

Madura Coats G-4

Name: Prem Nitin P

RegNo: 19S027

Exercise 1:

2) Criteria for Cluster Nomenclature:

The clusters can definitely be segmented into

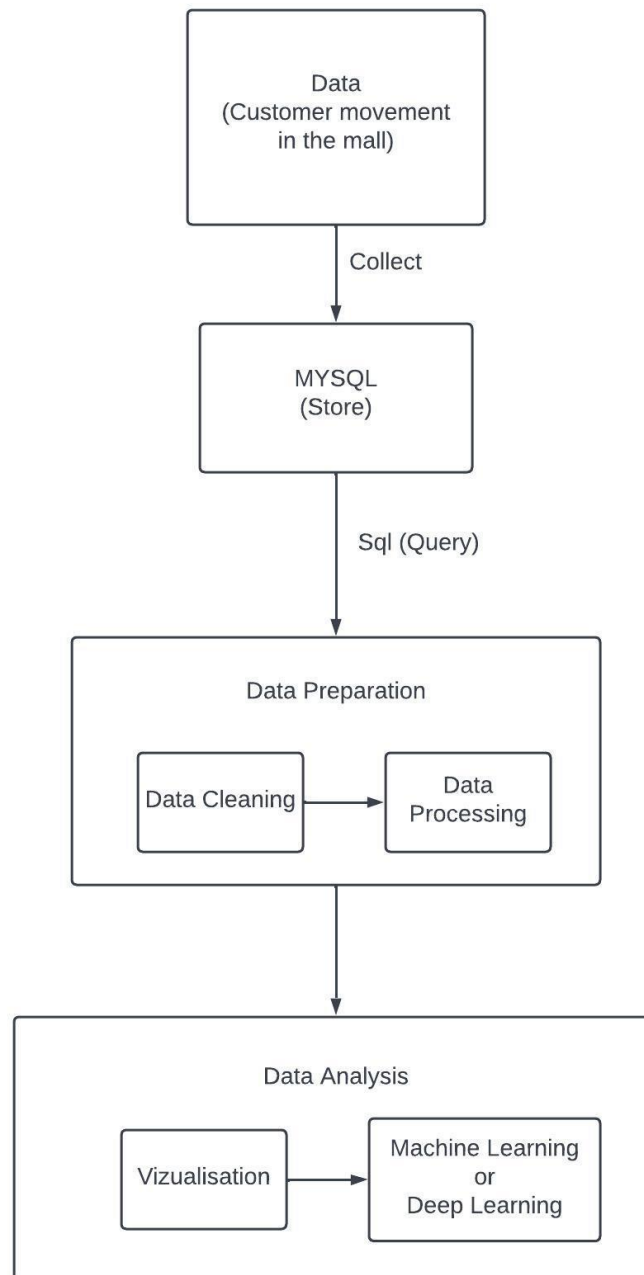
- High spenders (Cluster 0)
 - Low spenders (Cluster 1)
 - Medium spenders (Cluster 2)
-
- Low Spenders are old people and they have a moderate income
 - Medium Spenders are youngsters and they have a low income
 - High Spenders are Middle aged people and they have high income

The above statements are validated by finding the mean for the each of the clusters with respect to each feature and the values are visualized thus, providing the insights

So, spending score is the criteria that is used to name the clusters which is found through the large deviation between means of this cluster for this feature

Exercise 4:

1) Pipeline Architecture:



- First the location data is collected using the smart phones of customers
- The collected data is stored in a database (Mysql)
- To fetch the collected data sql queries is used
- Before Analysing the stored data, it is cleaned and processed. The data is fetched using sql and cleaned using python.
- The data processing is done using sql to fetch required data from the database

- The data is visualized to gain insights and required machine learning or deep learning model is applied for predictions.

