in order to increase optimize their business outcomes and also have a relevant presence online. Different plots that have helped in the analysis are also highlighted in this report.

1. DETAILED EXPLANATION OF ANALYSIS PROCESS/METHOD

Before proceeding to the phase 2 analysis of the data, here is the dataset that was collected ahead of the mid term phase of the project.

Product/Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total Sales
Television-all	189	107	85	248	230	225	286	305	118	139	147	205	2284
Refrigerator	201	157	196	303	308	115	157	119	145	108	221	157	2187

Table 1: Sales Data of Television and Refrigerator for 2022

Model by inches	Yearly Sales	Inventory	Average Price/unit (in INR)
32 Inches	623	1300	15000
43 inches	821	1800	25000
47 inches	551	1100	38000
50 inches	184	500	45000
>50 inches	105	250	70000
Total Sales	2284	4950	

Table 2: Yearly sales and Inventory stock of different models of Televisions with their average price

Phase 1 Analysis Summary:

Television set analysis:

- The total sales for televisions in the given year were 2,284 units.
- The mean monthly sales for televisions were approximately 190 units, with a standard deviation of 70 units.
- The highest sales month for televisions was August, with 305 units sold, and the lowest sales month was September, with 118 units sold.
- The range of monthly sales for televisions was from 85 units (in March) to 305 units (in August).
- The median monthly sales for televisions were 225 units.

Refrigerator Set analysis:

- The total sales for refrigerators in the given year were 2,187 units.
- The mean monthly sales for refrigerators were approximately 182 units, with a standard deviation of 68 units.

- The highest sales month for refrigerators was April, with 303 units sold, and the lowest sales month was October, with 108 units sold.
- The range of monthly sales for refrigerators was from 115 units (in June) to 308 units (in May).
- The median monthly sales for refrigerators were 157 units.

Explanation for the above analysis:

The above analysis was performed using the default in built functions that are already present in Microsoft Excel. The computing of various Statistical measures like Mean and Median monthly sales helps in conducting analysis at an advanced stage and also gives a rough idea about the patterns and trends involved in the sales.

Phase 2 analysis Summary:

1) The above table (table-1) is converted into a .csv file and is named as data.csv. There will be a column called "month" and two features called Television set and Refrigerator which are basically other two columns. The rows would obviously be different months.

2) Google Collaboratory or any other equivalent Python IDE is opened and certain statistical libraries are imported just like the following.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
import seaborn as sns
```

3) The data set would now be read using the pd.read_csv function. We shall be displaying only the first few rows of the data set which is why the data.head() function is used here.

```
[ ] data = pd.read_csv("D:\DriveE\Code_files\python\\varun\data.csv")
    data.head()
```

	month	Television-all	Rerigerator
0	January	189	201
1	February	107	157
2	March	85	196
3	April	248	303
4	May	230	308

4) A new list containing the months of the year would be created with the names of all the months and then it would be mapped to corresponding number (like 1 for January, 6 for June etc.) using the lambda function.

```
[ ] month_of_year = ['January',
                     'February',
                     'March',
                     'April',
                     'May',
                     'June',
                     'July',
                     'August',
                     'September',
                     'October',
                     'November',
                     'December']
```

```
[ ] data['month'] = data['month'].apply(lambda x: month_of_year.index(x)+1)
    data
```

On printing, the output should appear something like this:

	month	Television-all	Rerigerator
0	1	189	201
1	2	107	157
2	3	85	196
3	4	248	303
4	5	230	308
5	6	225	115
6	7	286	157
7	8	305	119
8	9	118	145
9	10	139	108
10	11	147	221
11	12	205	157

5) Three more columns are added to the data frame which contains the sales of the previous three months also. They are named as prev1, prev2 and prev3.

