Software Requirements Specification

For

**Identifying and promoting various traditional endangered cuisines of different states using data analysis**

09/12/2021

Prepared by

|  |  |  |
| --- | --- | --- |
| **Specialization** | **SAP ID** | **Name** |
| CSE-BAO | 500076037 | Saumya Sinha |
| CSE-BAO | 500075316 | Divyanshu Chopra |
| CSE-BAO | 500075273 | Shashwat Mishra |
| CSE-BAO | 500078207 | Tarun Kumar |



Department of Informatics School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007. Uttarakhand

# Table of Contents

|  |  |  |
| --- | --- | --- |
| **Topic** | | **Page No** |
| Table of Content | |  |
| Revision History | | 3 |
| 1 | Introduction | 4 |
|  | 1.1 Purpose of the Project | 4 |
|  | 1.2 Target Beneficiary | 4 |
|  | 1.3 Project Scope | 4 |
|  | 1.4 References | 4 |
| 2 | Project Description | 5 |
|  | 2.1 Reference Algorithm | 5 |
|  | 2.2 Data/ Data structure | 7 |
|  | 2.3 SWOT Analysis | 7 |
|  | 2.4 Project Features | 8 |
|  | 2.5 User Classes and Characteristics | 9 |
|  | 2.6 Design and Implementation Constraints | 10 |
|  | 2.7 Design diagrams | 10 |
|  | 2.8 Assumption and Dependencies | 11 |
| 3 | System Requirements | 12 |
|  | 3.1 User Interface | 12 |
|  | 3.2 Software Interface | 12 |
|  | 3.3 Database Interface | 12 |
|  | 3.4 Protocols | 12 |
| 4 | Non-functional Requirements | 12 |
|  | 4.1 Performance requirements | 12 |
|  | 4.2 Security requirements | 12 |
|  | 4.3 Software Quality Attributes | 13 |
| 5 | Other Requirements | - |
| Appendix A: Glossary | | 14 |
| Appendix B: Analysis Model | |  |
| Appendix C: Issues List | |  |

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Change** | **Reason for Changes** | **Mentor Signature** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## INTRODUCTION

* 1. **PURPOSE OF THE PROJECT**

Customer segmentation can significantly impact client management because it allows a company to appeal to distinct groups of consumers with similar demands and focus on what each type of customer requires at any given time. Depending on the company's resources or goals, specific consumer segments can be targeted, whether large or small.

Whatever segmentation models marketers choose, they all require them to create customer groups as a first step in segmenting the customer base. Marketers will usually have a series of levels for each form of segmentation model as a result of this. Marketers can then combine tiers from multiple models to create more defined groupings. Combining the top tier of customers based on an RFM model with a low longevity tier, for example, will result in a sector of highly engaged, freshly acquired customers for marketers.

## TARGET BENEFICIARY

Marketers can use customer segmentation to reach out to each consumer in the most efficient way feasible. A customer segmentation study makes use of the massive quantity of data on customers (and potential customers) to accurately identify discrete groups of customers based on demographic, behavioral, and other factors. Because the marketer's goal is to maximize the value (revenue and/or profit) from each client, it's crucial to know how each marketing action will affect the customer ahead of time. Ideally, such "activity-centric" customer segmentation will focus on the long-term customer lifetime value (CLV) impact of a marketing action rather than the short-term value of marketing activities. You can increase customer loyalty by better understanding the customer and thus being able to target them more effectively.  You've enhanced consumer loyalty by interacting with the firm more frequently, even though each basket is smaller .

Little and frequently is far more beneficial than a single event. It's also a better predictor of behavior, which can help businesses make better judgments. It will not only promote customer loyalty, but it will also increase the customer's value, hence increasing the customer's lifetime value.

## PROJECT SCOPE

## Customer segmentation is the process of grouping customers together based on common characteristics. These customer segments can help with marketing campaigns, identifying potentially profitable clients, and building customer loyalty. You can develop the correct product, arrange the right distribution and positioning, and match the right sales motion to each consumer after you have all of the segments, while also refining your segments over time. When done correctly, it's a model that allows anyone in your firm to understand your clients right away. All strategic concerns, as well as changes in product, pricing, and packaging over time, cannot be accounted for in segmentation. Instead, it's a continuous effort to identify relevant differences among your clients.