2) Components and its purpose:

- **Data Collection:**
It is used to collect the data of customers which can be analysed to provide offers thus increasing the sales.
- **Data Storage:**
This component is used to store the collected data which can be used for further analysis
- **Data Preparation:**
In this module, the data is retrieved from the database using “sql” and cleaned and transformed to make it suitable for analysis
- **Data Analysis:**
In this module, the prepared data is used to make predictions and recommendations using machine learning and deep learning. Visualizations are done to gain insight

3) Data Flow:

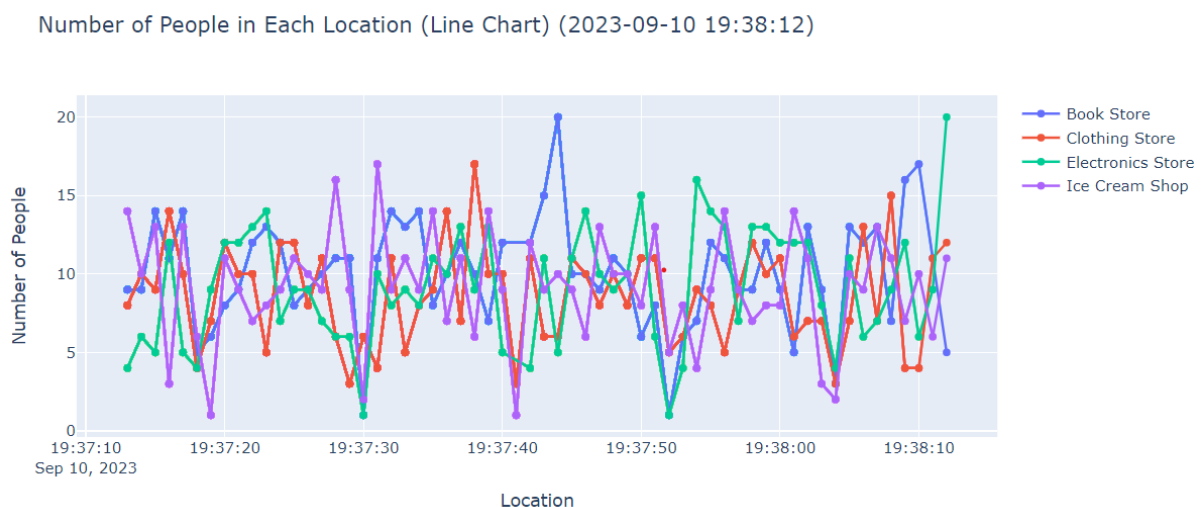
- The data is collected from the using the simulation. One simulation is done to generate the customer details and the details are stored in MySQL.
- Another simulation is used to simulate the movement of customers in the mall and these data are also stored in the database.
- The data from the database is fetched using python and cleaned (Removing duplicated, removing empty cells, Checking Integrity)
- The cleaned data is then transformed (Scaling the numerical features, encoding the categorical features, extracting year, month from timestamp)
- Required data of movement of customers are extracted using specific queries to place offers.

4) Business Logic Used:

- Location based offers: When the customers are near to the pizzario multiple offers can be provided
- Offers with respect to their age (youngsters, Middle Age, Old aged)

- Frequency of visits of the user: Can provide reward points which can be used for special offers
- Provide offers regarding the local events that are happening if any
- Offers with respect to the time of the day
- Offers considering the weather (hot or cold weather)
- Seasonal Offers (festivals)

Data Collect time series chart:



5) Queries Used:

- **Average Age by Location using window functions**

```
select distinct(ld.location), avg(cd.age) over (partition BY ld.location) AS avg_age
from customer_details cd
inner join location_details ld ON cd.customerid = ld.customerid;
```

- **Customers who are in either Electronics Store or Book Store**

```
select cd.customerid, cd.customername, cd.age
from customer_details cd
inner join location_details ld on cd.customerid = ld.customerid
where ld.location = "Electronics Store" or ld.location = "Book Store"
```

- **Number of customers in each location between 3 pm to 3:30 pm**

```
select ld.location, count(distinct(cd.customerid)) as count
from customer_details cd
```

inner join location_details ld on cd.customerid = ld.customerid

where ld.timestamp_ >= "2023-09-09 15:00:00" and

ld.timestamp_ <= "2023-09-09 15:30:00"

group by ld.location

- **Phonenumber and name of customers of age less than or equal to 26**

select cd.phonenumber, cd.customername

from customer_details cd

inner join location_details ld on cd.customerid = ld.customerid

where cd.age <= 26

Demo Link :

https://drive.google.com/file/d/1_y7WhsmDUHF8Ev4CUiVqLtTiMZM50wm/view?usp=sharing