

Creating Regression and Clustering Models

Presented by:
Shania Widianingrum Puspitasari



Background Story

The inventory team gave the task of **predicting daily sales quantity** of Kalbe Nutritionals products. The aim of this project is to help the inventory team plan sufficient and adequate daily stock inventory.

The Marketing Team is tasked with creating **customer segmentation** (clusters). The goal of this project is to create groups of customers who share similar characteristics. This customer segmentation will be used by the Marketing Team to carry out customized promotions and sales each segment



Company Overview



447 Customers

The company have 447 customers over 2022



14 Stores

The company have 14 stores over 2022



40 years old

The average customer is 40 years old



\$162,043,000

Total income company in 2022



5,020 transaction

Transactions in companies in 2022



10,057 female and 8,239 male

Total customer male and female



Case Study

1

The average age with **Married** status is **43 years old** and **Single** is **29 years old**

```
SELECT p."product_name", sum (t.totalamount) as sum_amount
FROM product as p
JOIN transaction as t
ON p.productid = t.productid
GROUP BY p."product_name"
ORDER BY sum_amount DESC
LIMIT 1;
```

	product_name	sum_amount
1	Cheese Stick	27615000

2

The average age of **women** is **40 years old** and **men** is **39 years old**

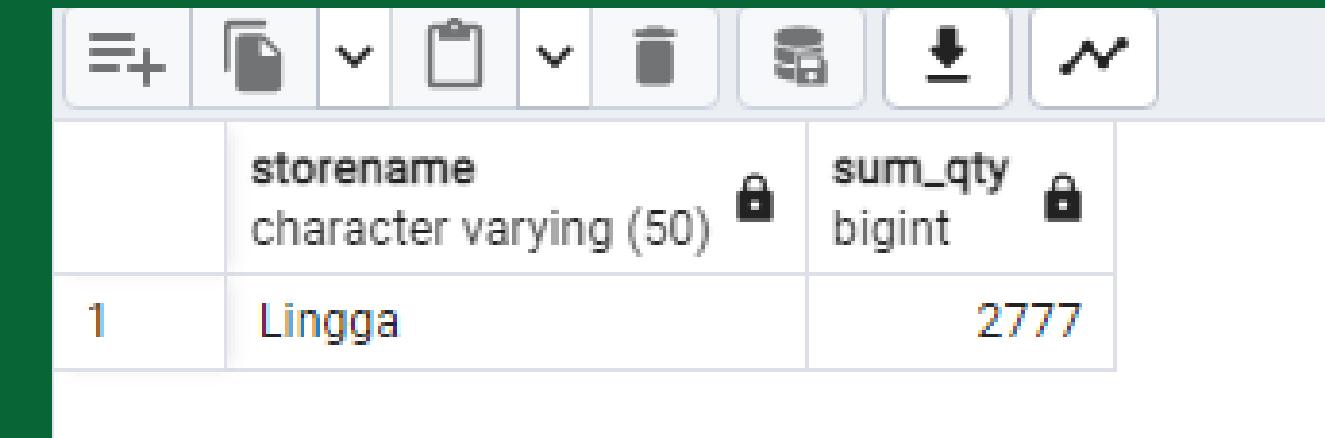
```
4 SELECT "gender", avg(age) from customer
5 group by "gender";
6
7
```

	gender	avg
1	0	40.3264462809917355
2	1	39.1414634146341463

Case Study

3

```
9 |SELECT s.storename, sum(t.qty) as sum_qty
10| FROM store as s
11| JOIN transaction as t
12| ON s.storeid = t.storeid
13| GROUP BY s.storename
14| ORDER BY sum_qty desc
15| LIMIT 1;
```

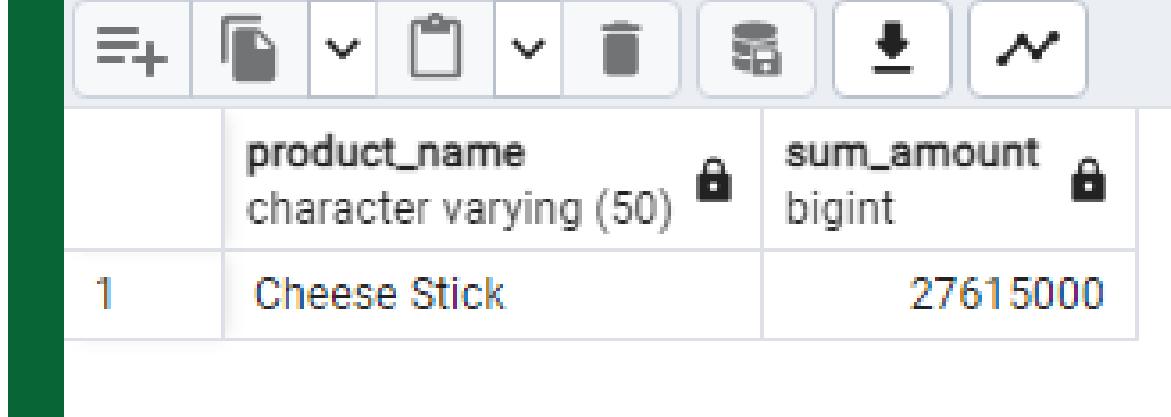


A screenshot of a database management system interface showing a results grid. The grid has two columns: 'storename' and 'sum_qty'. There is one row of data with values 'Lingga' and '2777'. The interface includes a toolbar with various icons above the grid.

