

# Brand Analysis from the perspective of Retail User

Submitted By

**Vishal Sheth**

**18MCEC11**



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
INSTITUTE OF TECHNOLOGY  
NIRMA UNIVERSITY

AHMEDABAD-382481

MAY 2020

---

# Brand Analysis from the perspective of Retail User

---

## Major Project

Submitted in fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering (CSE)

Submitted By

**Vishal sheth**

(18MCEC11)

Guided By

**Dr. Jaiprakash Verma**



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
INSTITUTE OF TECHNOLOGY  
NIRMA UNIVERSITY  
AHMEDABAD-382481

MAY 2020

# Certificate

This is to certify that the major project entitled ”**Brand Analysis from the perspective of Retail User** ” submitted by **Vishal Sheth (Roll No: 18MCEC11 )**, towards the fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project part-II, to the best of my knowledge, haven’t been submitted to any other university or institution for award of any degree or diploma.

Dr. Jaiprakash Verma  
Guide & Associate Professor,  
CSE Department,  
Institute of Technology,  
Nirma University, Ahmedabad.

Dr. Priyanka Sharma  
Professor,  
Coordinator M.Tech - CSE  
Institute of Technology,  
Nirma University, Ahmedabad

Dr. Madhuri Bhavsar  
Professor and Head,  
CSE Department,  
Institute of Technology,  
Nirma University, Ahmedabad.

Dr. R. N. Patel  
I/C Director,  
Institute of Technology,  
Nirma University, Ahmedabad,

## Statement of Originality

---

I, **Vishal Sheth, 18MCEC11.**, give undertaking that the Major Project entitled ”**Brand Analysis from the perspective of Retail User**” submitted by me, towards the fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

---

Signature of Student

Date:

Place:

Endorsed by  
Dr. Jaiprakash Verma  
(Signature of Guide)

## Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Dr. Jaiprakash Verma**, Professor, Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Madhuri Bhavsar**, Hon'ble Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for her kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr. R.N. Patel**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation she has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Science and Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Vishal Sheth  
18MCEC11

# Abstract

In today's world *"Customer anticipates that retailers should know it about the things they buy, regardless of whether this is data on availability, ingredients or the manufacturing procedure."* Being retailer endeavors to satisfy customer's needs due to an escalating the complexity of vaster spans of product, trial avors, and convenient ready-to-go products. Also, the product offered at a choice of price points and with an appreciation of more complex dietary requirements. Brand Analysis from the perspective of Retail User aims to protect Brand integrity and enable brand owners to collaborate with their supply chain in the sourcing, development, marketing, and quality control of their products. To enable the stack holder to know source product compositions, labeling, and dietary advice, track and manage. It provides end-to-end life-cycle management by enabling the capture of detailed product information during the sourcing and selection of suppliers and the manufacturing process. Consolidating all necessary product detail promotes transparency throughout the supply chain. Having this product, users can access and the interrogate data in seconds, allowing brand owners to immediately reply to the product or industry conflicts and secure continued user confidence in their trademark

To increase their market it becomes a compulsion for retailers of any order to support their loyal society of customers and expands their customer relevance. Also to analyze and predict the repeat customer purchase. Input to our problem is the collection of obfuscate data taken from open source databases. Also, this application provides the retailer and supplier an efficient way of process to handle a complex situation. It provides a maximum return on investment efficiently.

# Abbreviations

<b>SVM</b>	Support Vector Machine
<b>RE</b>	Regular Expression
<b>LSTM</b>	Long Short Term Memory
<b>SCN</b>	Supply Chain Network
<b>SIM</b>	Sales Inventory Management
<b>SRS</b>	Software requirement system
<b>CRM</b>	Customer Relationship Management
<b>SOA</b>	Service Oriented Architectures
<b>API</b>	Application User Interface
<b>UI</b>	User Interface
<b>GUI</b>	Graphical User Interface

---

—

# Contents

<b>Certificate</b>	<b>iii</b>
<b>Statement of Originality</b>	<b>iv</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Abstract</b>	<b>vi</b>
<b>Abbreviations</b>	<b>vii</b>
<b>List of Figures</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction To The Definition . . . . .	1
1.2 Challenges And Benefits To The Retail Industry . . . . .	1
1.2.1 Problem Statement . . . . .	1
1.2.2 Understanding The Problem . . . . .	2
1.2.3 Market Need To The Industry . . . . .	2
1.3 Solution To The Problem . . . . .	3
1.4 Business Values . . . . .	4
1.4.1 Reliability . . . . .	4
1.4.2 Efficiency . . . . .	5
1.4.3 Accessibility . . . . .	5
1.4.4 Protection . . . . .	5
1.5 Environment . . . . .	5
1.5.1 System Environment . . . . .	5
1.5.2 Development Environment . . . . .	6
1.5.3 Deployment Environment . . . . .	6
1.6 Key Modules . . . . .	7
1.7 Client Side Requirements . . . . .	7
<b>2 Literature Survey</b>	<b>8</b>
2.1 Market Requirement Survey . . . . .	8
2.2 Product Background . . . . .	8
2.3 Software Requirement Specification . . . . .	9
2.4 The Future Of Retailing: . . . . .	10
2.5 Value-Based Consumer Clustering Of Generous Local Data-sets: . . . . .	11
2.5.1 Cluster Visualization . . . . .	12
2.6 Purpose of K-Means Algorithm for Dynamic Consumer Segmentation: . . . . .	13



2.7	An effective approach to customer segmentation. . . . .	14
<b>3</b>	<b>Architecture and Workflow</b>	<b>16</b>
3.1	Application Workflow . . . . .	17
3.2	Application Programming Interface . . . . .	17
3.2.1	SOAP and RESTful APIs . . . . .	18
3.2.2	Request and Response XML . . . . .	18
3.2.3	Observations and Findings . . . . .	18
3.2.4	Security Related Observations . . . . .	18
3.3	Coding Mechanism . . . . .	19
3.3.1	Entry Point Method . . . . .	19
3.3.2	Event Handling Mechanism . . . . .	19
3.3.3	Listening events . . . . .	19
3.4	Debugging . . . . .	19
3.4.1	Compilation . . . . .	20
3.5	Limitation . . . . .	20
<b>4</b>	<b>API Overview and Architecture</b>	<b>21</b>
4.1	Accessing the APIs . . . . .	21
4.1.1	Availability . . . . .	21
4.1.2	Deployment Model . . . . .	21
4.1.3	Authentication . . . . .	21
4.1.4	Password Management . . . . .	22
4.1.5	Rich Text Data . . . . .	22
4.1.6	Excluded Fields . . . . .	22
4.1.7	Secured Connections . . . . .	22
4.1.8	URL Encoding . . . . .	23
4.2	Key Features of the API Process . . . . .	23
<b>5</b>	<b>Customer Segmentation - POC</b>	<b>26</b>
5.1	Pre-requisites . . . . .	26
5.1.1	Data set Requirement . . . . .	26
5.1.2	Data set Description . . . . .	26
5.1.3	The features extracted are: . . . . .	26
5.1.4	Pre-processing steps . . . . .	27
5.2	Approach To Implement Customer Segment. . . . .	27
5.3	Analysis from Implementation . . . . .	28
<b>6</b>	<b>Subject System</b>	<b>30</b>
6.1	Integration Application Testing . . . . .	30
<b>7</b>	<b>Functional Enhancements</b>	<b>31</b>
7.1	Productisation . . . . .	31
7.2	Glossary Import/Export . . . . .	32
7.3	Supplier And Product Upload . . . . .	32
7.4	System Text . . . . .	33
7.5	Scorecard . . . . .	33
7.6	Product Specification . . . . .	33

