

# Analisis Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis dan K-Means Clustering

Shinta Permata Sari\*, Raissa Amanda Putri

Fakultas Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatra Utara, Medan, Indonesia

Email: <sup>1</sup>\*shintapermatasari71@gmail.com, <sup>2</sup>raissa.ap@uinsu.ac.id

Email Penulis Korespondensi: shintapermatasari71@gmail.com

Submitted: 17/12/2023; Accepted: 29/12/2023; Published: 30/12/2023

**Abstrak**—Banyaknya inovasi di sektor ritel sangat erat kaitannya dengan digitalisasi dan kemajuan teknologi. Nilai data penjualan semakin meningkat sebagai sumber daya bagi bisnis di era digital saat ini. Data penjualan terdiri dari detail tentang barang yang dijual, klien yang mereka layani, pola penjualan, dan elemen lain yang memengaruhi seberapa baik bisnis beroperasi. Namun, menangani dan memahami data penjualan yang luas dan rumit secara efektif bisa jadi sulit dan membingungkan. Maka dengan permasalahan tersebut penulis akan menentukan tingkat penjualan produk pertahun yang paling banyak dipasarkan dengan menerapkan Exploratory Data Analysis (EDA) dan Algoritma K-Means Clustering metode yang digunakan dengan menentukan tingkat penjualan berdasarkan attribute yang digunakan, tingkat penjualan nantinya akan dibagi menjadi 3 yaitu cluster 1 tinggi, cluster 2 sedang dan cluster 3 rendah. Berdasarkan hasil dari metode EDA dan K-Means mendapatkan hasil perbandingan penjualan selama 4 tahun nilai rata-rata yang tinggi menunjukkan bahwa penjualan pada tahun 2019 memiliki variasi dan nilai rata-rata yang besar diantara tahun 2020, 2021, dan 2022. Pada hasil visualisasi dapat disimpulkan bahwa produk yang masuk ke dalam kategori cluster-1 terdapat 27 produk, cluster-2 terdapat 10 produk, dan cluster-3 terdapat 5 produk.

**Kata Kunci:** Analisis; Visualisasi Data; Exploratory Data Analysis ; K-Means Clustering; Penjualan

**Abstract**—Many innovations in the retail sector are closely related to digitalization and technological advances. The value of sales data is increasing as a resource for businesses in today's digital era. Sales data consists of details about the items sold, the clients they serve, sales patterns, and other elements that influence how well a business operates. However, effectively handling and understanding vast and complex sales data can be difficult and confusing. So with this problem the author will determine the annual sales level of the products that are most widely marketed by applying Exploratory Data Analysis (EDA) and the K-Means Clustering Algorithm, the method used is to determine the sales level based on the attributes used, the sales level will later be divided into 3, namely clusters 1 is high, cluster 2 is medium and cluster 3 is low. Based on the results of the EDA and K-Means methods, the results of a sales comparison for 4 years have an average value, indicating that sales in 2019 have large average values between 2020, 2021 and 2022. From the visualization results it can be concluded that there are 27 products in the cluster-1 category, 10 products in the cluster-2 category, and 5 products in the cluster-3 category.

**Keywords:** Analysis; Data Visualization; Exploratory Data Analysis; K-Means Clustering; Sale

## 1. PENDAHULUAN

Banyaknya inovasi di sektor ritel sangat erat kaitannya dengan digitalisasi dan kemajuan teknologi. Namun, menangani dan memahami data penjualan yang luas dan rumit secara efektif bisa jadi sulit. Pada gilirannya, mereka dapat menggunakan analisis data belanja pelanggan untuk mengembangkan tujuan dan tindakan yang selaras dengan arah strategis organisasi dan mengoptimalkan operasi mereka secara berkelanjutan [1].

Sebagai perusahaan Shukaku menawarkan beragam produk dengan cabang produk berbeda, seperti Lightings, Circuit Breakers, Bakelite, Electrical Appliances, Accessories dan Stabilizer. Selain reputasinya dalam memproduksi barang-barang hemat energi, Shukaku juga memiliki sertifikasi SNI dan CE, sehingga mampu bersaing dalam negeri maupun global. Shukaku ini memiliki banyak data yang masih berupa data mentah di format excel yang dimana mereka masih perhitungannya manual dan tidak ada gambaran visual untuk melihat tiap keuntungan pertahun, maka peneliti ingin membantu dalam menghitung cepat hasil analisis perhitungan dan juga hasil visual dari data yang diperoleh untuk memudahkan perusahaan mengetahui pola penjualan produk setiap tahun yang lebih menguntungkan mereka selanjutnya yang akan menjadi gambaran strategi untuk kedepannya [2].

Analisis dan visualisasi data penjualan menggunakan Exploratory Data Analysis (EDA) dan K-means adalah teknik yang berguna untuk menarik kesimpulan mendalam dari data penjualan. EDA merupakan teknik yang efisien untuk mengeksplorasi dan menganalisis data secara visual untuk mengidentifikasi tren, koneksi, dan pola dalam data penjualan. Sedangkan data dikelompokkan menjadi banyak kelompok menggunakan algoritma clustering K-Means berdasarkan fitur bersama [3].

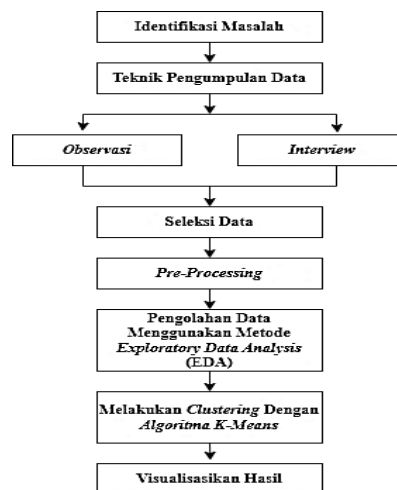
Perusahaan dapat lebih memahami pola penjualan, preferensi pelanggan, dan variabel yang mempengaruhi kinerja penjualan dengan menggunakan pendekatan EDA dan K-Means dalam analisis dan visualisasi data penjualan. Hal ini memungkinkan meningkatkan kepuasan pelanggan, efisiensi operasional, dan optimalisasi strategi pemasaran. Penelitian ini akan mengelompokkan data menggunakan algoritma K-Means dan menggunakan kedua teknik tersebut untuk menganalisis dan mendeskripsikan data penjualan secara visual.[4]Maka dengan permasalahan tersebut penulis menentukan tingkat penjualan produk pertahun yang paling banyak dipasarkan dengan menerapkan Exploratory Data Analysis (EDA) dan Algoritma K-Means Clustering mendapatkan tingkat penjualan akan dibagi menjadi 3 yaitu cluster 1 tinggi, cluster 2 sedang dan cluster 3 rendah.

