



Machine Learning Project

Kalbe Nutritionals Data Scientist
Project Based Internship Program

Bagas Shalahuddin Wahid

Challenge 1

Rata rata umur berdasarkan marital status

ABC maritalstatus	123 rata_rata_umur
Married	43
Single	31

rata-rata umur customer berdsarkan dari gender

123 gender	123 rata_rata_umur
0	41
1	40

Nama store dengan total quantity terbanyak

ABC storename	123 total_qty
Lingga	2,715
Sinar Harapan	2,523
Prestasi Utama	1,373
Prima Kota	1,339
Prima Tendean	1,276
Prima Kelapa Dua	1,271
Buana	1,266
Bonafid	1,262
Harapan Baru	1,261
Gita Ginara	1,220
Priangan	1,219
Buana Indah	1,174

Nama produk terlaris dengan total amount terbanyak

ABC productname	123 total_amount
Cheese Stick	27,030,000
Choco Bar	21,040,800
Coffee Candy	19,345,200
Yoghurt	19,260,000
Oat	15,072,000
Crackers	13,035,000
Potato Chip	12,672,000
Thai Tea	11,751,600
Cashew	11,106,000
Ginger Candy	8,227,200

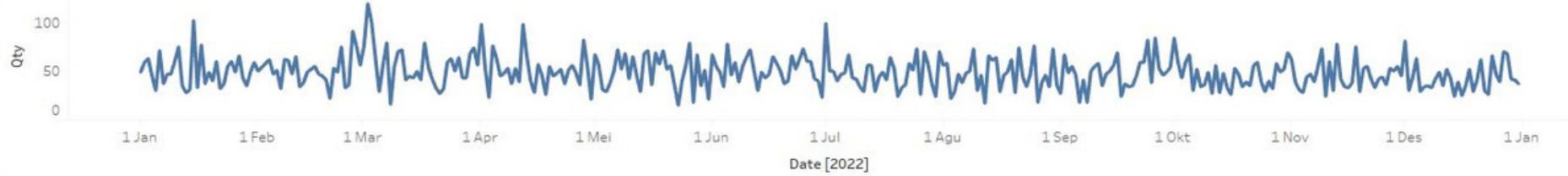
Challenge 2

Dasboard Sales

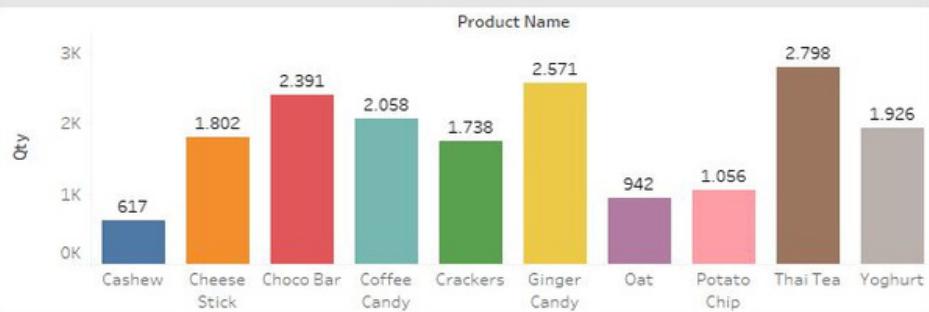
Total Quantity by Moths



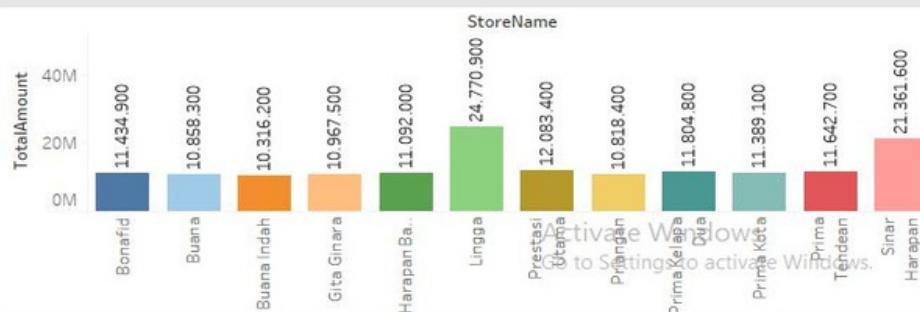
Total Amount by Days



Total Quantity by Product



Total Amount by Store



Activate Windows
Go to Settings to activate Windows.

Challenge 3

forecasting

Stasionery test

Hasil Uji ADF untuk Data:

ADF Statistic: -19.018782802299725

P-value: 0.0

Critical Values: {'1%': -3.4484434475193777, '5%': -2.869513170510808, '10%': -2.571017574266393}

Interpretasi hasil ini:

~ADF Statistic: Nilai ADF Statistic yang sangat negatif menunjukkan bahwa data latihan cenderung stasioner.

