Rakamin Academy X KALBE Nutritionals

ML Project: Regression and Clustering

Rakamin Academy x Kalbe Nutritionals

Created by:
Yashinta Nilam Sari
On September 2023

Yashinta Nilam Sari

About

An enthusiastic undergraduate Statistics student at Universitas Padjadjaran who has strong interest in data science. Constantly look forward to learn new things in order to improve self-quality.

Rakamin Academy

Experiences

Machine Learning Cohorts
(Feb - Jul 2023)
Bangkit Academy by Google, GoTo, Traveloka

Head of Multimedia
(Feb - Dec 2022)
Forum Kajian Statistika Unpad

Staff of Finance Department
(Feb - Dec 2022)
BE Himpunan Mahasiswa Statistika Unpad

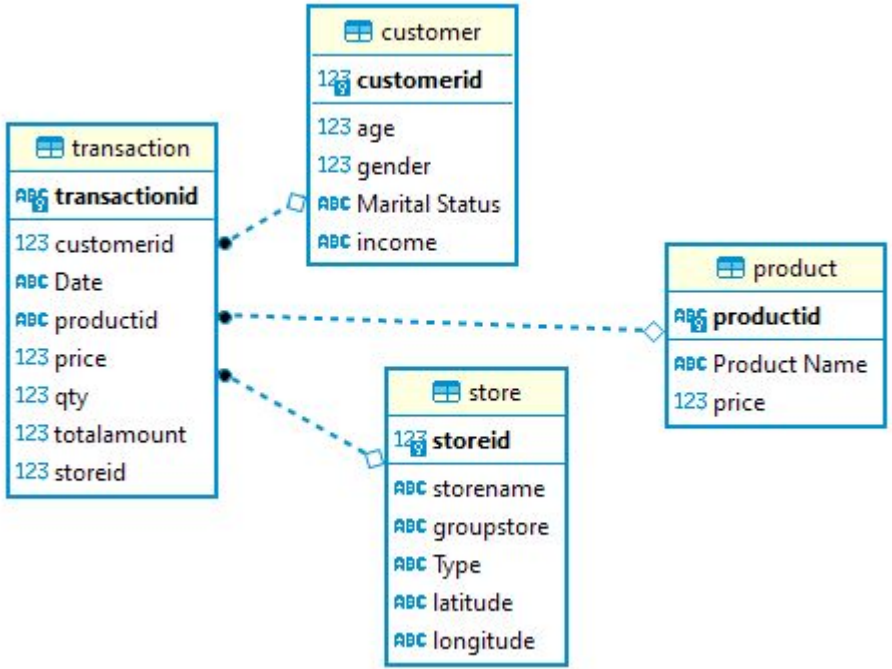
KALBE Nutritionals

CHALLENGE

1

Data Ingestion and EDA

Tools:


customer
customerid
age
gender
Marital Status
income
transaction
transactionid
customerid
Date
productid
price
qty
totalamount
storeid
product
productid
Product Name
price
store
storeid
storename
groupstore
Type
latitude
longitude

ER-Diagram

```sql
select
	"Marital Status",
	round(avg(age),2) as "Average Age"
from
	customer
group by
	"Marital Status";
```

Query 1: Berapa rata-rata umur *customer* jika dilihat dari *marital status*-nya?

	Marital Status	Average Age
1		31.33
2	Married	43.04
3	Single	29.38

```sql
select
	case gender
		when 0 then 'Wanita'
		when 1 then 'Pria'
	end "Gender",
	round(avg(age),2) as "Average Age"
from
	customer
group by
	"Gender";
```

Query 2: Berapa rata-rata umur *customer* jika dilihat dari *gender*-nya ?

	Gender	Average Age
1	Wanita	40.33
2	Pria	39.14

```sql
select
	s.storename,
	sum(t.qty) as "Total Quantity"
from
	"transaction" as t
inner join
	"store" as s
on
	t.storeid = s.storeid
group by
	s.storename
order by
	"Total Quantity" desc
limit 3;
```

Query 3: Tentukan nama *store* dengan *total quantity* terbanyak!