Adapun penelitian terdahulu yang menjadi acuan yaitu: Menurut penelitian oleh Desi Aryanti dan Johan Setiawan, 2018 dengan judul “Visualisasi Data Penjualan dan Produksi PT Nitto Alam Indonesia” dengan metode Exploratory Data Analysis mendapatkan hasil dari eksplorasi data penjualan dan produksi yaitu dapat diketahui bahwa pola penjualan dalam lima tahun terakhir tidak memiliki pola trend, cyclic, maupun seasonal dengan puncak penjualan tertinggi terletak pada kuartar keempat (Q4) tahun 2017 sebesar 6,552% [5]. Pada penelitian yang dilakukan oleh Aditya Hari Wibowo, Khatin Faisah dan Yudo Devianto, 2022, Dengan Judul “Analisa Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis Pada PT. Telkominfra”. Sistem informasi yang dikembangkan dalam penelitian ini dapat melakukan pre-processing data, analisis data dengan EDA hingga menampilkan hasil visualisasi dalam bentuk grafik, sehingga dapat memberikan informasi kepada manajemen terkait kontrak mana yang perlu di evaluasi, selain itu juga memiliki benefit adanya improvement pada proses pengolahan data dari 6 jam menjadi 30 menit [6]. Penelitian yang dilakukan oleh Mawaddah Harahap, Yusniar Lubis dan Zakarias Situmorang, 2022, dengan Judul “Analisis Pemasaran Bisnis dengan Data Science: Segmentasi Kepribadian Pelanggan berdasarkan Algoritma K-Means Clustering” Studi ini melakukan Kerangka kerja Data Science (ilmu data) diterapkan untuk memberikan pemahaman bisnis, pemahaman data, analisis data dan pemodelan. Kemudian algoritma K-Means untuk segmentasi pelanggan dengan menggunakan metode elbow dan silhouette yang menghasilkan nilai k=4 yang paling optimal [7]. Pada penelitian yang dilakukan oleh Susi Oktaviani dan Agus Bahtiar, 2023 dengan judul “Implementasi Algoritma K-Means Dalam Pengelompokan Data Penjualan CV. Widuri Menggunakan Orange” Studi ini melakukan pengolahan data penjualan dan Hasil clustering dengan menggunakan aplikasi orange menghasilkan 7 cluster dengan rincian, cluster 1 (C1) 104 transaksi, cluster 2 (C2) 1 transaksi, cluster 3 (C3) 139 transaksi, cluster 4 (C4) 95 transaksi, cluster 5 (C5) 102 transaksi, cluster 6 (C6) 76 transaksi, cluster 7 (C7) 97 transaksi [8]. Dengan demikian, Tujuan penelitian ini bisa menerapkan metode Exploratory Data Analysis dengan aplikasi orange yang akan meringkas waktu pre-processing dari data awal yang berjumlah ribuan yang akan menghasilkan beberapa data saja, dan juga mendapatkan visual dari data yang sudah di olah tersebut dan menampilkan gambar berupa grafik tiap tahun setiap produknya. Jika menerapkan K-Means Clustering nantinya juga akan mendapatkan hasil tiap perhitungan produk ke beberapa kelompok dan menampilkan hasil visual untuk jumlah produk yang paling menguntungkan perusahaan berdasarkan kelompoknya.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan metode kuantitatif, yaitu penelitian yang menggunakan data berupa angka-angka kuantitatif. Hasil penelitian kuantitatif dapat digeneralisasikan karena diperoleh melalui analisis statistik [9]. Pada penelitian ini melalui beberapa tahapan yaitu sebagai berikut:



**Gambar 1.** Tahap Penelitian

Pada Gambar 1 diatas merupakan gambar kerangka penelitian yang akan di gunakan pada penelitian ini. Berikut penjelasan tentang tahapan penelitian:

- a. Identifikasi Masalah  
Langkah Pertama yang dilakukan pada penelitian ini adalah mengidentifikasi masalah, peneliti berupaya untuk mengamati dan memahami permasalahan dengan adanya permasalahan data penjualan.
- b. Pengumpulan Data  
Untuk memperoleh data, peneliti melakukan observasi yaitu data primer bentuk verbal atau kata-kata yang diucapkan secara lisan, sehingga penulis memerlukan pengolahan data tambahan. untuk memperoleh hasil yang akurat. [10] Sedangkan segi interview, peneliti mewawancari Pegawai PT yaitu Bapak Mhd Taufik

- c. Seleksi Data  
Proses ini adalah penentuan data yang akan digunakan sesuai dengan permasalahan yang berkaitan. Data yang akan digunakan adalah data penjualan dari PT. Shukaku dari tahun 2019 sampai dengan 2022
- d. Pre-Processing  
Tahap untuk melakukan sebuah proses awal dalam pengolahan data. Pada tahap ini data yang akan diolah bertujuan untuk menghindarkan dari data yang mengganggu atau data yang tidak konsisten [11].
- e. Pengolahan Data Menggunakan Metode Exploratory Data Analysis (EDA)  
Pada EDA proses analisis statistik yang digunakan untuk menyelidiki dan memahami data secara khusus yang bertujuan untuk mengidentifikasi pola, trend, dan hubungan dalam dataset [12].
- f. Melakukan Clustering dengan Algoritma K-Means  
Setelah menentukan jumlah kluster, terapkan algoritma K-Means Clustering pada data. Proses ini akan mengelompokkan data ke dalam kluster berdasarkan kemiripan atribut.
- g. Visualisaikan Hasil  
Pada tahapan ini, visualisasi hasil dari penggabungan metode EDA dan K-means menggunakan orange sangat berguna karena memungkinkan kita untuk memvisualisasikan hasil clustering dengan cara yang lebih baik.

## 2.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) ialah teknik pencarian heuristik untuk menemukan relasi signifikan antara variabel pada dataset yang besar. Kita bisa mendapatkan pemahaman yang lebih baik tentang keadaan kumpulan data kita dengan menggunakan EDA [13].