~P-value: Nilai p-value sangat kecil 0.0, yang menunjukkan bahwa data memiliki akar unit non-stasioner. Semakin kecil nilai p-value, semakin kuat bukti bahwa data adalah stasioner.

~Critical Values: Dalam kasus ini, karena ADF Statistic lebih kecil dari semua critical values, hal ini menunjukkan bahwa kita dapat menolak hipotesis nol dan data adalah stasioner.

Challenge 3

forecasting

Dalam melakukan forecasting dilakukan dengan 2 cara yaitu dengan auto arima (pmdarima) dan manual arima (ARIMA)

Model auto arima
Model manual arima

Best model: ARIMA(5,1,1)(0,0,0)[0]

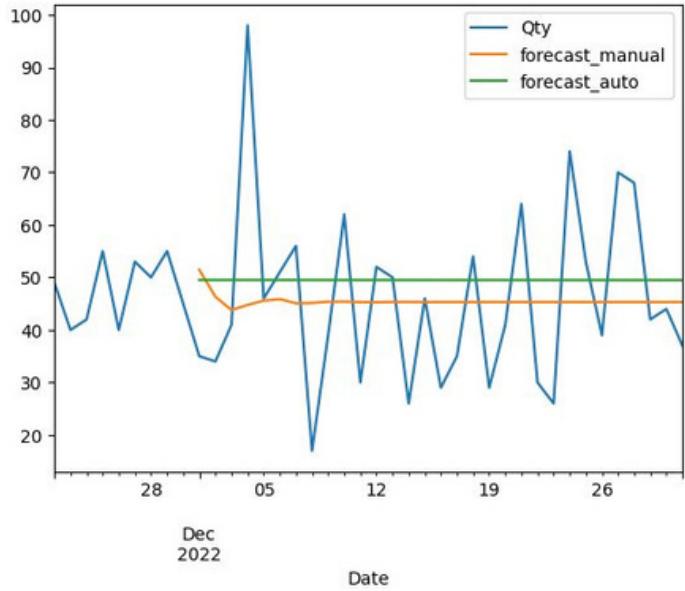
Total fit time: 55.835 seconds

	param	mae	mape	rmse
2	(70, 2, 1)	15.045552	0.349530	18.992447
0	(50, 2, 1)	15.423963	0.413216	20.005044
1	(60, 2, 1)	16.176883	0.386983	19.952589

Dalam model auto diperoleh model arima dengan p,d,q (5,1,1) sedangkan untuk model manual arima model terbaik (70,2,1) dengan rmse terendah 18,9

