**A MINI PROJECT REPORT**

**ON**

**“Supermarket Product Recommendation”**

Submitted to

SAVITRIBAI PHULE PUNE UNIVERSITY

in completion of

**Data Mining and Warehousing**

**&**

**Software Testing and Quality Assurance**

**(B.E Computer Engineering)**

**BY**

Param Sonawane Exam No: B150234391

Atharva Lonkar Exam No: B150234306

Abhishek Deshmukh Exam No: B150234250

Soham Sagade Exam No: B150234388



Department of Computer Engineering

Sinhgad College of Engineering, Pune-41

**Accredited by NAAC with grade ‘A’**

**YEAR 2020-2021**

**CERTIFICATE**

Sinhgad Technical Education Society,

Department of Computer Engineering

Sinhgad College of Engineering, Pune-41

**Accredited by NAAC with grade ‘A’**



**“SUPERMARKET PRODUCT ANALYSIS”**

Submitted to

SAVITRIBAI PHULE PUNE UNIVERSITY

in completion of

**Data Mining and Warehousing**

**&**

**Software Testing and Quality Assurance**

**(B.E Computer Engineering)**

BY

Param Sonawane 405A076 B150234391

Atharva Lonkar 405A044 B150234306

Abhishek Deshmukh 405C005 B150234250

Soham Sagade 405A074 B150234388

Prof. T. P. Vaidya Prof. M. Sadaphule

Internal Guide External Guide

Department of Computer Engineering Department of Computer Engineering

Dr.S.D.Lokhande

Principal

SCOE, Pune

**CONTENTS**

# TITLE PAGE NO

# 

Certificate 2

Acknowledgement 4

Abstract 5

List of Figures 6

Introduction 7

Project planning & management 8

Analysis & design 9-14

Implementation & coding 15-18

Conclusion 19

Reference 19

**ACKNOWLEDGEMENT**

“Gratitude is not only the greatest of virtues, but the parent of all others.”

Hence, it is great pleasure for me to acknowledge the assistance and contribution of number of individuals who helped me in developing “**Supermarket Product Recommendation**” project. First and Foremost, I wish to record my gratitude and thanks **to Prof. T. P. Vaidya** for her enthusiastic guidance and help in successful completion of seminar. I express my thanks to **Prof. S.D. Lokhande** (Principal), **Prof. M.P. Wankhade** (Head of Computer Department) and **Prof. M. Sadaphule** (External Guide) for their valuable guidance. I am also thankful to other teachers and non-teaching staff of Computer Engineering Department and Library for their co-operation and helping us and guiding us to complete our project accurately.

**ABSTRACT**

Supermarket retail industry is a very competitive and cutthroat field. This field requires analysis on a day to day basis. Every supermarket has a limited floor space and the goal is to use this limited space in an efficient manner. Executives or decision makers exploit the power of association rules to arrange products that are best sold together. Our project not only performs association rule mining, but takes it a step further by performing association rule mining on specific clusters of customers. This targeted analysis gives the supermarket ability to design policies and offers for a specific target group. Our project also provides a functionality that recommends relevant products to customers.

**LIST OF FIGURES**

1. Process Model.
2. Use Case Diagram
3. System Architecture Diagram
4. Sequence Diagram
5. Gui Interface
6. **INTRODUCTION** 
   1. **Background and Basics**

* Supermarket is a big chain, and it is difficult for them to predict which products they are more likely to buy. Associative mining rules can help them predict the likelihood of customers buying two products simultaneously.
* There are two data sets in the project. One is customer dataset. It contains fields such as customerID, Gender, Age, Annual Income, Spending Score (1-100). The other one is transaction dataset. It contains information about the what customers have bought in the timeframe and the frequency of buying a particular product.
* Each customer has a unique identity.
  1. **Problem Statement**

1. To recommend supermarket products to customers and perform analysis using apriori algorithm and k-means algorithm.
2. To perform unit testing on program modules using unit testing library.
3. To perform automated functional testing using Selenium testing tool.
   * 1. **Scope Statement**

For the scope of this project, knowing the trend in buying pattern was the main concern. We created this model to help the authorities understand which products customers are likely to buy and in which combinations. Also, we have used various testing tools and techniques to see if the system is giving us the desired result.

