



**Kampus
Merdeka**
INDONESIA JAYA

FINAL PROJECT

Business Analitik

"Analysis and Data Visualization Sales in Acme Sports Company"

**Dataset
1**

Anggota Kelompok:

Zulvan Avivi Akmal Firdaus

Vada Zafira

01111840000042

02411840000164

Dosen Pembimbing

Dr. Dra. Kartika Fithriasari, M.Si

Adatul Mukarromah, S.Si., M.Si

Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si.

Asisten Dosen

Tiza Ayu Virania

Hasri Wiji Aqsari

Muhammad Adlansyah Muda



----- *Halaman ini Sengaja Dikosongkan* -----

KATA PENGANTAR

Puji syukur kepada Allah SWT berkat rahmat dan karunianya-Nya Laporan *Final Project* ini dapat diselesaikan dengan lancar tanpa ada hambatan. Laporan berjudul “*Final Project Business Analytics: Analysis and Data Visualization Sales in Acme Sports Company*” merupakan laporan yang ditujukan untuk memenuhi syarat kelulusan mata kuliah Bisnis Analitik dalam program Kredensial Mikro Mahasiswa Indonesia (KMMI) di Institut Teknologi Sepuluh Nopember.

Penulis menyadari bahwa laporan ini dapat diselesaikan dengan baik atas bantuan dan bimbingan dari berbagai pihak. Maka dari itu, penulis berterimakasih kepada semua pihak yang secara langsung maupun tidak langsung memberikan kontribusi terhadap pengerjaan laporan ini. Kemudian secara khusus kami ucapkan terima kasih kepada Ibu Adatul Mukarromah, Bapak Dedy Dwi Prastyo, dan Ibu Kartika Fithriasari selaku pengajar mata kuliah Bisnis Analitik yang telah membimbing dan mengarahkan penulis dalam penyusunan laporan ini dari awal hingga laporan terselesaikan.

Dalam penulisan laporan yang dibuat masih terdapat beberapa kekurangan yang dilakukan, sehingga kami mengharapkan saran dan kritik dari semua pihak sebagai masukan terhadap perbaikan pengerjaan laporan selanjutnya. Semoga laporan ini dapat bermanfaat dan menginspirasi bagi para pembaca, terima kasih.

Surabaya, 25 September 2021

Penulis



----- *Halaman ini Sengaja Dikosongkan* -----

DAFTAR ISI

KATA PENGANTAR	II
DAFTAR ISI.....	IV
DAFTAR GAMBAR	VI
DAFTAR LAMPIRAN.....	VIII
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA	4
2.1 Bisnis Analitik.....	4
2.2 Analisa Deskriptif	4
2.3 Visualisasi Data.....	4
2.4 K-mean Clustering	5
BAB III METODOLOGI PENELITIAN	6
3.1 Sumber Data	6
3.2 Metode Penelitian.....	6
BAB IV ANALISIS DAN PEMBAHASAN.....	10
4.1 Statistik Deskriptif.....	10
4.1.1 Analisis Gambaran Umum Dataset.....	10
4.1.2 Analisis Total Penjualan	16
4.2 <i>Clustering K-Means</i>	17
4.2.1 Variabel Jumlah Order, Keuntungan Order dan Harga Produk	17
BAB VI PENUTUP	20
6.1 Kesimpulan.....	20
6.2 Saran.....	21
DAFTAR PUSTAKA	II
LAMPIRAN.....	IV



----- *Halaman ini Sengaja Dikosongkan* -----

DAFTAR GAMBAR

Gambar 1. 1 Acme Sport <i>Company</i> Logo.....	1
Gambar 3. 1 Data Penjualan Acme Sport Companu.....	6
Gambar 3. 2 Metodologi Pengerjaan Laporan.....	7
Gambar 3. 3 Metodologi <i>Clustering</i> K-means.....	8
Gambar 3. 4 Metodologi Statistika Deskriptif.....	9
Gambar 4. 1.Dataset yang akan dianalisa	10
Gambar 4. 2. Grafik produk paling laris	11
Gambar 4. 3. Grafik order sukses per negara.....	11
Gambar 4. 4. Grafik keterlambatan pengiriman dan negara	12
Gambar 4. 5 Grafik Total Order Tiap Jam.....	13
Gambar 4. 6 Grafik Total Order Tiap Bulan.....	13
Gambar 4. 8 Grafik Total Order Tiap Bulan.....	14
Gambar 4. 9 Grafik Total Order Tiap Bulan.....	15
Gambar 4. 8 Grafik Order <i>Canceled</i> Tiap <i>Customer Segment</i>	15
Gambar 4. 8 Grafik Order <i>Canceled</i> Tiap <i>Customer Segment</i>	15
Gambar 4. 8 Grafik Order Tiap <i>Customer Segment</i>	16
Gambar 4. 8 Grafik Order Tiap <i>Customer Segment</i>	16
Gambar 4. 7. Grafik total penjualan tiap waktu	16
Gambar 4. 10 WSS <i>Clustering</i>	17
Gambar 4. 11 Silhouette <i>Clustering</i>	17
Gambar 4. 12 <i>Clustering k-means</i> , n=2	18
Gambar 4. 13 <i>Clustering k-means</i> , n=3	18
Gambar 4. 14 <i>Clustering k-means</i> , n=4	18
Gambar 4. 15 <i>Clustering k-means</i> , n=5	18
Gambar 4. 16 <i>Confusing Matrix</i>	18



----- *Halaman ini Sengaja Dikosongkan* -----

DAFTAR LAMPIRAN

Gambar 1 <i>Source Code</i> Plot Negara Paling Banyak Order Sukses	IV
Gambar 2 <i>Source Code</i> Jumlah Barang Terjual Paling Banyak	IV
Gambar 3 <i>Source Code</i> Hubungan Keterlambatan Pengiriman Dengan Negara.....	IV
Gambar 4 <i>Source Code</i> Plot Time Series Total Penjualan Per Region	V
Gambar 5. <i>Source Code</i> 1 <i>Clustering K-Means</i>	V
Gambar 6. <i>Source Code</i> 2 <i>Clustering K-Means</i>	V
Gambar 7. <i>Source Code</i> 3 <i>Clustering K-Means</i>	VI
Gambar 8. <i>Source Code</i> 5 <i>Clustering K-Means</i>	VI
Gambar 9. <i>Source Code</i> 1 Analisis <i>Customer Segment</i>	VII
Gambar 10. <i>Source Code</i> 2 Analisis <i>Customer Segment</i>	VII
Gambar 11. <i>Source Code</i> Tipe Pembayaran	VIII
Gambar 12. <i>Source Code</i> 1 Analisis Waktu Order Produk.....	VIII
Gambar 13. <i>Source Code</i> 2 Analisis Waktu Order Produk.....	IX



----- *Halaman ini Sengaja Dikosongkan* -----

BAB I

PENDAHULUAN

Pada bab ini, akan dipaparkan latar belakang, rumusan masalah, tujuan proyek, manfaat proyek serta batasan dan asumsi yang digunakan dalam laporan “Final Project Business Analytics: Analysis and Data Visualization Data Sales in Acme Sports *Company*”.

1.1 Latar Belakang Masalah

Perkembangan teknologi menjadi semakin pesat terutama pada sektor industri, dimana era saat ini disebut dengan revolusi industri 4.0. Dalam revolusi industri 4.0 teknologi informasi menjadi faktor penting dalam aktivitas operasional dan perkembangan bisnis di berbagai industri. Dengan perkembangan ini, maka semakin ketat pula persaingan bisnis baik dalam skala kecil maupun skala besar. Oleh karena itu, beberapa industri menerapkan *data science* pada perusahaannya dengan harapan dapat membangun dan mengembangkan strategi bisnis sesuai tren masa mendatang, sehingga dapat bersaing dengan kompetitor lainnya.

Data science adalah gabungan dari ilmu matematika dan statistika dengan ilmu komputer. Tujuan dari *data science* sendiri adalah untuk menganalisis suatu data dari himpunan data, baik yang berskala kecil maupun berskala besar menjadi suatu informasi yang bermanfaat. Dalam suatu industri keberadaan *data science* telah menjadi kunci untuk menjawab kebutuhan perusahaan yang mampu membawa perusahaan pada jalur kesuksesan.



Gambar 1. 1 Acme Sport *Company* Logo

Sumber: <https://www.indiamart.com/acme-sports/aboutus.html>, 2021

Acme Sport *Company* merupakan salah satu industri yang bergerak di bidang peralatan olahraga. Acme Sport *Company* memproduksi peralatan olahraga multi dan cincin tinju dan memproduksi peralatan olahraga, cincin tinju dan papan skor LED di India sejak 2007 di bawah Nama Merek "Acme Sports". Acme Sports memiliki visi dan misi untuk dapat memperkuat posisinya sebagai pemimpin inovatif dalam produksi Peralatan Olahraga dan menyediakan

Peralatan Olahraga skala besar di India. Untuk dapat menggapai cita-cita tersebut maka perusahaan membutuhkan strategi bisnis yang baik dan sesuai. Oleh karena itu, dalam penelitian ini dilakukannya analisis dan visualisasi data penjualan perusahaan sehingga diharapkan dapat membantu memudahkan pengambilan keputusan mengenai strategi bisnis perusahaan.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas dapat dirumuskan permasalahan yang akan diselesaikan sebagai berikut ini:

1. Apa produk yang paling diminati oleh *customer* berdasarkan jumlah *order* produk?
2. Apa negara pengiriman yang melakukan *order* produk paling banyak, berdasarkan jumlah *order* produk?
3. Apakah lokasi negara pengiriman dengan tingkat keterlambatan pengiriman terbanyak?
4. Bagaimana waktu yang paling sering digunakan *customer* untuk melakukan *order* produk?
5. Apa media pembayaran yang paling sering digunakan *customer* untuk melakukan *order* produk?
6. Apakah segmen *customer* mempengaruhi total *order* dan membatalkan produk?
7. Bagaimana analisis penjualan produk dalam kurun waktu 3 tahun?
8. Bagaimana pengelompokan data penjualan berdasarkan jumlah *order*, keuntungan *perorder* dan harga produk?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah memecahkan permasalahan yang telah diidentifikasi sebelumnya, yaitu sebagai berikut:

1. Menganalisis yang paling diminati oleh *customer* berdasarkan jumlah *order* produk Acme Sport Company.
2. Menganalisis negara pengiriman yang melakukan *order* produk paling banyak, berdasarkan jumlah *order* produk Acme Sport Company.
3. Menganalisis lokasi negara pengiriman dengan tingkat keterlambatan pengiriman terbanyak di Acme Sport Company.
4. Menganalisis waktu yang paling sering digunakan *customer* untuk melakukan *order* produk di Acme Sport Company.
5. Menganalisis media pembayaran yang paling sering digunakan *customer* untuk melakukan *order* produk di Acme Sport Company.