<b>8 Observations And Findings</b>	<b>34</b>
<b>9 Conclusion and Future Scope</b>	<b>36</b>
9.1 Conclusion . . . . .	36
9.2 Future Work . . . . .	36
<b>Bibliography</b>	<b>38</b>

# List of Figures

1.1	Phase Workflow . . . . .	4
1.2	Workflow for Repeat Purchase User . . . . .	4
2.1	Snippet of future in retail. . . . .	10
2.2	Example . . . . .	12
2.3	Visualization Result . . . . .	12
2.4	Graph of customer segmented based on characteristics . . . . .	13
3.1	Snippet of Implementation Architecture/Work-flow Diagram . . . . .	16
5.1	Chart Shows customer segmentation. . . . .	28
5.2	The Analysis of customer segmentation. . . . .	29

# Chapter 1

## Introduction

### 1.1 Introduction To The Definition

Brand Analysis from the perspective of Retail User System consists of Supply Chain Planning and Monitoring Application and supporting other various application that can be changed easily and dynamically (depends on customer to customer). The product definition mainly focuses on Supply Chaining Monitoring and Planning Application rather than the system. Basically, 'Brand Analysis from the perspective of Retail User System' consists of two parts i.e. First phase is growth of current application which includes challenges which required research to overcome this challenges, steps to overcome challenges are Plan, Define, Source, Develop. Second phase consist of adding or redesigning the current application by adding required module(s) required by the consumer. So this project has to be carried out into two phases, Growth and Protection phase, both will be worked parallel and dependent on one other as follows:

### 1.2 Challenges And Benefits To The Retail Industry

#### 1.2.1 Problem Statement

Any product to come into market and become successful need a idea behind and a problem that it is going to solve. For this product main Challenge that came across was "consumer demand and trust".<sup>[1]</sup>

Consumers continue to drive complexity by increased demand for greater ranges, experimental flavors, and convenient ready to-go products. Wide product ranges are often offered at a choice of price points and with an appreciation of more complex requirements.<sup>[1]</sup>

In the modern retail environment, to keep confidence in their brands, retailers must prove due to persistence in their selection, recovery, and validation of supplier and commodity data. There is an understandable necessity for retailers to counter immediately to any concerns being challenged by consumers.[1]

### **1.2.2 Understanding The Problem**

Modern records reveal that consumers now notice more than ever before. A third of users are allergic to something and one in six has a form of food-related disease each year. This, along with a thirst for healthier living, healthier alternatives, and the media's support to know more, has strongly authenticated that if it's your label, it's your product and your problem. The retail production, and especially the food sector, has become a deeply tense market with consumers necessitating full clarity, correct labeling, quality control and rapid acknowledgment to events.

In the modern right environment, to conserve trust in their brands, retailers must prove due to persistence in their collection, acquisition, and validation of supplier and product data. There is an open obligation for retailers to react quickly to any problems being fronted by users.

### **1.2.3 Market Need To The Industry**

Retailers seek to meet local consumer needs by giving transparency of every product by giving details of ingredients, manufacturing, scheduling, supplier and unique promotions. From this they also maximizing profit and ROI. From every product requires appropriate planning and monitoring of audit and visit of sites for each category of product. So evaluation of suppliers can be done in periodic time. For this scorecard is maintained for the same. It helps to maintain and proper form work can be carried out by suppliers and sites. However, the lack of visibility into each individual site can be fixed. As each individual sites strives to achieve perfection and store the details of manufacturing process, it can be difficult to ensure that what it executes meets consumer needs'. Overall objectives is to maintain details of every process in mean time period. Furthermore, because retailer and supplier rarely have visibility into and an accurate understanding of the actual modifications performed at each individual product, Maintaining communication

becomes cumbersome

In addition, retailers often manage details of product, leading to significant disadvantages throughout the retail organization. For example, retailers sell product might be poisonous so this product give news to all retailer that this product is not to sell. Moreover, season restrictions such as equipment range must be taught and learned during the supply chain. If limitations are not carried downstream, there is a risk of incorrect inventory preparation, unequal service levels, and out-of-stocks on the racks.

### **1.3 Solution To The Problem**

The product aims at protecting Brand Integrity and it Allows brand owners to collaborate with their supply chain in the sourcing, development, marketing, and quality control of their outcomes using a scalable line of integrated modules. Specically designed to enable grocery retailers, restaurants, food service and manufacturers. It enables detailed capture of conformance during the sourcing and selection of suppliers and their manufacturing plants. It offers a range of characteristic reviews and traceability resolutions to review and test all products during production, delivery, and rack living. The solution quickly allows brand owners to reply to product or industry incidents ensuring continued consumer trust in their brand. Built on industry best practice advisory and supplier engagement programs to ensure approval and proceeded t for purpose. The third-party applications which are integrated with this are mainly SCN i.e. Supply Chain Network. External System is efficiently used for all the demanded data transfer that happens via API. Usage of Restful Web service stronger than the SOAP is one of the key bonuses in this.

For this project I am contributing, to solve major functionality bugs that is raised by internal QA cycle and also from the customer. Also working on Supplier Scorecard and product specication module. Externally also trying to analyze and predict the repeat customer purchase. For this I collect obfuscate data from various open source. It contains 10 million users data from all across the globe. This gives us surety for highly diverse of user data will be there and also the purchased product data. This is unprocessed raw data of product data,sales data, user data. To increase the accuracy of algorithm, the

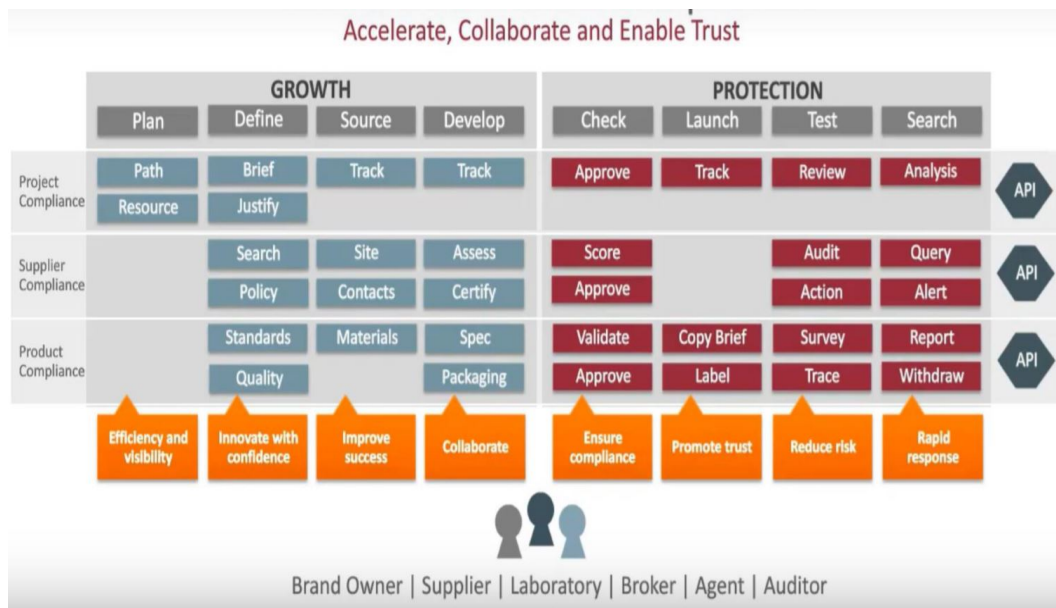


Figure 1.1: Phase Workflow

data pre-processing is done. Data contains various column characteristics. This small data contains spelling mistakes, local slang words. Also they do contain number, range of data, References(Handles) and punctuation's. To train a model for repeat purchase various machine learning approaches are suggested to use. we will use K-means, LSTM, Random Forest. For Validating the Quality of our model we do testing with our test data and calculate accuracy from it.



