

# Machine Learning Project

**Kalbe Nutritionals Data Scientist  
Virtual Internship Program**

Presented by  
Fadhilah Rizky Setiawan



# Fadhilah Rizky Setiawan

## About You

Hello,  
I'm a Fresh Graduate of Information Technology Student at Muhammadiyah University of Yogyakarta, majoring in Information Technology. I have interest in technology, especially in Data and Social Media.

## Insert Your Experience

-  Social Media Specialist  
PT Scriptiva Research Indonesia  
September 2022 - Agustus 2023
-  Assistant Lecturer  
Universitas Muhammadiyah Yogyakarta  
September 2022 - Juli 2023
-  PHP Web Developer (Internship)  
CV Creative Gama Studio  
Maret 2022 - April 2022

# Case Study



# Background Story

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

- Tim inventory minta bantuan untuk 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.
- Tim marketing minta bantuan 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

- Python
- Jupyter Notebook
- Tableau
- Dbeaver
- PostgreSQL



# Challenge

- Melakukan data ingestion dan exploratory data analysis di dbeaver
- Melakukan data ingestion dan membuat dashboard di tableau
- Membuat model prediktif menggunakan regresi dan membuat clustering

# Results



# Data Ingestion & EDA di Dbeaver

	Marital Status	Rata-Rata Umur
1		31
2	Married	43
3	Single	29

Rata-rata umur berdasarkan Marital Status

	Gender	Rata-Rata Umur
1	Female	40
2	Male	39

Rata-Rata umur berdasarkan Gender

	Store Name	Total Quantity
1	Lingga	2,777

nama store dengan total quantity terbanyak

	Product Name	Total Amount
1	Cheese Stick	27,615,000

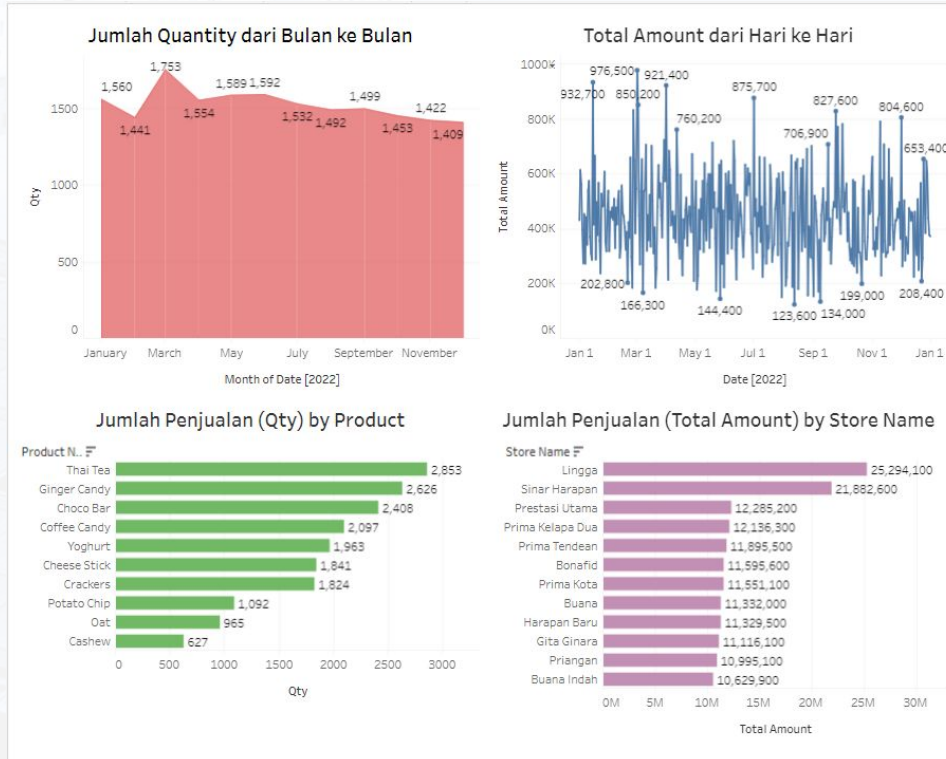
Produk terlaris dengan total amount terbanyak

Dari hasil Exploratory Data Analysis, dapat disimpulkan bahwa:

1. Rata rata umur customer yang sudah menikah adalah 43 tahun dan yang masih single berumur 29 tahun, dan ada beberapa customer yang tidak mengisi marital statusnya.
2. Rata rata umur customer perempuan adalah 40 tahun dan pria 39 tahun.
3. Store dengan total quantity terbanyak yaitu Lingga dengan total quantity 2,777 pcs
4. Produk terlaris adalah Cheese Stick dengan total amount Rp. 27.615.000.