	ABC storename	123 Total Quantity
1	Lingga	2,777
2	Sinar Harapan	2,588
3	Prestasi Utama	1,395

```sql
select
	p."Product Name",
	sum(t.totalamount) as "Total Amount"
from
	"transaction" as t
inner join
	"product" as p
on
	t.productid = p.productid
group by
	p."Product Name"
order by
	"Total Amount" desc
limit 3;
```

Query 4: Tentukan nama produk terlaris dengan *total amount* terbanyak!

	ABC Product Name	123 Total Amount
1	Cheese Stick	27,615,000
2	Choco Bar	21,190,400
3	Coffee Candy	19,711,800

KALBE Nutritionals

CHALLENGE

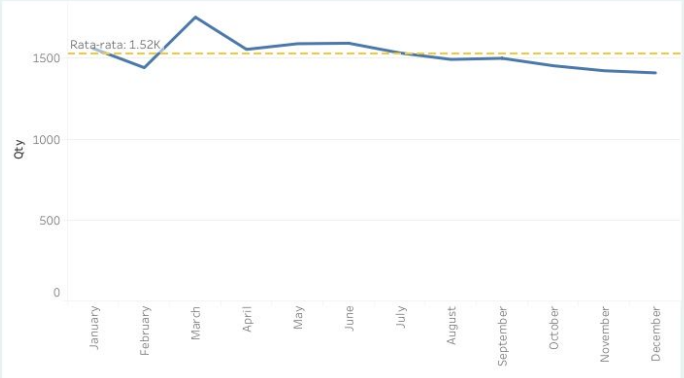
2

Data Visualization

Tools: tableau

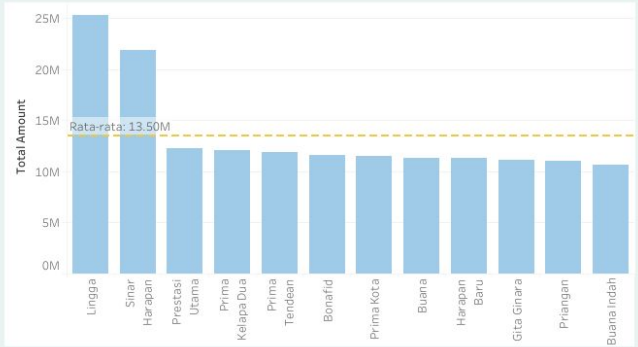
Total Produk Terjual Bulanan

Sejak bulan Agustus, total produk terjual menurun di bawa rata-rata tahunan.

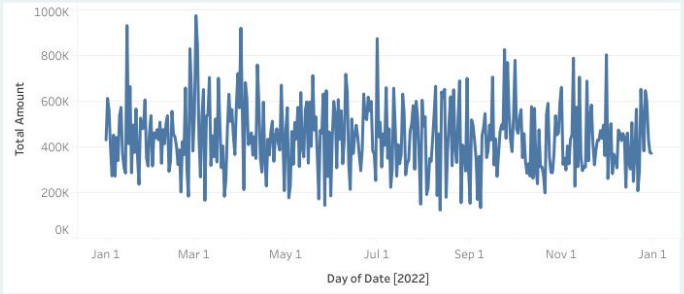


Total Penjualan Per Toko, 2022

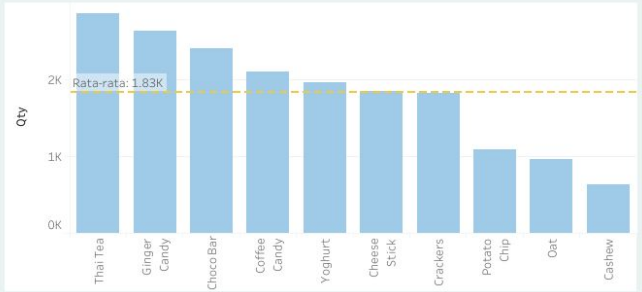
Toko Lingga dan Sinar Harapan memiliki total penjualan di atas rata-rata toko lain.