Figure 1.2: Workflow for Repeat Purchase User

## 1.4 Business Values

### 1.4.1 Reliability

Entering required information at the source of the data and flowing this through the supply chain maintains compliance, improves operational transparency and drives confidence and reassurance that a brand's labelling claims are accurate, helping retailers accomplished the request on their Brand Promise.

## **1.4.2 Efficiency**

As information on the sourcing or manufacturing process is created in the system by each supplier, Cloud Service validates this data against the ever-growing number of local regulatory or retailer requirements and different legislation standards. This improves the efficiency of the supply chain by enhancing data quality and speed-to-market, and reduces process duplication.

## **1.4.3 Accessibility**

Data entry at the beginning provides an immediately ready single 'truth' for product and supplier information, overcoming data re-entry, and human slip further down the supply chain. Convenience in the cloud puts users in control, enabling rapid and secure, 24/7 incident management, and improves reporting capabilities.

## **1.4.4 Protection**

Developed on manufacturing the best custom announcement and supplier engagement meetings, Cloud Service safeguards adoption and ensures continued fit-for-purpose. The solution provides a range of features and workflows to validate and manage suppliers, due diligence, and incident response and reporting.

# **1.5 Environment**

## **1.5.1 System Environment**

### **1. Hardware Specifications**

- (a) RAM : 1 GB or more
- (b) Hard Disk Space : 100 MB or more

### **2. Software Specifications**

- (a) Operating System : Windows 7(64 bit),Oracle BI environment.
- (b) Language : JAVA (1.8 or higher version)



- (c) Database : Oracle DB 12c
- (d) Web Browser : Firefox, IE explorer or google chrome.
- (e) IDE : Eclipse
- (f) Code Review Board
- (g) Oracle JET
- (h) Oracle Virtual Box.

### **1.5.2 Development Environment**

The software application is developed and tested using Eclipse in different modes. JUnit, integrated with eclipse, is an open source framework designed for the purpose of writing and running tests in the Java programming language. This framework builds a relationship between development and testing process. If user wants to test any application, in standalone, he/she needs to call the respective method's main() method but there are so many issues in testing the flow using the main() method. In a web apps, to test the flow we need to deploy it in the server so if there is a change then again we need to restart the server. Oracle Vm virtual box is used to import .ova image where all the development environment was present including Mysql, eclipse Mars, console , work space etc

### **1.5.3 Deployment Environment**

It is deployed on a Web Logic Server after its completion. Oracle Weblogic is a server a software application that works on a central tier and generally used with UNIX, Linux and Windows OS, that is, between the database and browser-based thin clients(A thin the client machine is going to interact with a primary processing server, meaning there are little hardware and software installed on the user's machine. At times, thin could also be defined as simply not needing the software or OS installed on the user machine. This provides all end-users' policies to be centrally maintained and software expanded on a central server location as against installed on each system. Thin clients are best-suited to environments during which a piece of equivalent information goes to be accessed by the clients, making it a far better solution for public environments.

For this reason, thin clients are often deployed in hotels and airports, were installing software to all systems wouldn't make sense. It would be a huge headache for IT to both deploy and maintain.

Oracle WebLogic Server sustains the deployment of the numerous varieties of shared applications and it is a correct foundation for fabricating applicability based on SOA. The beliefs of service-orientation are independent of any huckster, commodity or technology

## 1.6 Key Modules

1. Supplier - Supplier and site contact, Audits visits , Scorecard management for Suppliers.
2. Product - Product development and validation, Specifications , and quality surveillance management .
3. Project - Management of new and existing product developments , Activities corresponding to one product.
4. Library - Management of policy and guideline documentation, readership, and acceptance.
5. Reports - Cross module data query and reporting.

## 1.7 Client Side Requirements

The Following technology is supported.

1. The mentioned web browsers are supported on Microsoft Windows 7
  - o Microsoft Internet Explorer 11
  - o Mozilla Firefox 45.1.1 ESR –
  - o Google Chrome 50+
2. The following web browsers are supported on Oracle Linux 7:
  - o Mozilla Firefox 45.1.1 ESR –
  - o Google Chrome 50+ .

# Chapter 2

## Literature Survey

### 2.1 Market Requirement Survey

The Indian retail part is experiencing a change and this developing business sector is seeing a noteworthy change in its development and speculation labelling and manufacturing's design. Both existing and new players are trying different things with new retail arranges. At present two well known sectors - hypermarkets and grocery stores are becoming quick. Purchaser elements in India is changing and the retailers need to observe this and define their procedures and strategies to convey an incentive to the buyer.

The researches mention retail improvements and development needed. The difficulties raise and openings accessible to the retailers to prevail in country.[2].

### 2.2 Product Background

The Brand compliance module, which was also present in the previous releases of the product mainly gave an insight into the end to end product life-cycle which could provide each and every information related to any specific product.

The previous releases included the reports that were generated in the .jasper format which was later on found out to be a cumbersome technology to be used for generating the reports. The jasper report generation required the programmers to physically write the code including thousands of lines of code and prone to errors in the programming. Hence, there was an utter need to shift the report generation onto a much easier platform which was more programmer-friendly and consumed less time for generation. Also, these

reports were generated separately directly connecting to the database with no middle ground. There was no provision for integration of the application with the data and the analytics part.

Hence, in this project a framework is developed which allows the two applications (customer relationship management and business analytics) to integrate and communicate with each other without the reports using the data from the database directly. This project, therefore, includes a UI framework for reports, integration of two applications, template generation for reports and data model construction to be used by the templates.

The integration of the application uses a java based framework. The GWT libraries are used for UI creation which is a tool that provides many inbuilt functionalities to be implemented with a little changes. These libraries call the GWT services that fetches the report from the backend and displays it. The data from the POS stores are stored on the cloud based network and the template formation in .rtf format is done using an added plugin in Microsoft word. The generated templates are uploaded on the business analysis application which is a shared instance used by all the application users. The data model in the .xdo format of the respective template is then attached and the report is generated for further analysis. The report can be generated in any format such as pdf, html, etc. The low level security is implemented using the PL/SQL functions.

## 2.3 Software Requirement Specification

- **Reliability** : Provide trust in brand through proper labelling, transparent supply chain cycle.
- **Efficiency** : 80% decrease in rework, 50% increase in growth, 30% increase speed of market.
- **Accessibility** : Single point to buy and sell make supply chain trust,fast response.
- **Protection** : Provide Consumer trust, Provide depth details of ingredients, awareness driving improved quality[3]
- **Pricing** : No monopoly, provide relief to frequent member user in pricing.

