

Final Project

Rakamin X Kalbe Nutritionals

Berapa rata-rata umur customer jika dilihat dari gender nya ?

```
SELECT `Marital Status`, AVG(`Age`) AS Age FROM customer GROUP BY `Marital Status`
```

Marital Status	Age
Married	43,0382
Single	29,3846

Berapa rata-rata umur customer jika dilihat dari gender nya ?

```
SELECT CASE WHEN `Gender`= 0 THEN 'Wanita' WHEN `Gender`= 1 THEN 'Pria' END AS `Gender`, AVG(`Age`) AS Age FROM customer GROUP BY `Gender`
```

Gender	Age
Pria	39,2574
Wanita	40,3264

Tentukan nama store dengan total quantity terbanyak !

```
SELECT a.`StoreName`, subquery.`Qty` AS 'Qty' FROM store a INNER JOIN ( SELECT b.`StoreID`, SUM(b.`Qty`) AS 'Qty' FROM transaction b GROUP BY b.`StoreID` ) AS subquery ON a.`StoreID` = subquery.`StoreID` WHERE subquery.`Qty` = ( SELECT MAX(`Qty`) FROM ( SELECT SUM(b.`Qty`) AS 'Qty' FROM transaction b GROUP BY b.`StoreID` ) AS sub )
```

StoreName	Qty
Lingga	1.439

Tentukan nama produk terlaris dengan total amount terbanyak !

```
SELECT a.`Product Name`, subquery.`TotalAmount` AS 'TotalAmount' FROM product a INNER JOIN ( SELECT b.`ProductID`, SUM(b.`TotalAmount`) AS 'TotalAmount' FROM transaction b GROUP BY b.`ProductID` ) AS subquery ON a.`ProductID` = subquery.`ProductID` WHERE subquery.`TotalAmount` = ( SELECT MAX(`TotalAmount`) FROM ( SELECT SUM(b.`TotalAmount`) AS 'TotalAmount' FROM transaction b GROUP BY b.`ProductID` ) AS sub )
```

Product Name	TotalAmount
Cheese Stick	27.615.000

Merge data yang akan digunakan

```
SELECT a.`Date`, a.`TransactionID`, a.`CustomerID`, a.`ProductID`, a.`StoreID`, a.`Qty`, a.`TotalAmount`, b.`StoreName`, b.`GroupStore`, b.`Type`, b.`Latitude`, b.`Longitude`, c.`Age`, c.`Gender`, c.`Marital Status`, c.`Income`, d.`Product Name`, d.`Price` FROM transaction a INNER JOIN store b ON a.StoreID = b.StoreID INNER JOIN customer c ON a.CustomerID = c.CustomerID INNER JOIN product d ON a.ProductID = d.ProductID ORDER BY 1, 2, 3, 4, 5
```

Date	TransactionID	CustomerID	ProductID	StoreID	Qty	TotalAmount	StoreName	Price
01/01/2022	TR11369	328	P3	12	4	30.000	Prestasi Utama	7.500
01/01/2022	TR16356	165	P9	1	7	70.000	Prima Tendean	10.000
01/01/2022	TR1984	183	P1	4	4	35.200	Gita Ginara	8.800
01/01/2022	TR35256	160	P1	4	7	61.600	Gita Ginara	8.800
01/01/2022	TR41231	386	P9	4	1	10.000	Gita Ginara	10.000
01/01/2022	TR51675	283	P10	5	1	15.000	Bonafid	15.000
01/01/2022	TR54287	51	P8	2	2	32.000	Prima Kelapa Dua	16.000
01/01/2022	TR67455	49	P5	13	3	12.600	Buana	4.200
01/01/2022	TR73041	222	P9	4	6	60.000	Gita Ginara	10.000
01/01/2022	TR7596	270	P7	14	2	18.800	Priangan	9.400
01/01/2022	TR76340	287	P4	12	4	48.000	Prestasi Utama	12.000
01/01/2022	TR97336	233	P7	12	2	18.800	Prestasi Utama	9.400
01/01/2022	TR99839	123	P2	1	6	19.200	Prima Tendean	3.200
01/02/2022	TR33585	245	P2	11	5	16.000	Sinar Harapan	3.200
01/02/2022	TR39283	293	P2	8	5	16.000	Sinar Harapan	3.200
01/02/2022	TR40082	293	P3	7	3	22.500	Buana Indah	7.500
01/02/2022	TR50178	280	P10	5	2	30.000	Bonafid	15.000
01/02/2022	TR51488	61	P1	10	9	79.200	Harapan Baru	8.800
01/02/2022	TR51730	3	P5	5	4	16.800	Bonafid	4.200
01/02/2022	TR52040	289	P5	10	4	16.800	Harapan Baru	4.200
01/02/2022	TR57306	16	P5	13	2	8.400	Buana	4.200
01/02/2022	TR96028	141	P7	13	6	56.400	Buana	9.400