Total Penjualan Harian



Total Produk Terjual Per Produk, 2022



Dashboard Tableau: [Click Here]

KALBE Nutritionals

CHALLENGE 3

ML Regression

Tools:

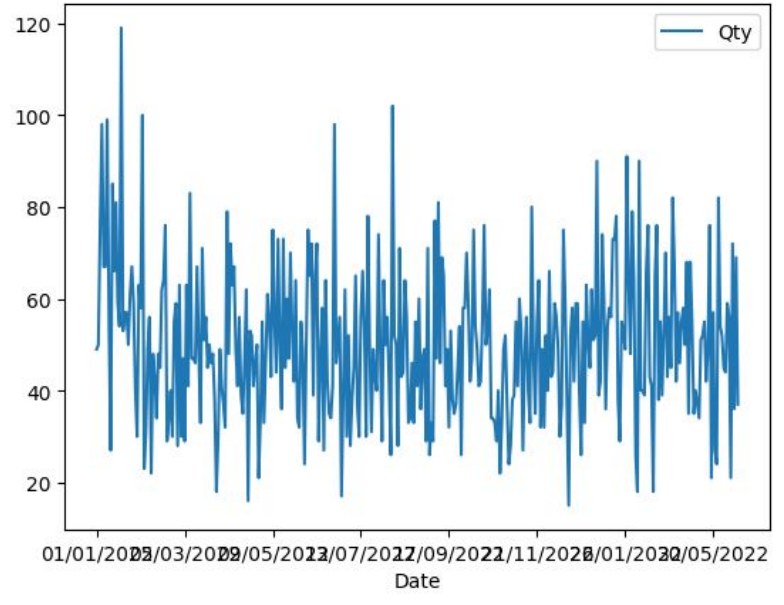
TensorFlow

Rakamin Academy

Problem Statements

Goal	Know the estimated quantity of products sold, so that the inventory team can make sufficient daily inventory stock.
Objective	Forecast 7-day-ahead total quantity of products sold using LSTM.
Dataset	Daily total quantity of products sold during 2022.

INPUT DATA



Visualization of daily total Quantity during 2022

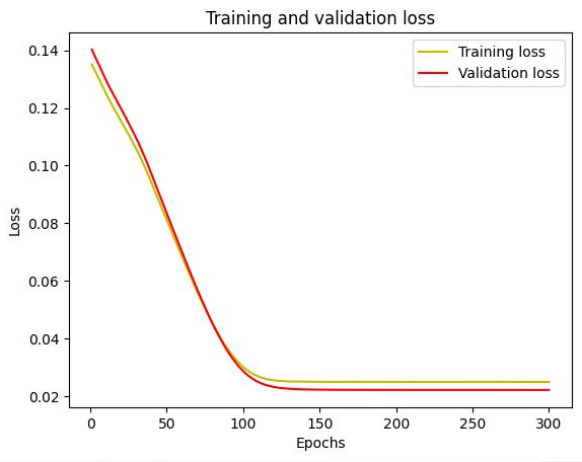
Dataset is splitted into training and validation set (85:15) and then normalized with Min Max Scaler.