- **ROI** : Less required resource, No complex installment, Faster access

## 2.4 The Future Of Retailing:

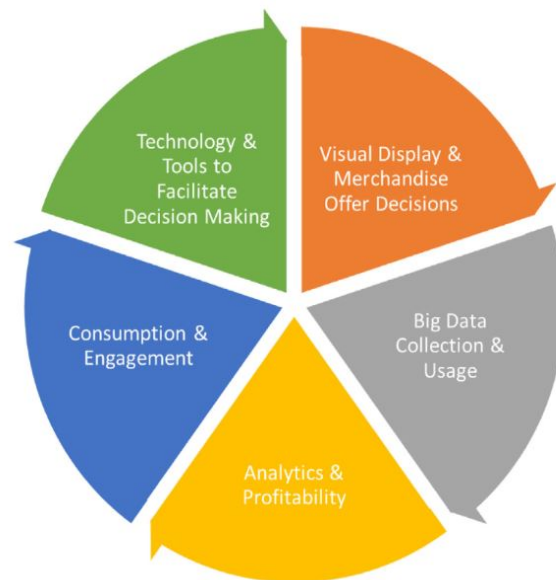


Figure 2.1: Snippet of future in retail.

### 1. Big Data Collection and Usage

Retail sector generate big data on daily basis which is collect from every source. It need to transform unstructured to structured data form. So the retailer can able to apply any algorithm and do analysis.

### 2. Analytics Profitability

When any user gets analytics in graphically form and visualize in well form manner. It helps retailer to take proper step to make growth of business.

### 3. Consumption Engagement

From historical data and proper analysis retailer can meet market demand and customer consumption effectively take place with proper engagement(CRM).

### 4. Technology Tools to Facilitate Decision Making

Every technology has both cons and pros. As per demand proper technology need to pick to get proper and effectively result. Like K-means clustering can give good result for segment the retailer on various characteristics. RFM Rec ency, Frequency, Monetary.

## 5. Display merchandise Offer Decisions

Brand perspective to user and market make a run whole monopoly. The way retailer display various brand to customer, different geographical location, customer group base on age, base on income, base on buying behaviors etc.

## 2.5 Value-Based Consumer Clustering Of Generous Local Data-sets:

This paper discussed about various approaches for creation of clusters from database. It provides comparison of various techniques and top-down clustering algorithm which generates cluster more efficiently than bottom up clustering in term of number of steps and number of cluster. Main objective of any good clustering technique is "maximum intra cluster object similarity and minimum inter cluster similarity" "in retail world retail analytic maintains customer's buying history to analyse their customer and base on that make future decision to improve business. When we try to create cluster from retail data then it face following problems[2]:

1. High Dimensionality : Number of dimension of product are more compare to samples. We need to maintain it in database. Which may cause high dimensionality.
2. Sparsity : Many features of particular product are zero because of that customer-product matrix is sparse which affect in calculation of similarity.
3. Discrete Feature Levels : Feature has ordinal values we need to convert them in to standard form and normalize them to create standard scale for comparison.
4. Feature Normalization : There are large number of features which need to be normalize to specific range and scale otherwise product with large volume would be considered as much more important then low volume product because of it's features has large variation possible in their values.
5. Significant Outerliers : Big corporate customer are considered as outlier in small retail customer cluster group removing them is not significant as they has high influence on business.

## 2.5.1 Cluster Visualization

Since human eye is capable of processing only 2d we need to convert high dimensional data which are considered in segmentation to more suitable format which can be represented in 2d plot using similarity measures, we can generate n by n similarity matrix using n by d data matrix then we can use quadratic gray-level image to represent similarity metrics having entries where black pixel represents minimal and white pixel represents maximum similarity.

Quantity	Milk (\$ 2)	Cereal (\$ 3)	Bike (\$ 200)	TV (\$ 300)	total in \$
Andrea (#1)	1	1	-	-	\$ 5
Bill (#2)	1	100	-	-	\$ 302
Chase (#3)	2	2	-	-	\$ 10
Diana (#4)	1	1	1	-	\$ 205
Eric (#5)	1	1	-	1	\$ 305
Fritz (#6)	-	-	-	1	\$ 300
Gina (#7)	2	2	2	-	\$ 410
total in \$	\$ 16	\$ 321	\$ 600	\$ 600	\$ 1537

Figure 2.2: Example

From above example we can say that diana and eric are very closely related if we consider quantity of product only. But base on total spend they are not at all related to each other because diana spend 205\$ and eric spent 305\$. If we consider similarity base on value of item then friz and eric are closely related then friz and diana even though they share less item in common.

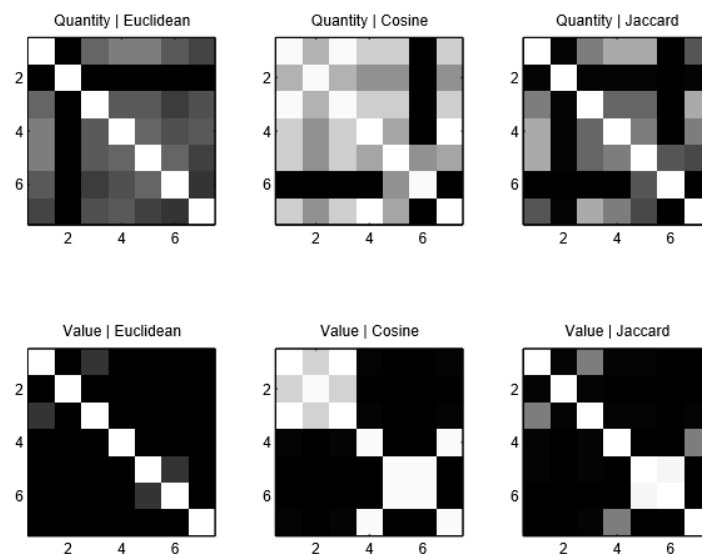


Figure 2.3: Visualization Result

1. Value Balanced : Base on value of features cluster should be balanced.
2. sample balanced : Each cluster should have approximately same number of samples(customer) in it.

It assigns weight to each customer and applies constrain on sum of weight in each cluster and it uses CLUSIONS clustering visualization to visualize clustering. In volume base clustering high volume buyer bill is successfully separated from low volume buyer.

In case of value based clustering eric and fritz are separated into single cluster to make value base balanced clusters.

In future work it is mentioned that we can use product meta data such as hierarchies and department information to improve cluster performance and online version of OPOSSUM which updates cluster data point as new data added.

