Rakamin Academy X KALBE Nutritionals

Final Project Data Modeling

Kalbe Nutritionals Data Scientist Virtual Internship Program

Presented by
Dinar Wahyu Rahman

Isi Konten

Dinar Wahyu Rahman

About You

I am a sixth-semester undergraduate student in Mathematics from the Universitas Terbuka and Political Science from UPN Veteran Jakarta. I am interested in coding and aspire to be a Machine Learning/AI engineer in the future.

From the understanding of programming language skills and these skills, I was able to produce several small and medium-scale projects related to data science, machine learning, AR applications, Game, and website programming. I like to learn new things and develop them.

Rakamin Academy

Experience

Machine Learning Student
Bangkit Academy led by Google, Tokopedia, Gojek, & Traveloka • Apprenticeship
Feb 2023 - Jul 2023 • 6 mos

Project-Based Intern : Big Data Analytics Virtual Internship Experience KimiaFarma
PT. KIMIA FARMA, TBK • Internship
Aug 2022 • 1 mo
Jakarta, Jakarta, Indonesia • Remote

Case Study

Rakamin Academy

Background Story

Kamu adalah seorang Data Scientist di Kalbe Nutritionals dan sedang mendapatkan project baru dari tim inventory dan tim marketing.

Dari tim inventory, kamu diminta untuk dapat membantu memprediksi jumlah penjualan (quantity) dari total keseluruhan product Kalbe

- Tujuan dari project ini adalah untuk mengetahui perkiraan quantity product yang terjual sehingga tim inventory dapat membuat stock persediaan harian yang cukup.
- Prediksi yang dilakukan harus harian.

Dari tim marketing kamu diminta untuk membuat cluster/segment customer berdasarkan beberapa kriteria.

- Tujuan dari project ini adalah untuk membuat segment customer.
- Segment customer ini nantinya akan digunakan oleh tim marketing untuk memberikan personalized promotion dan sales treatment.

Tools yang akan kamu gunakan dalam project ini adalah

- Python
- Jupyter Notebook
- Tableau
- Dbeaver
- PostgreSQL

Exploration with PostgreSQL and Dbeaver

Rakamin Academy

Average Customer Age based on Marital Status:

```sql
--Rata-rata umur customer berdasarkan Marital Staatus

SELECT "MaritalStatus", ROUND(AVG("Age"), 2) AS rata_rata_umur
FROM (
    SELECT c."Age", c."MaritalStatus"
    FROM public."transaction" AS t
    LEFT JOIN public.customers AS c ON t."CustomerID" = c."CustomerID"
) AS tb
GROUP BY "MaritalStatus";
```

customers 1

Enter a SQL expression to filter results (use Ctrl+Space)