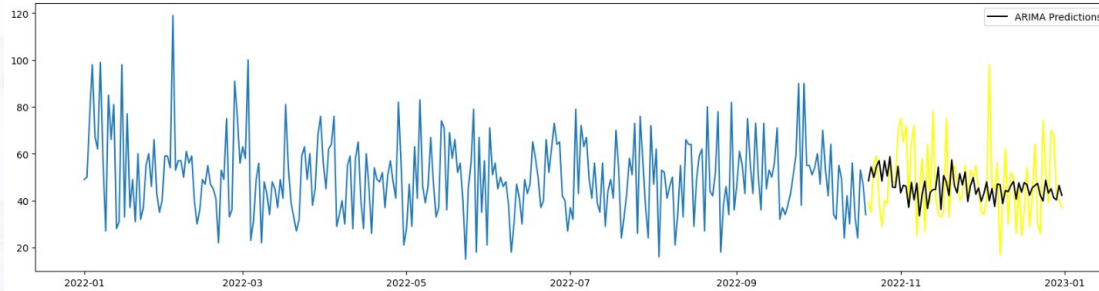
# Dashboard di tableau



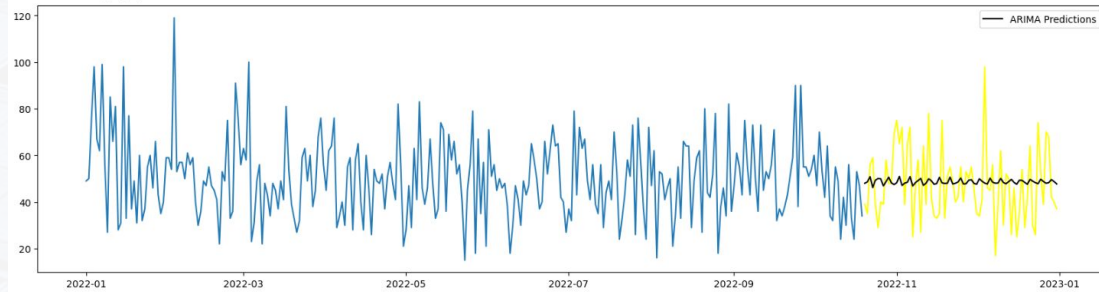
Berdasarkan dashboard disamping, dapat disimpulkan bahwa:

1. Jumlah Quantity tertinggi ada pada Q1 yaitu bulan Maret dan cenderung menurun hingga akhir tahun.
2. Jumlah total amount tertinggi mencapai 976.500 dan terendah 123.600.
3. Produk dengan penjualan tertinggi adalah Thai Tea dengan 2.853 dan terendah cashew dengan 627
4. Store dengan penjualan tertinggi adalah Lingga dengan Rp. 25.294.100 dan terendah Buana Indah dengan Rp. 10.629.900. Dari data tersebut terlihat bahwa store Lingga dan Sinar Harapan punya selisih penjualan yang jauh dengan toko lainnya.