## 2.6 Purpose of K-Means Algorithm for Dynamic Consumer Segmentation:

Retailer analysis and extracts useful information, that convey the message of historical market and the behavior of customer according to interest. From this business improving strategy can apply to particular customer based on RFM. Special coupons, offers, recommendation of product to particular segment of customer and make profit the approach is called Target Position. K-means clustering algorithm is good approach that segment the consumer based on similar characteristics

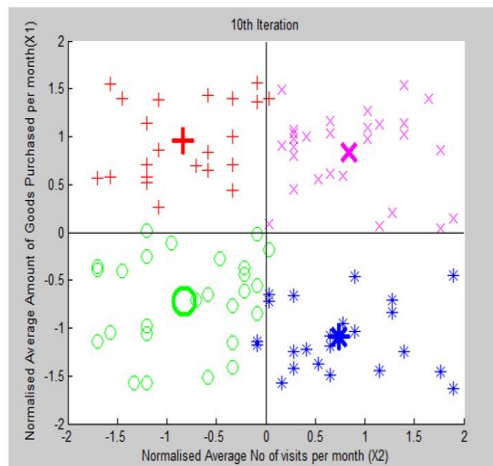


Fig. 3. The centroids converge after 100th iteration

Figure 2.4: Graph of customer segmented based on characteristics



**Details of K-means Algorithm :** K-means is classified into clustering and unsupervised learning algorithm. It works with unlabeled data it also performs more complex task compare to supervised learning. K means is hard clustering algorithm means at a time data point take part into only one group. k-means is partitioning clustering it is computationally faster when variables are huge compare to hierarchical clustering and it creates tighter clusters compare to hierarchical clustering.in hierarchical clustering clusters have tree like structure which shows relationship between parent and child cluster. While partitioning cluster performs partitions of object into two different cluster. The K-Means method are numerical, unsupervised, non-deterministic and iterative [4] top down, bottom up clustering ,partitioning etc.

**Approach:** Each customer has there different choice, requirement ,needs ,wants, demography ,geography etc. So retailer can't treat them all as equal this challenge has motivated an idea of customer segmentation. We can use different attribute to segment customer like base on lifestyle, base on purchases ,base on income etc.

## 2.7 An effective approach to customer segmentation.

In many sector customer segmentation will more useful i.e. in (CRM), Retail and Finance sector. this process is called Segmented, Targeted and Positioning

**Recency** — It describe that how recently the customer buys the product?

**Frequency** — It describe that how frequent they purchase?

**Monetary Value** — It describe how much they spend on particular product?

Summary :

In this paper they mention the best way to find initial centroid in K-means with the help of RFM. The scores of three factors Rec ency, Frequency and Monetary directly proportional to customer retention and lifetime. Different approach are reviewed that both customer segmentation and behavior of buying are important to improve the marketing performances [5]. Also discuss different approach like Three-dimensional approach is enable to improve customer retention, Customized recommendation to identify customer behaviour using weighted value of trend, Identify the pattern from online purchasing, K-means partition method, Customer churn prediction, Direct clustering approach[5]. In

these paper 3 different algorithm are analyzed namely Repetitive Median based K-Means, Fuzzy C-Means and K-Means.

#### Algorithm approach

After finding RFM (Recency, Frequency and Monetary) score, then K-means algorithm are applied to the validated score to get the sub-group based on purchase and frequency. [5]. A new repetitive median-based K-means algorithm is offered to decrease the number of repetitions than the conventional clustering algorithm. [5]

# Chapter 3

## Architecture and Workflow

This project follows a layered architecture and the user requests going from layer to layer. The architecture given below depicts the working of the projects in brief and how the requests travel from user to the database and how the response is generated and communicated back to the user with the desired results. The above architecture is used in the project and depicts the request and response from the user end. Also, it depicts the layered architecture as discussed above.

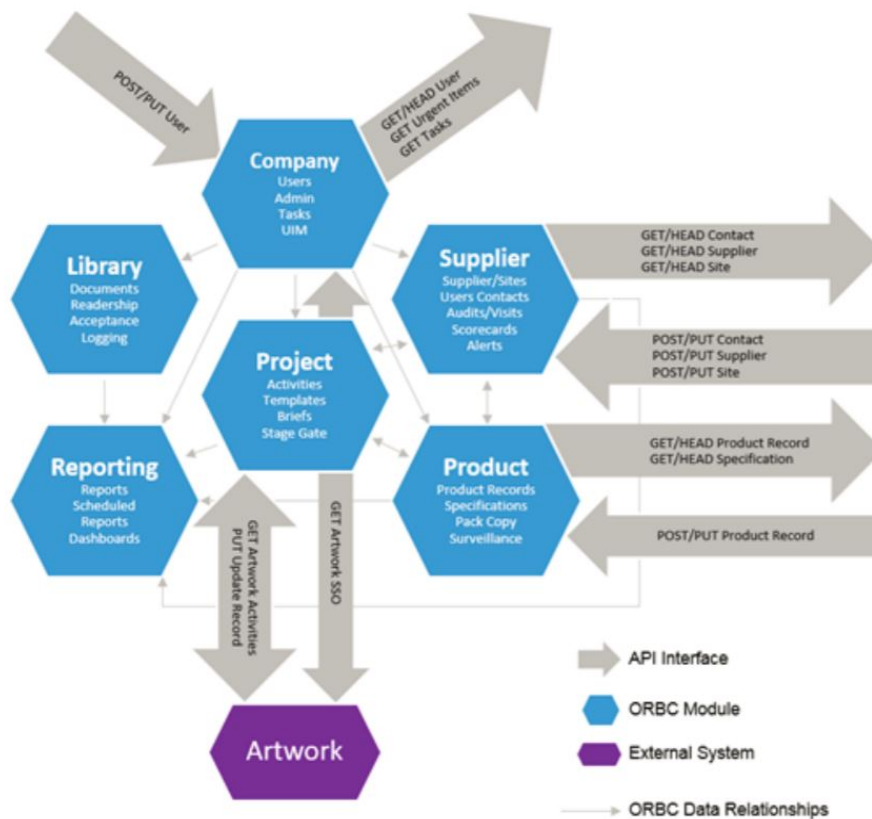


Figure 3.1: Snippet of Implementation Architecture/Work-flow Diagram

## 3.1 Application Workflow

The following points outline how the application integration and data retrieval is done in the project and shows the workflow of the application.

1. The user logs on to the Brand compliance application which is the main UI Entry point to the application.
2. On the application, under the tab Company, lies the different categories of users that can be generated by the system to look into the specific data . There can be different type of users.
  - (a) Retailer
  - (b) Supplier
  - (c) System administrator
  - (d) Power user
3. There are no. of authority permissions that can be given to only a specific user. This is the part where all roles are defined.
4. UIM layer is encountered in Supplier. It enables a user to deal with any urgent issues that require the user's attention.

## 3.2 Application Programming Interface

It includes several exposed APIs that can be called by external systems. This document details all of the disclosed APIs and the knowledge needed to obtain them. An API is a way of transferring data between computer systems. The partners of an outside system can interact with Brand Compliance by receiving the required functionality to trigger the web service obedience, and handle the repeated reports from Brand Compliance.

To access Brand Compliance using its APIs, the portal owner must award the external system admittance. This is achieved by creating an External Systems account in the Admin module and assigning the unique access credentials in the form of a login ID and password. The APIs exposed by Brand Compliance are completed in one of two formats, SOAP web services or RESTful web services. In both cases, requests and responses are normally in an XML format.

### **3.2.1 SOAP and RESTful APIs**

The Brand Compliance APIs are web services in both SOAP and RESTful architectural styles. In general, the SOAP services are developed for specific requirements or are historical services. Except there is a specific requirement for a SOAP implementation of a setting, any new services or new variants of services are implemented utilizing the RESTful architectural style.