**2**. **PROJECT PLANNING & MANAGEMENT**

**2.1 Software requirement**

**2.1.1 Python**

Python is a powerful high-level, object-oriented programming language created by Guido van Rossum. It has simple easy-to-use syntax, making it the perfect language for someone trying to learn computer programming for the first time. Python is a general-purpose programming language created in the late 1980s, and named after Monty Python, that’s used by thousands of people to do things from testing microchips at Intel, to powering Instagram, to building video games with the Py Game library. It’s small, very closely resembles the English language, and has hundreds of existing third-party libraries. Python is a dynamic, interpreted (byte code-compiled) language. There are no type declarations of variables, parameters, functions, or methods in source code. This makes the code short and flexible, and you lose the compile-time type checking of the source code. Python tracks the types of all values at runtime and flags code that does not make sense as it runs.

**2.1.2 JavaScript**

**JavaScript** (**JS**) is a lightweight, interpreted, or just-in-time compiled programming language with [first-class functions](https://developer.mozilla.org/en-US/docs/Glossary/First-class_Function). While it is most well-known as the scripting language for Web pages, [many non-browser environments](https://en.wikipedia.org/wiki/JavaScript" \l "Uses_outside_Web_pages) also use it, such as [Node.js](https://developer.mozilla.org/en-US/docs/Glossary/Node.js), [Apache CouchDB](https://couchdb.apache.org/) and [Adobe Acrobat](http://www.adobe.com/devnet/acrobat/javascript.html). JavaScript is a [prototype-based](https://developer.mozilla.org/en-US/docs/Glossary/Prototype-based_programming), multi-paradigm, dynamic language, supporting object-oriented, imperative, and declarative (e.g. functional programming) styles. Read more [about JavaScript](https://developer.mozilla.org/en-US/docs/Web/JavaScript/About_JavaScript). This section is dedicated to the JavaScript language itself, and not the parts that are specific to Web pages or other host environments. For information about [APIs](https://developer.mozilla.org/en-US/docs/Glossary/API) specific to Web pages, please see [Web APIs](https://developer.mozilla.org/en-US/docs/Web/API) and [DOM](https://developer.mozilla.org/en-US/docs/Glossary/DOM). The standard for JavaScript is [ECMAScript](https://developer.mozilla.org/en-US/docs/Web/JavaScript/Language_Resources). As of 2012, all [modern browsers](https://kangax.github.io/compat-table/es5/) fully support ECMAScript 5.1. Older browsers support at least ECMAScript 3. On June 17, 2015, [ECMA International](https://www.ecma-international.org/) published the sixth major version of ECMAScript, which is officially called ECMAScript 2015, and was initially referred to as ECMAScript 6 or ES6. Since then, ECMAScript standards are on yearly release cycles. This documentation refers to the latest draft version, which is currently [ECMAScript 2020](https://tc39.github.io/ecma262/).