## REFERENCES

## PROJECT DESCRIPTION 2.1 REFERENCE ALGORITHM

K Means Clustering Algorithm:

One of the most basic and often used unsupervised machine learning algorithms is K-means clustering.

A cluster is a collection of data points that have been grouped together due to particular similarities.

You'll set a target number, k, for the number of centroids required in the dataset. A centroid is a fictional or real location that represents the cluster's center.

By lowering the in-cluster sum of squares, each data point is assigned to one of the clusters.

To put it another way, the K-means algorithm finds k centroids and then assigns each data point to the closest cluster while keeping the centroids as small as possible.

For data cluster analysis, K-means clustering is a widely used technique. However, because little alterations in the data might lead to considerable variance, its performance is usually not as good as that of other complex clustering algorithms.

Furthermore, clusters are considered to be spherical and uniform in size, which could lower the precision of the K-means clustering Python findings.

## CHARACTERISTICS OF DATA

PRIMARY SOURCE: We have taken a dataset of over 500k entries from UCI machine learning repository. This Online Retail II data set contains all the transactions occurring for a UK-based and registered, non-store online retail between 01/12/2009 and 09/12/2011.The company mainly sells unique all-occasion gift-ware. Many customers of the company are wholesalers.

Source of data: Dr. Daqing Chen, Course Director: MSc Data Science. chend **'@'** lsbu.ac.uk, School of Engineering, London South Bank University, London SE1 0AA, UK.

ACCURACY:

In this mathematical analysis model, we are using a dataset from UCI data repository having columns like Invoice number, stock code, Quantity, invoice date etc. As we are using data from well-reputed repository, we can say that the data is highly accurate and in our project we have included static data which is only applicable to a selected state.

RELIABILITY:

The source of our data in this dataset is UCI repository which can be considered a reliable source as professionals use data for their projects .

RELEVANCE:

As we are doing customer segmentation the customer ID entries aligned with the product bought would be highly relevant to our project.

TIMELINESS:

Data used to build our mathematical model is static as it is collected till a particular day.

## SWOT ANALYSIS

### STRENGTHS-

* + 1. User & group permissioning
    2. Scalability.
    3. Document management.
    4. Easy-to-use.
    5. Advantage in making marketing related decisions.
    6. High ratings on review.
    7. Customer “love” for the brand.
    8. Coupling and recommendations of product to specific target audience.

### WEAKNESSES-

1. Deployment.
2. No offline mode.

### OPPORTUNITIES-

1. Emerging "Open Courseware" market.
2. Calendaring
3. Research features .
4. Improved document management.
5. Rating/marking of content.
6. Better collaboration
7. Customer Proximity

### THREATS-

1. "Old school" IT attitudes
2. Entrenched proprietary domain-specific platforms (e.g. Blackboard)

## Project Features

Customer segmentation is important for businesses to understand their target audience..

Based on their demographic profile, interests, and socioeconomic level, different adverts can be selected and sent to different audience segments. There are numerous unsupervised machine learning techniques that can assist businesses in identifying their user base and segmenting their customers. The goal of this method is to increase the similarity of data inside a cluster while decreasing the dissimilarity between clusters. To categorise or 'cluster' all available data into non-overlapping sub-groups that are unique from one another.

## User Classes and Characteristics

The influence of customer segmentation on customer management is enormous. Customers can be divided into multiple groups based on common features and needs, making it easier to promote to each segment separately and efficiently.

