**Introduction to Data Mining**

**Project Title**: Online Retail Segmentation.

**Introduction:**

MySQL is one of the most widely used database management systems due to its speed, reliability, and ease of use. It powers large web platforms like Facebook, Twitter, and YouTube, as well as countless smaller websites and applications.

In summary, MySQL is a powerful, flexible, and popular database management system that is commonly used for managing structured data across various applications.

**Overview:**

Customer segmentation is a typical strategy used by organizations to categories clients based on their demographics, shopping patterns, or other traits.

Customer segmentation constitutes a widely adopted strategic approach within organizations, aimed at systematically categorizing clients by analyzing a spectrum of factors such as demographics, shopping patterns, and other discernible traits. By effectively segmenting the customer base, businesses gain the ability to tailor their offerings, communications, and marketing efforts with a higher degree of precision, ultimately leading to enhanced customer satisfaction, targeted engagement, and improved overall business performance.

**Meta-Data:**

 Look for a dataset that contains information about customers such as demographic information, purchasing history, and customer interactions. The data set contains the following variables:

Invoice No: The invoice number for each transaction

Stock Code: The unique code for each product sold

Description: The description of each product sold

Quantity: The quantity of each product sold in each transaction

Invoice Date: The date and time of each transaction

Unit Price: The price of each product sold

Customer ID: The unique identifier for each customer

Country: The country where each transaction occurred.

To do this task given named “Online Retail” in MySQL software having the queries given below.

**Beginner Queries.**

* Define meta data in mysql workbench or any other SQL tool
* What is the distribution of order values across all customers in the dataset?
* How many unique products has each customer purchased?
* Which customers have only made a single purchase from the company?
* Which products are most commonly purchased together by customers in the dataset?

**Advance Queries**

1. **Customer Segmentation by Purchase Frequency**

Group customers into segments based on their purchase frequency, such as high, medium, and low frequency customers. This can help you identify your most loyal customers and those who need more attention.

1. **Average Order Value by Country**

Calculate the average order value for each country to identify where your most valuable customers are located.

1. **Customer Churn Analysis**

  Identify customers who haven't made a purchase in a specific period (e.g., last 6 months) to assess churn.

1. **Product Affinity Analysis**

Determine which products are often purchased together by calculating the correlation between product purchases.

1. **Time-based Analysis**

  Explore trends in customer behavior over time, such as monthly or quarterly sales patterns.

**Main Body:**

To do the queries in MySQL one by one with attached screenshot of code as well as the result given in same screenshot are given below.

**Database:**

The main function of a MySQL database is to provide a structured environment for storing, managing, and retrieving data efficiently. It is used for managing relational data and supporting operations required in web applications, enterprise systems, and other data-driven solutions.

**Use Database:**

Use command is a command that sets the working database for the current MySQL session, allowing you to execute queries without needing to specify the database name every time.

**Uploading Dataset:**

After that uploading the data set from downloaded file and then creating table automatically in MySQL.

Beginners Queries:

First query is actually uploading the data set in workbench of MySQL by uploading the data set at workbench and view the dataset by writing the code given below

Select \* from online\_retail;

For the 1st query the code written is:

Create database retailshop;

Use retailshop;

The result for the code of 1st query is given below:

For the second query the code written is:

SELECT CustomerID,

SUM(Quantity) AS TotalOrderValue

FROM online\_retail;

The result for the code of 2nd query is given below:

For the third query the code written is:

SELECT customerid, COUNT(DISTINCT stockcode) AS unique\_products

FROM online\_retail

GROUP BY customerid;

The result for the code of 3rd query is given below:

For the Fourth query the code written is:

SELECT CustomerID,

SUM(Quantity) AS TotalOrderValue

FROM online\_retail

group by customerid

having TotalOrderValue=1;

The result for the code of 4th query is given below:

For the fifth query the code written is:

SELECT stockcode

FROM online\_retail

group by stockcode

having count(quantity)=max(quantity);

The result for the code of 5th query is given below:

For the 1st Advance query the code written is:

SELECT customerid, count(\*) as ordervalue,

case

when count(\*) >10 then "high frequency"

when count(\*) between 5 and 10 then "medium frequency"

else "low frequency"

end as frequency\_category

FROM online\_retail

group by customerid;

The result for the code of 1st query of advance queries given below:

For the 2nd Advance query the code written is:

SELECT country, AVG(quantity) as average\_sales

FROM online\_retail

GROUP BY country;

The result for the code of 2nd query of advance is given below:

For the third Advance query the code written is:

select \*from online\_retail

where invoicedate != date\_add(curdate(), interval 6 month);

The result for the code of 3rd query of advance is given below:

**Conclusion**:

MySQL is a powerful, flexible and popular database management system that is commonly used for managing structured data across various applications. Data mining is widely used in fields like marketing, finance, healthcare, and business for tasks such as customer segmentation, fraud detection, and predicting trends. MySQL software using OLAP (Online Analytical Processing) for analyzing and visualizing data of given data set as project given. The dataset named “Online\_retail.csv”, by using this dataset we perform the beginner queries as well as advance queries step by step given in main body of report along with the screenshots of result.