**2.1.3 HTML**

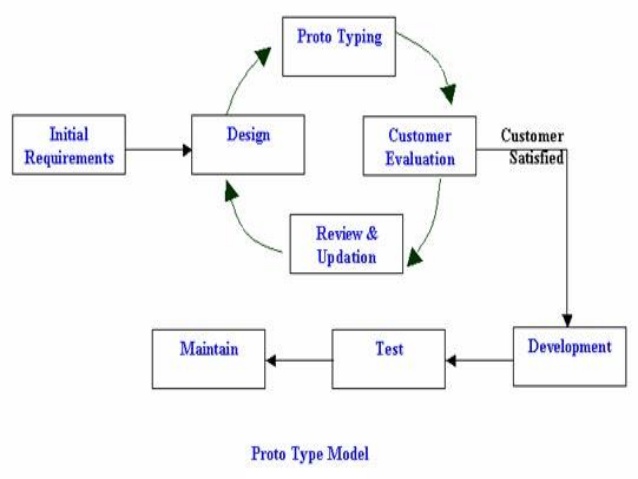
**Hypertext Mark-up Language** (**HTML**) is the standard mark-up for documents designed to be displayed in a [web browser](https://en.wikipedia.org/wiki/Web_browser). It can be assisted by technologies such as [Cascading Style Sheets](https://en.wikipedia.org/wiki/Cascading_Style_Sheets) (CSS) and [scripting languages](https://en.wikipedia.org/wiki/Scripting_language) such as [JavaScript](https://en.wikipedia.org/wiki/JavaScript). [Web browsers](https://en.wikipedia.org/wiki/Web_browser) receive HTML documents from a [web server](https://en.wikipedia.org/wiki/Web_server) or from local storage and [render](https://en.wikipedia.org/wiki/Browser_engine) the documents into multimedia web pages. HTML describes the structure of a web page [semantically](https://en.wikipedia.org/wiki/Semantic_Web) and originally included cues for the appearance of the document. [HTML elements](https://en.wikipedia.org/wiki/HTML_element) are the building blocks of HTML pages. With HTML constructs, [images](https://en.wikipedia.org/wiki/HTML_element" \l "Images_and_objects) and other objects such as [interactive forms](https://en.wikipedia.org/wiki/Fieldset) may be embedded into the rendered page. HTML provides a means to create [structured documents](https://en.wikipedia.org/wiki/Structured_document) by denoting structural [semantics](https://en.wikipedia.org/wiki/Semantics) for text such as headings, paragraphs, lists, [links](https://en.wikipedia.org/wiki/Hyperlink), quotes and other items. HTML elements are delineated by *tags*, written using [angle brackets](https://en.wikipedia.org/wiki/Bracket" \l "Angle_brackets). Tags such as <**img** /> and <**input** /> directly introduce content into the page. Other tags such as <**p**> surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags, but use them to interpret the content of the page.

**2.1.4. Associative Mining Rules (K-Means and Apriori).**

Association rule mining is a technique to identify underlying relations between different items. Apriori algorithm is one of the most widely used method to mine association rules. Apriori algorithm uses various measures to select the most interesting and useful association rules like **Support, Confidence, Lift**. There are several applications of apriori algorithm like market basket analysis, recommendation systems, etc. K means clustering is one of the most popular unsupervised algorithms. K means identifies k number of clusters, and then allocates every point in the dataset to the nearest cluster. A cluster is a collection of data points that are similar in some way.