	storename	sum_qty
1	Lingga	2777

4

```
SELECT p."product_name", sum (t.totalamount) as sum_amount
FROM product as p
JOIN transaction as t
ON p.productid = t.productid
GROUP BY p."product_name"
ORDER BY sum_amount DESC
LIMIT 1;
```

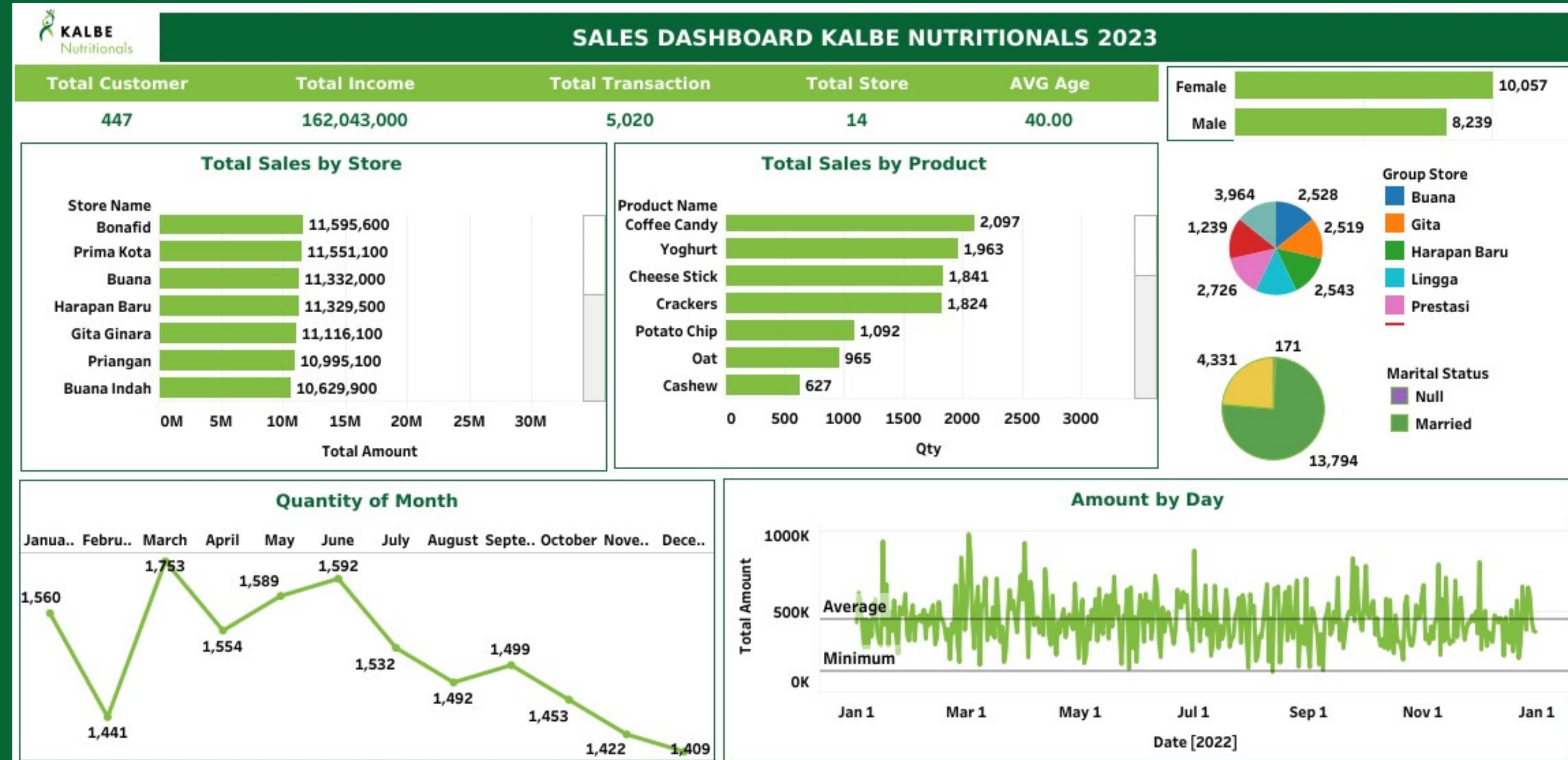


A screenshot of a database management system interface showing a results grid. The grid has two columns: 'product_name' and 'sum_amount'. There is one row of data with values 'Cheese Stick' and '27615000'. The interface includes a toolbar with various icons above the grid.

	product_name	sum_amount
1	Cheese Stick	27615000

Sales Dashboard

Link Dashboard: bit.ly/DS_VIX_KalbeNutritionals_2023



Clustering

	CustomerID	TransactionID	Qty	TotalAmount
cluster_label				
0	165	11.066667	41.551515	346976.363636
1	194	11.587629	40.613402	409496.391753
2	81	10.777778	40.395062	287344.444444
3	4	7.250000	29.500000	150100.000000

- **Cluster 1:** largest number of customers. Recommendation: build **good relationships** with customers and conduct **surveys** to develop customer interest.
- **Cluster 0:** second highest number. Recommendation: provide **regular promotions** to increase transactions and **upsell** products at high prices.
- **Cluster 2:** small number of customers. Recommendations: provide **significant discounts** to increase customer transactions, offer **promotions** on higher quantity transactions, and conduct **surveys** to identify potential product development.
- **Cluster 3:** lowest number of customers. Recommendations offer **loyalty promotions** to maintain transactions, conduct customer **satisfaction surveys**, and **upsell** products at higher prices.

THANK YOU



[Project link on Github](#)

