# EDA ON SUPERMARKET DATASET USING SOL

**STEP 1:** Create database in MYSQL
**STEP 2:** Use the database:
**STEP 3:** Import the dataset
**STEP 4:** Analyze the dataset STEP 5: Study the Dataset

**Fields of Dataset**

**Invoice ID:** Computer generated sales slip-Invoice identification number

**Branch:** Branch of superCentre 3 branches are available identified by A. B, C.

**City:**Location of the superCenter

**Customer Type:** Type of customer member the customers using membership card normal the customer without member Card

**Gender:** Gender type of customer

**Product line:** General line categorization accessories groups- electronic accessories, Health and beauty, home & lifestyle, fashion accessories, sports & travel, food & beverages.
Unit Price: Price of Product in S

**Quantity**: number of product purchased by customer

**Tax 5%:** tax foes for customer buying

**Total:** Total price including taxes

**Date:** date of purchase(record available from jan 2019 to mar 2019)

**Time:** time of purchasing

**Payment:** Payment used by customer for purchase-3 Payments methods available Ewallet Cash, Credit Card.

**Cogs:** is implies that cost of goods sold

**gross margin percentage:**gross margin percentage

**Gross income:** gross income

**Rating:-** customer satisfaction rating on their overall shopping experience on a scale 1 to 10

**STEP 6:** solve some queries related to dataset
**1. Display the first 5 rows from the dataset.**

CMD: Select * from super limit 5
Output-The output will be the first 5 rows with all the columns.

**2. Display the last 5 rows from the dataset.**

CMD: Select * from super order by Invoice id desc limit 5;
Output: The output will be the last 5 rows with all the columns in the dataset

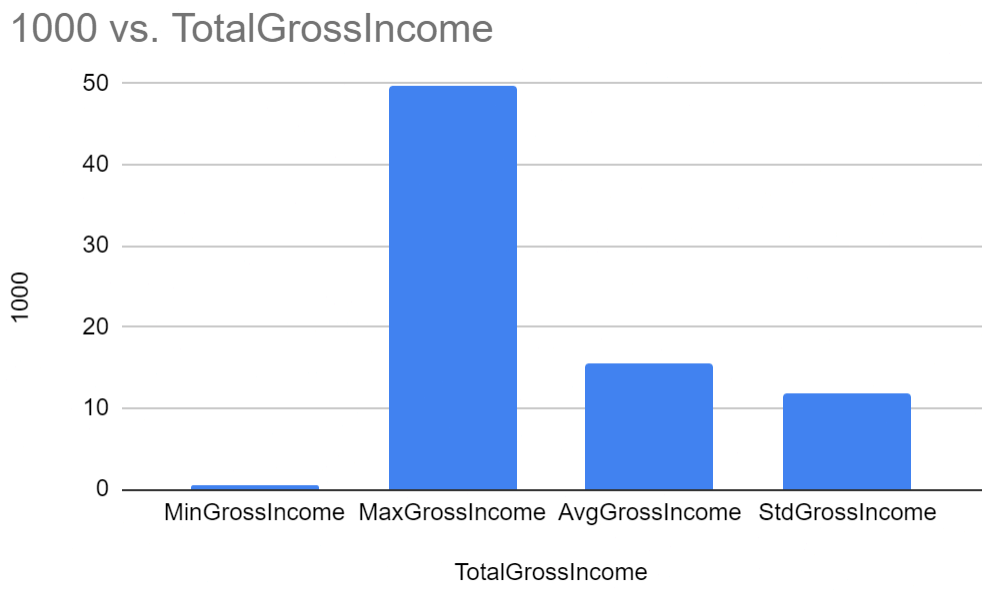
**3. Display random 5 rows from the dataset.**

CMD: Select * from super order by rand() limit 5;
Output The output will be the Random 5 rows with all the columns in the dataset

**4. Display count, min, max, avg and std values for a column(gross income) in the dataset.**

CMD: select count(`gross income`)as TotalGrossIncome,min(`gross income`)as MinGrossIncome,max(`gross income`)as MaxGrossIncome,avg(`gross income`)as AvgGrossIncome ,std(`gross income`)as StdGrossIncome from supermarket ;
Output: Here the min, max std & count values will be printed of gross income column