## 2.3 Algoritma K Means Clustering

K-Means Clustering merupakan salah satu metode analisis data atau data mining yang melakukan proses pemodelan tanpa pengawasan (unsupervised). K dimaksudkan sebagai jumlah konstan cluster yang diinginkan. diartikan sebagai cluster. Sedangkan Clustering memiliki kemampuan untuk mengenali bentuk dalam data, yang kemudian dapat diterapkan pada berbagai aplikasi umum, termasuk pengenalan pola, pemrosesan gambar, dan klasifikasi [14][15]. Proses algoritma K-Means menggunakan rumus Euclidean distance sebagai berikut:

1. Menentukan jumlah kelompok
2. Data dialokasikan ke dalam kelompok secara acak
3. Menghitung pusat dari data kelompok (centroid/rata-rata) pada masing-masing kelompok menggunakan rumus Euclidean Distance yaitu:

$$D(i,j) = \sqrt{(X_{1i} - Y_{1j})^2 + (X_{2i} - Y_{2j})^2 + \dots + (X_{ki} - Y_{kj})^2} \tag{1}$$

Dimana

D (i,j) = Jarak data ke I ke pusat cluster j

X<sub>ki</sub> = Data ke i pada atribut data ke k

X<sub>kj</sub> = Titik pusat ke j pada atribut ke k

4. Alokasikan setiap data ke centroid/rata-rata yang paling mendekati menggunakan rumus mean yaitu:

$$X = \frac{\sum x}{n} \tag{2}$$

5. Kembali ke Langkah 3, apabila masih terdapat data yang berpindah kelompok, atau apabila ada perubahan nilai centroid [16].

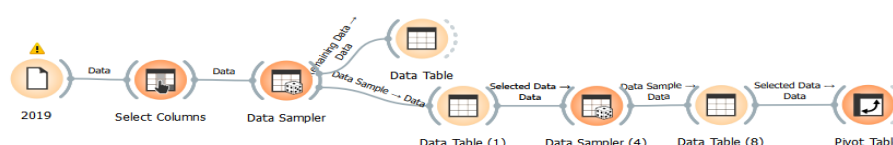
## 2.4 Orange

Orange ialah perangkat lunak untuk pemrograman visual yang berbasis komponen dan digunakan untuk analisis data. Setiap widget di Orange memiliki tujuan unik dan memiliki kemampuan untuk menerima input atau menghasilkan output [17][18].

# 3. HASIL DAN PEMBAHASAN

## 3.1 Preprocessing

Pada tahap ini setelah data awal didapatkan sebanyak 10342 data yang berhasil dikumpulkan, kemudian data akan di proses dahulu menggunakan Aplikasi Orange dan menggunakan metode Exploratory Data Analysis (EDA).



**Gambar 2.** Tahapan Preprocessing

Pada gambar 2 diatas adalah salah satu tahapan data dari tahun 2019 dan tahapan ini dilakukan sampai tahun 2022, tahap awal dalam prosesnya yaitu pertama pada bagian widget File dimasukkan data awal dari tahun 2019 sampai 2022 sebanyak 10342, lalu dimasukkan ke dalam widget Select Columns untuk kita melihat data yang mana yg diperlukan beserta menghilangkan atribut yang berlebih dan juga tidak terlalu penting, setelah itu beralih ke Data Sampler untuk memilih dengan opsi Fixed Proportion of data yang akan dibuat di 1% dan menghasilkan total data baru keseluruhan yang berjumlah 1132 data, kita bisa melihatnya di Data Table kedua. Namun data ini masih terlalu banyak, sehingga peneliti melakukan Sampling Type lagi untuk mengurangi jumlahnya dengan menaikkan ke 4% dan mendapatkan hasil akhir berjumlah 42 data.

Selanjutnya data penjualan dianalisis untuk mengidentifikasi jenis informasi di setiap kolom dan untuk mengatasi nilai yang hilang atau tidak lengkap. Setiap kolom diperiksa, format tanggal diubah ke format yang sesuai dan data numerik disesuaikan untuk memenuhi persyaratan analisis. Setelah pembersihan data, langkah selanjutnya adalah menerapkan analisis data eksplorasi (EDA) dan mengekstrak fitur terkait. Selanjutnya data penjualan grup dikenai algoritma K-Means [19] [20].

### 3.2 Exploratory Data Analysis (EDA)

#### 3.2.1 Statistik Deskriptif

Statistik deskriptif ialah statistik yang menggambarkan atau merangkum data tanpa berusaha menarik kesimpulan guna memberikan informasi yang luas. Banyak informasi penting, termasuk mean, median, modus, deviasi standar, varians, dan kecekungan, dapat ditampilkan menggunakan statistik deskriptif. Ada beberapa cara untuk menyajikan statistik deskriptif ini, antara lain tabel, grafik, dan diagram. Pada tahap ini selanjutnya menggunakan salah satu widget orange yaitu pivot table, peneliti disini hanya akan menghitung jumlah hasil dan juga nilai rata-rata setiap tahun yang sudah di proses sebelumnya, untuk hasilnya sebagai berikut.

Aggregate	PIB-241	PIB-422	PIB-723	PIB-727	PIB-730	PIB-1087	PIB-1334	PIG-50902	PIG-54019	Total		
<b>ABDIAN</b>	Count	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0		
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	10000.0	0.0	10000.0		
	Mean	?	?	?	?	?	?	10000.0	?	10000.0		
<b>CABANG UTARA - BANGKA</b>	Count	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	0.0	0.0	6000.0	0.0	0.0	6000.0		
	Mean	?	?	?	?	?	6000.0	?	?	6000.0		
<b>CAROLINA - BENGKULU</b>	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
	Sum	0.0	17000.0	0.0	0.0	0.0	0.0	0.0	0.0	17000.0		
	Mean	?	17000.0	?	?	?	?	?	?	17000.0		
<b>ELANGGURU NEGARAKEMENTERIAN SPORO</b>	Count	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
	Mean	?	?	?	?	?	?	?	?	?		
<b>MAJLIS AKADAMIYAH KH 17</b>	Count	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0		
	Sum	0.0	0.0	10000.0	0.0	0.0	0.0	0.0	0.0	10000.0		
	Mean	?	?	10000.0	?	?	?	?	?	10000.0		
<b>SIRA</b>	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0		
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	32000.0	32000.0		
	Mean	?	?	?	?	?	?	?	32000.0	32000.0		
<b>SINAR UTARA INDRAMATI - KESAMPAT</b>	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	75000.0	75000.0		
	Mean	?	?	?	?	?	?	?	75000.0	75000.0		
<b>SINAR UTAMA / KODE KE MIE - KEM 81</b>	Count	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	4200.0	0.0	0.0	0.0	0.0	4200.0		
	Mean	?	?	?	4200.0	?	?	?	?	4200.0		
<b>SINCA BOLA - PANGKALAN</b>	Count	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	34150.0	0.0	0.0	0.0	0.0	34150.0		
	Mean	?	?	?	34150.0	?	?	?	?	34150.0		
<b>SUMBER LESTARI (DONGRA EKSTRA) - BENTEN</b>	Count	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	0.0	3000.0	0.0	0.0	0.0	3000.0		
	Mean	?	?	?	?	3000.0	?	?	?	3000.0		
<b>Total</b>	Count	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	2.0		
	Sum	0.0	17000.0	15000.0	4200.0	34150.0	3000.0	6000.0	13000.0	32000.0	75000.0	210700.0
	Mean	?	17000.0	15000.0	4200.0	34150.0	3000.0	6000.0	13000.0	32000.0	75000.0	210700.0