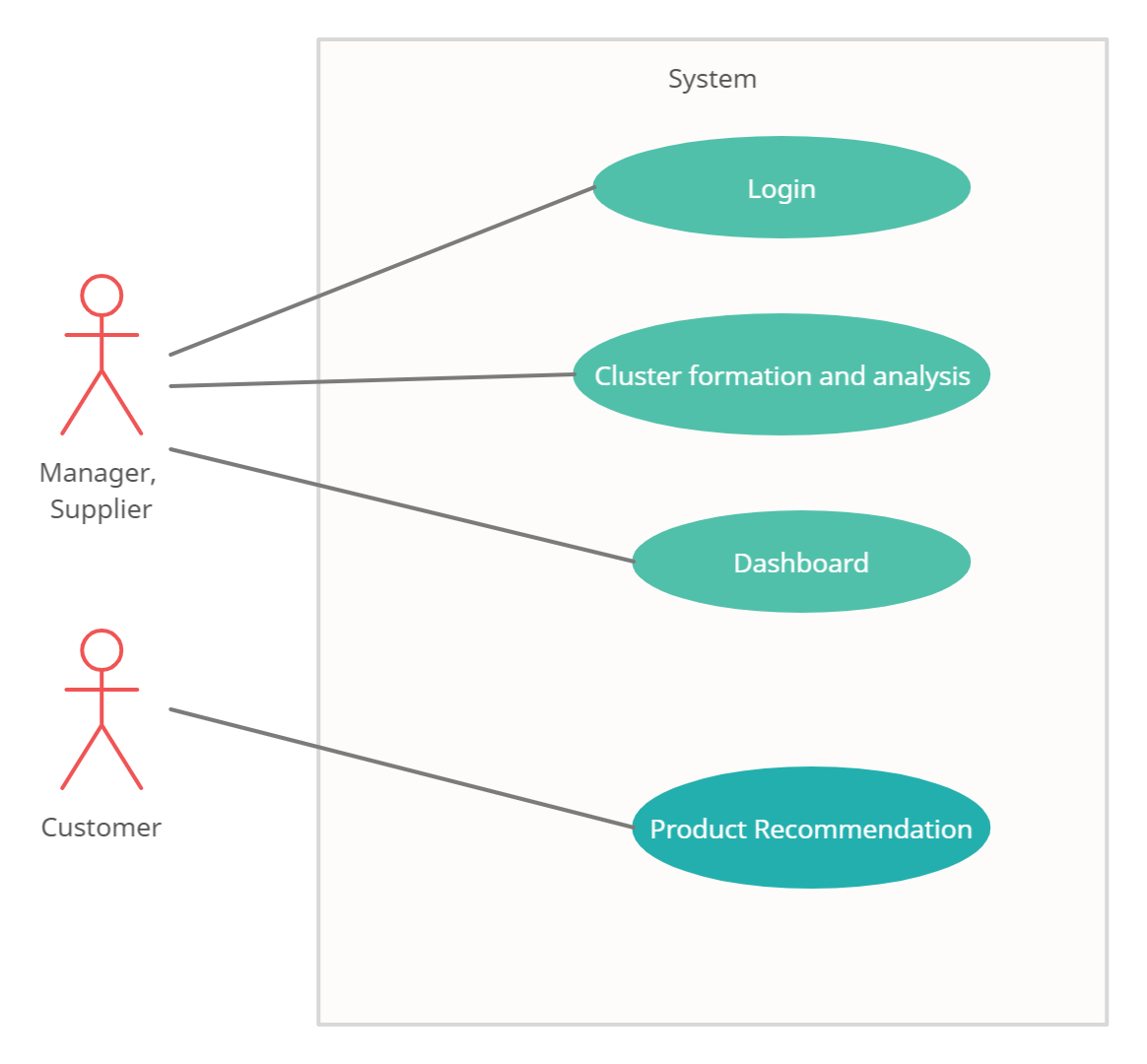
There are various applications of clustering in data mining like patter recognition, image processing, clustering help marketers find distinct patterns or behaviors within their customer base.