6. Menganalisis segmen *customer* mempengaruhi total *order* dan membatalkan produk di Acme Sport Company.
7. Menganalisis penjualan produk dalam kurun waktu 3 tahun di Acme Sport Company.
8. Pengelompokkan data penjualan Acme Sport Company berdasarkan karakteristik berdasarkan jumlah *order*, keuntungan perorder dan harga produk di Acme Sport Company.

1.4 Manfaat Penelitian

Manfaat yang dari penyusunan laporan ini adalah memberikan informasi mengenai asal negara terbanyak yang melakukan *order*, produk paling diminati, waktu *favorite customer* melakukan *order*, serta pengelompokkan data penjualan dari Acme Sport Company. Informasi tersebut dapat digunakan untuk membantu dan memudahkan pengambilan keputusan mengenai strategi bisnis perusahaan.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah bahwa data yang digunakan adalah data penjualan Acme Sport Company dibatasi pada tahun 2015 – 2018.

BAB II

TINJAUAN PUSTAKA

Pada bab ini, akan dipaparkan beberapa tinjauan pustaka yang berkaitan dan dapat membantu dalam pengerjaan laporan *final project* ini.

2.1 Bisnis Analitik

Bisnis analitik dimulai dari sebuah *data set* (kumpulan dari koleksi beberapa data) atau bisa juga berasal dari *database*. Analitik dapat didefinisikan sebagai proses yang meliputi teknik – teknik statistika, perangkat lunak sistem informasi, dan operasi beberapa metodologi untuk mengeksplor, memvisualisasikan, menemukan dan menghubungkan pola atau tren dalam data tersebut. Secara singkatnya analitik mengubah data menjadi suatu informasi yang berguna. Ada berbagai teknik untuk analitik bisnis, namun yang umum digunakan 3 jenis kategori berikut, yakni deskriptif, prediktif dan preskriptif. (Marc J. Schniederjans dkk, 2014).

Ketiga kategori tersebut yang membedakan antara analitik biasa dengan bisnis analitik. Dimana analitik berfokus pada upaya untuk menghasilkan suatu informasi yang berguna dari sebuah *data set* dengan mengandalkan salah satu dari tiga kategori di atas, bisnis analitik mengambil langkah yang lebih jauh dengan menambahkan analitik menggunakan ketiganya secara langsung sehingga dapat menghasilkan sebuah informasi yang baru, unik, dan berharga untuk membantu sebuah organisasi bisnis dalam pengambilan keputusannya. (Marc J. Schniederjans dkk, 2014).

2.2 Analisa Deskriptif

Dengan menggunakan aplikasi dari teknik statistika sederhana untuk mendeskripsikan atau menjelaskan informasi apa saja yang terdapat pada sebuah *data set*. Selain itu analisis deskriptif juga dapat digunakan untuk mencari tren – tren yang mungkin ada pada *data set* tersebut, yang mana tujuannya adalah untuk mendapat gambaran kasar bagaimana bentuk data tersebut dan kriteria apa yang berpotensi untuk mengidentifikasi tren atau perilaku bisnis di masa depan. (Marc J. Schniederjans dkk, 2014).

2.3 Visualisasi Data

Data merupakan sekumpulan nama dan besarannya. Berupa pengelompokan, deskripsi dan pengukuran. Bisa pula berupa tanggal dan lokasi. Data biasanya ditampilkan pada sebuah tabel, dengan berbagai label kolom dan baris, beserta besaran – besaran dan variabel yang telah ada di dalamnya. Data juga merupakan unsur penting dalam melakukan pengambilan keputusan pada berbagai bidang. Oleh karena itu diperlukan visualisasi data tidak pada setiap data yang ada, melainkan pada sebagian besarnya. (Andy Kirk, 2019).

Data dalam bentuk tabel memang dapat menampilkan keseluruhan data dengan lengkap, namun akan sangat kesulitan untuk menjawab cakupan yang lebih dari sekedar berapa nilai sebuah variabel dalam data. Representasi sebuah data dalam visualisasi harus dikaitkan dengan pengambilan keputusan tentang bagaimana bentuk visualisasinya dan juga aspek mana yang ingin kita tonjolkan dalam suatu data. Hal ini bertujuan agar audiens yang kita tuju dapat dengan tepat dan cepat menangkap maksud atau *insight* yang kita angkat dari data yang kita analisa. (Andy Kirk, 2019).

2.4 K-mean Clustering

Metode ini merupakan salah satu metode clustering atau klasifikasi yang mengelompokkan data menjadi grup – grup K, dimana K merupakan jumlah kuantitatif grup yang diinginkan. Proses algoritma yang dilakukan mengidentifikasi kandidat awal untuk grup K dan kemudian secara interaktif akan mencari anggota lain pada data set untuk dirata – rata menjadi grup K tertentu. Proses pengambilan anggota grupnya didasarkan dengan memaksimalkan jarak dari anggota K awal dengan keseluruhan iterasi tiap data pada *data set*. (Marc J. Schniederjans dkk, 2014).

Proses K-mean clustering ini dapat memberikan cara yang cepat untuk mengklasifikasikan data ke tiap – tiap grup yang berbeda. Apalagi di masa sekarang yang banyak sekali keputusan – keputusan cepat harus dibuat berdasarkan data set yang sangat besar ukurannya, maka metode ini dapat dijadikan salah satu alternatif cepat, dan efisien untuk mendapat *insight* dari *data set* yang ada. (Marc J. Schniederjans dkk, 2014).

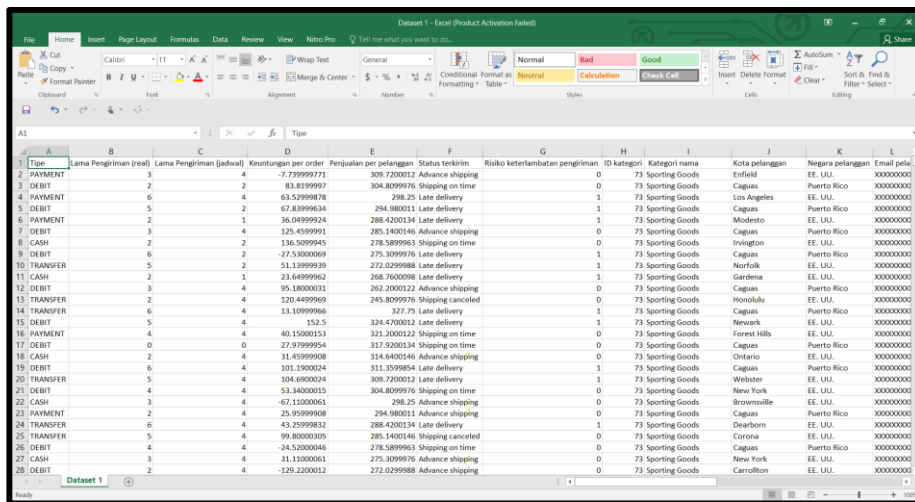
BAB III

METODOLOGI PENELITIAN

Pada bab ini, akan dijelaskan metodologi penelitian yang digunakan dalam pengerjaan laporan *final project* ini.

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang berbentuk dokumentasi penjualan pada tahun 2015-2018 perusahaan Acme Sport *Company*. Dokumen data penjualan tersebut tertulis dalam *software* Ms. Excel sebagai berikut ini:



Type	Lama Pengiriman (real)	Lama Pengiriman (jadwal)	Keuntungan per order	Penjualan per pelanggan	Status terkirim	Risiko keterlambatan pengiriman	ID kategori	Kategori nama	Kota pelanggan	Negara pelanggan	Email pelab
PAYMENT	3	4	-7.79999771	309.7200012	Advance shipping	0	73	Sporting Goods	Enfield	EE, UU.	XXXXXXXXXX
DEBIT	2	2	83.8199997	304.8099976	Shipping on time	0	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
PAYMENT	6	4	63.5299978	298.25	Late delivery	1	73	Sporting Goods	Los Angeles	EE, UU.	XXXXXXXXXX
DEBIT	5	2	67.82999634	294.9800011	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
PAYMENT	2	1	36.0499924	288.4200134	Late delivery	1	73	Sporting Goods	Modesto	EE, UU.	XXXXXXXXXX
DEBIT	3	4	125.4599991	285.1400146	Advance shipping	0	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
CASH	2	2	136.5099945	278.5899963	Shipping on time	0	73	Sporting Goods	Irvington	EE, UU.	XXXXXXXXXX
DEBIT	6	2	-27.53000069	275.3099976	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
TRANSFER	5	2	51.1399939	272.0299988	Late delivery	1	73	Sporting Goods	Norfolk	EE, UU.	XXXXXXXXXX
CASH	2	1	23.6499962	268.7600008	Late delivery	1	73	Sporting Goods	Gardena	EE, UU.	XXXXXXXXXX
DEBIT	3	4	95.18000031	262.2000122	Advance shipping	0	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
TRANSFER	2	4	120.4499969	245.8099976	Shipping canceled	0	73	Sporting Goods	Honolulu	EE, UU.	XXXXXXXXXX
TRANSFER	6	4	13.10999966	327.75	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
DEBIT	5	4	152.5	324.4700012	Late delivery	1	73	Sporting Goods	Newark	EE, UU.	XXXXXXXXXX
PAYMENT	4	4	40.15000153	321.2000122	Shipping on time	0	73	Sporting Goods	Forest Hills	EE, UU.	XXXXXXXXXX
DEBIT	0	0	27.9799954	317.9200134	Shipping on time	0	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
CASH	2	4	31.40999908	314.6400146	Advance shipping	0	73	Sporting Goods	Ottawo	EE, UU.	XXXXXXXXXX
DEBIT	6	4	101.1900024	311.3599854	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
TRANSFER	5	4	104.6900024	309.7200012	Late delivery	1	73	Sporting Goods	Webster	EE, UU.	XXXXXXXXXX
DEBIT	4	4	53.34000015	304.8099976	Shipping on time	0	73	Sporting Goods	New York	EE, UU.	XXXXXXXXXX
CASH	3	4	67.11000061	298.25	Advance shipping	0	73	Sporting Goods	Brenneville	EE, UU.	XXXXXXXXXX
PAYMENT	2	4	25.95999908	294.9800011	Advance shipping	0	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
TRANSFER	6	4	43.25999832	288.4200134	Late delivery	1	73	Sporting Goods	Dearborn	EE, UU.	XXXXXXXXXX
TRANSFER	5	4	99.80000305	285.1400146	Shipping canceled	0	73	Sporting Goods	Corona	EE, UU.	XXXXXXXXXX
DEBIT	4	4	-24.52000046	278.5899963	Shipping on time	0	73	Sporting Goods	Caguas	Puerto Rico	XXXXXXXXXX
CASH	3	4	31.11000061	275.3099976	Advance shipping	0	73	Sporting Goods	New York	EE, UU.	XXXXXXXXXX
DEBIT	2	4	-129.2200012	272.0299988	Advance shipping	0	73	Sporting Goods	Carrollton	EE, UU.	XXXXXXXXXX