# Machine Learning Regression (Time Series)



a. ARIMA Base Model



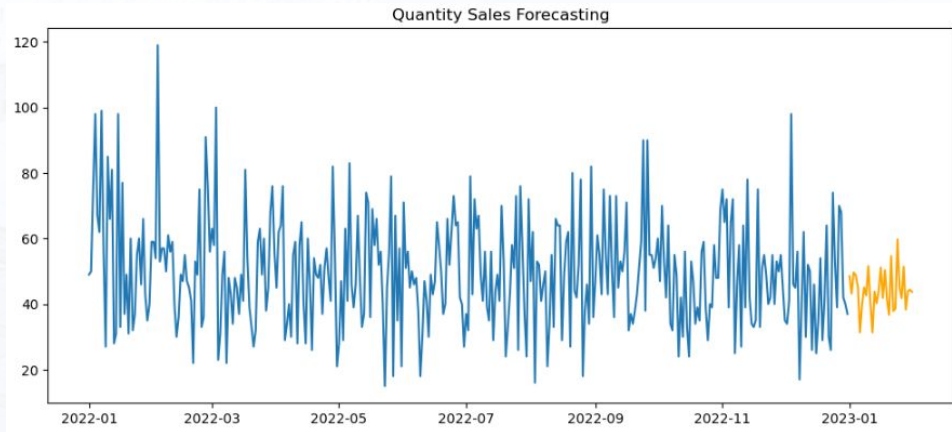
b. ARIMA Hypertuning

Dari hasil train menggunakan ARIMA base model didapatkan hasil MAE : 12.63

Dari hasil train menggunakan ARIMA Hypertuning didapatkan hasil MAE : 12.64

ARIMA Base Model menunjukkan hasil yang lebih baik dibandingkan ARIMA Hypertuning

# Machine Learning Regression (Time Series)



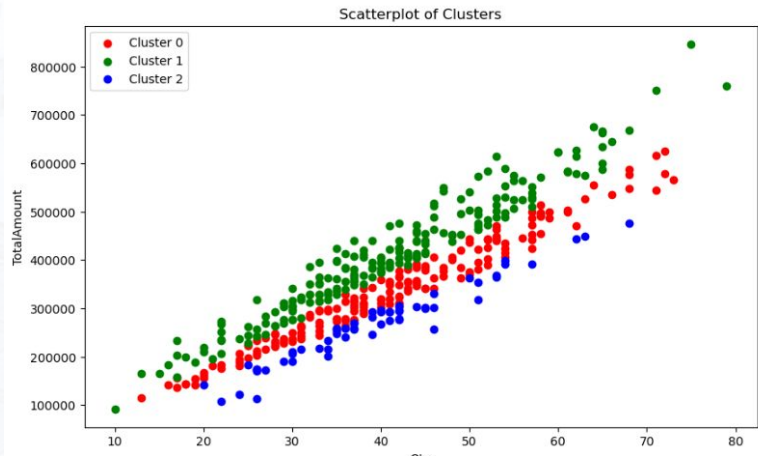
```
[42]: forecast.describe()
```

```
[42]: count    31.000000  
      mean     44.083597  
      std       6.275007  
      min     31.356339  
      25%     40.347365  
      50%     43.801777  
      75%     48.660629  
      max     59.752672  
      Name: predicted_mean, dtype: float64
```

Dari hasil prediksi tersebut didapatkan jumlah rata rata penjualan per hari untuk 1 bulan ke depan (1 Jan 2023 - 31 Jan 2023) adalah 44 pcs/hari



# Machine Learning Clustering



[55]:

	CustomerID	TransactionID	Qty	TotalAmount
cluster				
1	204	11.642157	40.838235	409532.843137
0	179	11.005587	41.346369	338749.720670
2	61	10.344262	39.229508	268652.459016

- Cluster 2**  
 Merupakan segment customer dengan jumlah transaksi dan total quantity terendah, serta menghabiskan uang paling sedikit  
  
**Rekomendasi** : Memberikan voucher diskon produk Kalbe Nutritionals kepada customer cluster 2 untuk transaksi selanjutnya, sehingga terjadi pembelian berulang dan memberikan promo buy 1 get 1 atau bundling produk Kalbe Nutritionals untuk meningkatkan jumlah Quantity sales
- Cluster 1**  
 Merupakan segment customer dengan jumlah transaksi tertinggi dan menghabiskan uang paling banyak  
  
**Rekomendasi** : Memberikan voucher diskon untuk transaksi produk Kalbe Nutritionals selanjutnya
- Cluster 0**  
 Merupakan segment customer dengan total quantity terbanyak dan jumlah transaksi yang sedikit lebih kecil dibanding cluster 1  
  
**Rekomendasi** : Memberikan promo buy 1 get 1 atau bundling produk Kalbe Nutritionals untuk meningkatkan jumlah Quantity sales



# LINK

[Drive](#)

[GitHub](#)

[Youtube](#)

# Thank You



**Rakamin**  
Academy



**KALBE**  
Nutritional