Gambar 3. Pivot Table 2019

Pada gambar 3 terdapat item count yang berjumlah 10.0 yang berarti ada 10 item tertera yang didapat dari proses sebelumnya, item sum yang berjumlah 205700.0 dan item mean berjumlah 22855.566

Aggregate	1436	PIB-540	PIB-616	PIB-886	PIB-1064	PIB-1180	PIG-10004	PIG-10014	PIG-10017	PIG-10019	Total
<b>EDUNIA 118</b>	Count	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
	Sum	0.0	0.0	0.0	0.0	0.0	10000.0	0.0	0.0	0.0	10000.0
	Mean	?	?	?	?	?	10000.0	?	?	?	10000.0
<b>MEGA FOTO</b>	Count	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
	Sum	10000.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10000.0
	Mean	10000.0	?	?	?	?	?	?	?	?	10000.0
<b>PANORAMA BUNING WANGU - PLEMAN</b>	Count	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
	Sum	0.0	6000.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6000.0
	Mean	?	6000.0	?	?	?	?	?	?	?	6000.0
<b>PELAGI - BUNING LINDANGAN</b>	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	3700.0	0.0	0.0	3700.0
	Mean	?	?	?	?	?	?	3700.0	?	?	3700.0
<b>PURWA LESTARI / EMEN - BENGKULU</b>	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	34000.0	0.0	34000.0
	Mean	?	?	?	?	?	?	?	34000.0	?	34000.0
<b>PURWA SILIVESTRA / DAMAR LINDANGAN DAMAR - PURWA SILIVESTRA</b>	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	21000.0	21000.0
	Mean	?	?	?	?	?	?	?	?	21000.0	21000.0
<b>SARASAT LESTARI - BENGKULU</b>	Count	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
	Sum	0.0	0.0	17000.0	0.0	0.0	0.0	0.0	0.0	0.0	17000.0
	Mean	?	?	17000.0	?	?	?	?	?	?	17000.0
<b>SILIVESTRA - KOTA PARI</b>	Count	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0
	Sum	0.0	0.0	0.0	0.0	10000.0	0.0	0.0	0.0	0.0	10000.0
	Mean	?	?	?	?	10000.0	?	?	?	?	10000.0
<b>SINAR (MELUBER) - BATAKAMA</b>	Count	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
	Sum	0.0	0.0	0.0	12350.0	0.0	0.0	0.0	0.0	0.0	12350.0
	Mean	?	?	?	12350.0	?	?	?	?	?	12350.0
<b>SINAR UTARA INDRAMATI - TL. KESAMPAT</b>	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Mean	?	?	?	?	?	?	?	?	?	0.0
<b>SINAR UTAMA - PANGKALAN BARAI</b>	Count	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	2.0
	Sum	0.0	0.0	0.0	6700.0	0.0	0.0	0.0	0.0	0.0	6700.0
	Mean	?	?	?	3350.0	?	?	?	?	?	3350.0
<b>Total</b>	Count	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	10.0
	Sum	10000.0	17000.0	15000.0	4200.0	34150.0	3000.0	6000.0	13000.0	32000.0	210700.0
	Mean	10000.0	17000.0	15000.0	2100.0	17075.0	3000.0	6000.0	13000.0	32000.0	21070.0

Gambar 4. Pivot Table 2022

Pada gambar 4 terdapat item count yang berjumlah 12.0 yang berarti ada 12 item tertera yang didapat dari proses sebelumnya, item sum yang berjumlah 225850.0 dan item mean berjumlah 18820.833

Aggregate	PLB-252	PLB-529	PLB-673	PLB-82002	PLG-CP001	PLG-HB003	PLG-ID001	PLG-PO023	PLG-R1001	PLG-S1004	Total	
BM JAYA - MARIANA	Count	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	
	Sum	0.0	0.0	0.0	1000.0	0.0	0.0	0.0	0.0	0.0	1000.0	
	Mean	7	7	7	7500.0	7	7	7	7	7	7500.0	
CAHYA PEREJABA - PASAR 16	Count	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	
	Sum	0.0	0.0	0.0	0.0	17000.0	0.0	0.0	0.0	0.0	17000.0	
	Mean	7	7	7	7	17000.0	7	7	7	7	17000.0	
DUNIA BUDAJANJOS - LITKOL	Count	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	
	Sum	18000.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18000.0	
	Mean	18000.0	7	7	7	7	7	7	7	7	18000.0	
HEM KAMBAR - LAHAT	Count	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	
	Sum	0.0	0.0	0.0	0.0	10000.0	0.0	0.0	0.0	0.0	10000.0	
	Mean	7	7	7	7	10000.0	7	7	7	7	10000.0	
ISTANA CAHYA - REGAL BIRANGUN	Count	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	13000.0	0.0	0.0	13000.0	
	Mean	7	7	7	7	7	7	13000.0	7	7	13000.0	
NUSA BANGSA - MAREKAPUSA	Count	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	
	Sum	0.0	8000.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8000.0	
	Mean	7	8000.0	7	7	7	7	7	7	7	8000.0	
PERMAI - MANIKKUNINGASA	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	27500.0	0.0	27500.0	
	Mean	7	7	7	7	7	7	7	27500.0	7	27500.0	
RAHAYU ELEKTRONIK	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10700.0	10700.0	
	Mean	7	7	7	7	7	7	7	7	10700.0	10700.0	
SEDERHANA 3 - ARUSKAMAH	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11200.0	11200.0	
	Mean	7	7	7	7	7	7	7	7	11200.0	11200.0	
SIMAR BARU II / BENEY ANHAY - BEMFO	Count	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	
	Sum	0.0	0.0	8100.0	0.0	0.0	0.0	0.0	0.0	0.0	8100.0	
	Mean	7	7	8100.0	7	7	7	7	7	7	8100.0	
<b>Total</b>	Count	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	10.0	
	Sum	18000.0	8000.0	8100.0	2500.0	17000.0	10000.0	13000.0	27500.0	10100.0	11200.0	140350.0
	Mean	18000.0	8000.0	8100.0	2500.0	17000.0	10000.0	13000.0	27500.0	10100.0	11200.0	14035.0