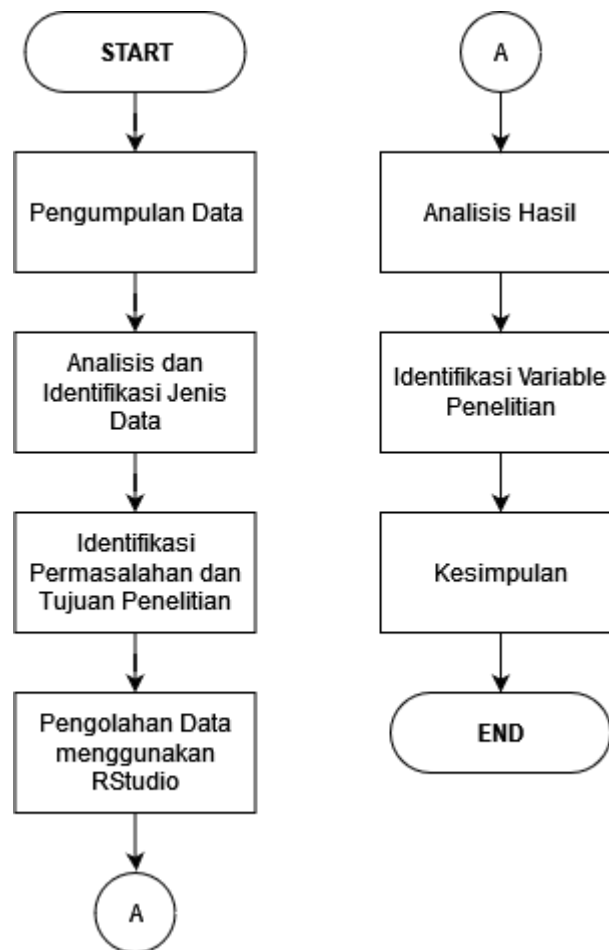
Gambar 3. 1 Data Penjualan Acme Sport Company

Sumber: Acme Sport *Company*, 2021

Data dokumentasi penjualan tersebut merupakan data sekunder yang diperoleh secara tidak langsung yang diberikan pada perkuliahan Bisnis Analitik program Kredensial Mikro Mahasiswa Indonesia (KMMI) di Institut Teknologi Sepuluh Nopember.

3.2 Metode Penelitian

Penelitian ini dilakukan dengan mengelola dan menganalisis suatu data penjualan yang dimiliki oleh perusahaan Acme Sport. Dengan pengelolaan dan analisis Informasi tersebut diharapkan dapat digunakan untuk membantu dan memudahkan pengambilan keputusan mengenai strategi bisnis perusahaan. Oleh karena itu dilakukan beberapa metode penelitian terhadap data yaitu metode *descriptive analysis* dan *clustering K-means*. Berikut ini adalah gambaran struktur metodologi penelitian dalam laporan ini:



Gambar 3. 2 Metodologi Pengerjaan Laporan

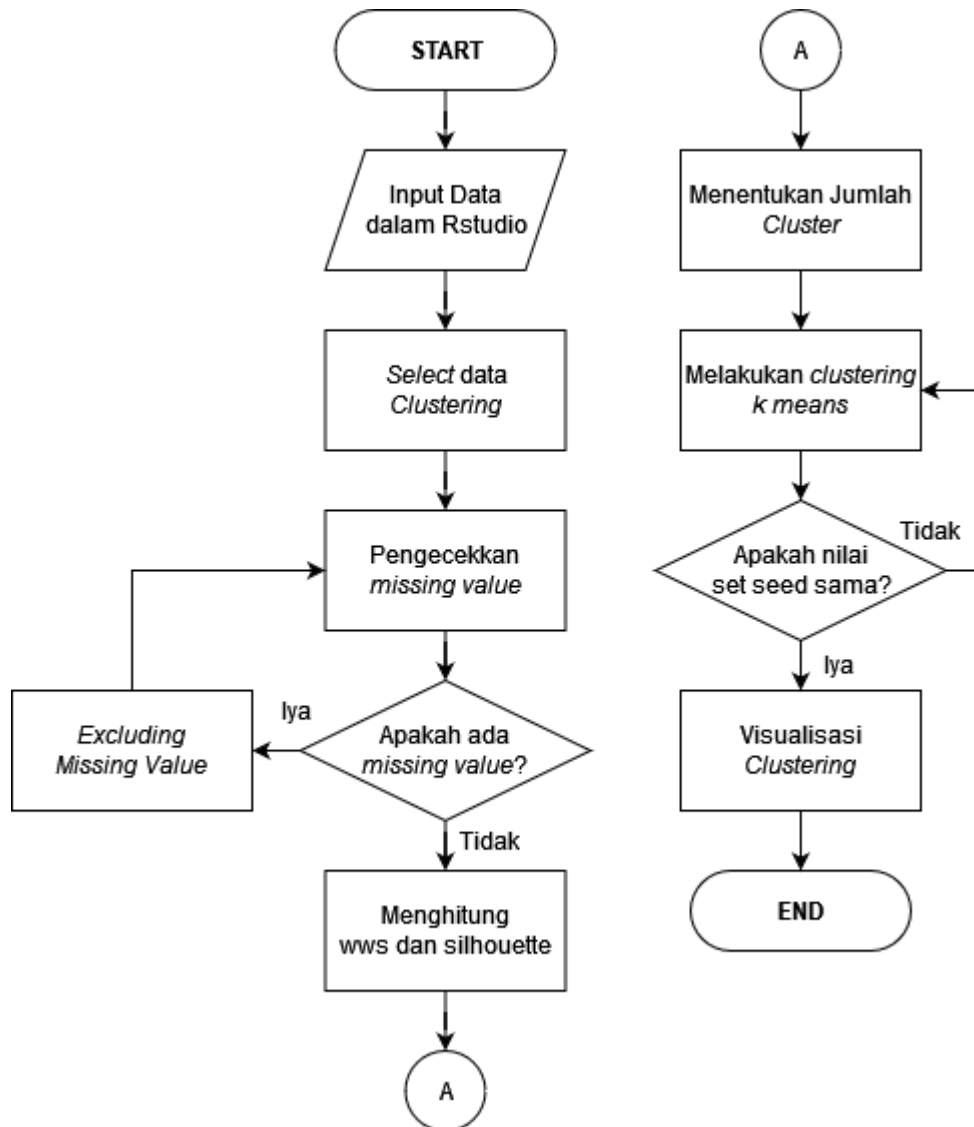
Sumber: Dokumen Pribadi, 2021

Variable merupakan karakteristik atau atribut yang ditetapkan terhadap objek amatan. Penentuan variable penelitian dilakukan agar analisis data dapat terfokuskan dengan konsisten dan berorientasi kepada tujuan. Variable data yang digunakan adalah sebagai berikut:

1. Variabel data nama produk dan jumlah penjualan/*order*
2. Variabel data nama negara *customer* dan jumlah *order* produk.
3. Variabel data negara *customer* dan keterlambatan pengiriman.
4. Variabel data waktu (jam) dan jumlah *order* produk
5. Variabel data waktu (bulan) dan jumlah *order* produk.
6. Variabel data tipe pembayaran *customer* dan jumlah *order* produk.
7. Variabel data *customer segment* dan jumlah *order* produk.
8. Variable data waktu (tahun) dan total penjualan.
9. Variabel jumlah *order* produk, keuntungan *order* produk, dan harga produk/

Metode yang digunakan dalam analisis dan visualisasi data penjualan Acme Sport Company adalah metode statistik yaitu *descriptive analysis* dan *clustering unsupervised learning method*. *Unsupervised learning method* adalah metode pengelompokkan data tanpa adanya label data, dimana pengelompokkan data tersebut dapat dibuat menjadi 2, 3 dan

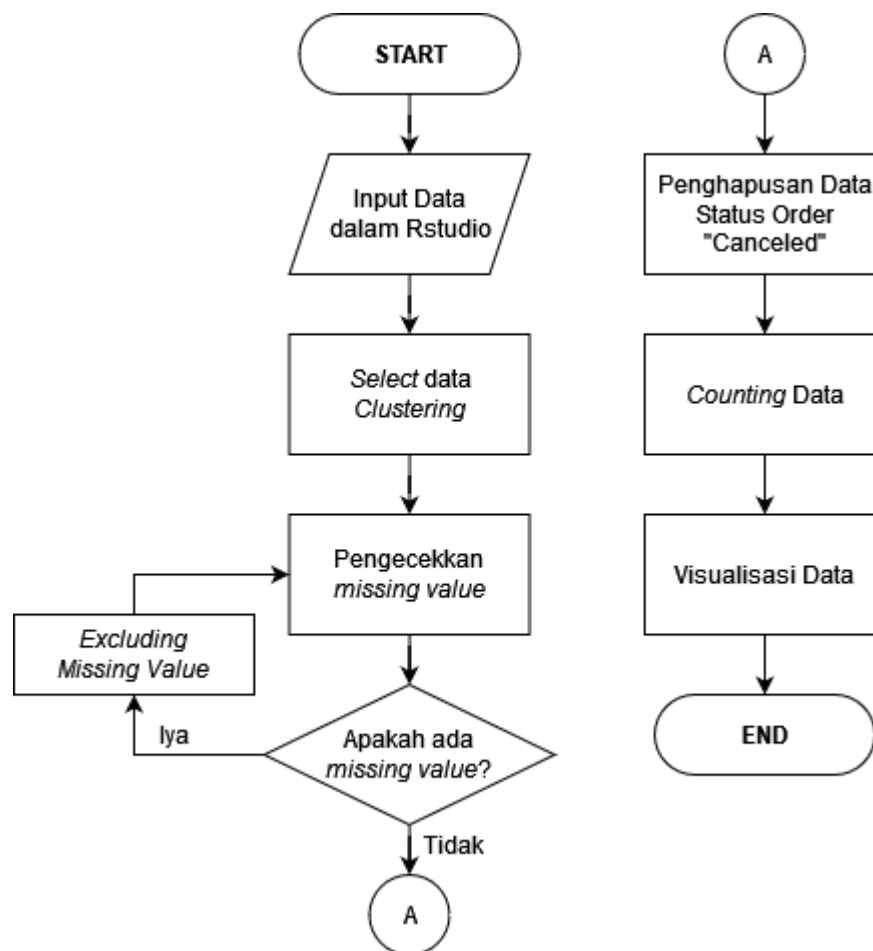
seterusnya. Pengelompokkan tersebut terjadi berdasarkan data yang memiliki kesamaan karakteristik didalamnya. Penelitian dengan metode tersebut dilakukan dengan bantuan *software* R Studio. Rstudio adalah perangkat lunak yang terintegrasi dan dirancang untuk pemrograman komputasi statistik dan grafik. Berikut ini adalah langkah implementasi metode *clustering* k-means dengan menggunakan Rstudio:



Gambar 3. 3 Metodologi *Clustering* K-means

Sumber: Dokumen Pribadi, 2021

Statistik deskriptif adalah analisis statistik yang dilakukan dengan memberikan gambaran secara umum mengenai karakteristik dari masing-masing variabel penelitian yang digunakan. Dimana karakteristik tersebut meliputi seperti *mean*, modus, median, maksimal, minimum dst. Berikut ini adalah langkah implementasi metode statistik deskriptif dengan menggunakan Rstudio:



Gambar 3. 4 Metodologi Statistika Deskriptif

Sumber: Dokumen Pribadi, 2021

BAB IV

ANALISIS DAN PEMBAHASAN

Pada bab ini, akan dijelaskan analisis dan pembahasan dari pengerjaan laporan *final project* ini.