**2.2 Process Model**

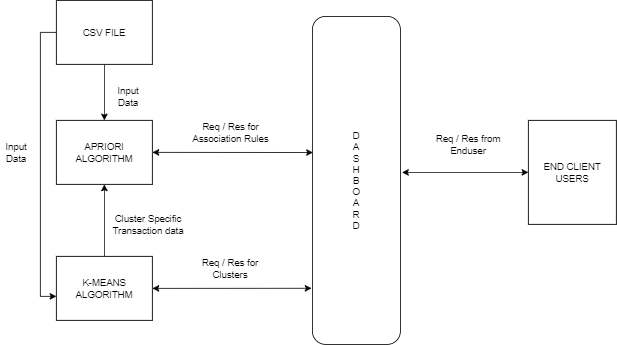


**3. ANALYSIS & DESIGN**

**3.1 Use-Case Diagram**

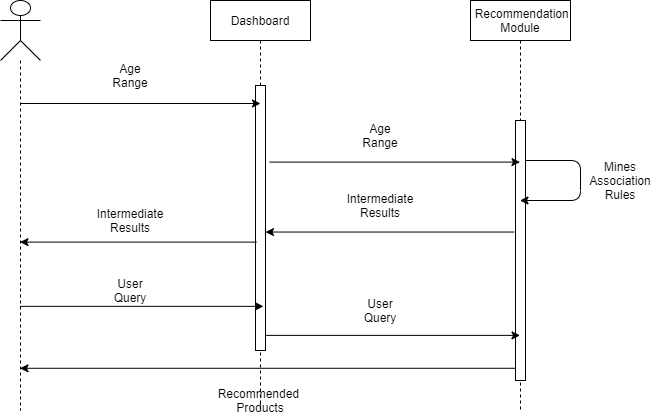


**3.6 System Architecture Diagram**

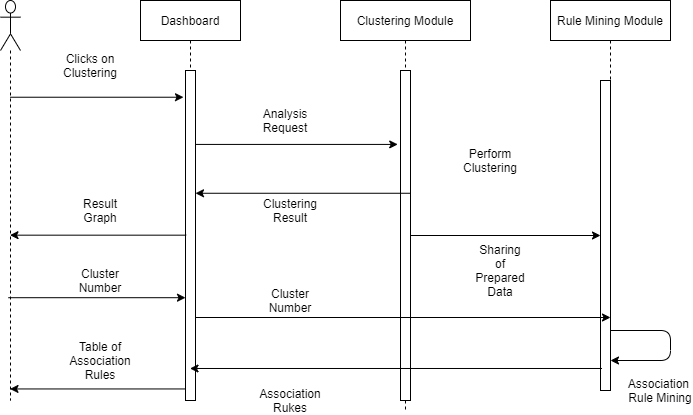


**3.7 Sequence Diagrams**

1. **Product Recommendation**



1. **Cluster Specific Rule Mining**



**4. IMPLEMENTION & CODING**

**4.1 Methodology**

We did the quantitative research by visiting various websites.

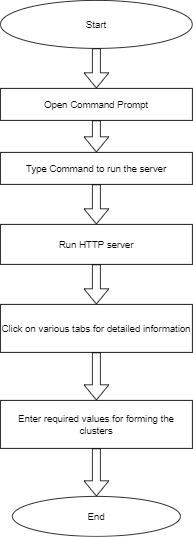
We collected our Dataset from Kaggle.com.

Before analysis the gathered data was prepared. The dataset was checked for missing data and outliers.