**Gambar 5. Pivot Table 2021**

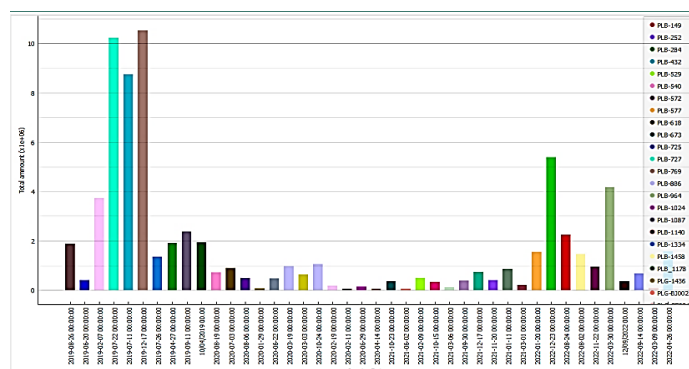
Pada gambar 5 terdapat item count yang berjumlah 10.0 yang berarti ada 10 item tertera yang didapat dari proses sebelumnya, item sum yang berjumlah 140350.0 dan item mean berjumlah 14035.0

Aggregate	PLB-140	PLB-572	PLB-577	PLB-564	PLG-CB001	PLG-CW001	PLG-MB001	PLG-RB001	Total		
CAHYA ABADI / HESABATO - DUMANG	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
	Sum	0.0	0.0	18250.0	0.0	0.0	0.0	0.0	18250.0		
	Mean	7	7	18250.0	7	7	7	7	18250.0		
CAHYA BAHAGIA / EDDY SUSANTO - SIKARANGUN	Count	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	0.0	21000.0	0.0	0.0	21000.0		
	Mean	7	7	7	7	21000.0	7	7	21000.0		
CAHYA ELEKTRIK - EL HAMER	Count	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0		
	Sum	0.0	0.0	0.0	0.0	40000.0	0.0	0.0	40000.0		
	Mean	7	7	7	7	40000.0	7	7	40000.0		
DELTA - PRABANDIHUN	Count	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
	Mean	7	7	7	7	7	7	7	7		
ENW ZIL - TIGU MUKYO	Count	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0		
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	10000.0	10000.0		
	Mean	7	7	7	7	7	7	10000.0	10000.0		
JAYA MANGKABHUPITA SELAMATEKA - PL	Count	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	0.0	15000.0	0.0	0.0	15000.0		
	Mean	7	7	7	7	15000.0	7	7	15000.0		
MIL - EL BARONGIAN	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30000.0		
	Mean	7	7	7	7	7	7	7	30000.0		
PRATAMA MARI SWARI / SIDDHY AMANIKH SH - PLAJU	Count	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	9000.0	0.0	0.0	0.0	9000.0		
	Mean	7	7	7	9000.0	7	7	7	9000.0		
RUMAH TERANG - PRABANDIHUN	Count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
	Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18250.0		
	Mean	7	7	7	7	7	7	7	18250.0		
TUNGGAL INDA - TIGAL BIRANGUN	Count	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
	Sum	17000.0	0.0	0.0	0.0	0.0	0.0	0.0	17000.0		
	Mean	17000.0	7	7	7	7	7	7	17000.0		
<b>Total</b>	Count	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	10.0	
	Sum	18000.0	8000.0	18250.0	9000.0	15000.0	21000.0	40000.0	10000.0	30000.0	196500.0
	Mean	18000.0	8000.0	18250.0	9000.0	15000.0	21000.0	40000.0	10000.0	30000.0	19650.0

**Gambar 6. Pivot Table 2022**

Pada gambar 6 terdapat item count yang berjumlah 10.0 yang berarti ada 10 item tertera yang didapat dari proses sebelumnya, item sum yang berjumlah 196500.0 dan item mean berjumlah 21833.333

Pada tahap Statistik Deskriptif pada EDA ini kita bisa menghitung berbagai banyak hal mulai dari mean, median, modus, deviasi standar, varians, dan kecekungan. Namun karena peneliti ingin mengetahui yang paling penting saja seperti nilai hasil dan nilai rata-rata yang kita inginkan, disini peneliti masih menggunakan orange untuk mengetahui jumlahnya karena orange ini sangat memudahkan kita untuk menghitung otomatis tanpa memerlukan waktu lama, setelah semua sudah di masukkan per tiap satu tahun maka di orange akan ada pilihan untuk kita mau menghitung apa saja. Dari hasil proses tersebut dapat dilihat nilai tiap tahun mulai dari tahun 2019 sampai tahun 2022 didapatkan hasil yang berbeda setiap tahun. Maka bisa disimpulkan ialah jika kita mencari nilai rata-rata maka dibandingkan antar semua tahun nilai rata-rata tertinggi diperoleh dari pivot tabel tahun 2019 yang berjumlah 22855.566 diikuti tahun 2022 yang berjumlah 21.833.333, tahun 2020 berjumlah 18.820.833 dan yang terakhir diikuti tahun 2021 yang berjumlah 14035.0.



**Gambar 7. Bar Plot**

Pada gambar 7 terdapat bar plot yang ada di orange yang terdiri dari invoice date, code customer dan juga total nilai rata-rata tiap tahunnya. Bar plot ini didapat dari menggabungkan hasil dari 4 tahun yang sudah di proses di pivot table sebelumnya agar bisa lebih mudah untuk melihat perbandingannya. Dapat dilihat pada bar plot menunjukkan nilai rata-rata tertinggi ialah pada tahun 2019 diikuti 2022, 2020 dan 2021.

### 3.3 Algoritma K-Means

Setelah data dilakukan processing kemudian data bisa dilakukan clustering menggunakan algoritma k-means sebagai berikut:

a. Penentuan Clustering

Dalam menentukan 3 cluster untuk data ini, analisis k-means dilakukan dengan menggunakan beberapa fitur kunci, yaitu "Quantity" (kuantitas barang), "Unit Price" (harga per unit), "Discount Value" (nilai diskon), dan "Total Amount" (total nilai transaksi). Cluster pertama mungkin mencakup transaksi dengan kuantitas rendah dan nilai total yang terkait, yang mungkin mencerminkan pembelian produk dengan harga rendah. Cluster kedua mungkin mencakup transaksi dengan kuantitas dan nilai total sedang, sementara cluster ketiga mungkin mencakup transaksi dengan kuantitas tinggi dan nilai total yang lebih besar. Pemilihan 3 cluster dipilih setelah melihat analisis metode Elbow yang menunjukkan penurunan inersia yang signifikan pada jumlah cluster tersebut, memberikan wawasan yang cukup tanpa terlalu memecah data.

b. Menentukan Titik Pusat Centroid

Pada Tabel 1 berikut ini terdapat centroid cluster 1 yang meliputi Quantity, Unit Price, Discount Value dan Total amount

**Tabel 1. Centroid Cluster 1**

Feature	Centroid
Quantity	30
Unit Price	1000000
Discount Value	2.00%
Total Amount	500.000.00

Pada Tabel 2 berikut ini terdapat centroid cluster 2 yang meliputi Quantity, Unit Price, Discount Value dan Total amount

**Tabel 2. Centroid Cluster 2**

Feature	Centroid
Quantity	20
Unit Price	1500000
Discount Value	5.00%
Total Amount	300.000.00

Pada Tabel 3 berikut ini terdapat centroid cluster 3 yang meliputi Quantity, Unit Price, Discount Value dan Total amount.