4.1 Statistik Deskriptif

Statistik deskriptif adalah analisis statistik yang dilakukan dengan memberikan gambaran secara umum mengenai karakteristik dari masing-masing variabel penelitian yang digunakan. Dimana karakteristik tersebut meliputi seperti *mean*, modus, median, maksimal, minimum dst.

4.1.1 Analisis Gambaran Umum Dataset

Data yang diberikan berupa data penjualan dari perusahaan Acme Sport selama kurun waktu 3 tahun, data tersebut terdiri dari 53 kolom dan 179337 baris, seperti ditunjukkan pada gambar berikut,

```
> glimpse(data1)
Rows: 179,337
Columns: 53
$ Tipe                                <chr> "PAYMENT", "DEBIT", "PAYME~
$ Lama.Pengiriman..real.              <int> 3, 2, 6, 5, 2, 3, 2, 6, 5,~
$ Lama.Pengiriman..jadwal.            <int> 4, 2, 4, 2, 1, 4, 2, 2, 2,~
$ Keuntungan.per.order                <dbl> -7.74, 83.82, 63.53, 67.84~
$ Penjualan.per.pelanggan             <dbl> 309.72, 304.81, 298.25, 29~
$ Status.terkirim                     <chr> "Advance shipping", "Shipp~
$ Risiko.keterlambatan.pengiriman      <int> 0, 0, 1, 1, 1, 0, 0, 1, 1,~
$ ID.kategori                         <int> 73, 73, 73, 73, 73, 73, 73~
$ Kategori.nama                       <chr> "Sporting Goods", "Sportin~
$ Kota.pelanggan                      <chr> "Enfield", "Caguas", "Los ~
$ Negara.pelanggan                   <chr> "EE. UU.", "Puerto Rico", ~
$ Email.pelanggan                     <chr> "xxxxxxxx", "xxxxxxxx", ~
$ Customer.Fname                      <chr> "Whitney", "Aileen", "Came~
$ ID.pelanggan                        <int> 16750, 16749, 16748, 16747~
$ Customer.Lname                      <chr> "Lindsay", "Trevino", "Hen~
$ Password.pelanggan                  <chr> "xxxxxxxx", "xxxxxxxx", ~
$ Customer.Segment                    <chr> "Corporate", "Corporate", ~
$ Status.pelanggan                    <chr> "CT", "PR", "CA", "PR", "C~
$ Customer.Street                     <chr> "9089 Round Willow Route",~
$ Zipcode.pelanggan                   <int> 6082, 725, 90043, 725, 953~
$ Id.departemen                       <int> 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ Nama.departemen                     <chr> "Fitness", "Fitness", "Fit~
$ Latitude                            <dbl> 42.00019, 18.29552, 33.985~
$ Longitude                           <dbl> -72.59647, -66.03705, -118~
$ pasar                               <chr> "Pacific Asia", "Pacific A~
```

Gambar 4. 1.Dataset yang akan dianalisa

Sumber: Dokumen Pribadi, 2021

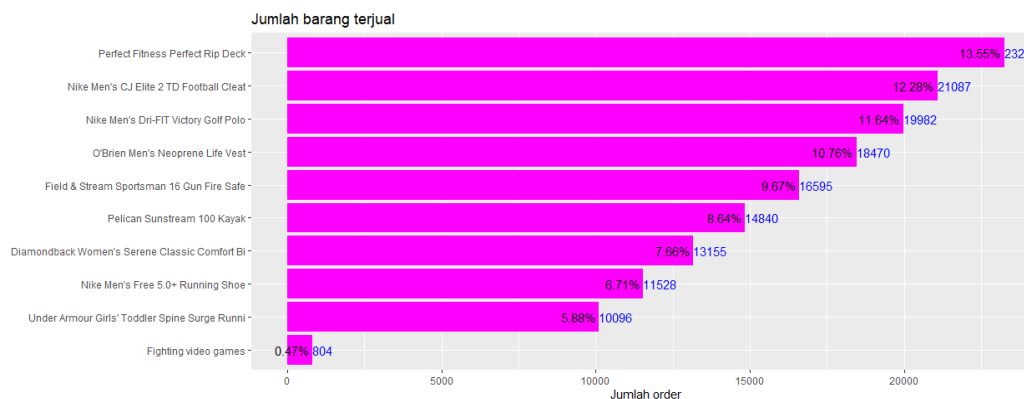
Data tersebut berisi berbagai variabel mulai dari lama waktu pengiriman, tanggal pengiriman, tipe pembayaran, daerah order dan lain – lain.

Dari data tersebut, kami ingin mengetahui beberapa variabel – variabel penjualan penting seperti barang yang paling laku, asal negara yang paling banyak sukses membeli barang, hubungan antara keterlambatan dengan letak geografis, dan jam – jam yang lalu lintas pembeliannya paling ramai. Diharapkan dari beberapa variabel tersebut akan dapat memberi gambaran awal bagaimana keadaan dari penjualan perusahaan tersebut.

Sebagai catatan, setiap analisa yang dilakukan disini diolah dengan bahasa pemrograman R, *source code* untuk setiap analisa akan ditampilkan pada bagian lampiran.

4.1.1.1 Jenis Barang Paling Laku

Merupakan analisa yang paling mendasar dari sebuah data hasil penjualan. Tujuannya adalah untuk melihat produk apa yang paling diminati oleh pasar, sehingga perusahaan dapat menentukan langkah selanjutnya untuk semakin mengembangkan barang yang diminati pasar. Dataset yang diberikan memberikan data sebanyak 118 produk unik yang terdata, dari sana, diambil 10 jumlah penjualan terbanyak per produk sebagai berikut,



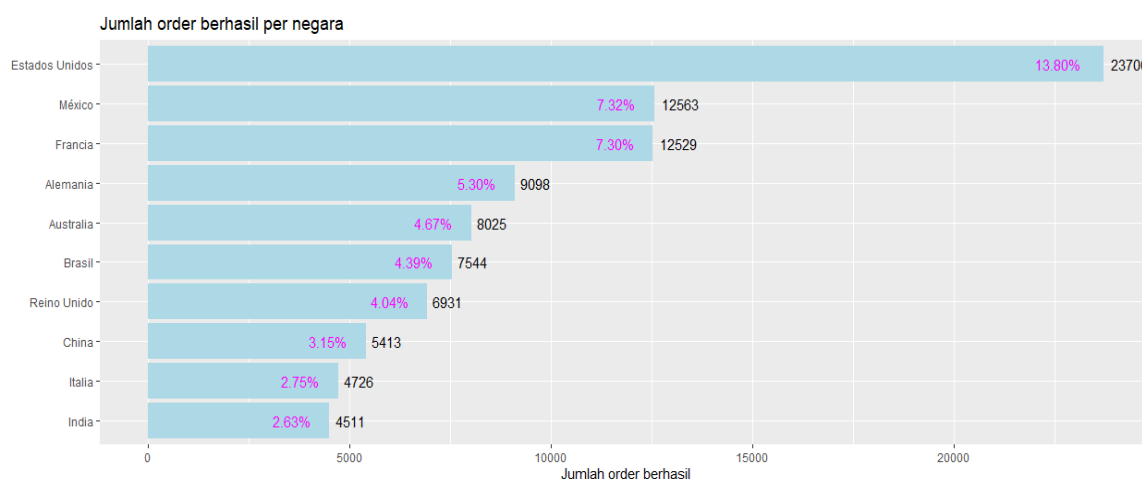
Gambar 4. 2. Grafik produk paling laris

Sumber: Dokumen Pribadi, 2021

Berdasarkan grafik tersebut diperoleh bahwa, produk paling tinggi penjualannya adalah “*Perfect Fitness Perfect Rip Deck*” dengan penjualan sejumlah 23.263 transaksi yang merupakan 13,55 % dari total seluruh penjualan produk.

4.1.1.2 Daftar Negara dan Pembelian Sukses Tertinggi

Sesuai namanya, analisis ini memberikan gambaran tentang dari negara mana paling banyak transaksi pembelian barang yang sukses terjadi. Analisis ini tentu untuk melihat, pasar terbesar dari perusahaan ini kira – kira ada pada penduduk di daerah mana. Dari pengolahan data diperoleh,



Gambar 4. 3. Grafik order sukses per negara

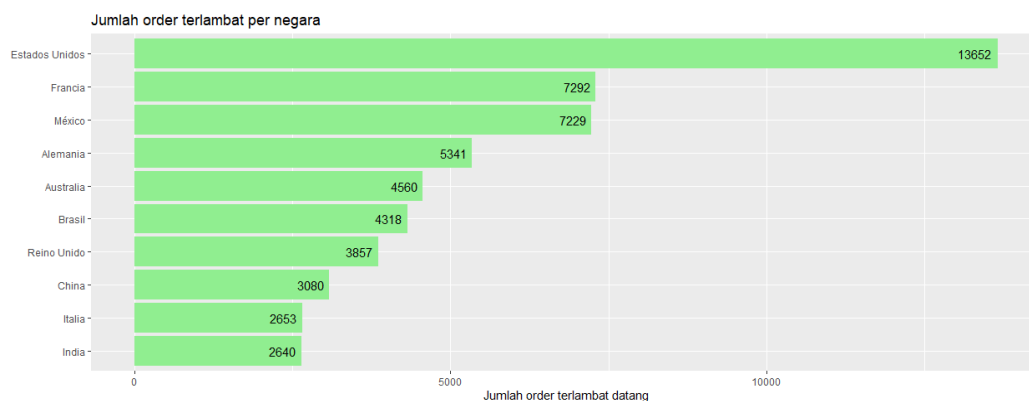
Sumber: Dokumen Pribadi, 2021

terlihat order sukses paling banyak dilakukan di Amerika Serikat dengan total penjualan 23.700 yang merupakan 13,8% dari total penjualan, diikuti oleh Meksiko dan seterusnya.