Challenge 3

forecasting

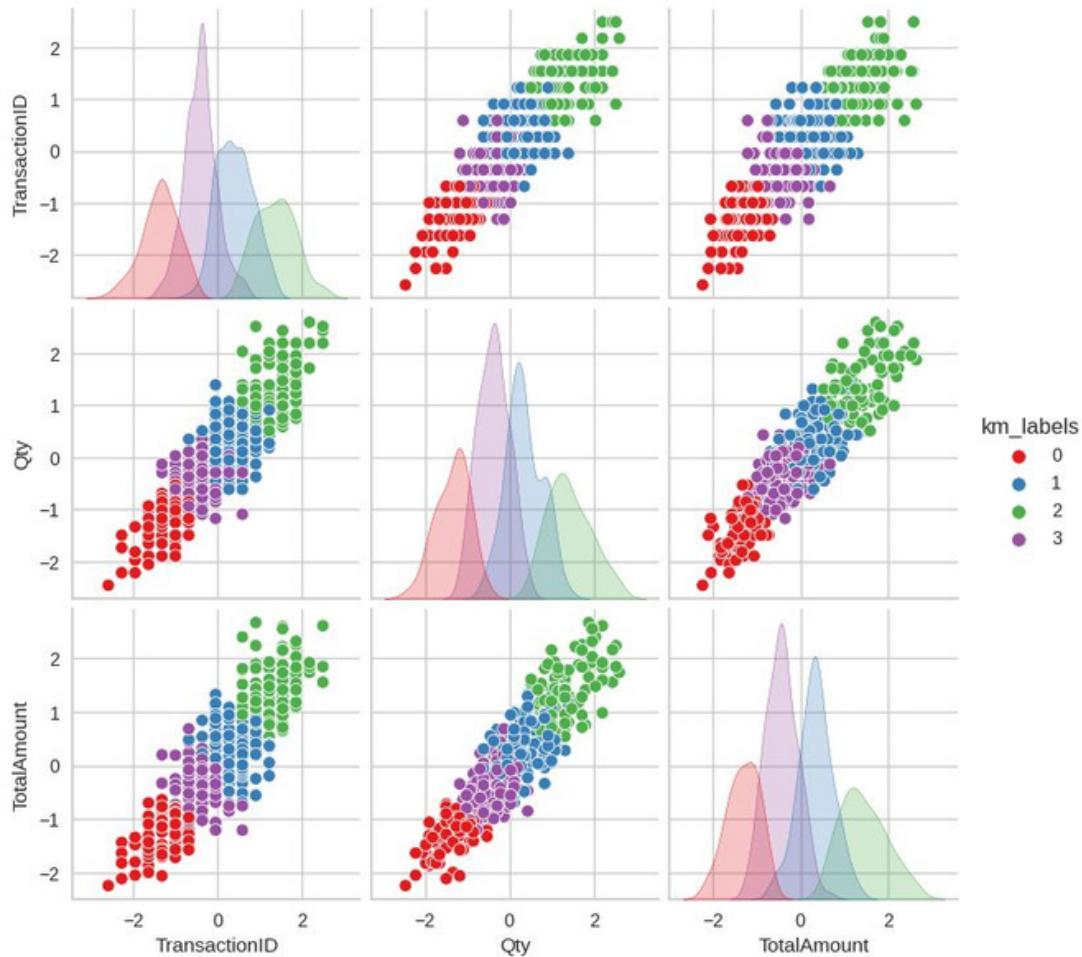


	Qty	forecast_manual	forecast_auto
count	40.000000	31.000000	31.000000
mean	46.175000	45.492021	49.503215
std	15.281776	1.166170	0.006355
min	17.000000	43.790246	49.492732
25%	36.500000	45.310685	49.497973
50%	44.500000	45.311778	49.503215
75%	53.250000	45.315380	49.508457
max	98.000000	51.451485	49.513700

Dari visualisasi di atas menunjukan bahwa model auto dan manual tidak jauh beda dalam memprediksi, prediksi menunjukan rata rata yang tidak jauh berbeda yaitu manual 45.4 dan auto 49.5. Namun yang perlu menjadi catatan adalah berdasarkan visualisasi di atas model masih kurang baik dalam memprediksi dan perlu dilakukan perbaikan untuk mencari model terbaik

Challenge 4

clustering



nilai silhouettescore :
0.613177607682284

Nilai silhouette score sebesar 0.613177607682284 mengindikasikan bahwa clustering yang dilakukan memiliki hasil yang relatif baik. Rentang nilai silhouette score berkisar dari -1 hingga 1. Semakin dekat nilai silhouette score dengan 1, semakin baik kualitas clusteringnya. Dalam hal ini, nilai silhouette score yang mendekati 1 menunjukkan bahwa objek dalam cluster memiliki jarak yang lebih dekat dengan objek dalam cluster yang sama dibandingkan dengan objek dalam cluster lain. Ini mengindikasikan bahwa pengelompokan objek dalam cluster sangat baik.

Challenge 4

clustering

km_labels	Total_Customer	Total_Transaction	TotalQty	TotalAmount
0	0	83	83	23.8 201980.7
1	1	127	127	44.8 399892.9
2	2	89	89	58.2 531419.1
3	3	144	144	35.8 307320.1

Interpretasi:

- Cluster 0: Pelanggan dalam cluster ini cenderung memiliki frekuensi dan nilai transaksi yang rendah, tetapi masih memberikan kontribusi yang berarti terhadap TotalAmount.
- Cluster 1: Pelanggan dalam cluster ini memiliki frekuensi dan nilai transaksi yang tinggi, serta memberikan kontribusi signifikan terhadap TotalAmount.
- Cluster 2: Pelanggan dalam cluster ini adalah pelanggan dengan frekuensi dan nilai transaksi yang sangat tinggi, dan memberikan kontribusi besar terhadap TotalAmount.
- Cluster 3: Pelanggan dalam cluster ini memiliki frekuensi transaksi sedang dengan kontribusi TotalAmount yang cukup baik.

Rekomendasi Bisnis:

- Cluster 0: Fokus pada strategi retensi pelanggan untuk meningkatkan frekuensi transaksi dan nilai transaksi mereka.
- Cluster 1: Tingkatkan layanan pelanggan dan tawarkan insentif untuk mendorong pelanggan dalam cluster ini untuk melakukan lebih banyak transaksi.
- Cluster 2: Pertahankan kualitas layanan dan tawarkan insentif khusus untuk meningkatkan loyalitas pelanggan dalam cluster ini.
- Cluster 3: Tingkatkan nilai transaksi pelanggan dalam cluster ini dengan menghadirkan penawaran produk atau layanan yang baru.

Thank You