**Tabel 3. Centroid Cluster 3**

Feature	Centroid
Quantity	40
Unit Price	8.000.00
Discount Value	1.00%
Total Amount	600.000.00

c. Menghitung Jarak Data Dengan Centroid Pusat

Setelah titik pusat centroid awal ditentukan, maka langkah berikutnya yang akan dilakukan adalah menghitung jarak masing-masing data terhadap pusat centroid. Proses menghitung masing-masing jarak menggunakan rumus Euclidean distance.

1. Jarak Titik Data dengan Centroid Cluster 1:

$$\sqrt{(25 - 30)^2 + (12000 - 10000)^2 + (3.5\% - 2.00\%)^2 + (450000 - 500000)^2} = 50151.32$$

2. Jarak Titik Data dengan Centroid Cluster 2:

$$\sqrt{(25 - 20)^2 + (12000 - 15000)^2 + (3.5\% - 5.00\%)^2 + (450000 - 300000)^2} = 91067.16$$

3. Jarak Titik Data dengan Centroid Cluster 3:

$$\sqrt{(25 - 40)^2 + (12000 - 8000)^2 + (3.5\% - 1.00\%)^2 + (450000 - 600000)^2} = 204249.13$$

4. Jarak Titik Data dengan Centroid Cluster 1:

$$\sqrt{(35 - 30)^2 + (9000 - 10000)^2 + (1.75\% - 2.00\%)^2 + (550000 - 500000)^2} = 50151.32$$

5. Jarak Titik Data dengan Centroid Cluster 2:

$$\sqrt{(35 - 20)^2 + (9000 - 15000)^2 + (1.75\% - 5.00\%)^2 + (550000 - 300000)^2} = 111254.02$$

6. Jarak Titik Data dengan Centroid Cluster 3:

$$\sqrt{(35 - 40)^2 + (9000 - 8000)^2 + (1.75\% - 1.00\%)^2 + (550000 - 600000)^2} = 165329.69$$

Pada Tabel 4 berikut ini terdapat Iterasi 1 yang didapatkan nilai total cluster untuk setiap data pada Iterasi 1 di bawah yang berjumlah 42 data.

**Tabel 4.** Iterasi 1

No.	Quantity	Unit Price	Discount Value	Total Amount	Cluster
1	12	60,500.00	0.00%	653,400.00	1
2	96	7,000.00	5.00%	638,400.00	2
3	12	13,900.00	5.00%	158,460.00	3
4	24	7,350.00	5.00%	167,580.00	3
5	72	6,850.00	5.00%	468,540.00	2
6	12	12,000.00	5.00%	136,800.00	3
7	24	13,800.00	10.00%	298,080.00	1
8	24	17,850.00	10.00%	385,560.00	1
9	12	22,600.00	10.00%	244,080.00	1
10	24	7,900.00	0.00%	189,600.00	3
11	10	34,500.00	0.00%	345,000.00	3
12	12	21,500.00	3.00%	250,260.00	3
13	12	21,500.00	3.00%	250,260.00	3
14	24	20,500.00	3.00%	477,240.00	3
15	12	20,500.00	3.00%	238,620.00	3
16	10	4,900.00	0.00%	49,000.00	3
17	10	6,250.00	0.00%	62,500.00	3
18	10	6,750.00	0.00%	67,500.00	3
19	10	8,250.00	0.00%	82,500.00	3
20	3	4,500.00	0.00%	13,500.00	3
21	3	11,250.00	0.00%	33,750.00	3
22	3	15,650.00	0.00%	46,950.00	3
23	3	21,000.00	0.00%	63,000.00	3
24	6	68,250.00	5.00%	389,025.00	3
25	2	775,000.00	0.00%	1,550,000.00	3
26	10	7,650.00	0.00%	76,500.00	3
27	10	9,800.00	0.00%	98,000.00	3
28	3	29,000.00	0.00%	87,000.00	3
29	3	39,000.00	0.00%	117,000.00	3
30	3	57,000.00	0.00%	171,000.00	3
31	120	3,750.00	0.00%	450,000.00	1
32	12	17,500.00	0.00%	210,000.00	3
33	12	19,750.00	0.00%	237,000.00	3
34	60	9,250.00	0.00%	555,000.00	1
35	12	10,500.00	0.00%	126,000.00	1
36	3	47,250.00	5.00%	134,662.50	2
37	3	68,250.00	5.00%	194,512.50	2
38	1	915,000.00	0.00%	915,000.00	1
39	1	1,950,000.00	0.00%	1,950,000.00	1
40	1	3,100,000.00	0.00%	3,100,000.00	1
41	96	3,500.00	5.00%	319,200.00	1
42	48	3,200.00	5.00%	145,920.00	1

**Tabel 5.** Total Cluster Iterasi 1

Cluster	Total
1	12
2	4
3	26

Pada Tabel 6 berikut terdapat Centroid Baru setelah dilakukan Iterasi 1

**Tabel 6.** Centroid Baru

Cluster	Quantity	Unit Price	Discount Value	Total Amount
1	40.57	10,898.70	1,085.87	95,915.22
2	2.67	45,833.33	3,125.00	53,333.33
3	16.50	18,666.67	1,833.33	193,500.00

Pada Tabel 7 berikut ini terdapat Iterasi 2 yang didapatkan nilai total cluster untuk setiap data pada Iterasi 2 di bawah yang berjumlah 42 data.