Dari data tersebut maka perusahaan dapat mengambil langkah selanjutnya dalam mengoptimalkan sumber dayanya agar lebih efisien dan terfokus, sehingga dapat mengurangi beban kerja perusahaan itu sendiri.

4.1.1.3 Daftar Negara dan Tingkat Keterlambatan Pengiriman

Analisis ini ditujukan untuk melihat bagaimana alur pengiriman dari barang yang telah diorder oleh pembeli. Dengan melihat analisis ini perusahaan dapat memantau apakah jasa pengiriman yang mereka gunakan sudah baik, dan juga apakah ada masalah dengan pengiriman di suatu negara tertentu. Dari hasil analisis diperoleh,



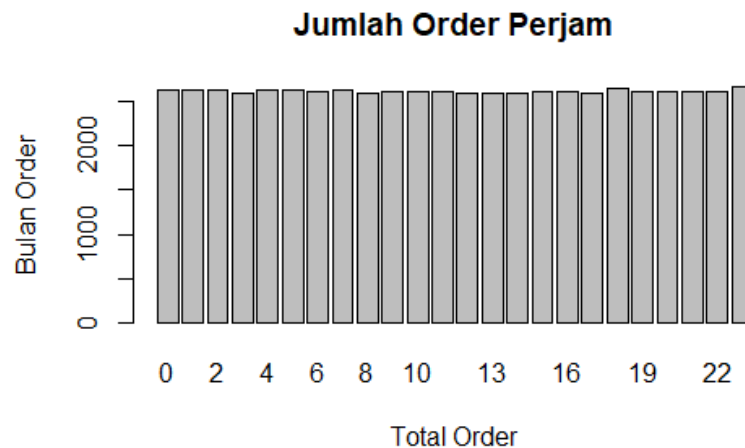
Gambar 4. 4. Grafik keterlambatan pengiriman dan negara

Sumber: Dokumen Pribadi, 2021

Berdasarkan hasil analisis yang diperoleh ternyata didapat tingkat keterlambatan pengiriman lebih dipengaruhi oleh banyaknya jumlah order dan bukan oleh letak geografis tujuan pengiriman.

4.1.1.4 Jam Penjualan Tertinggi

Pada bagian ini tujuannya adalah lebih spesifik untuk mengetahui bagaimana penjualan yang telah dilakukan perusahaan berdasarkan persebaran jumlah order sukses pada tiap jamnya. Total order sukses akan dianalisis perjamnya, sehingga dapat dilihat pada jam berapakah perusahaan harus memfokuskan *customer service*-nya. Setelah dilakukan analisis maka diperoleh gambar sebagai berikut ini:



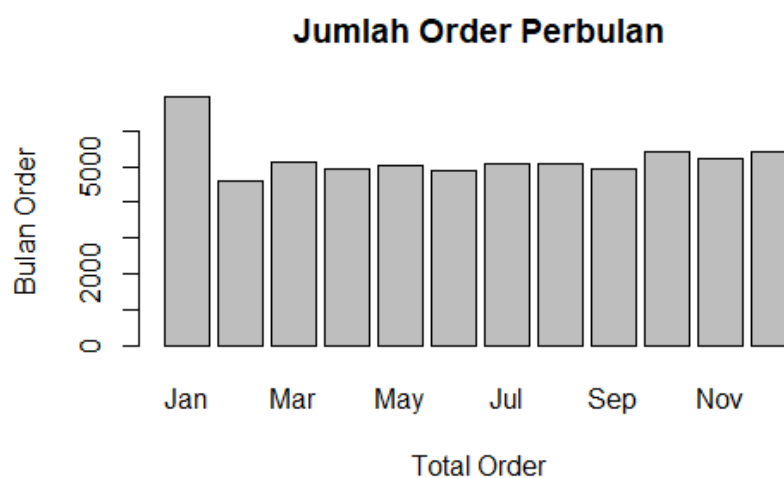
Gambar 4. 5 Grafik Total Order Tiap Jam

Sumber: Dokumen Pribadi, 2021

Berdasarkan grafik diatas, maka diketahui bahwa tidak ada waktu spesifik dimana *customer* melakukan *order* terhadap produk perusahaan. Maka dapat dikatakan bahwa waktu pembelian tidak mempengaruhi jumlah order yang diterima oleh perusahaan dari *customer*.

4.1.1.5 Bulan Penjualan Tertinggi

Pada bagian ini tujuannya adalah lebih spesifik untuk mengetahui bagaimana penjualan yang telah dilakukan perusahaan berdasarkan persebaran jumlah order sukses pada tiap bulannya. Total order sukses akan dianalisis perbulan dalam kurun waktu 3 tahun, sehingga dapat dilihat pada jam berapakah perusahaan harus memfokuskan strategi bisnisnya. Setelah dilakukan analisis maka diperoleh gambar sebagai berikut ini:



Gambar 4. 6 Grafik Total Order Tiap Bulan

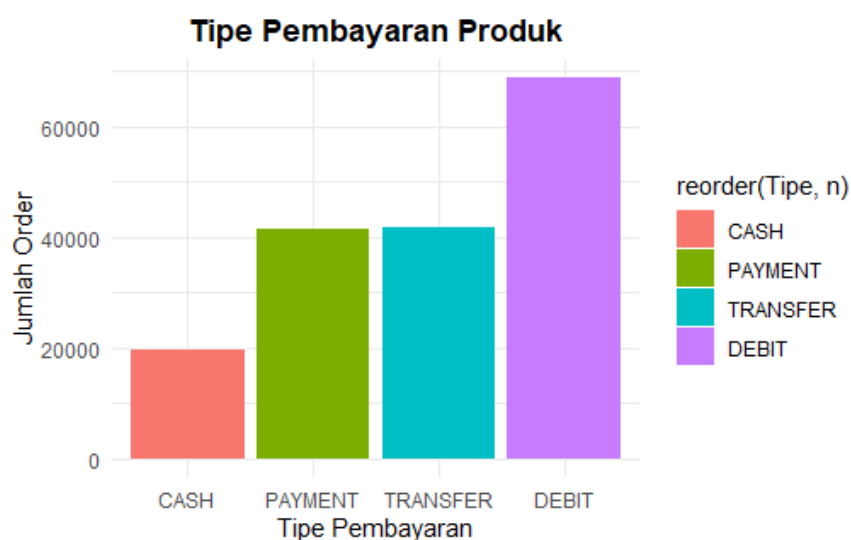
Sumber: Dokumen Pribadi, 2021

Berdasarkan grafik diatas, maka diketahui bahwa bulan Januari, merupakan bulan dimana banyak *customer* melakukan pembelian produk perusahaan. Kemudian pada bulan

februari, merupakan bulan dimana *customer* melakukan paling sedikit *order* terhadap produk perusahaan. Hal ini tentu saja dapat menjadi suatu bahan diskusi dalam pengambilan keputusan maupun pembelajaran bagi masa depan perusahaan.

4.1.1.6 Tipe Pembayaran Terbanyak

Pada bagian ini tujuannya adalah lebih spesifik untuk mengetahui apa media pembayaran yang paling sering digunakan oleh *customer*, sehingga perusahaan dapat dengan mudah mengambil strategi pemasara berupa pemberian diskon, iklan, event dan lain sebagainya. Total order sukses akan dianalisis dalam kurun waktu 3 tahun, yang kemudian dilakukan penjumlahan total terhadap tipe pembayaran yang dilakukan. Setelah dilakukan analisis maka diperoleh gambar sebagai berikut ini:

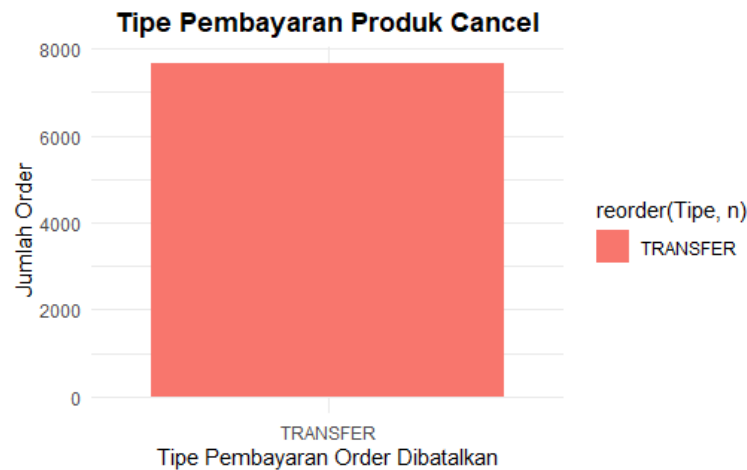


Gambar 4. 7 Grafik Total Order Tiap Bulan

Sumber: Dokumen Pribadi, 2021

Berdasarkan grafik diatas, maka diketahui bahwa tipe media pembayaran yang paling banyak digunakan oleh *customer* dalam mengorder produk perusahaan adalah debit. Sedangkan tipe media pembayaran yang paling sedikit digunakan oleh *customer* saat mengorder produk perusahaan adalah *cash*/kas. Hal ini tentu saja dapat menjadi suatu bahan diskusi dalam pengambilan keputusan maupun pembelajaran bagi masa depan perusahaan.

Selain itu, juga dilakukan analisis terhadap tipe pembayaran yang dilakukan oleh *customer* saat *order* dibatalkan. Hal ini dilakukan untuk membantu perusahaan dalam penyusunan strategi bisnisnya. Berikut gambaran dari analisis tipe pembayaran oleh order yang dibatalkan:



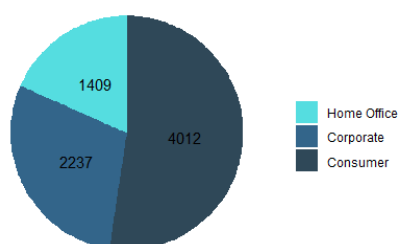
Gambar 4. 8 Grafik Total Order Tiap Bulan

Sumber: Dokumen Pribadi, 2021

Berdasarkan grafik diatas, maka diketahui bahwa tipe media pembayaran yang paling banyak digunakan oleh *customer* yang membatalkan order produk perusahaan adalah *transfer*. Hal ini tentu saja dapat menjadi suatu bahan diskusi dalam pengambilan keputusan maupun pembelajaran bagi masa depan perusahaan.