**Visualization:**



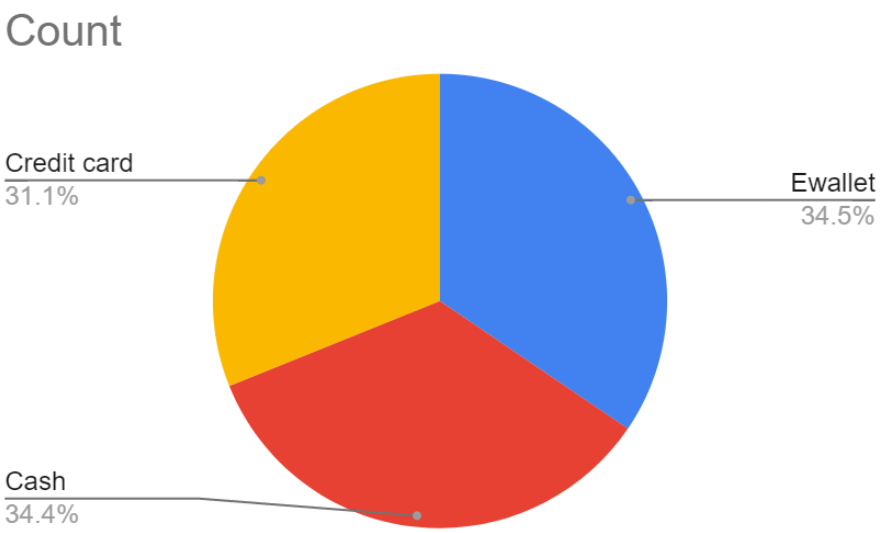
**5. Find the number of missing Values.**

CMD: select count(*) from super where branch is null;
Output: There are no missing values in the dataset. Our dataset is clean

**6.Count the number of occurrences of each city.**

CMD: select city, count(city) from super group by city;
Output: It gives the occurrence of each cities, as how many times a particular city has Occurred

**Visualization:**



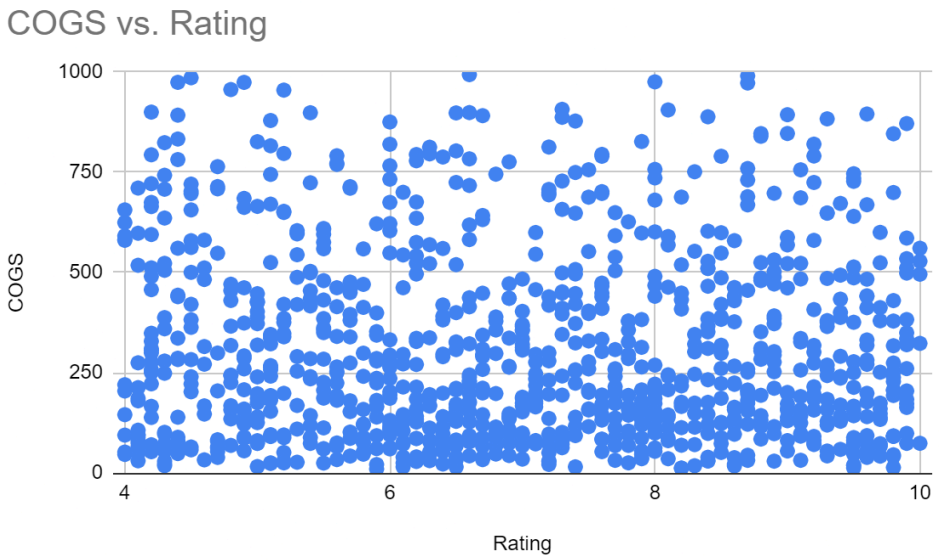

**7. Find the most frequently used payment method.**

CMD: select payment, count(*) from supermarket group by payment order by count('') desc,
Output it shows which payment method has been used most

**8. Does The Cost of Goods Sold Affect The Ratings That The Customers Provide?**

CMD: select rating.cogs from supermarket;
Output the rating does not affect the cost of good sold

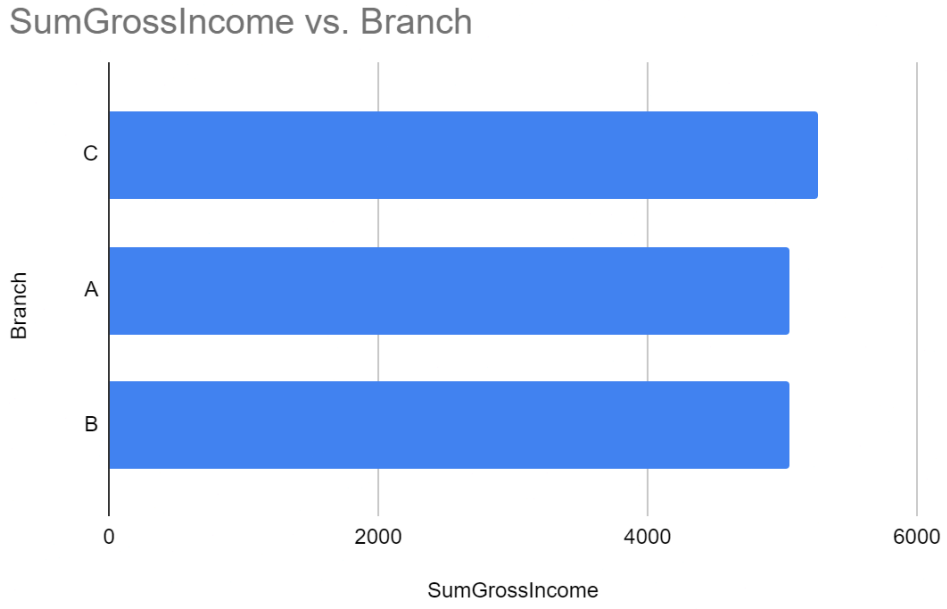
**Visualization:**




**9. Find the most profitable branch as per gross income.**

CMD: select branch, round(sum(gross income),2) as sum gross income from supermarket group by branch order by sum gross income desc;
Output shows the most profitable branch as per gross income