### **3.2.2 Request and Response XML**

In general, APIs exposed by Brand Compliance as the web services accomplished XML messages for both requests and responses. Exceptions to this are RESTful services where the use of an XML request is replaced by the use of URI parameters in a GET call. The APIs do not support other message formats, such as JSON.

### **3.2.3 Observations and Findings**

The observations and findings while working on this project mainly depict why one technology is preferable than the other one or why there was a shift from one technology to another. Also, it shows a comparison between the two and helps the programmer decide which of the two is better under what scenarios.

- It provides end-to-end life cycle management by enabling the capture of detailed product information during the sourcing and selection of suppliers and the manufacturing process.
- Consolidating all necessary product detail promotes transparency throughout the supply chain.
- Having this product, users can access and interrogate data in seconds, allowing brand owners to quickly respond to product or industry incidents and maintain continued consumer trust in the brand

### **3.2.4 Security Related Observations**

There are some of the .class files whose Data can't be modified but can be seen. Function calls made from .class file can be modified but can be used in such a way to serve the whole purpose.

## **3.3 Coding Mechanism**

### **3.3.1 Entry Point Method**

The entry point method is used for the application to know where to start from. The entry point is `template-coreLauncherOnMysql()`. It contains the code that executes when you launch the application.

Typically, the types of things you do in the `template-coreLauncherOnMysql()` are:

1. All Xml request for login page
2. Set up handlers for events

### **3.3.2 Event Handling Mechanism**

GWT is an event-based technology. This indicates that the code achieves in reply to some event happening. Most often, that event is triggered by the user, who uses the mouse or keyboard to interact with the application interface. GWT provides several different event handler interfaces. To examine the `onClick` functions of Add and Remove buttons, the `ClickHandler` interface is used. To handle keyboard functions of the input box, the `KeyPressHandler` interface is used. Events in GWT use the handler model. A handler interface defines one or more ways that the widget calls to publish functions. A class wishing to sustain functions of a special type performs the associated handler interface and then relinquishes a relating to itself to the widget to contribute to a set of circumstances.

### **3.3.3 Listening events**

To subscribe to an event, a particular event handler interface is passed to the appropriate widget. An event handler interface defines one or more methods that the widget then calls to announce (publish) an event.

## **3.4 Debugging**

In this phase the Java source code is debugged before compiling it into JavaScript. Debugging is done in the development mode. In this phase the following activities are performed:

Set break points.

Step through the code line by line.

Drill down in the code.

Inspect the values of variables.

Display the stack frame for suspended threads If a bug is found, after some analysis it is fixed and then again the bug fix is tested in the development mode.

### **3.4.1 Compilation**

Using the JVM compiler the source code is compiled..

After the application is compiled, the application is run in production mode by forcing the HTML file in a new browser window. The application looks and behaves just as it did in development mode. The real difference is hidden under the covers. When you associate with the purpose now, it's executing JavaScript code in the browser, not Java bytecode in the JVM.

## **3.5 Limitation**

The retail clients will are willing to move to cloud so this application should be modified according to a cloud environment. Although the product is cloud enabled, but still there are many more features that need to be added to the product is make it ready for cloud environment. This shift will create multiple kinds of security issues, these will need to be handled.

# Chapter 4

## API Overview and Architecture

### 4.1 Accessing the APIs

#### 4.1.1 Availability

If enabled within the portal, an API is accessible while the portal is online. The support for a portal's availability is subject to the client's service level agreement. Updates to the APIs are applied as per changes and fixes to the application. Each release involves a scheduled period of downtime. Details of the change and impact are covered in the supporting Release Notes.

#### 4.1.2 Deployment Model

The APIs, made available by Brand Compliance, are all web services implemented using either a SOAP or RESTful architecture. The web services are exposed over HTTPS and require HTTP Basic authentication in order to access the service requested.

#### 4.1.3 Authentication

In order to authenticate with Brand Compliance when accessing an API deployed as a web service, a user name and password must be supplied for HTTP basic authentication. Within Brand Compliance, the External System record defines the user names and passwords that have access to the web services.

To see the list of existing External System records, log in to Brand Compliance as a Power



User, navigate to Company *i* Admin *i* Roles and Permissions, and select the External Systems option.

#### **4.1.4 Password Management**

Unlike interactive logins for users, the use of web services does not permit warnings to be included in the returned messages. Also, there is no guarantee that there is someone handling the call in order to interpret any warnings about password expiry. Therefore, Brand Compliance does not currently apply the password lifetime rules to External Systems. The passwords used to access the Brand Compliance APIs should be stored securely to prevent potential attackers from being able to masquerade as an authorized system.

#### **4.1.5 Rich Text Data**

In Brand Compliance, certain fields allow the display of HTML-formatted data. Such data is cleansed before display in Brand Compliance to ensure that any attempts at Cross-Site Scripting (XSS) attacks are mitigated, but this cleansing is not applied within the APIs. Therefore, it is advisable to cleanse any HTML-formatted data before it is displayed to users.

#### **4.1.6 Excluded Fields**

Many of the APIs published by Brand Compliance provide access to manage all of the records of a given type. With certain records, the fields accessible through the API are a subset of the fields within Brand Compliance. Any updates leave those fields unchanged. An example of this is the password-related fields on the User record, none of which are exposed in the APIs.

#### **4.1.7 Secured Connections**

As all of the APIs are exposed as web services over HTTPS, it is necessary for callers to communicate over SSL/TLS using the ciphers that the Brand Compliance servers accept. To ensure optimal security, it is advised that standard certificate checks are in place and not disabled. In the case of Apache CXF, the Client TLS Parameter `disableCNCheck`

should be left as false.

### 4.1.8 URL Encoding

Certain characters within URLs are reserved for a special meaning. In order to avoid calls that contain these characters being blocked by the API's security, the characters must be encoded and passed as plain text. The reserved characters include ";", "/", "?", ":", "@", and "=", all of which could be present in the parameters passed to Brand Compliance APIs - such as within email addresses or within the names of users, products, or suppliers. UTF-8 URL encoding should therefore be applied to all calls to the Brand Compliance APIs. The APIs will automatically decode the values on receipt. For example, an un-encoded call of: `.../services/rest/user/byKey/john.smith@example.com` should be encoded as: `.../services/rest/user/byKey/john.smith%40example.com` If a call comprises multiple values, each must be encoded individually.

## 4.2 Key Features of the API Process

This section summarizes the key aspects of the API process:

- The external system makes a call to Brand Compliance using HTTPS. The external system must be registered in Brand Compliance, and the user ID and password must match.
- Calls to RESTful services use the GET, POST, PUT, and DELETE methods to retrieve, create, update, and delete records respectively. Calls to SOAP services use a predefined set of methods to perform the operations.
- Each call is logged in the Web Service Log in Brand Compliance. The request and response XML messages are attached to the logs.
- A predefined set of parameters is available to apply filters to the call, to retrieve a specific record or set of records.

- Data is returned as XML, as defined by the associated WADL or WSDL schema. Only elements that contain data are returned; empty elements are omitted.
- For RESTful APIs, an initial GET call to the service returns the unique internal key ID of the returned records in either the recordId or the id element. The record ID is then used in a subsequent call using a URI to retrieve the record's full XML data set, or to perform an update operation. Update functions require the full set of fields and their values to be passed, as the full contents of the record are replaced. If all fields are not present in the external system, the update process would need to first retrieve the current contents of the record from Brand Compliance before applying the changes to that data set, and submitting the full data set as the update.
- There could be several initiators for the data exchange, including the following:
  - Human initiation, where an individual makes the request, such as by clicking a button to run the request.
- End user action, where the completion of an activity in the external system triggers the exchange.
- Scheduled request (or polling), where the external system sends or checks for data on a regular basis. For example, it may be looking for newly created Product Records or Specifications with a change in status. For security purposes, the APIs rely on the external system polling, retrieving or submitting calls to the APIs. Brand Compliance does not push data externally.