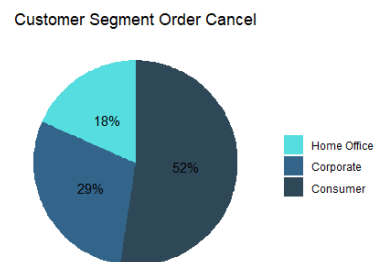
4.1.1.7 Segmentasi *Customer*

Pada bagian ini tujuannya adalah lebih spesifik untuk mengetahui segmen *customer* mana yang paling sering melakukan order produk perusahaan. Total order sukses akan dianalisis dalam kurun waktu 3 tahun, yang kemudian dilakukan penjumlahan total terhadap order yang dilakukan. Setelah dilakukan analisis maka diperoleh gambar sebagai berikut ini:



Gambar 4. 9 Grafik Order *Canceled* Tiap *Customer Segment*

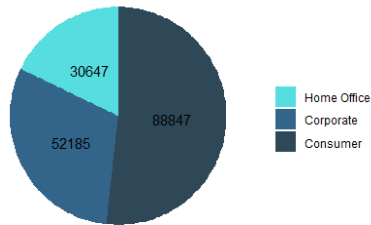
Sumber: Dokumen Pribadi, 2021



Gambar 4. 10 Grafik Order *Canceled* Tiap *Customer Segment*

Sumber: Dokumen Pribadi, 2021

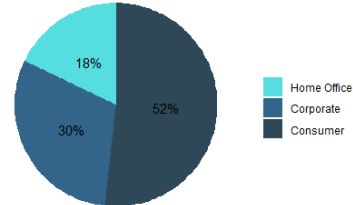
Customer Segment Order Success



Gambar 4. 11 Grafik Order Tiap Customer Segment

Sumber: Dokumen Pribadi, 2021

Customer Segment Order Success



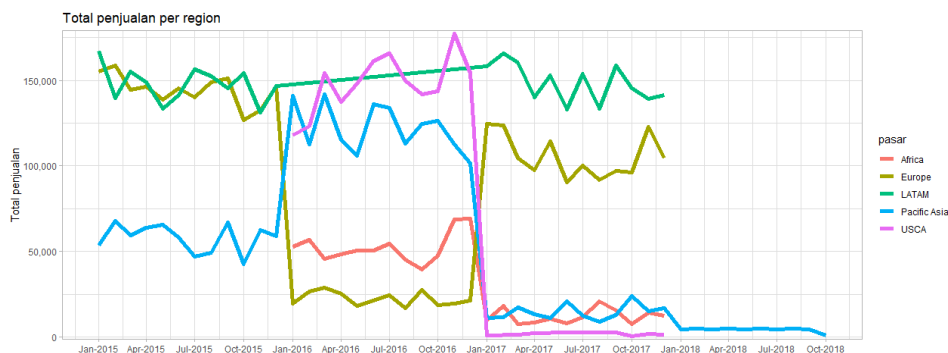
Gambar 4. 12 Grafik Order Tiap Customer Segment

Sumber: Dokumen Pribadi, 2021

Berdasarkan gambar *chart* diatas, diketahui presentase *customer segment* pada *order* yang dibatalkan dengan *order* yang tidak dibatalkan adalah sama dengan perbedaan nilai kuantitatif didalamnya. Maka diketahui bahwa *customer segment* tidak mempengaruhi pembatalan *order* produk perusahaan.

4.1.2 Analisis Total Penjualan

Pada bagian ini tujuannya adalah lebih spesifik untuk mengetahui bagaimana penjualan yang telah dilakukan perusahaan selama kurun waktu 3 tahun ini. Total penjualan akan dianalisis per region dimana transaksi dilakukan, hal ini untuk melihat apakah ada suatu peristiwa khusus yang berdampak pada penjualan produk di daerah tersebut di suatu waktu tertentu. Setelah dilakukan analisis diperoleh,



Gambar 4. 13. Grafik total penjualan tiap waktu

Sumber: Dokumen Pribadi, 2021

Dari grafik tersebut terlihat jumlah penjualan untuk region USCA, Asia pasifik, dan Afrika mengalami penurunan cukup drastis di Januari 2017, sebaliknya, pada region Eropa terjadi kenaikan drastis di Januari 2017 setelah terjadi penurunan di Januari 2016. Sementara untuk

region LATAM tergolong memiliki tingkat penjualan yang stabil dengan melihat grafik yang naik-turun tidak terlalu jauh.

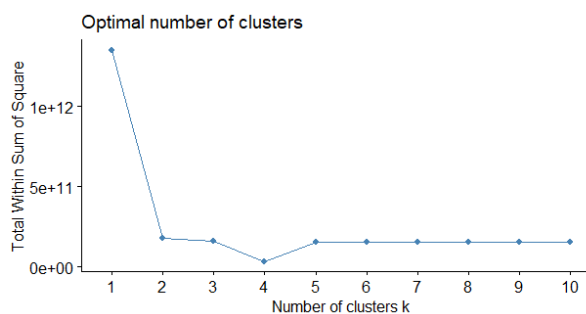
Data tersebut bisa diakibatkan oleh 2 hal, adanya kesalahan pada input data atau memang ada suatu peristiwa khusus di tahun 2016. Alasan tersebut dilandasi karena sepanjang tahun 2016 untuk semua region terdapat perubahan nilai yang cukup signifikan (ada yang naik dan ada yang turun drastis), terkecuali untuk LATAM dimana grafiknya tiba – tiba menjadi stabil dan naik secara linear. Hal ini tentu saja dapat menjadi suatu bahan diskusi dalam pengambilan keputusan maupun pembelajaran bagi masa depan perusahaan.

4.2 Clustering K-Means

Pengklasteran *k-means* dilakukan menggunakan Rstudio dengan membuat algoritme untuk membagi n data pengamatan menjadi k kelompok dengan jumlah yang sesuai, dimana tiap kelompok memiliki rata-rata nilai terdekat.

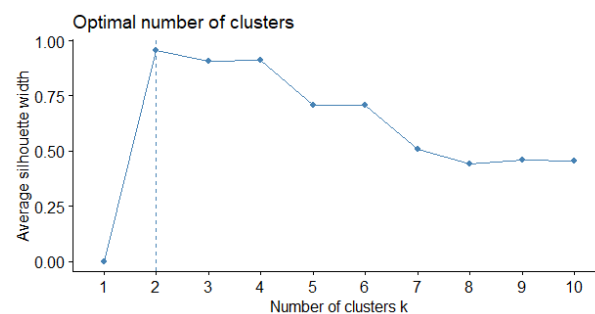
4.2.1 Variabel Jumlah Order, Keuntungan Order dan Harga Produk

Klustering ini dilakukan dengan harapan dapat membantu perusahaan untuk mengelompokkan produknya, sehingga dapat memudahkan perusahaan mengambil keputusan strategi. Langkah awal untuk penentuan jumlah n kluster adalah dengan memperhitungkan dan menganalisis WSS serta silhouette yang dimiliki. Berikut adalah hasil Rstudio WSS dan silhouette data:



Gambar 4. 14 WSS Clustering

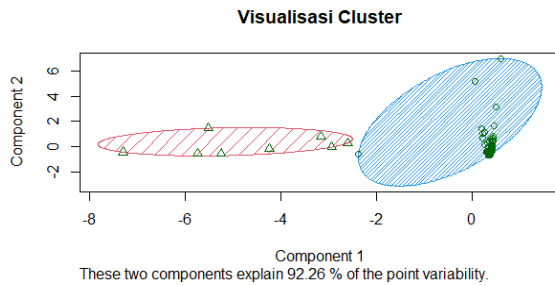
Sumber: Dokumen Pribadi, 2021



Gambar 4. 15 Silhouette Clustering

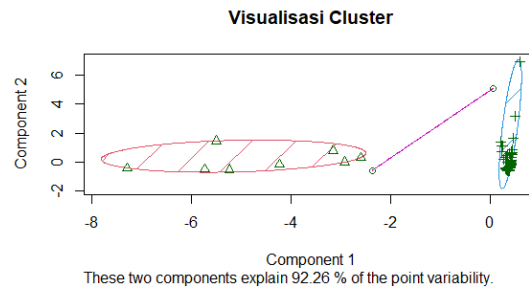
Sumber: Dokumen Pribadi, 2021

Berdasarkan grafik diatas maka didapatkan bahwa n -kluster yang paling sesuai adalah $n=2$, namun untuk melakukan pemastian kembali maka dilakukan beberapa *clustering k-means*. Berikut adalah hasil Rstudio *clustering k-means*:



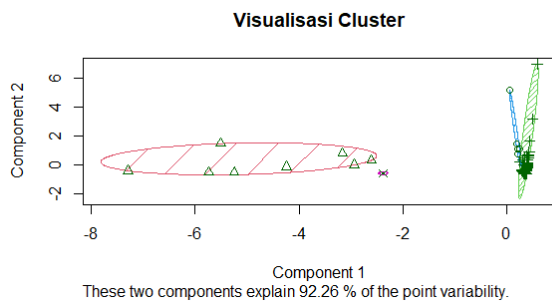
Gambar 4. 16 Clustering *k-means*, n=2

Sumber: Dokumen Pribadi, 2021



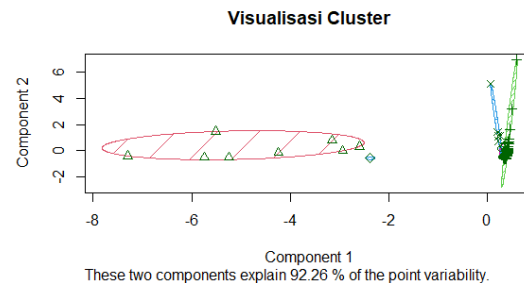
Gambar 4. 17 Clustering *k-means*, n=3

Sumber: Dokumen Pribadi, 2021



Gambar 4. 18 Clustering *k-means*, n=4

Sumber: Dokumen Pribadi, 2021



Gambar 4. 19 Clustering *k-means*, n=5

Sumber: Dokumen Pribadi, 2021

Berdasarkan gambar persebarang kluster diatas, maka diketahui bahwa *clustering k-means* yang paling sesuai adalah dengan menggunakan n sebesar 2 kelompok. Untuk dapat mengevaluasi kinerja pengelompokkan, maka dibuatlah *confusion matrix* sebagai berikut ini:

Console	Terminal	Jobs	Console	Terminal	Jobs	Console	Terminal	Jobs
<pre>R 4.1.1 ~ / > table(data6[,1],k.means.fit\$cluster)</pre>			<pre>R 4.1.1 ~ / 79.98999786 3 0 80 1 0 84.40000153 1 0 89.98999786 2 0 94.98999786 1 0 99 1 0 99.94999695 1 0 99.98999786 4 1 100 1 0 108 1 0 109.9499969 1 0 109.9899979 1 0 119.9899979 1 0 124.9899979 1 0 129.9900055 2 1 134.9900055 1 0 149.9900055 1 0 159.9900055 1 0 164.3800049 1 0 169.9900055 1 0 179.9900055 1 0 189 1 0 199 1 0 199.9900055 2 1 209.9900055 1 0 210.8500061 1 0 215.8200073 1 0 249.9900055 2 0 252.8800049 1 0 260.6499939 1 0 293.0400085 1 0 299.980011 0 1 299.9899902 5 0 327.75 1 0 329.9899902 1 0 349.9899902 1 0 357.1000061 1 0 399.980011 0 1 399.9899902 2 0 452.0400085 1 0 461.480011 1 0 532.5800171 1 0 599.9899902 2 0 999.9899902 1 0 1500 1 0 1999.98999 1 0</pre>			<pre>R 4.1.1 ~ / 129.9900055 2 1 134.9900055 1 0 149.9900055 1 0 159.9900055 1 0 164.3800049 1 0 169.9900055 1 0 179.9900055 1 0 189 1 0 199 1 0 199.9900055 2 1 209.9900055 1 0 210.8500061 1 0 215.8200073 1 0 249.9900055 2 0 252.8800049 1 0 260.6499939 1 0 293.0400085 1 0 299.980011 0 1 299.9899902 5 0 327.75 1 0 329.9899902 1 0 349.9899902 1 0 357.1000061 1 0 399.980011 0 1 399.9899902 2 0 452.0400085 1 0 461.480011 1 0 532.5800171 1 0 599.9899902 2 0 999.9899902 1 0 1500 1 0 1999.98999 1 0</pre>		

Gambar 4. 20 Confusing Matrix

Sumber: Dokumen Pribadi, 2021

Confusion Matrix adalah pengukuran dari performa suatu klasifikasi *machine learning*/clustering dimana keluaran algoritma dapat berupa dua kelas atau lebih. Dalam



matriks konfusi, label kelas yang diprediksi (0, 1, 2) ditulis di bagian atas (nama kolom). Label kelas yang sebenarnya (Iris-setosa, dll.) ditulis di sepanjang sisi kanan. Setiap sel dalam matriks adalah hitungan berapa banyak contoh kelas yang benar di mana diklasifikasikan sebagai masing-masing kelas yang diprediksi. Klasifikasi dikatakan sempurna ketika terdapat banyak angka “0” didalamnya. Maka berdasarkan gambar diatas diketahui bahwa data *clustering* memiliki performa yang bagus atau mendekati sempurna.

BAB VI

PENUTUP

Pada bab ini, akan dipaparkan kesimpulan dan saran yang didapatkan dari pengerjaan laporan *final project* ini.

6.1 Kesimpulan

Berdasarkan rumusan masalah, maka kesimpulan dari pengerjaan laporan ini adalah sebagai berikut:

1. Produk dengan penjualannya paling tinggi adalah produk “*Perfect Fitness Perfect Rip Deck*” dengan penjualan sejumlah 23.263 transaksi yang merupakan 13,55 % dari total seluruh penjualan produk.
2. Negara yang melakukan *order* pembelian terbanyak adalah negara Amerika Serikat dengan total penjualan 23.700 yang merupakan 13,8% dari total penjualan.
3. Tingkat keterlambatan pengiriman tidak dipengaruhi oleh letak geografis tujuan pengiriman, namun lebih dipengaruhi oleh banyaknya jumlah order yang dilakukan *customer*.
4. Tidak terdapat waktu jam spesifik dimana *customer* paling aktif melakukan *order* terhadap produk perusahaan. Kemudian bulan yang paling ramai dengan *order demand customer* adalah bulan Januari, dimana banyak *customer* melakukan pembelian produk perusahaan. Kemudian pada bulan februari, merupakan bulan dimana *customer* melakukan paling sedikit *order* terhadap produk perusahaan.
5. Tipe media pembayaran yang paling banyak digunakan oleh *customer* dalam mengorder produk perusahaan adalah debit. Kemudian tipe media pembayaran yang paling banyak digunakan oleh *customer* yang membatalkan order produk perusahaan adalah *transfer*.
6. Berdasarkan gambar chart diatas, diketahui presentase customer segment pada order yang dibatalkan dengan order yang tidak dibatalkan adalah sama dengan perbedaan nilai kuantitatif didalamnya. Maka diketahui bahwa customer segment tidak mempengaruhi pembatalan order produk perusahaan.
7. Pada bulan Januari tahun 2017 penjualan untuk region USCA, Asia pasifik, dan Afrika mengalami penurunan cukup, sebaliknya penjualan pada region Eropa mengalami kenaikan drastis setelah terjadi penurunan di Januari 2016. Sementara untuk region LATAM tergolong memiliki tingkat penjualan yang stabil.
8. *Clustering k-means* yang paling sesuai adalah dengan menggunakan n sebesar 2 atau dengan mengelompokkan data menjadi 2 kelompok. Kemudian berdasarkan *confusion*



matrix yang dilakukan klasifikasi dikatakan sempurna atau memiliki performa yang bagus, hal ini dikarenakan terdapat banyak angka “0” didalamnya.

6.2 Saran

Berdasarkan pengerjaan pengolahan dan analisis yang telah dilakukan, maka diberikannya saran terhadap database perusahaan *Acme Company* sebagai berikut:

1. Perlunya melakukan input data dan perbaruan pada sistem *database* yang lebih baik, teliti dan hati – hati, mengingat banyak sekali nilai – nilai pada data yang berupa NA dan nilainya kurang sesuai dengan variabel yang ditampilkan.
2. Diperlukan perbaikan pada sistem pengiriman maupun pemberian estimasi barang sampai yang lebih akurat, karena banyak sekali kasus keterlambatan terutama di daerah yang memiliki tingkat order yang tinggi. Dikhawatirkan hal ini mengurangi tingkat kepuasan pada konsumen.
3. Ada baiknya untuk konsisten menjaga kualitas beberapa produk – produk terlaris, disertai dengan promosi terhadap produk – produk yang kurang diminati, agar semakin dapat meningkatkan penjualan secara keseluruhan.



DAFTAR PUSTAKA

- Schniederjans, M. J, dkk. 2014. Business Analytics Principles, Concepts, and Applications What, Why and How. New Jersey: Pearson Education, Inc.
- Kirk, A. 2019. Data Visualisation A Handbook for Data Driven Design. London: SAGE Publications Ltd.



----- *Halaman ini Sengaja Dikosongkan* -----

LAMPIRAN

```
1 data1 <- read.csv("D:/kulyah/kmmi bisnis analitik/Dataset 1.csv",sep=",")
2 library(dplyr)
3 library(ggplot2)
4
5 negara.status <- select(data1, Negara.order, Status.terkirim)
6 negara.terkirim <- negara.status[~grep("canceled",negara.status$Status.terkirim),]
7
8 jumlah.negara.terkirim <- negara.terkirim %>%
9   count(Negara.order)
10
11 jumlah.negara.terkirim <- jumlah.negara.terkirim[order(-jumlah.negara.terkirim$n),]
12 jumlah.negara.terkirim <- jumlah.negara.terkirim %>%
13   mutate(persen=label_percent(accuracy=0.01)(jumlah.negara.terkirim$n/sum(jumlah.negara.terkirim$n)))
14
15 ggplot(head(jumlah.negara.terkirim,10),aes(x=reorder(Negara.order,n),y=n)) +
16   geom_bar(stat = "identity",fill="lightblue") +
17   coord_flip() +
18   geom_text(aes(label=n),hjust=-0.2) +
19   geom_text(aes(label=persen),hjust=1.5,colour="magenta") +
20   labs(x="",
21        y="Jumlah order berhasil",
22        title="Jumlah order berhasil per negara")
23 )
```

Gambar 1 Source Code Plot Negara Paling Banyak Order Sukses

```
1 data1 <- read.csv("D:/kulyah/kmmi bisnis analitik/Dataset 1.csv",sep=",")
2 produk.terjual <- data1[~grep("canceled",data1$Status.terkirim),]
3 library(dplyr)
4 library(ggplot2)
5 library(scales)
6
7 jumlah.produk.terjual <- as.data.frame(table(produk.terjual$Product.Name))
8 jumlah.produk.terjual <- jumlah.produk.terjual[order(-jumlah.produk.terjual$Freq),]
9 jumlah.produk.terjual <- jumlah.produk.terjual %>%
10   mutate(persen=label_percent(accuracy=0.01)(jumlah.produk.terjual$Freq/sum(jumlah.produk.terjual$Freq)))
11
12 ggplot(head(jumlah.produk.terjual,10),aes(x=reorder(Var1,Freq),25,y=Freq)) +
13   geom_bar(stat="identity",fill="magenta") +
14   coord_flip() +
15   geom_text(aes(label=Freq),hjust=0,colour="blue") +
16   geom_text(aes(label=persen),hjust=1.1,colour="black") +
17   labs(x="",
18        y="Jumlah order",
19        title="Jumlah barang terjual")
20 )
```

Gambar 2 Source Code Jumlah Barang Terjual Paling Banyak

```
1 data1 <- read.csv("D:/kulyah/kmmi bisnis analitik/Dataset 1.csv",sep=",")
2 library(dplyr)
3 library(ggplot2)
4
5 data.negara.terlambat <- select(data1, Negara.order, Status.terkirim)
6 data.negara.terlambat <- data.negara.terlambat[grep("Late",data.negara.terlambat$Status.terkirim),]
7
8 jumlah.negara.terlambat <- data.negara.terlambat %>%
9   count(Negara.order)
10
11 jumlah.negara.terlambat <- jumlah.negara.terlambat[order(-jumlah.negara.terlambat$n),]
12
13 ggplot(head(jumlah.negara.terlambat,10),aes(x=reorder(Negara.order,n),y=n,fill=Negara.order)) +
14   geom_bar(stat="identity",fill="lightgreen") +
15   coord_flip() +
16   geom_text(aes(label=n),hjust=1.2) +
17   labs(x="",
18        y="Jumlah order terlambat datang",
19        title="Jumlah order terlambat per negara")
20 )
```

Gambar 3 Source Code Hubungan Keterlambatan Pengiriman Dengan Negara

```
1 data1 <- read.csv("D:/kulyah/kmmi bisnis analitik/Dataset 1.csv",sep=",")
2 library(ggplot2)
3 library(dplyr)
4 library(lubridate)
5 library(tidyr)
6 library(hrbrthemes)
7 data <- select(data1,tanggal.order..DateOrders.,pasar)
8
9 data <- separate(data, tanggal.order..DateOrders., c("date", "time"), sep = " ")
10 data$jumlah.penjualan <- as.numeric(ave(data$date, data$date, FUN=length))
11 data <- data %>% drop_na()
12 data$date <- as.Date(data$date,format='%d/%m/%Y')
13
14 data %>%
15   group_by(Monthly=floor_date(date,unit = "month"),pasar) %>%
16   summarise(Total_Sales=sum(jumlah.penjualan)) %>%
17   ggplot(aes(Monthly>Total_Sales))+
18   geom_line(aes(color=pasar),size=1.5)+
19   theme_light()+
20   scale_y_comma()+
21   scale_x_date(date_labels = "%b-%Y",breaks="3 month") +
22   labs(title = "Total penjualan per region",y="Total penjualan",x="")
```