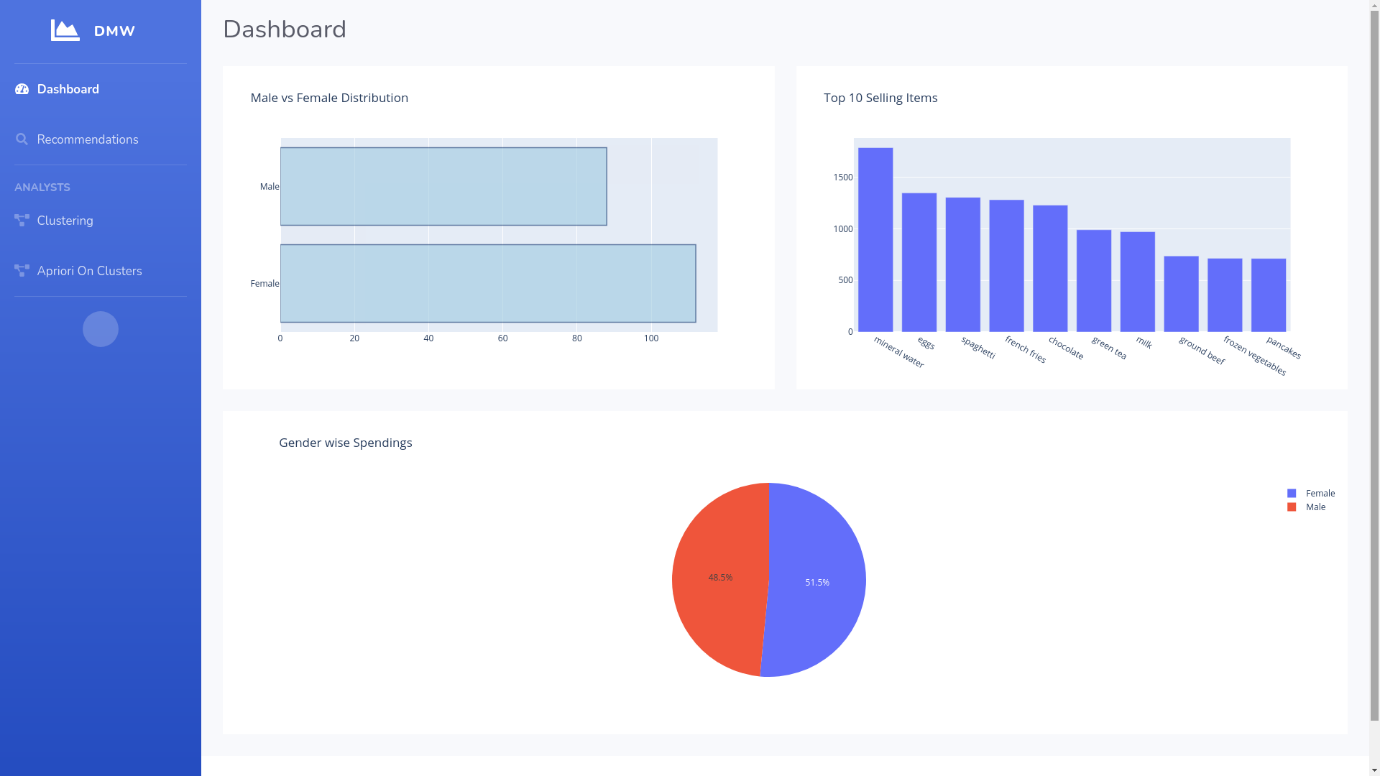
* 1. **Algorithms**

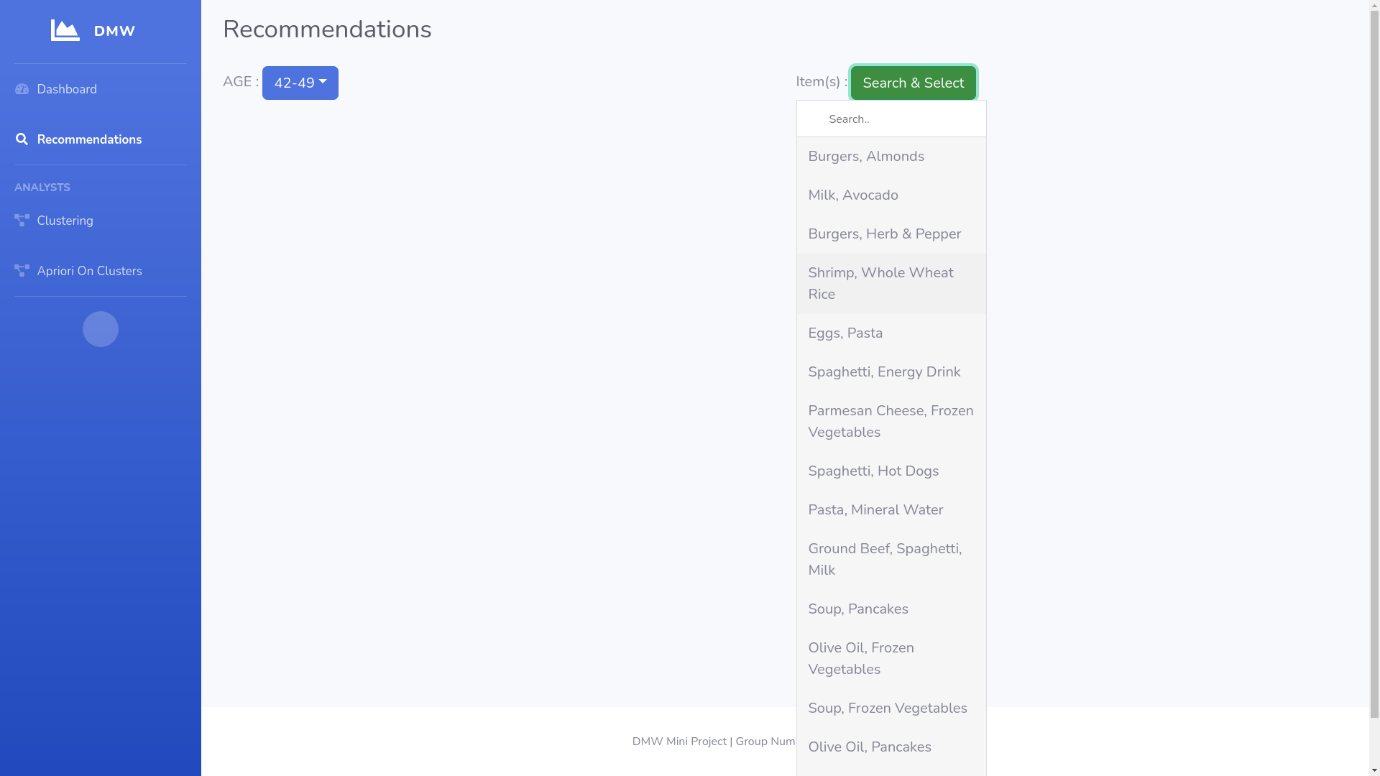
1. Start
2. Open Command Prompt
3. Type Command “python manage.py runserver” and press Enter
4. Once you get an URL- [http://127.0.0.1.8000](http://127.0.0.1.8000/) copy the URL in browser and run
5. Click on various tabs and sidebars for detailed analysis.
6. Enter required values for recommendations and cluster analysis.
7. End.