MODELLING

Rakamin Academy

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 7, 128)            66560

 lstm_1 (LSTM)               (None, 64)                49408

 dense (Dense)               (None, 128)               8320

 dense_1 (Dense)             (None, 64)                8256

 dense_2 (Dense)             (None, 1)                 65

=================================================================
Total params: 132,609
Trainable params: 132,609
Non-trainable params: 0
_________________________________________________________________
```

Model architecture

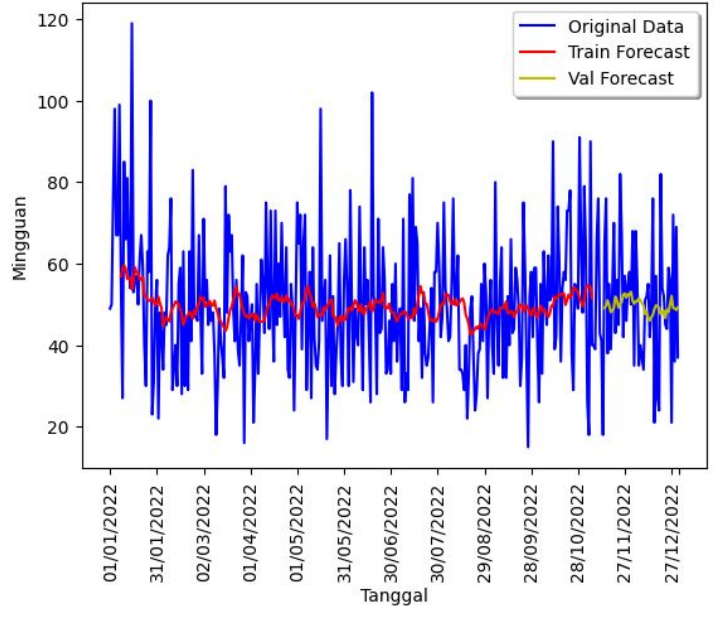


Training and validation loss

Training hyperparameters

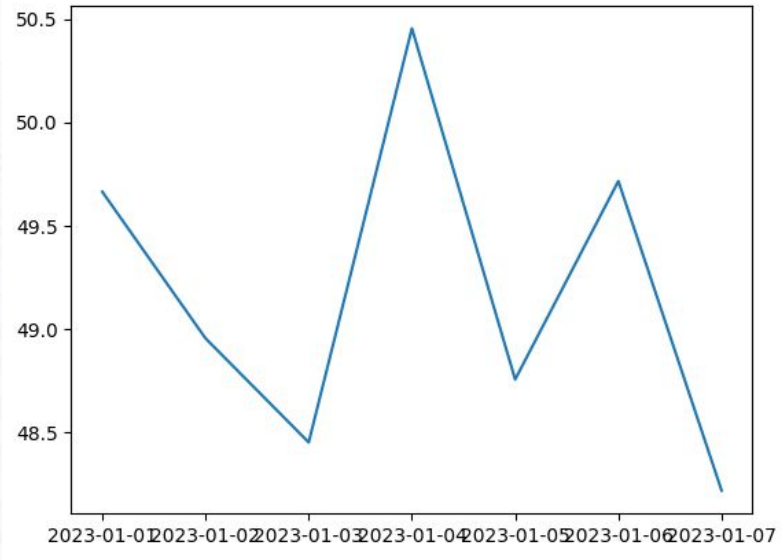
- Optimizer: Adam
- Learning rate: 1e-6
- Epoch: 300
- Batch size: 3

RESULTS



Actual vs Predicted Value

MAE for validation set: **12.56 units**



Forecasting 7-day ahead daily total quantity of products sold

KALBE Nutritionals

CHALLENGE 4

ML Clustering

Tools:

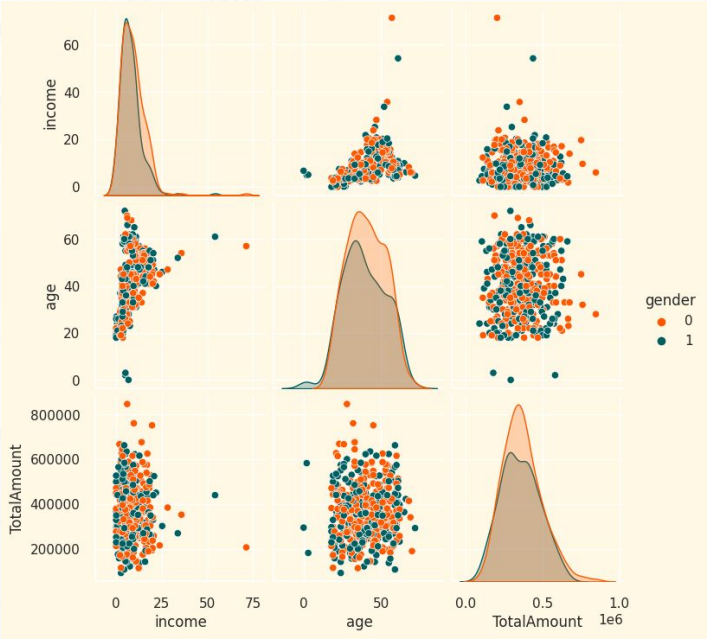
Rakamin Academy

Problem Statements