```
[ ] ''' I want to use previous data of three months as well as current month for input'''

    ' I want to use previous data of three months as well as current month for input'

[ ] data_tv = data.drop(['Rerigerator'], axis=1)

[ ] temp_list = data_tv['Television-all']

[ ] data_tv['prev1'] = data_tv['month'].apply(lambda x: np.nan if x-1<3 else temp_list[x-2])
    data_tv['prev2'] = data_tv['month'].apply(lambda x: np.nan if x-1<3 else temp_list[x-3])
    data_tv['prev3'] = data_tv['month'].apply(lambda x: np.nan if x-1<3 else temp_list[x-4])
```

The mapping using lambda function shall be used again.

6) This is the crucial part of the analysis. The data shall be split into test and train based upon the three new columns that have been created. This is crucial for the Decision Tree Classifier.

```
xtrain = data_tv[['month','prev1', 'prev2', 'prev3']]
ytrain = data_tv['Television-all']

xtrain, ytrain
```

```
(     month  prev1  prev2  prev3
 3        4   85.0  107.0  189.0
 4        5  248.0   85.0  107.0
 5        6  230.0  248.0   85.0
 6        7  225.0  230.0  248.0
 7        8  286.0  225.0  230.0
 8        9  305.0  286.0  225.0
 9       10  118.0  305.0  286.0
 10      11  139.0  118.0  305.0
 11      12  147.0  139.0  118.0,
 3      248
 4      230
 5      225
 6      286
 7      305
 8      118
 9      139
 10     147
 11     205
 Name: Television-all, dtype: int64)
```

7) For the extremely crucial part of the analysis, we shall import the relevant decision tree's library again and then fit the model for the training and the testing data using the fit() function.

```
[ ]  from sklearn import tree
```

```
[ ]  model = tree.DecisionTreeClassifier()
```

```
[ ]  model.fit(xtrain,ytrain)

     DecisionTreeClassifier()
```

8) We can now compute the RMSE (Root Mean Square Error) value and the MSE (Mean Square error) value by creating relevant functions for the same.

```
[ ]  def mse(ytest, ypred):
         return np.mean((ytest-ypred)**2)
```

```
[ ]  def rmse(ytest, ypred):
         return np.sqrt(mse(ytest,ypred))
```

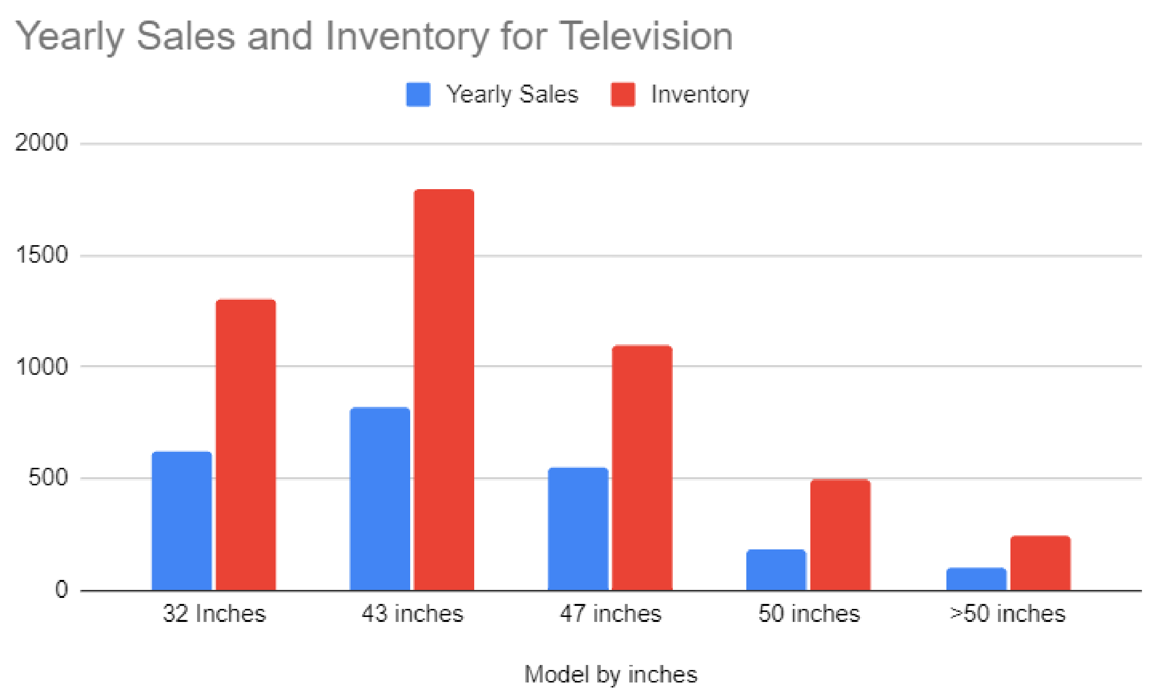
```
[ ]  rmse(ytest, model.predict(xtest))