The major user classes which will get best benefit from this project are:

* + 1. Increased competitiveness: Customer segmentation techniques enable a firm to become more active, assertive, and even aggressive. The notion is that if a corporation understands its consumers' characteristics, it would be better able to determine what they require.
    2. Ability to Expand: Customer segmentation enables development not just in terms of space but also in terms of strategy. Because one will be able to cater to the correct client and turn to places populated by those who suit one's current customer base, it will make spatial/physical sense.
    3. Increased customer retention: After the sale, successful customer segmentation will allow you to keep in touch with your clients. It becomes easier to formulate what further can be supplied (additionally) to express the company's passion for meeting their wants and needs by segmenting clients into exact subgroups.
    4. Price optimization: The project will help various restaurants & café within the state and outside to attract large number of people to try traditional cuisines and this will also make their cuisine and recipes different from others.

## DESIGN AND IMPLEMENTATION CONSTRAINTS

Programming Language Used: Python

Other Tools & Technologies: PyCharm, Tableau, MS Excel

The data is collected UCI data repository. Data must be stored in a proper manner using MS Excel as a spreadsheet management tool. PyCharm Must be installed on the device. The project is build using an Object-Oriented Programming Language so we can reuse various portions of code and it also provides flexibility. There are reduced security risks as there are explicit pointers. The code is platform independent, can run on any python supported machine and offers a feature of Write Once Run Anywhere. There is no feature to keep a backup of our data.

## Design Diagrams

Use Case Diagram

Activity Diagram

**Code:**

ASSUMPTIONS AND DEPENDENCIES

## 3.SYSTEM REQUIRMENTS

## 3.1 USER INTERFACE

We are using Microsoft Windows Operating System, various PYTHON supported IDEs like PyCharm and Google Colab. So, it has various GUI functions like Editor, Action indicators and Action list, Navigation Bar, Status Bar, Tool windows etc. These GUI applications provide Popup menus, Context menus, etc allowing easy controls with the help of various input devices like mouse and keyboard.

## SOFTWARE INTERFACE

Software interfaces (programming interfaces) are the languages, codes and messages that programs use to communicate with each other and to the hardware. Examples are the Microsoft Windows Operating Systems. In this project we are using PYTHON as programming language..

## DATABASE INTERFACE

Instead of using a database management system we are using CSV format to store our data in a strictly organized tabular manner. Here, information is stored in the form of records separated by a comma (,). The major advantages of storing data in CSV format are – it is human readable, easy to edit manually, smaller in size and faster way to handle tabular structured data and is quite easy to generate..

## PROTOCOLS

No such communication, security or encryption issue as of now.

## 4.1PERFORMANCEREQUIREMEN

The proposed machine learning model will be based on PYTHON programming language and has to be implemented from IDE. The model will take initial load time depending on local hardware components like RAM & CPU cycles. The performance of the machine learning model implemented in PYTHON may also depend on the users’ operating system like Microsoft Windows, MacOS or Linux..

## SECURITY REQUIREMENTS

The PC on which the Customer Segmentation machine learning model resides will have its own security. Physical access to the system and the application on it will be limited to the user. The system has no extra security features..

## SOFTWARE QUALITY ATTRIBUTES

RELIABILITY- The machine learning model will dependably produce the proper results. Model’s reliability is determined by how well a project performs in various working environments and conditions.

MAINTAINABILITY - It is easy to add code to the existing machine learning system, the program is easy to upgrade for new features from time to time.

USABILITY – The machine learning model is quite easy to use and navigation is quite simple. The analysis program is easy to learn and user friendly.

CORRECTNESS - The application's functionality, internal calculations, and navigation should all be proper.

TESTABILITY – It is easy to test and find defects in the system. If required, the system will be easy to divide into smaller modules for the purpose of testing.

FLEXIBILITY – The regression model is flexible enough to modify.

REUSEABILITY – The machine learning model can be reused, if required by changing some of the values and collecting appropriate data.

INTEROPERABILITY - It should be simple for a product to communicate data or services with other systems. Different system modules should be compatible with various operating systems (MacOS, Windows, Linux) & IDEs..

## APPENDIX A: GLOSSARY

SRS: Software Requirement Specification IDE: Integrated Development Environment GUI: Graphical User Interface

CSV: Comma Separated Values JVM: JAVA Virtual Machine