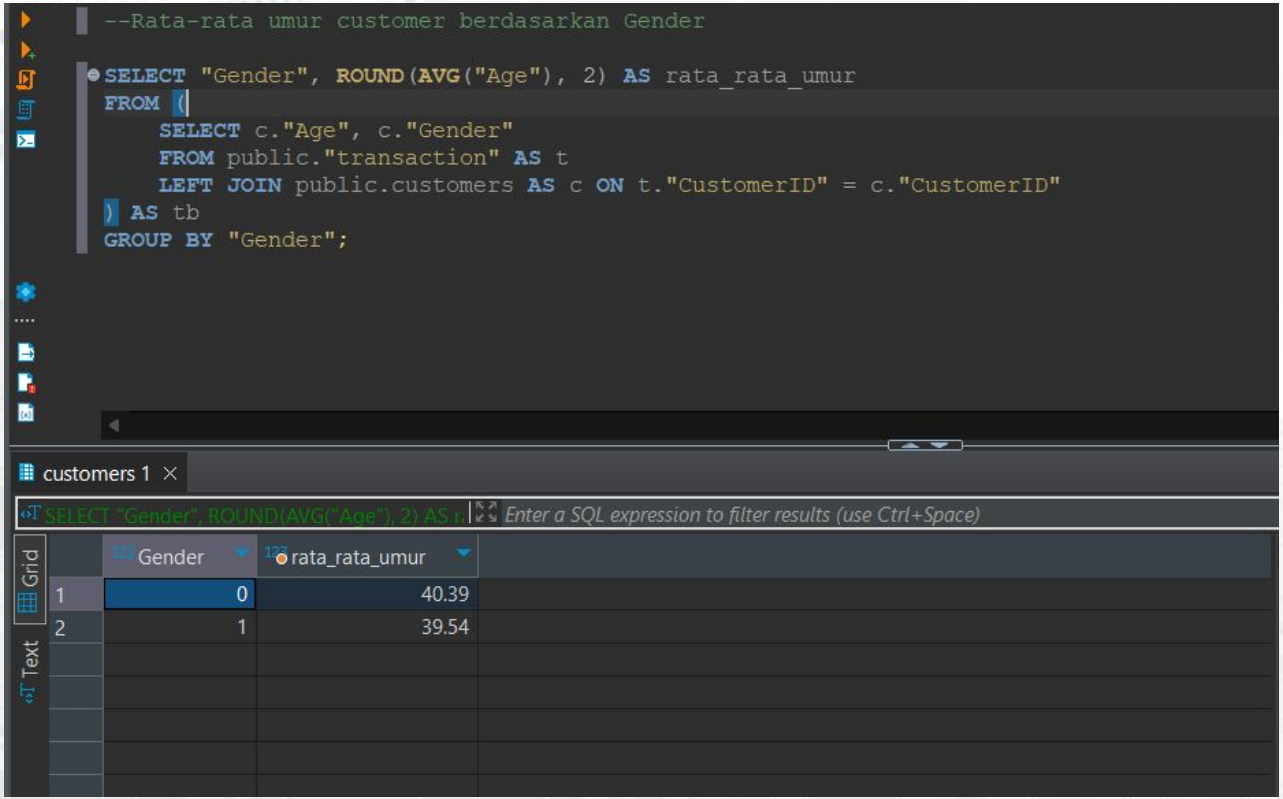
Grid | Text

	MaritalStatus	rata_rata_umur
1	Unknown	31.41
2	Married	43.37
3	Single	29.69

Exploration with PostgreSQL and Dbeaver

Rakamin Academy

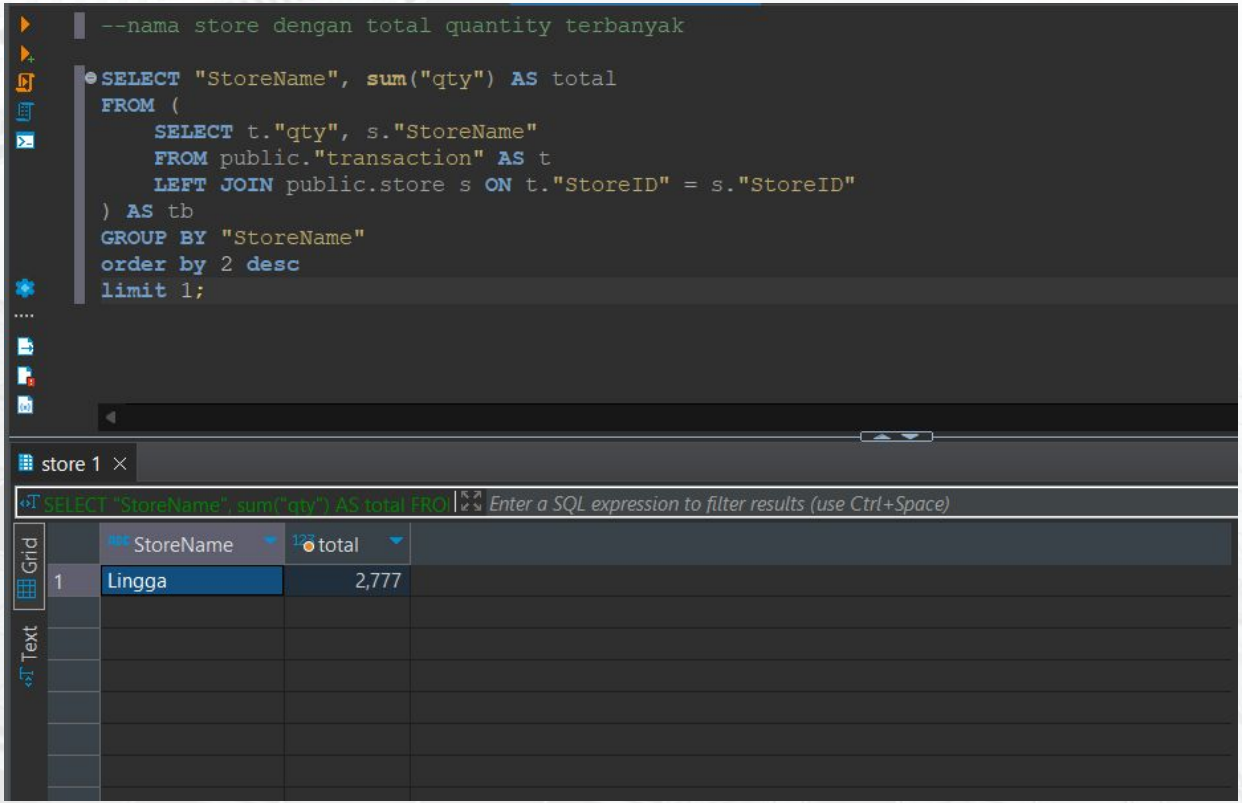
Average Customer Age based on Gender:



Exploration with PostgreSQL and Dbeaver

Rakamin Academy

Store with Highest Total Qty:



Exploration with PostgreSQL and Dbeaver

Rakamin Academy

Top Product by Total Amount:

```sql
-- nama produk terlaris dengan total amount terbanyak

select p."ProductName", sum(t."TotalAmount") as total
from product p
inner join "transaction" t on p."ProductID" = t."ProductID"
group by p."ProductName"
order by total desc
limit 1;
```

product 1

Enter a SQL expression to filter results (use Ctrl+Space)