     62.245481763739285
```

The further results and key points of the Phase 2 analysis are emphasized in the Results and Findings and the Interpretation of Results Section.
Link to Python notebook:
https://drive.google.com/file/d/1Lqn8DzhMo_3p7ZXibeSbuqzFNQ0kwtHx/view?usp=sharing

2. RESULTS AND FINDINGS

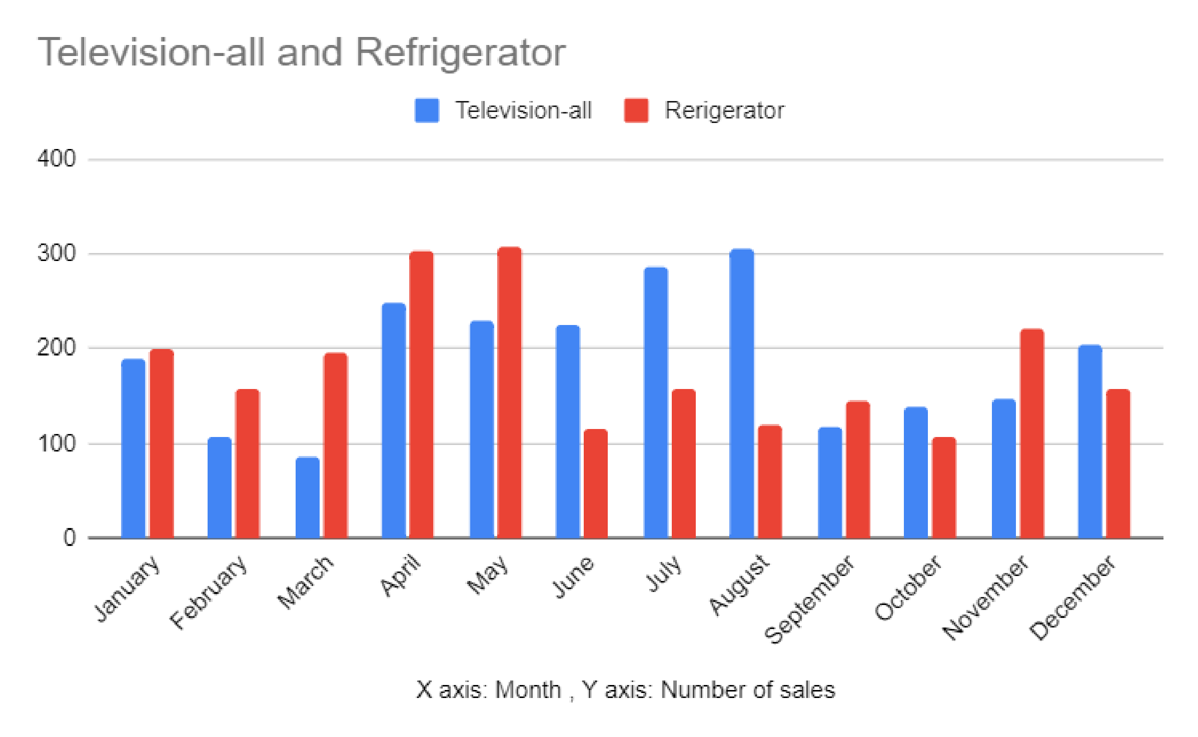
The phase 1 of the Analysis in this Business Data Management Case study project gave the following results:



This bar chart clearly explains the Inventory vs Sales of different models of television. Clearly close to only 50% of the inventory gets sold out.
One of the suggestions here can be: Reduce the Television Set inventory to 1.5 times the selling stock instead of always keeping it at twice the selling stock.

Some of the other key points to note and interpret can be:

1) 43 inches model is the best-selling model
2) The price of models of 50 inches and above should be kept at least 10% lower for better sales, else inventory will have to be reduced significantly.
3) The price of 32 inch model can be raised by about 5% for a better business outcome and then discounts can be offered at a later stage.
4) The models greater than 50 inches should also be given extended warranty to increase the number of sales as these seem to be making good revenue (again this won't be affected by the recommendation in point 2).