The nature of the APIs may require multiple requests to be made, for example:

- Request a list of Product Records with their IDs.
- Request the timestamp reference for each Product Record.
- Identify any Product Records changed or created within a given time scale.
- Request the full details for that Product Record.

- These types of requests can be automated. The external system is altered to enable the operations to take place in sequence according to a pre-defined algorithm. Certain modifications can also be made to the external system to enable the storage of data returned from Brand Compliance. For example, if the external system requests a record to be created in Brand Compliance, a confirmation response is sent, containing the record ID. This ID may then be stored in the external system for future reference.

# Chapter 5

## Customer Segmentation - POC

### 5.1 Pre-requisites

#### 5.1.1 Data set Requirement

For this I needed data containing two main attributes Text and Number. As our target application is for Markets the data must have been of the form such that it contains Number, Regex and Range of Income and length of the text should be around 25 characters. These data collected are in obfuscated form we can get this data from Search API ,Rest API and UCI open repository.

#### 5.1.2 Data set Description

The data contains 1,00,000 sell product details. I have used only some characteristic parameters only.

Age - Customers age.

Income - From scholar paper mentioned approach, good to categorize the income.

Sub-category - Products the features and comments extracted using scrapping.

#### 5.1.3 The features extracted are:

Total Money and Total number of Frequency to visits

- The Customer spent.
- Spent in the last one, two, three, six months.
- The Customer shopped.

- Shopped done in the last one, two, three, six months.
- Last shopped customers month.

The targeted classes are indicated as 0 = Recency in months more, 1 = spends per customer, 2 = frequency per person.

### 5.1.4 Pre-processing steps

1 features normalization:- Before performing any kind of analysis, it mandatory to clean all data element in to one standard scale normalization performed on it. In data set to represent customer's income BRACKET\_DESC attribute is used which contains range of income which is normalized to continues numerical value during this phase.

---

```
# In model.py
# regex to extract numbers from income bracket and average them as income
# Maybe better to use continuous income instead of a categorical variable for
# Income
customers['INCOME']=customers['BRACKET_DESC'].
apply(lambda x:np.array(list(map(int, re.findall(r'\d+', x)))).mean())
cust=customers[['CUSTOMER_ID','INCOME']]
train=pd.merge(train,cust,on='CUSTOMER_ID')
```

---

## 5.2 Approach To Implement Customer Segment.

First of all merge customer detail with order detail to get which Customerid and particular product id then merge it with item data set to get information about that particular ordered product and its subcategory. Here discount is 0 and promotion id is same so we can't use this attribute in analysis. Also emp\_id has no impact on analysis. So drop unnecessary column from data set like 'ORDER\_ID', 'EMP\_ID', 'PROMOTION\_ID', 'QTY\_SOLD', 'DISCOUNT' then I used sales data from January to October to train model and i used Novemebr months data to test model.

Then i calculated frequency of customer in past 1 month,3 month and 6 month and also calculated money spend by particular customer in last 1 month,3 month and 6 month.

Then I find when was the last visit of customer to buy any product. In data set there is one field called BRACKET\_DESC which contains range of customer's income I converted it into numeric continuous values by finding average of it which is also considered in analysis. here value of k is 3 which means it will generate 3 cluster group of customer.

### 5.3 Analysis from Implementation

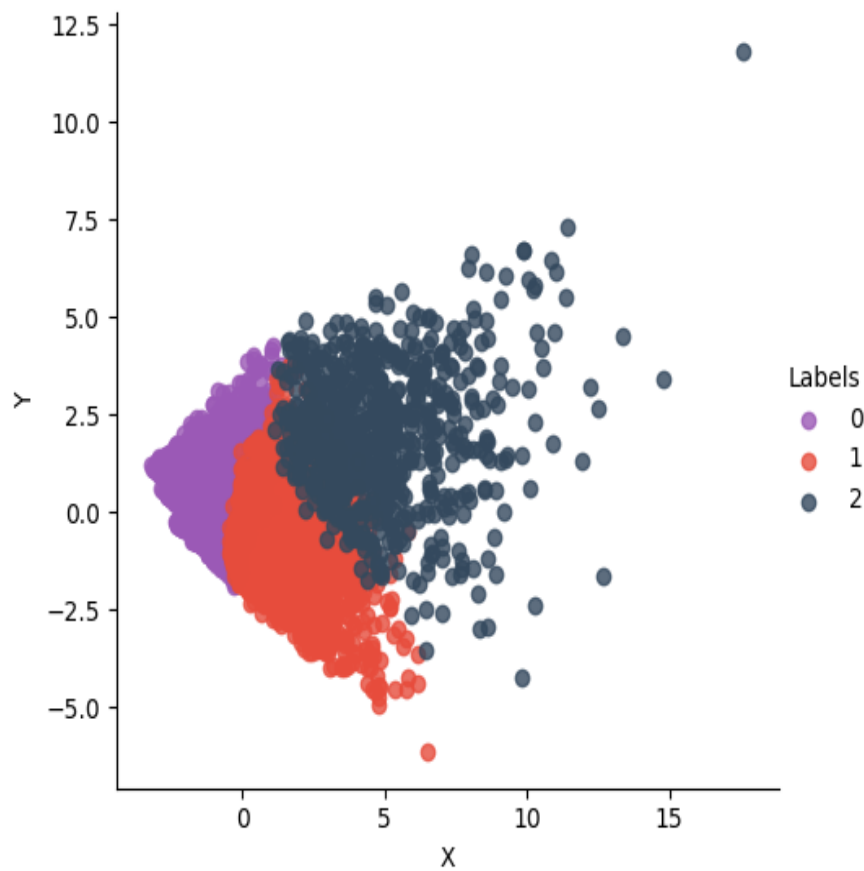


Figure 5.1: Chart Shows customer segmentation.

CLUSTER COLOR	CUSTOMERS SHARE	SPENDS PER CUSTOMER	REGENCY IN MONTHS	FREQUENCY PER PERSON
	45.75%	675	1.65	30
	15.07%	1718	1.15	39
	39.18%	935	1.01	40

Figure 5.2: The Analysis of customer segmentation.



# Chapter 6

## Subject System

The subject system for this project includes nothing but the applications themselves. Since, the project is completely software based, it does not require any kind of hardware equipment to test the applications. All that is needed are the applications themselves.

### 6.1 Integration Application Testing

For the testing of the project following components have been used:

1. Integrated applications (Customer Relationship Management and Business Analytics Tool).
2. Use Case generated by testers to test the functionalities.
3. Functional Testing.

For checking whether the project is successfully working according to the requirements, in this project, just the manual testing is used. No tool for automatic testing is used in this project.