Goal	Create customer segmentation that will be used by the marketing team to provide personalized promotion and sales treatment.
Objective	Apply K-means clustering algorithm to create customers segmentation based on their similarities.
Dataset	Transaction data which is merged with customer and product data. (*447 rows, 1 unique identifier, 6 numerical columns, 2 categorical columns*)

DATA VISUALIZATION

Rakamin Academy

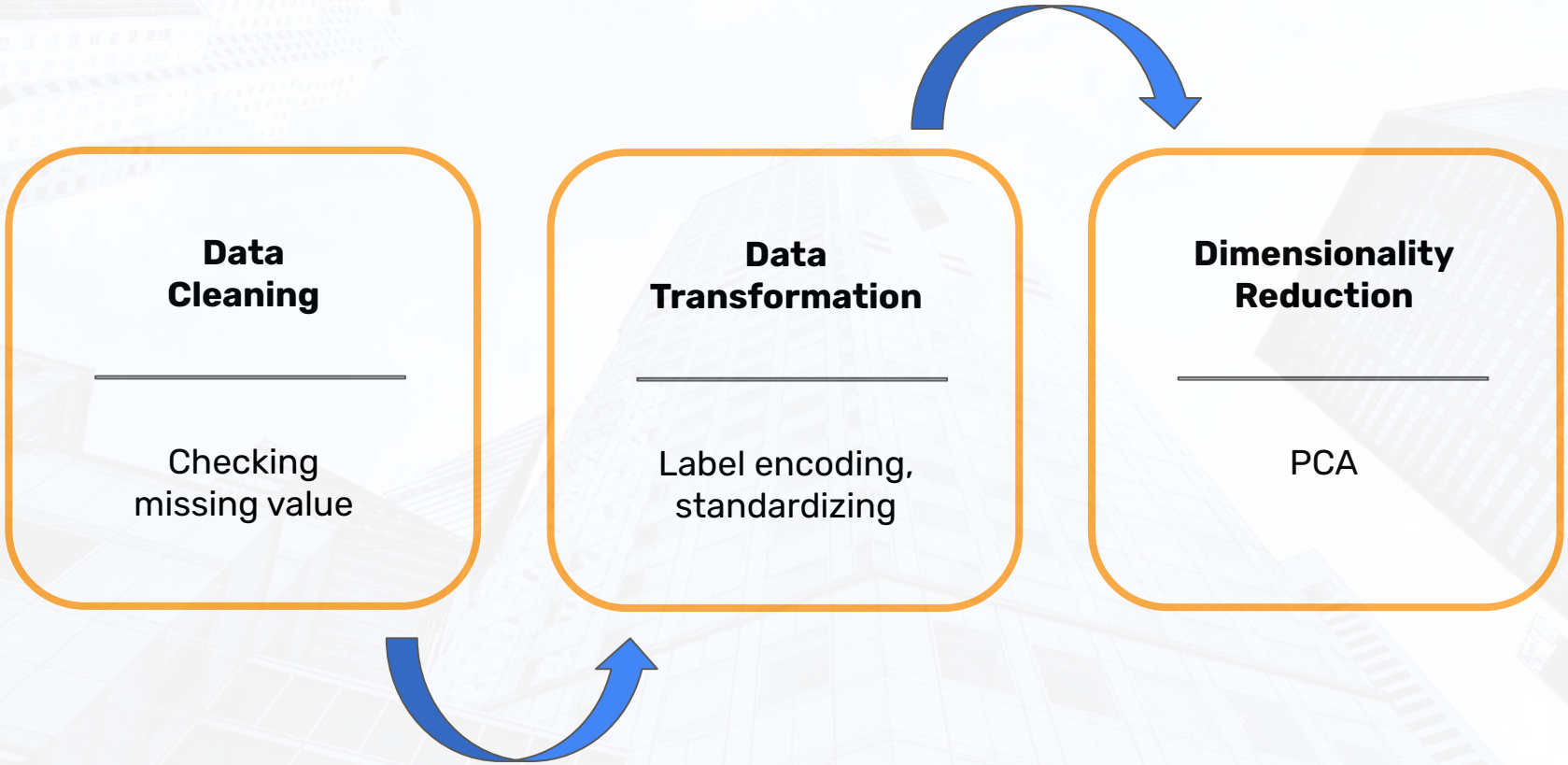


Distribution and scatter plots among several features

Correlation matrix

DATA PREPROCESSING

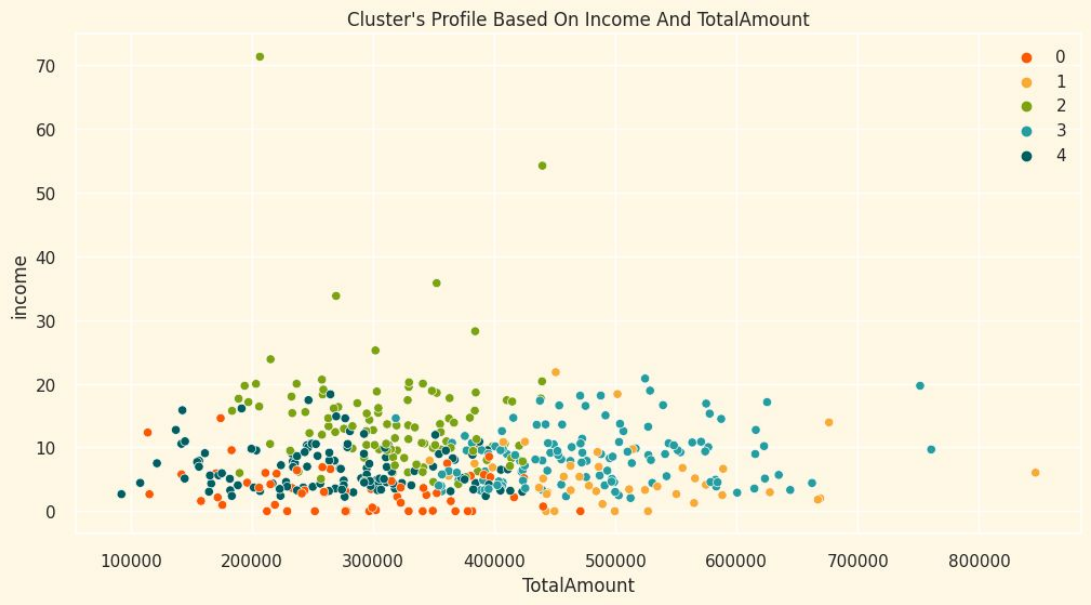
Rakamin Academy



K-MEANS CLUSTERING

The optimal number of clusters using Elbow Methods is **5 clusters** with WCSS value of 2363.679.

K-MEANS CLUSTERING



- Low income, low to medium spending (**cluster 0**)
- Low to medium income, low to medium spending (**cluster 4**)
- Medium to high income, medium spending (**cluster 2**)
- Low income, high spending (**cluster 1**)
- Medium income, high spending (**cluster 3**)

Rakamin Academy

CLUSTER 1 PROFILE

TRANSACTION CHARACTERISTICS

- **Number of transactions**: 15 (± 2) times
- **Total Quantity**: 55 (± 9) products