5) Around 25% of the inventory can be put up on the online website for sales.

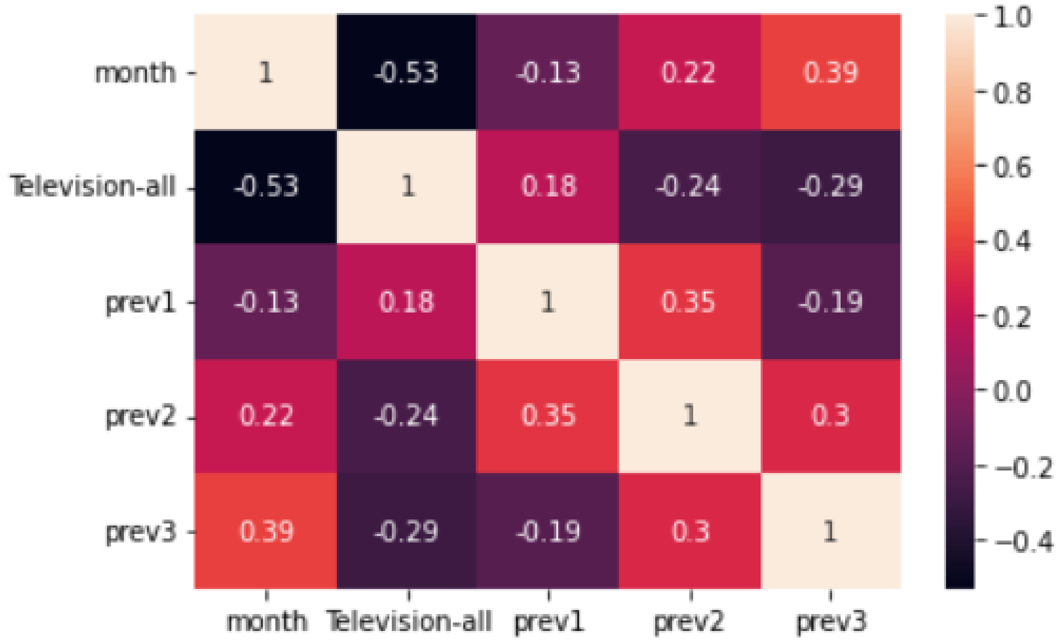


This bar chart clearly shows the comparison in number of sales of television and refrigerator. The aim here is to not compare them, as they are two different products and it might become a case of comparing apples to oranges. The aim here is to instead give highlights on the best-selling months and which months the inventory can be optimized.

1) The months of April, May and August seem to be performing well. Hence the inventory in these months don't have to be changed much and can be kept almost the same.
2) It is recommended to reasonable discounts in these months and grab customer satisfaction so that they can come back in future.
3) January and December, even though they are festival seasons have slightly lesser sales, so in these months the advertising, marketing and discounts should be accelerated.
4) A good portion of the inventory should be allotted to the online websites for the months of February, March, October and November.

The phase 2 of the Analysis in this Business Data Management Case study project gave the following results:

<AxesSubplot:>



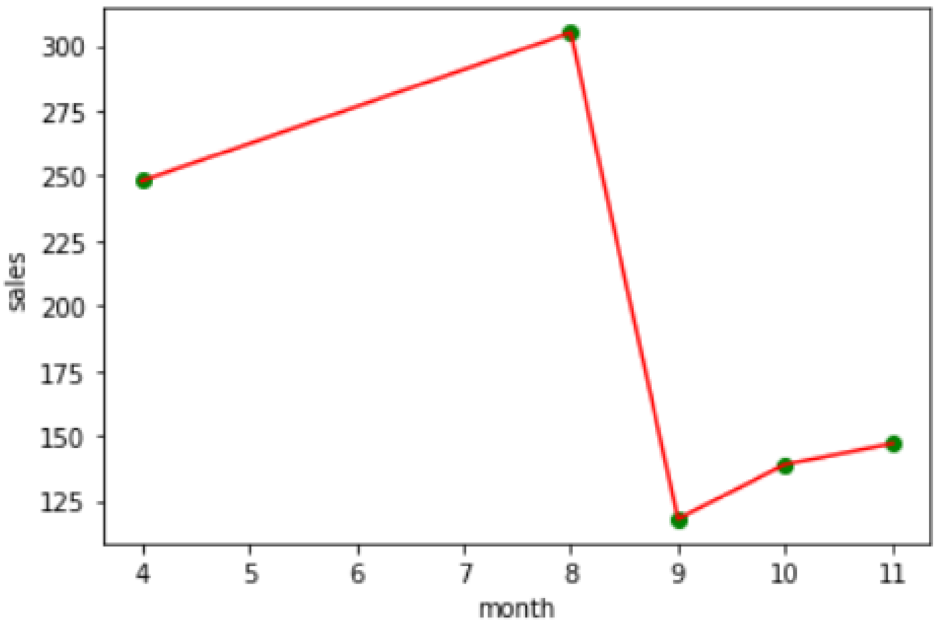
The intensity or magnitude of a variable across various categories or levels is visualised using colour in a heat map, which is a graphical representation of data.

Here are few potential key factors that the above heat map indicates:

1) Identifying patterns of the sales of television for the current month and previous 3 months

2) Highlighting the outliers in the sales of televisions.

3) Correlations between the variables (the monthly sales in our case)



\

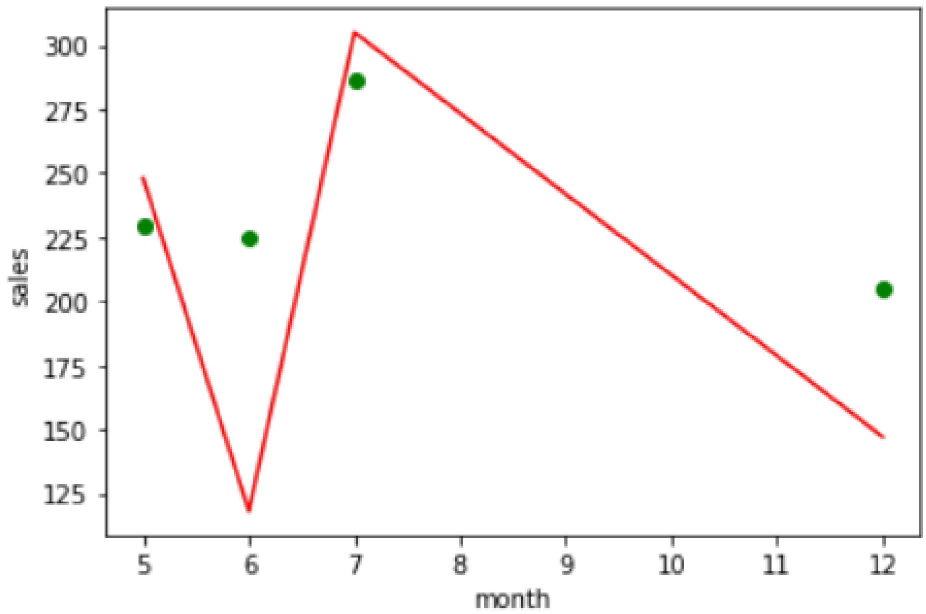
Even though the Data belongs strictly to 2022, this line chart that was generated through the Decision Tree model can help to predict the number of sales per month of the television and help the store plan their inventory accordingly.

Here are few Interpretations/Suggestions:

1) There is significant downfall from the 8th to the 9th month (i.e., from August to September). This clearly Indicated that the online presence and optimizing inventory is important for the Month of September.

2) The store can introduce a random scheme called "End of the year scheme" for better sales from November to January.

Note: The above graph is restricted to the Television sales only. Refrigerator sales, however follows a similar pattern.



The above scatter plot also gives a similar analysis and interpretation as the previous one. The red line gives the prediction whereas the actual values are the green dots. Even though there might be a huge gap for December, it should help the store plan the inventory for the other months.

3. INTERPRETATION OF RESULTS AND RECOMMENDATIONS

Interpretations of Results:

```
[ ] def mse(ytest, ypred):
        return np.mean((ytest-ypred)**2)
```

```
[ ] def rmse(ytest, ypred):
        return np.sqrt(mse(ytest,ypred))
```

```
rmse(ytest, model.predict(xtest))

62.245481763739285
```

We have obtained a Root mean square value of close to 62.24% for the Decision Tree model that was built. A lower RMSE score, which means that the predicted values are more closely aligned with the actual values, often signifies that the model is working better.

Recommendations for optimizing the Inventory

1. <u>Implement Inventory Management Software</u>: Implementing inventory management software can help offline stores track their inventory levels, monitor sales trends, and manage stock more efficiently. By having a real-time view of their inventory, they can make data-driven decisions about when to order new stock, which products are selling well, and which ones are not.

2. <u>Analyse the sales data frequently</u>: Finding out which products are popular and which are not can be done by analysing sales data. Understanding sales trends allows retailers to decide which products to order more of and which to mark down or stop carrying.

3. <u>Conduct regular audits</u>: Stores can find problems and stop inventory shrinkage by conducting routine inventory audits. They can locate goods that are defective, out of stock, or overstocked by conducting routine audits.

4. Use cross selling and up selling techniques: Techniques for cross-selling and up-selling can help retailers move sluggish inventory and boost sales. Stores can raise the average order value and move products that may have been sitting on the shelf by proposing improved versions of a product or suggesting complimentary products.

5. Partner with other stores or Liquidators: Partnerships with other retailers or liquidators might aid stores in fast moving overstock. They can free up space in their store and make money from things that might not have been doing well in their own store by selling their inventory to other retailers or liquidators. Partnership even with other branches of their own company is one recommendation

6. Offer Special Deals and Discounts: Providing exclusive offers and discounts can aid retailers in moving merchandise rapidly. Stores can increase interest in their products and encourage customers to make purchases by providing discounts on slow-moving inventory.

Recommendations for increasing presence on online sites/own e-commerce website:

1. Active website: Any offline business that wants to enter the online market must first establish an online footprint. They can accomplish this by creating a website, adding products to online stores like Amazon or Flipkart, or creating a social media profile.

2. Leverage social media: They can use social media sites like Facebook, Instagram, and Pinterest to advertise their goods and connect with prospective customers. To interact with their target audience, promote their goods, and display their products, they can develop a social media marketing strategy.

3. Offer In-Store Pickup: The ability to offer in-store pickup is one of the greatest advantages that physical stores have over their online competitors. Customers can purchase products online and pick them up in-store, saving time and money on shipping, thanks to this service.

4. Provide Personalized Service: For offline retailers wishing to enter the e-commerce market, personalised service can be a crucial differentiator. They can provide services that improve the online shopping experience, such as personalised recommendations and online chat assistance.

5. Run online Promotions: On their e-commerce platforms, offline retailers can run promotions to entice new consumers and boost sales. These promos may offer discounts, free shipping, or unique offers for brand-new clients.

6. Utilize Email Marketing: For offline retailers, email marketing can be a potent instrument for maintaining contact with customers and encouraging return visits. They can communicate with their audience through newsletters, marketing emails, and other means to keep them interested in their business.

4. CONCLUSION

The Business Data Management Analysis Case Study for “Helping Offline Stores through E-commerce” has been successfully performed and the appropriate recommendations and suggestions for the same have also been mentioned in the report.