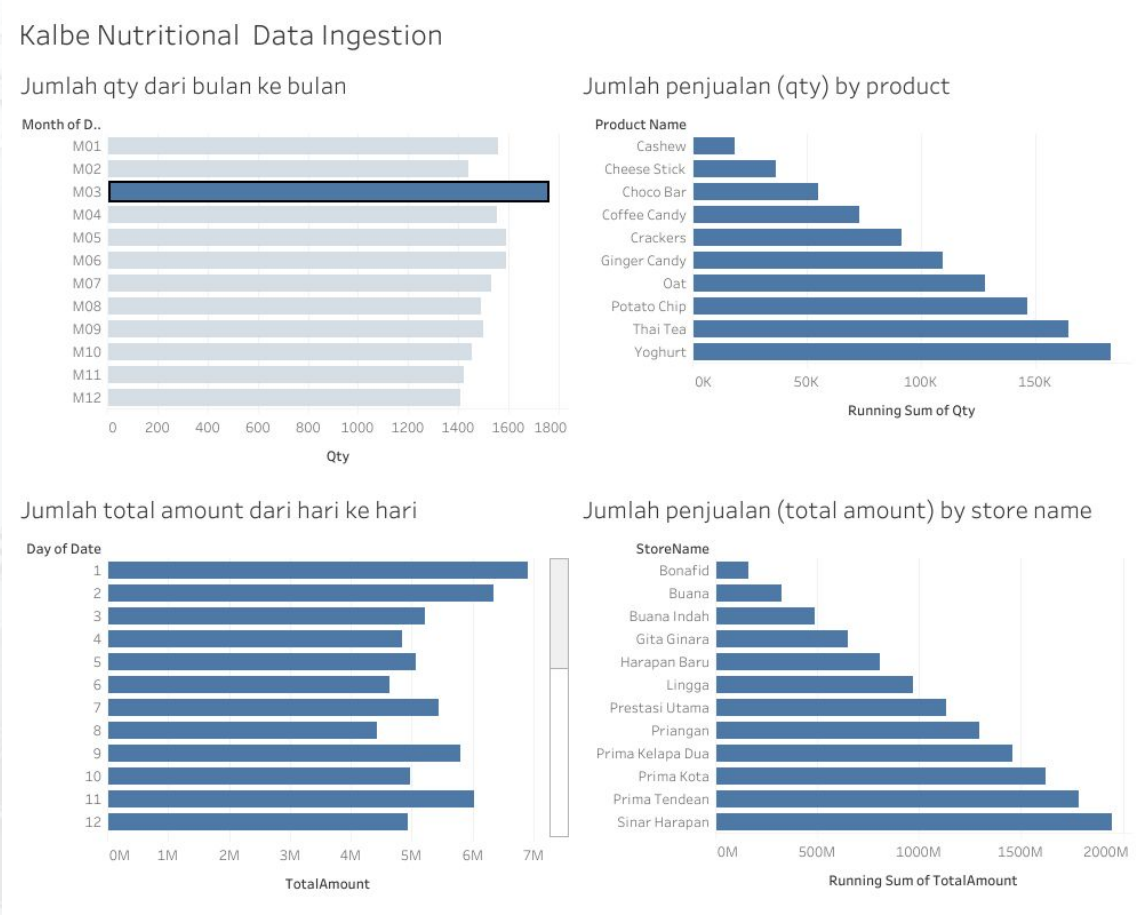
	ProductName	total
1	Cheese Stick	27,615,000

Grid

Text

Build Dashboard with Tableau

Rakamin Academy



Model Regression Forecasting in Jupyter

Rakamin Academy

```
#Products Quantity forecast
round(dfp.describe().T['mean'],0)
```

```
Crackers         5.0
Oat              2.0
Thai Tea         6.0
Choco Bar        4.0
Coffee Candy     6.0
Yoghurt          4.0
Ginger Candy     6.0
Cheese Stick     6.0
Cashew           2.0
Potato Chip      3.0
Name: mean, dtype: float64
```

perkiraan quantity product yang terjual sehingga tim inventory dapat membuat stock persediaan harian yang cukup.

Model Machine Learning Clustering in Jupyter

Rakamin Academy

```
In [44]: df_clust = df.groupby('CustomerID').agg({'TransactionID':'count','Qty':'sum','TotalAmount':'sum'})
         df_clust
```

Out[44]:

CustomerID	TransactionID	Qty	TotalAmount
1	17	60	623300
2	13	57	392300
3	15	56	446200
4	10	46	302500
5	7	27	268600
...	...	...	...
443	16	59	485100
444	18	62	577700
445	18	68	587200
446	11	42	423300
447	13	42	439300

447 rows × 3 columns

Model Machine Learning Clustering in Jupyter

Rakamin Academy

clustering with Kmeans

In [47]:
```
# Kmeans n_cluster = 3
#Clustering Kmeans
kmeans_3 = KMeans(n_clusters=3,init='k-means++',max_iter=300,n_init=10,random_state=0)
kmeans_3.fit(X_std)

#Tambah clusters label pada dataset
df_cl3 = pd.DataFrame(data=X_std,columns=df_clust.columns)
df_cl3['cluster'] = kmeans_3.labels_
df_cl3.sample(3)
```

Out[47]:

	TransactionID	Qty	TotalAmount	cluster
14	0.237401	0.240877	0.355739	2
35	-0.379565	-0.700861	-0.402525	2
134	0.854367	0.162399	0.207942	2