1. Various use-scenario and test scenario are developed by the tester and those cases are then checked with the working of the project. If all the test are passed by the application, without generating any error, then the project is ready to be deployed on the client's system or for the next release.
2. If there arises some errors, then the feedbacks are raised which are sent back to the developers to be fixed. Until the bugs or the feedbacks are fixed completely, the circle keeps going between the testers and the developers.

# Chapter 7

## Functional Enhancements

### 7.1 Productisation

The Branding Theming enhancement provides the administrator with the ability to revoke the core ORBCM Cloud Service stylesheets and images with portal-specific custom favorites. This is accomplished through a new choice in the Admin area where the override images and CSS files are uploaded. The area also allows the content of the User and supplier Terms Conditions text to be configured (along with any language translations which the user may now choose to switch to when accepting the Terms Conditions).

The Supplier Site Coding enhancement provides the administrator with the ability to set the customs for the code format of new suppliers and sites. New system parameters hold settings for whether the codes are auto-assigned. 1) Oracle Retail believes that the retailer has assured its OS has been covered with all the relevant Windows updates. 2) Where new fields or (UI) settings are introduced as part of a revolution, the portal owner may require to set up any new system text and associated adaptations to support it. 3) An account is created, and if they may subsequently be changed. The environments also control the format of the site code whether numeric or alpha guide is used (length) if the site code must be individual across the portal, and if the site code is to insert the supplier code as a prefix.

The API Security characteristic contributes more granular power over access to the APIs by external systems, enabling access to be limited to specific web services as well as their endpoints or requests. This flexibility allows for a single account to be created for an external system to access multiple APIs, a separate account for each API the external

system needs to access. The accounts comprise login ID and password credentials for authentication with Cloud Service, the passwords continuing subject to the same format, and expiry rules as the user passwords, with extra expiry notifications.

## **7.2 Glossary Import/Export**

The productisation of ORBCM Cloud Service also provides a facility for the administrator or portal implementer to load the contents of the portal's glossaries without the need for development effort. This is achieved by providing a glossary import and export facility, available to users that have been granted the new Glossary Administrator authority profile. Glossaries may be exported and imported either individually or using a bulk export option. The import/export facility has options to include language translations. An Excel spreadsheet file is used to contain the contents of the glossary which can be exported, amended offline, and imported back into the system.

## **7.3 Supplier And Product Upload**

A further aspect of the productisation of ORBCM Cloud Service provides a facility for the administrator or portal implementer to load new products, suppliers sites and retailer users en masse, without the need for development effort. This is achieved by providing supplier, product, and user upload options within the Admin area, available to users that have been granted the new Upload Administrator authority profile. The Supplier Site upload uses an Excel spreadsheet to create new Supplier and Site accounts, subject to the appropriate validation checks. The spreadsheet template can be downloaded to ensure the correct format is used for the upload. The creation of a new Supplier account will also create an initial User account; the upload includes the option to automatically notify the designated user on creation of the account. The Product upload uses an Excel spreadsheet to create new Product Records (not Product Specifications), subject to the appropriate validation checks. The spreadsheet template can be downloaded to ensure the correct format is used for the upload. The Retailer User upload uses an Excel spreadsheet to create new retailer/portal owner user accounts (not supplier users or contacts), subject to the appropriate validation checks. The spreadsheet template can be down-

loaded to ensure the correct format is used for the upload.

## **7.4 System Text**

Improvements are introduced to the administration of system text (field labels, error messages, and so on), help text, and their language translations. The core default system text can be overridden with portal-specific custom text, by either editing within the Admin area or exporting for offline editing prior to importing back into the system. The improvements provide a more efficient means of managing text changes, in particular where there are large scale changes as part of a version release. Prior to 16.0, the ability to configure system text was restricted to Oracle users, with the changes being applied through the upload of spreadsheets. The 16.0 improvements provide a new module in the Admin area for administrators with the Power 10 Administrator authority profile to maintain the overrides directly within the system, or alternatively through an export/import option.

## **7.5 Scorecard**

## **7.6 Product Specification**

# Chapter 8

## Observations And Findings

The observations and findings while working on this project mainly depict why one technology is preferable than the other one or why there was a shift from one technology to another. Also, it shows a comparison between the two and helps the programmer decide which of the two is better under what scenarios.

- It provides end-to-end lifecycle management by enabling the capture of detailed product information during the sourcing and selection of suppliers and the manufacturing process.
- Consolidating all necessary product detail promotes transparency throughout the supply chain.
- Having this product, users can access and interrogate data in seconds, allowing brand owners to quickly respond to product or industry incidents and maintain continued consumer trust in their brand.

### **Coding Mechanism Related Observations:**

1. Application Programming Interface(API) is a better way to retrieve data from the server when compared to the traditional retrieval of complete html page since only the relevant data is loaded instead of loading the entire page.
2. The Event handling mechanism when implemented through event bus, this allows broadcasting of events and thus specific registered handlers handle the fired events.
3. Many java files are written actually to accomplish code reusability. Functions call with appropriate arguments serve purpose.

4. It follows MVC model, So it contains all service class along with Dao classes which follow a proper mechanism for even declaring objects.

**Security Related Observations:** There are some of the .class files whose Data can't be modified but can be seen. Function calls made from .class file can be modified but can be used in such any process to serve the whole purpose.

# Chapter 9

## Conclusion and Future Scope

### 9.1 Conclusion

The Proposed systems handle the tedious task of supply chain monitoring and analysis. It analysis raw-material, ingredients proportion, product specification type etc. This is handled with the help of proper and well defined approach i.e. audit and visit, supplier scorecard, product specification type, other labelling copy.

Also I tried to predict the repeat customer purchase that give accurate result based on unsupervised learning. It is used to to recognize the potential customers that frequently purchased. The system handles easily management's process with less manpower, less maintenance cost.

The retail trade, and individually the food area, has become a deeply delicate business with consumers compelling full clarity, correct labeling, quality ownership, and rapid response to conflicts.

### 9.2 Future Work

In future this application will be shifted to be Oracle cloud Infrastructure and ready to handle bulk data. Right now the product is going into the transition phase. The product is shift to Oracle JET.

Till now only particular kind of product is to be taken care of that to be store for long time i.e Wine,juice, horlicks, dairy products etc. But as demand is increased now need to handle the product which is used in daily life like variety of dairy product and its version of different quantity etc.

Every retailer want profit that can be enable by good quality of product. So supplier play a major role in handling quality of product. The module supplier scorecard helps retailer to give score and comments. In future ML and DP approach is a way to recommended best supplier out of given one.

In future customer segmentation can be done with LSTM and Random Forest to predict and recommend the product to consumer for better result and better accuracy.



# Bibliography

- [1] “<https://blogs.oracle.com/retail/new-stats-about-brand-compliance-adoption>,”
- [2] J. Joydeep Ghosh, Er. Strehl, “Value-based customer grouping from large retail datasets,” *Proceedings of SPIE - The International Society for Optical Engineering*, 2017.
- [3] “Oracle retail brand compliance management cloud service,”
- [4] S. O. Chinedu Pascal Ezenkwu University of Uyo, “Application of k-means algorithm for efficient customer segmentation: A strategy for targeted customer services,” *International Journal of Advanced Research in Artificial Intelligence*, 2015.
- [5] A. U. L. P. A. N. A. Joy Christya, , “Rfm ranking – an effective approach to customer segmentation,” *Journal of King Saud University – Computer and Information Sciences*, 2018.