Data terdiri 4976 baris dan 18 kolom

Memperbaiki tipe data yang tidak sesuai

Date	object	datetime64[ns]
TransactionID	object	object
CustomerID	int64	object
ProductID	object	object
StoreID	int64	object
Qty	int64	int64
TotalAmount	int64	int64
StoreName	object	object
GroupStore	object	object
Type	object	object
Latitude	object	float64
Longitude	object	float64
Age	int64	int64
Gender	int64	bool
Marital Status	object	category
Income	object	float64
Product Name	object	object
Price	int64	int64
dtype: object		dtype: object

Membuat data baru untuk Model Clustering

CustomerID	NumberOfTransaction	Qty	TotalAmount	Gender	Marital Status	Income	Age
1	17	60	623300	Laki-Laki	Married	5.120	55
2	13	57	392300	Laki-Laki	Married	6.230	60
3	15	56	446200	Laki-Laki	Married	9.170	32
4	10	46	302500	Laki-Laki	Married	4.870	31
5	7	27	268600	Laki-Laki	Married	3.570	58

Data terdiri 444 baris

Melakukan Clustering menggunakan KMeans Clustering dengan 5 kLuster berdasarkan Elbow Method

Cluster	NumberOfTransaction	Qty	TotalAmount	Income	Age	Gender	Marital Status
0	73	23.726	191141.096	8.724	37.507	Wanita	Married
1	38	62.816	611818.421	7.744	38.711	Wanita	Married
2	124	42.508	381419.355	8.330	41.323	Wanita	Married
3	127	34.843	288907.087	8.801	40.441	Wanita	Married
4	82	52.561	480232.927	9.050	39.268	Laki-Laki	Married

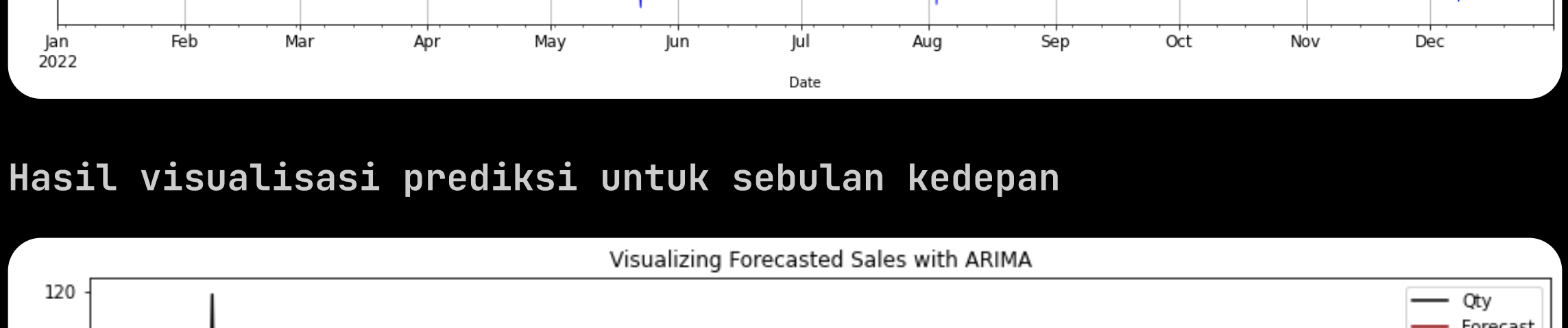
- **Cluster 0** ⇒ Customer yang berbelanja di toko dengan harga produk yang paling murah dengan jumlah belanja yang paling sedikit
- **Cluster 1** ⇒ Customer yang berbelanja di toko dengan harga produk yang paling mahal dengan jumlah belanja yang paling banyak
- **Cluster 2** ⇒ Customer yang berbelanja di toko dengan harga produk yang sedikit mahal dengan jumlah belanja yang lumayan banyak
- **Cluster 3** ⇒ Customer yang berbelanja di toko dengan harga produk yang lumayan murah dengan jumlah belanja yang lumayan sedikit
- **Cluster 4** ⇒ Customer yang berbelanja di toko dengan harga produk yang lumayan mahal dengan jumlah belanja yang banyak

Rata-rata customer adalah wanita dengan status sudah menikah yang berada pada rentang umur 37 hingga 42 tahun yang memiliki income 7 hingga 9

Membuat data baru untuk Model Time Series

Qty	Data terdiri 444 baris
Date	
2022-01-01	49
2022-01-02	50
2022-01-03	76
2022-01-04	98
2022-01-05	67

Persebaran data Qty berdasarkan harinya



Hasil visualisasi prediksi untuk sebulan kedepan



Melakukan forecasting menggunakan ARIMA dengan order (28,0,28)

	MSE	RMSE	MAE	R2	Execution Time
ARIMA	235.917	15.360	12.054	15.49 %	26.191