- **Total Amount**: Rp502,081.4 (± Rp96,591.8)
- **Favorite Product**: Cheese Stick (12/43), Thai Tea (11/43)

CUSTOMER CHARACTERISTICS

- **Income**: 5.72 (± 4.63) million IDR
- **Age**: 30.07 (± 9.87) years
- **Marital Status**: Single (40/43)
- **Gender**: 25 Female, 18 Male

Rakamin Academy

CLUSTER 3 PROFILE

TRANSACTION CHARACTERISTICS

- **Number of transactions**: 15 (± 3) times
- **Total Quantity**: 54 (± 9) products
- **Total Amount**: Rp480,642.62 (± Rp85,498.98)
- **Favorite Product**: Thai Tea (21/122), Cheese Stick (20/122)

CUSTOMER CHARACTERISTICS

- **Income**: 8.72 (± 4.36) million IDR
- **Age**: 42.20 (± 11.40) years
- **Marital Status**: Married (122/122)
- **Gender**: 69 Female, 53 Male

Rakamin Academy

References

- https://stackoverflow.com/questions/70420155/how-to-predict-actual-future-values-after-testing-the-trained-lstm-model/70421046#70421046
- https://www.kaggle.com/code/karnikakapoor/customer-segmentation-clustering
- https://www.mikulskibartosz.name/pca-how-to-choose-the-number-of-components/

GitHub Repository

https://github.com/yashintans/VIXKalbe-final-task

Thank You