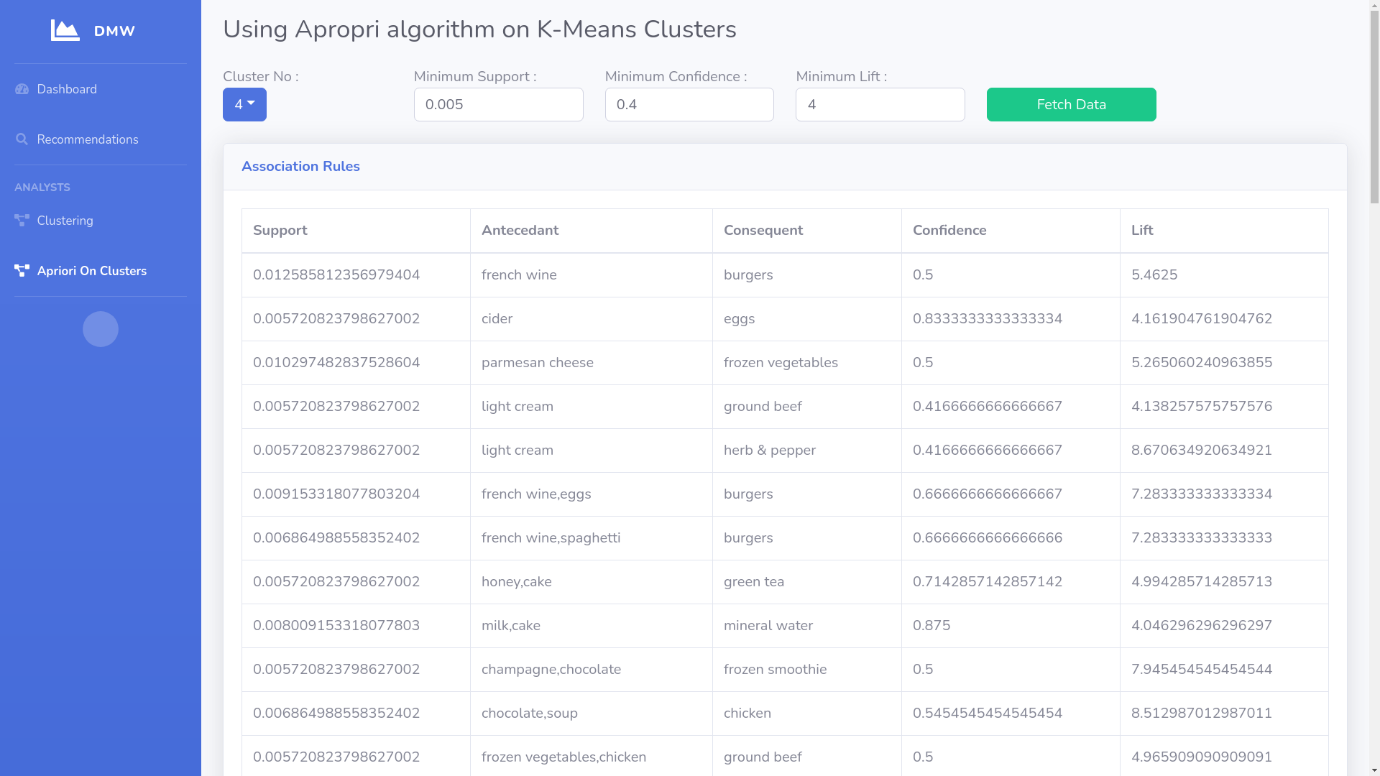
**4.3 Flowchart**



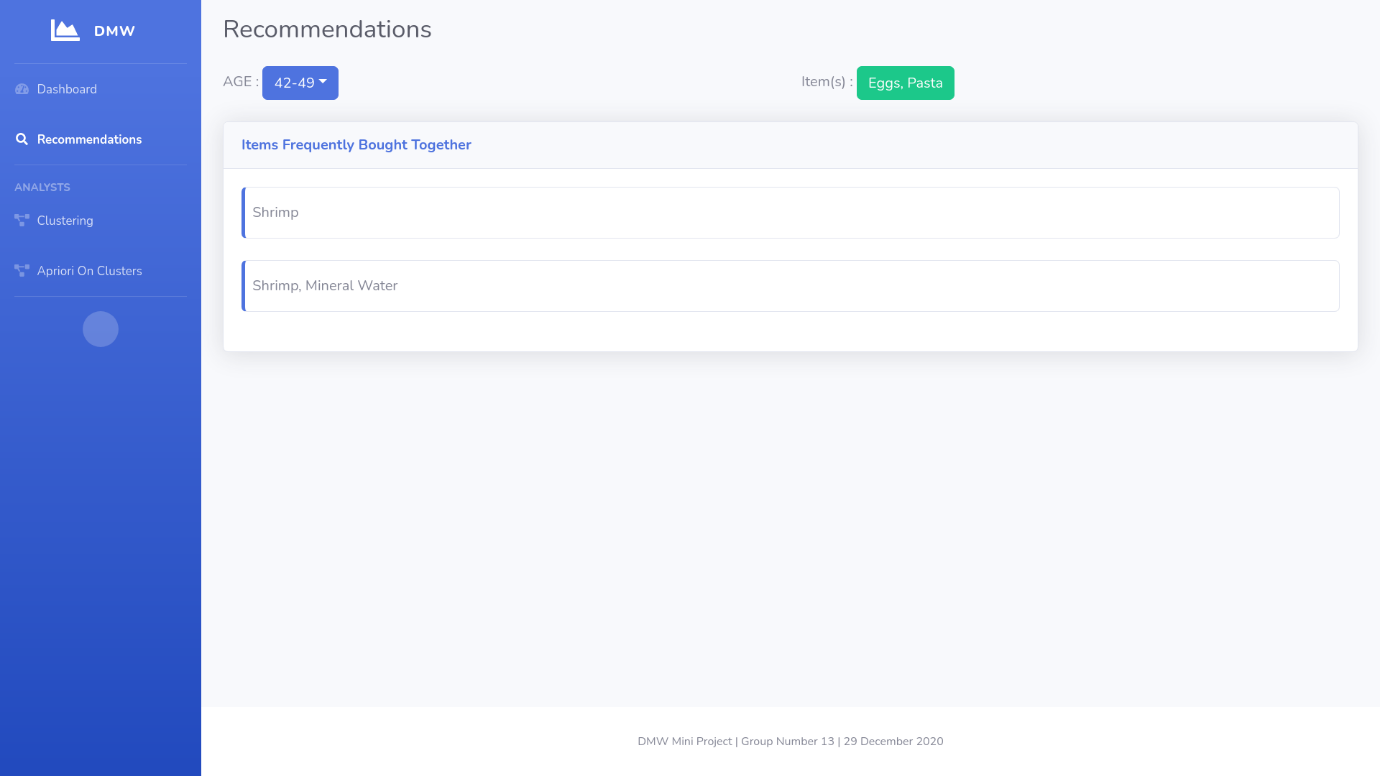
* 1. **GUI Design /screen shots**











**5. CONCLUSION**

* Hence, we conclude that by using the combination of the proposed algorithms we can perform cluster specific analysis and recommend products to customers.
* The extensive testing techniques and tools we used in this project proved to be helpful in making the project error-free.

**6. REFERENCES**

* https://scikit-learn.org
* https://stackoverflow.com
* https://www.wikipedia.org
* https://www.kaggle.com
* https://colorlib.com
* https://www.w3schools.com
* https://www.bootstrapdash.com