**Tabel 7.** Iterasi 2

No.	Quantity	Unit Price	Discount Value	Total Amount	Cluster
1	12	60,500.00	0.00%	653,400.00	1
2	96	7,000.00	0.00%	638,400.00	1
3	12	13,900.00	0.00%	158,460.00	1
4	24	7,350.00	0.00%	167,580.00	1
5	72	6,850.00	0.00%	468,540.00	1
6	12	12,000.00	0.00%	136,800.00	1
7	24	13,800.00	0.00%	298,080.00	1
8	24	17,850.00	0.00%	385,560.00	1
9	12	22,600.00	0.00%	244,080.00	1
10	24	7,900.00	0.00%	189,600.00	1
11	10	34,500.00	0.00%	345,000.00	1
12	12	21,500.00	0.00%	250,260.00	1
13	12	21,500.00	0.00%	250,260.00	1
14	24	20,500.00	0.00%	477,240.00	1
15	12	20,500.00	0.00%	238,620.00	1
16	10	4,900.00	0.00%	49,000.00	1
17	10	6,250.00	0.00%	62,500.00	1
18	10	6,750.00	0.00%	67,500.00	1
19	10	8,250.00	0.00%	82,500.00	1
20	3	4,500.00	0.00%	13,500.00	1
21	3	11,250.00	0.00%	33,750.00	2
22	3	15,650.00	0.00%	46,950.00	2
23	3	21,000.00	0.00%	63,000.00	2
24	6	68,250.00	5.00%	389,025.00	2
25	2	775,000.00	0.00%	1,550,000.00	2
26	10	7,650.00	0.00%	76,500.00	2
27	10	9,800.00	0.00%	98,000.00	2
28	3	29,000.00	0.00%	87,000.00	2
29	3	39,000.00	0.00%	117,000.00	2
30	3	57,000.00	0.00%	171,000.00	2
31	120	3,750.00	0.00%	450,000.00	1
32	12	17,500.00	0.00%	210,000.00	1
33	12	19,750.00	0.00%	237,000.00	1
34	60	9,250.00	0.00%	555,000.00	1
35	12	10,500.00	0.00%	126,000.00	1
36	3	47,250.00	5.00%	134,662.50	3
37	3	68,250.00	5.00%	194,512.50	3
38	1	915,000.00	0.00%	915,000.00	3
39	1	1,950,000.00	0.00%	1,950,000.00	3
40	1	3,100,000.00	0.00%	3,100,000.00	3
41	96	3,500.00	5.00%	319,200.00	1
42	48	3,200.00	5.00%	145,920.00	1

**Tabel 8.** Total Cluster Iterasi 2

Cluster	Total
1	27
2	10
3	5

Pada tabel 9 terdapat Centroid Baru (Iterasi 2)

**Tabel 9.** Centroid baru Iterasi 2

Cluster	Quantity	Unit Price	Discount Value	Total Amount
1	36.8	9,048.33	2.33%	318,294.00
2	5.33	845,416.67	1.17%	1,503,333.33
3	2.618	300	2.00%	1,668,835

Pada Tabel 10 berikut ini terdapat Iterasi 3 yang didapatkan nilai total cluster untuk setiap data pada Iterasi 3 di bawah yang berjumlah 42 data. Data yang dimiliki Iterasi 3 ini sama seperti iterasi sebelumnya, sehingga tidak perlu dilakukan Iterasi selanjutnya.

**Tabel 10.** Iterasi 3

No.	Cluster	Quantity	Unit Price	Discount Value	Total Amount	Label
1	1	12	7,000.00	350.00	84,000.00	C1
2	1	96	7,725.00	3,615.00	674,550.00	C1
3	1	12	13,900.00	1,158.33	187,200.00	C1
4	1	24	6,800.00	1,680.00	163,200.00	C1
5	1	72	6,425.00	3,003.75	442,800.00	C1
6	1	12	11,600.00	1,160.00	130,800.00	C1
7	1	24	13,450.00	2,043.75	324,000.00	C1
8	1	24	17,525.00	3,711.00	472,200.00	C1
9	1	12	22,600.00	2,226.67	212,400.00	C1
10	1	24	7,900.00	0.00	189,600.00	C1
11	1	10	34,500.00	0.00	345,000.00	C1
12	1	12	21,500.00	1,842.00	208,800.00	C1
13	1	12	21,500.00	1,842.00	208,800.00	C1
14	1	24	20,500.00	2,280.00	254,400.00	C1
15	1	12	20,500.00	1,764.00	190,800.00	C1
16	1	10	4,900.00	0.00	49,000.00	C1
17	1	10	6,250.00	0.00	62,500.00	C1
18	1	10	6,750.00	0.00	67,500.00	C1
19	1	10	8,250.00	0.00	82,500.00	C1
20	1	3	4,500.00	0.00	13,500.00	C1
21	2	3	11,250.00	0.00%	33,750.00	C2
22	2	3	15,650.00	0.00%	46,950.00	C2
23	2	3	21,000.00	0.00%	63,000.00	C2
24	2	6	68,250.00	5.00%	389,025.00	C2
25	2	2	775,000.00	0.00%	1,550,000.00	C2
26	2	10	7,650.00	0.00%	76,500.00	C2
27	2	10	9,800.00	0.00%	98,000.00	C2
28	2	3	29,000.00	0.00%	87,000.00	C2
29	2	3	39,000.00	0.00%	117,000.00	C2
30	2	3	57,000.00	0.00%	171,000.00	C2
31	1	120	3,750.00	0.00%	450,000.00	C1
32	1	12	17,500.00	0.00%	210,000.00	C1
33	1	12	19,750.00	0.00%	237,000.00	C1
34	1	60	9,250.00	0.00%	555,000.00	C1
...	...	...	...	...	...	...
42	1	48	3,200.00	5.00%	145,920.00	C1

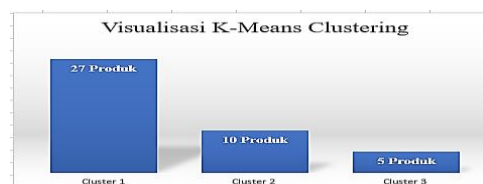
Pada tabel 11 berikut ini terdapat Total Cluster Iterasi 3

**Tabel 11.** Total Cluster Iterasi 3

Cluster	Total
1	27
2	10
3	5

### 3.4 Visualisasi Algoritma K-Means

Diagram berikut merupakan hasil dari visualisasi data yang terdapat dari data penjualan produk pada rentang tahun 2019 sampai dengan tahun 2022. Visualisasinya dapat dilihat pada gambar 8 berikut.



**Gambar 8.** Visualisasi K-Means Clustering Penjualan Produk

Pada gambar 8 di atas didapatkan visual bahwa produk apa saja yang masuk ke dalam kategori cluster 1 yaitu ada 27 produk, selanjutnya ada 10 produk yang termasuk ke dalam cluster 2. Sedangkan terakhir ada 5 produk masuk ke dalam cluster 3. Dari jumlah bobot terakhir, maka cluster 1 ialah cluster penjualan yang tinggi diikuti dengan cluster 2 penjualan sedang, kemudian cluster 3 penjualan rendah.