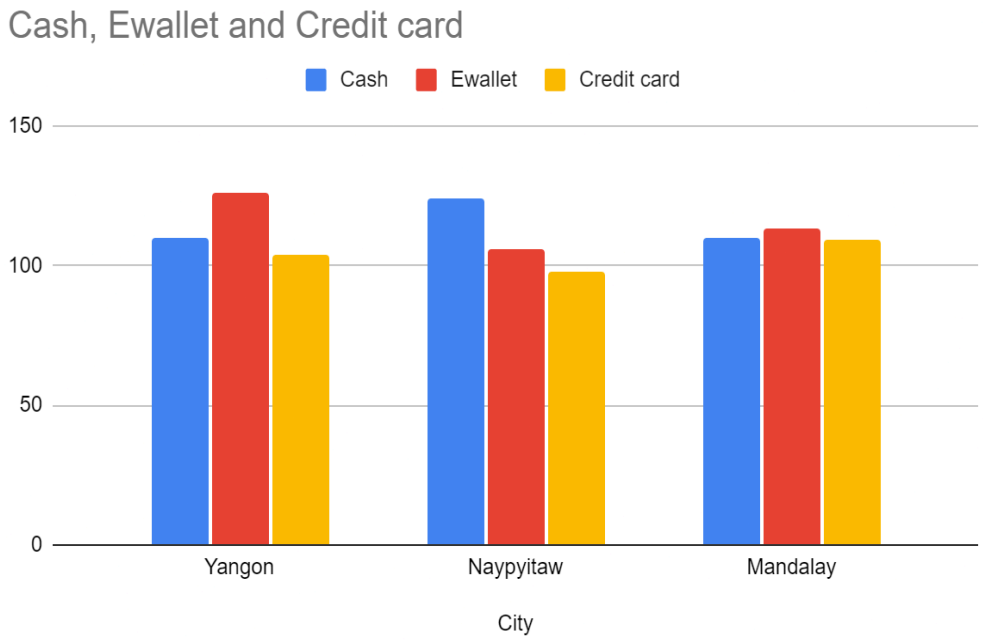
**Visualization:**



**10. Find the most used payment method city-wise.**

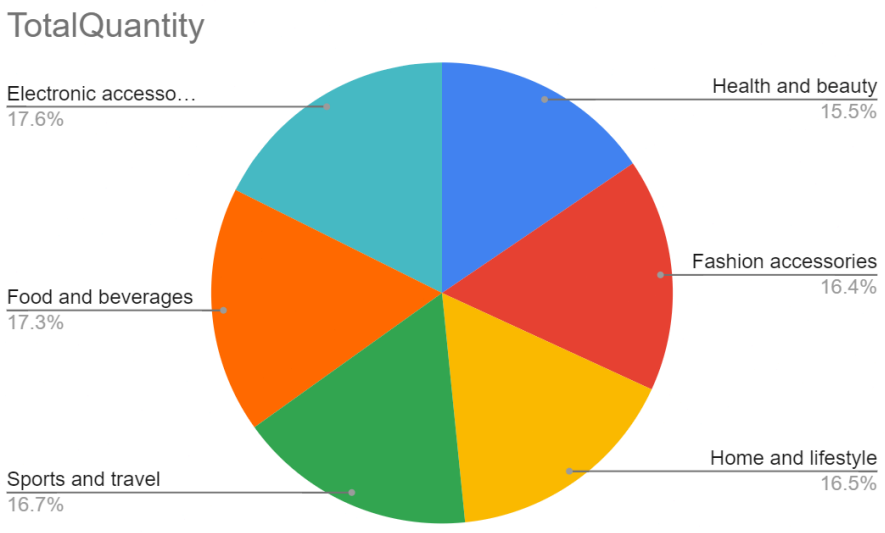
CMD:select City,
sum(case when Payment= 'Cash' then 1 else 0 end) as 'Cash',
sum(case when Payment=  'Ewallet' then 1 else 0 end) as 'Ewallet',
sum(case when Payment= 'Credit card' then 1 else 0 end) as 'Credit card'
from supermarket group by City;
Output: It shows the most used payment method city-wese

**Visualization:**




**11. Find the product line purchased in the highest quantity.**

CMD:select `Product line`,sum(Quantity) as TotalQuantity from supermarket group by `Product line`order by TotalQuantity;
Output:ll show which type of product purchased in highest quantity
**Visualization:**



**13. Find the month with the highest sales.**

CMD:select monthname(Date) as name, month(date) as month, round(sum(total).2) as total from supermarket group by name.month order by total desc:
Output: it shows the month with the highest sales

**14. Find the time at which sales are highest.**

CMD:select hour(time) as hour, round(sum (total),2) as total from supermarket group by hour order by total desc;
Output Shows the time of highest sales

**Visualization:**