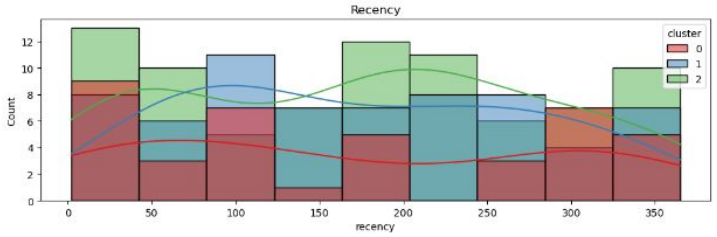
Model Machine Learning Clustering in Jupyter

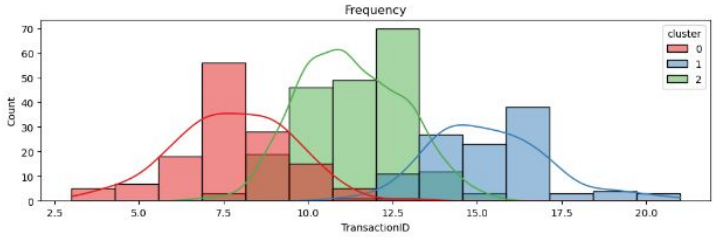
Rakamin Academy

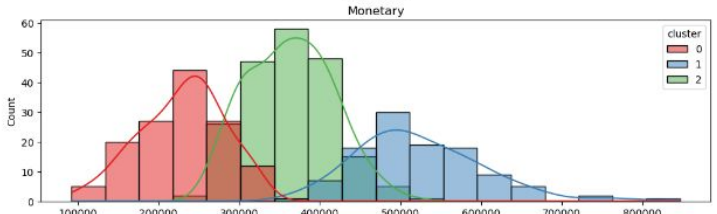
Cluster Analysis

```
In [57]: df_clust['cluster'] = kmeans_3.labels_

         df['recency'] = (pd.to_datetime('2023-01-01') - df_['Date']).dt.days.astype('int')
         df_r = df.groupby('CustomerID').agg({'recency':'min'})

         df_rfm = df_clust.copy()
         df_rfm['recency'] = df_r['recency']
```

```
In [59]: fig, ax = plt.subplots(3,1,figsize=(10,10))
         sns.histplot(data=df_rfm,x='recency',hue='cluster',palette='Set1',ax=ax[0],kde=True)
         ax[0].set_title('Recency')
         sns.histplot(data=df_rfm,x='TransactionID',hue='cluster',palette='Set1',ax=ax[1],kde=True)
         ax[1].set_title('Frequency')
         sns.histplot(data=df_rfm,x='TotalAmount',hue='cluster',palette='Set1',ax=ax[2],kde=True)
         ax[2].set_title('Monetary')

         plt.tight_layout()
         plt.show()
```







0 = New Customer

strategy -> because it is a new customer, the right company strategy is to provide attractive offers that can increase loyalty in the form of:
* gift discounts
* good customer support, and
* create new customer satisfaction surveys.

1 = Potensial Customer

strategy -> because they are potential customers, the right company strategy is to need attractive offers that can convert them into loyal customers in the form of:
* gift discounts
* provide proactive communication
* identify their needs by providing good service, and
* carry out regular follow-ups if the potential customer has not decided to buy.

2 = Loyal Customer

strategy -> because they are loyal customers, the right company strategy is to need attractive offers that can maintain and improve good relations with them in the form of offering loyalty programs specifically for loyal customers such as:
* exclusive product offers
* gift discounts, and
* creating special surveys for loyal customers.

Rakamin Academy

Link Github

Github: Final Project Kalbe Nutrition Data Modeling and Model Machine Learning

Video Presentation

the video on YouTube: https://youtu.be/ILocDwiv68s

Thank You