#### 4. KESIMPULAN

Berdasarkan hasil dari metode EDA dan K-Means mendapatkan hasil perbandingan penjualan selama 4 tahun nilai ragam rata-rata yang tinggi menunjukkan bahwa penjualan pada tahun 2019 memiliki nilai rata-rata yang besar diantara tahun 2020,2021, dan 2022. Pada hasil visualisasi dapat disimpulkan bahwa produk apa saja yang masuk ke dalam kategori cluster 1 seperti Cable Ties, Stop Kontak, Cable chips, Fitting Gantung, Raket Nyamuk, Kabel Setrika, Steker, Lampu Putih, Test Pen, Pvc Insulation, Universal Adaptor, Sekring Otomatis, CB, Fitting Flapon, Cable Clips, Test Pen Sku, Mcb Hitam, Ceiling Mini Fan, LED, OB Seri, Minyak goreng, Desk Fan, GZ/Ohm Saklar, Sekring Otomatis, Fitting, Downlight, Cable Ties Putih. Sedangkan produk Isolasi BG, Cable Ties, Kabel Setrika, Mini Fan t, IB Engkel, IB Stop, Lampu LX, , Ib Dus, MCB Ip, Mcb Biru termasuk ke dalam cluster 2. Lalu Stabilizer, AC 1/2 PK, Socket ,Campaign Setrika Maspion HA 110/220 V, Wall Fan Maspion MWF41K masuk ke dalam cluster 3. Dari jumlah bobot terakhir, maka cluster 1 ialah cluster penjualan yang tinggi diikuti dengan cluster 2 penjualan sedang, kemudian cluster 3 penjualan rendah.

#### REFERENCES

- [1] W. Hastomo, B. S. A. Karno, Sudjiran, D. Arif, and E. S. Moreta, "Exploratory Data Analysis Untuk Data Belanja Pelanggan dan Pendapatan Bisnis," *Infotekmesin*, vol. 13, no. 2, pp. 314-321, 2022, doi: 10.35970/infotekmesin.v13i2.1547.
- [2] P. Shukaku, "Sekilas Shukaku," 2014. [http://www.shukaku.co.id/about\\_sekilasshukaku.html](http://www.shukaku.co.id/about_sekilasshukaku.html).
- [3] M. E.-K. Kesuma and R. Iskandar, "Analisis Toko dan Asal Toko Fashion Pria di Shopee Menggunakan Data Scrapping dan Exploratory Data Analysis," *Maj. Ilm. Teknol. Elektro*, vol. 21, no. 1, p. 127, 2022, doi: 10.24843/mite.2022.v21i01.p17.
- [4] M. Harahap, F. Rozi, Y. Yennimar, and S. D. Siregar, "Analisis Wawasan Penjualan Supermarket dengan Data Science," *Data Sci. Indones.*, vol. 1, no. 1, pp. 1-7, 2021, doi: 10.47709/dsi.v1i1.1173.
- [5] D. Aryanti and J. Setiawan, "Visualisasi Data Penjualan dan Produksi PT Nitto Alam Indonesia Periode 2014-2018," vol. IX, no. 2, pp. 86-91, 2018.
- [6] A. Wibowo, K. Faisah, and Y. Devianto, "Analisa Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis Pada PT. Telkominfra," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2292-2304, 2022, doi: 10.35957/jatisi.v9i3.2737.
- [7] M. Harahap, Y. Lubis, and Z. Situmorang, "Analisis Pemasaran Bisnis dengan Data Science : Segmentasi Kepribadian Pelanggan berdasarkan Algoritma K-Means Clustering," *Data Sci. Indones.*, vol. 1, no. 2, pp. 76-88, 2022, doi: 10.47709/dsi.v1i2.1348.
- [8] S. Oktaviani and A. Bahtiar, "Implementasi Algoritma K-Means Dalam Pengelompokan Data," vol. 2, no. 1, pp. 188-196, 2023.
- [9] A. Mukhid, *METODOLOGI PENELITIAN PENDEKATAN KUANTITATIF*. Surabaya: CV. Jakad Media Publishing, 2021.
- [10] J. Beno, A. Silen, and M. Yani, "DAMPAK PANDEMI COVID-19 KEGIATAN EKSPOR IMPOR," *Saintek Marit.*, vol. 22, no. 2, pp. 120-125, 2022.
- [11] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75-81, 2021, doi: 10.33884/jif.v9i02.3755.
- [12] A. M. Sinaga and D. B. Arianto, "Analisis Consumer Behaviour Dalam Pembelian Wines Di Perusahaan Ifood," *J. Pijar*, vol. 2, pp. 63-74, 2023, [Online]. Available: <https://e-journal.naurendigiton.com/index.php/pmb/article/view/745%0Ahttps://e-journal.naurendigiton.com/index.php/pmb/article/download/745/281>.
- [13] F. V. P. Samosir, P. M. Loudry, and D. A. Erik, "Exploratory Data Analysis terhadap Kepadatan Penumpang Kereta Rel Listrik," vol. 7, pp. 449-467, 2021.
- [14] F. Amin, D. S. Anggraeni, and Q. Aini, "Penerapan Metode K-Means dalam Penjualan Produk Souq.Com," *Appl. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 7-14, 2022, doi: 10.15408/aism.v5i1.22534.
- [15] N. Silalahi, "Penentuan Strategi Promosi Universitas Budi Darma Menggunakan Algoritma K-Means Clustering," *TIN Terap. Inform. Nusan.*, vol. 1, no. 1, pp. 40-46, 2020.
- [16] I. M. Rian Yuliawan, D. Care Khrisne, and P. Arya Mertasana, "Penerapan Algoritma K-Means Clustering dalam Penentuan Nilai Huruf pada Permainan Susun Kata Bahasa Bali," *J. SPEKTRUM*, vol. 6, no. 3, p. 87, 2019, doi: 10.24843/spektrum.2019.v06.i03.p12.
- [17] M. S. Nawawi, F. Sembiring, and A. Erfina, "Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris," *Semin. Nas. Teknol. Inf. dan Komunikasi-2021*, pp. 789-797, 2021, [Online]. Available: <http://prosiding.unipma.ac.id/index.php/SENATIK/article/view/1837%0Ahttp://prosiding.unipma.ac.id/index.php/SENATIK/article/viewFile/1837/1723>.
- [18] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021, doi:



10.21107/nero.v6i2.237.

- [19] D. P. Hidayatullah, R. I. Rokhmawati, and ..., “Analisis Pemetaan Pelanggan Potensial Menggunakan Algoritma K-Means dan LRFM Model Untuk Mendukung Strategi Pengelolaan Pelanggan (Studi Pada ...,” ... *Inf. dan Ilmu* ..., vol. 2, no. 8, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1769>.
- [20] B. D. Mudzakkir, “Pengelompokan Data Penjualan Produk Pada Pt Advanta Seeds Indonesia Menggunakan Metode K-Means,” *J. Mhs. Tek. Inform.*, vol. 2, no. 2, pp. 34–40, 2019.