Gambar 4 Source Code Plot Time Series Total Penjualan Per Region

```
1 data1 <- read.csv("C:/Users/Lenovo/Documents/Vada Zafira/TUGAS/Semester 7/Business Analytics/Dataset 1.csv",sep=",")
2 head(data1)
3 library(dplyr)
4 library(ggplot2)
5 library(ggsci)
6 library(RColorBrewer)
7 library(factoextra)
8 library(factoextra)
9 library(hms)
10
11 data2 <- select(data1, Product.Name, Status.terkirim, jumlah.Order.Item, keuntungan.Order.Per.Order, Product.Price)
12
13 Order_sukses <- data2[~grep("canceled",data2$Status.terkirim),]
14 Order_cancel <- data2[grep("canceled",data2$Status.terkirim),]
15
16 data3 = Order_sukses[, -2]
17 head(data3)
18
19 data4 <- data3 %>%
20   group_by(Product.Name, Product.Price) %>%
21   summarise(total_order = sum(jumlah.Order.Item, na.rm = TRUE))
22
23 View(data4)
24
25 data5 <- data3 %>%
26   group_by(Product.Name) %>%
27   summarise(keuntungan_order = sum(keuntungan.Order.Per.Order, na.rm = TRUE))
28
29 View(data5)
30
31
32
```

Gambar 5. Source Code 1 Clustering K-Means

```
33 DFMerged <- merge(data4, data5, by=0)
34
35 View(DFMerged)
36
37 data6 = DFMerged[, -c(1,2,5)]
38 head(data6)
39
40 colnames(data6) = c("harga", "order", "untung")
41 head(data6)
42
43 View(data6)
44
45 summary(data6)
46 sum(is.na(data6))
47
48 fviz_nbclust(data6, kmeans, method = "wss")
49 fviz_nbclust(data6, kmeans, method = "silhouette")
50
51 #kustering 2
52 set.seed(123)
53 k.means.fit = kmeans(data6, iter.max = 10000, 2)
54 k.means.fit$centers
55 k.means.fit
56
57 library(cluster)
58 jarak <- as.matrix(dist(data6)); jarak
59 score <- mean(silhouette(k.means.fit$cluster,dmatrix=jarak)[,3])
60 print(paste("silhouette score = ", round(score, 3)))
61
62
63
64 clusplot(data6, k.means.fit$cluster,
```

Gambar 6. Source Code 2 Clustering K-Means

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Project:
Klustering_Final Project.R* Customer Segment.R* Tipe Pembayaran_Final Project.R* data6 Waktu_Final Project.R*
Source on Save
64 clusplot(data6, k.means.fit$cluster,
65 main = "Visualisasi Cluster",
66 color = TRUE, shade = TRUE, lines = 0)
67
68 confusion
69 table(data6[,1],k.means.fit$cluster)
70
71
72 ##kustering 3
73 set.seed(123)
74 k.means.fit = kmeans(data6, iter.max = 1000, 3)
75 k.means.fit$centers
76 k.means.fit
77
78
79 library(cluster)
80 jarak <- as.matrix(dist(data6)); jarak
81 score <- mean(silhouette(k.means.fit$cluster,dmatrix=jarak)[,3])
82 print(paste("silhouette score = ", round(score, 3)))
83
84
85 clusplot(data6, k.means.fit$cluster,
86 main = "Visualisasi Cluster",
87 color = TRUE, shade = TRUE, lines = 0)
88
89 ## clustering 4
90 set.seed(123)
91 k.means.fit = kmeans(data6, iter.max = 1000, 4)
92 k.means.fit$centers
93 k.means.fit
94
95
68:10 (Top Level) R Script
```

Gambar 7. Source Code 3 Clustering K-Means

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Project:
Klustering_Final Project.R* Customer Segment.R* Tipe Pembayaran_Final Project.R* data6 Waktu_Final Project.R*
Source on Save
93 k.means.fit
94
95
96 library(cluster)
97 jarak <- as.matrix(dist(data6)); jarak
98 score <- mean(silhouette(k.means.fit$cluster,dmatrix=jarak)[,3])
99 print(paste("silhouette score = ", round(score, 3)))
100
101
102 clusplot(data6, k.means.fit$cluster,
103 main = "Visualisasi Cluster",
104 color = TRUE, shade = TRUE, lines = 0)
105
106 ## clustering 5
107 set.seed(123)
108 k.means.fit = kmeans(data6, iter.max = 1000, 5)
109 k.means.fit$centers
110 k.means.fit
111
112
113 library(cluster)
114 jarak <- as.matrix(dist(data6)); jarak
115 score <- mean(silhouette(k.means.fit$cluster,dmatrix=jarak)[,3])
116 print(paste("silhouette score = ", round(score, 3)))
117
118
119 clusplot(data6, k.means.fit$cluster,
120 main = "Visualisasi Cluster",
121 color = TRUE, shade = TRUE, lines = 0)
122
123
124
68:10 (Top Level) R Script
```

Gambar 8. Source Code 5 Clustering K-Means

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Kluster_Final Project.R Customer Segment.R Tipe Pembayaran_Final Project.R data6 Waktu_Final Project.R
1 data1 <- read.csv("C:/Users/Lenovo/Documents/vada Zafira/TUGAS/Semester 7/Business Analytics/Dataset 1.csv",sep=",")
2 head(data1)
3 library(dplyr)
4 library(ggplot2)
5 library(ggsci)
6 library(RColorBrewer)
7
8 View(data1)
9
10 Customer <- select(data1, Status.terkirim, Customer.Segment)
11
12 Order_sukses <- Customer[grep("canceled",Customer$Status.terkirim),]
13 Order_cancel <- Customer[grep("canceled",Customer$Status.terkirim),]
14
15 Jumlah.order_sukses <- Order_sukses %>%
16   count(Customer.Segment)
17
18 Jumlah.order_cancel <- Order_cancel %>%
19   count(Customer.Segment)
20
21 View(Jumlah.order_sukses)
22
23 df = data.frame(head(Jumlah.order_sukses,3),"Customer.Segment" = c("Home Office", "Corporate", "Consumer"),
24   "n" = c(30647, 52185, 88847))
25
26
27 View(df1)
28 ggplot(df,aes(x=" ",y=n,fill=reorder(Customer.Segment,n))) +
29   geom_bar(stat="identity", width = 1) +
30   coord_polar("y", start = 0) +
31   geom_text(aes(label = paste0(round((n/sum(n))*100), "%"), position = position_stack(vjust = 0.5)) +
32   scale_fill_manual(values=c("#550000", "#33658A", "#2F4858")) +
33   labs(x = NULL, y = NULL, fill = NULL, title = "Customer Segment Order Success") +
34   theme_classic() + theme(axis.line = element_blank(),
35     axis.text = element_blank(),
36     axis.ticks = element_blank(),
37     plot.title = element_text(hjust = 0.5, color = "black"))
38
39 View(df1)
40 ggplot(df,aes(x=" ",y=n,fill=reorder(Customer.Segment,n))) +
41   geom_bar(stat="identity", width = 1) +
42   coord_polar("y", start = 0) +
43   geom_text(aes(label = paste0(n)), position = position_stack(vjust = 0.5)) +
44   scale_fill_manual(values=c("#550000", "#33658A", "#2F4858")) +
45   labs(x = NULL, y = NULL, fill = NULL, title = "Customer Segment Order Success") +
46   theme_classic() + theme(axis.line = element_blank(),
47     axis.text = element_blank(),
48     axis.ticks = element_blank(),
49     plot.title = element_text(hjust = 0.5, color = "black"))
50
51
52
53 View(Jumlah.order_cancel)
54 df1 = data.frame(head(Jumlah.order_cancel,3),"Customer.Segment" = c("Home Office", "Corporate", "Consumer"),
55   "n" = c(30647, 52185, 88847))
56
57 ggplot(df1,aes(x=" ",y=n,fill=reorder(Customer.Segment,n))) +
58   geom_bar(stat="identity", width = 1) +
59   coord_polar("y", start = 0) +
60   geom_text(aes(label = paste0(round((n/sum(n))*100), "%"), position = position_stack(vjust = 0.5)) +
61   scale_fill_manual(values=c("#550000", "#33658A", "#2F4858")) +
62   theme_classic() + theme(axis.line = element_blank(),
63     axis.text = element_blank(),
64     axis.ticks = element_blank(),
65     plot.title = element_text(hjust = 0.5, color = "black"))
66
67
68 ggplot(df1,aes(x=" ",y=n,fill=reorder(Customer.Segment,n))) +
69   geom_bar(stat="identity", width = 1) +
70   coord_polar("y", start = 0) +
71   geom_text(aes(label = paste0(n)), position = position_stack(vjust = 0.5)) +
72   scale_fill_manual(values=c("#550000", "#33658A", "#2F4858")) +
73   labs(x = NULL, y = NULL, fill = NULL, title = "Customer Segment Order Cancel") +
74   theme_classic() + theme(axis.line = element_blank(),
75     axis.text = element_blank(),
76     axis.ticks = element_blank(),
77     plot.title = element_text(hjust = 0.5, color = "black"))
78
79
80
```

Gambar 9. Source Code 1 Analisis Customer Segment

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Kluster_Final Project.R Customer Segment.R Tipe Pembayaran_Final Project.R data6 Waktu_Final Project.R
38 View(df1)
39 ggplot(df,aes(x=" ",y=n,fill=reorder(Customer.Segment,n))) +
40   geom_bar(stat="identity", width = 1) +
41   coord_polar("y", start = 0) +
42   geom_text(aes(label = paste0(n)), position = position_stack(vjust = 0.5)) +
43   scale_fill_manual(values=c("#550000", "#33658A", "#2F4858")) +
44   labs(x = NULL, y = NULL, fill = NULL, title = "Customer Segment Order Success") +
45   theme_classic() + theme(axis.line = element_blank(),
46     axis.text = element_blank(),
47     axis.ticks = element_blank(),
48     plot.title = element_text(hjust = 0.5, color = "black"))
49
50
51
52 View(Jumlah.order_cancel)
53 df1 = data.frame(head(Jumlah.order_cancel,3),"Customer.Segment" = c("Home Office", "Corporate", "Consumer"),
54   "n" = c(30647, 52185, 88847))
55
56 ggplot(df1,aes(x=" ",y=n,fill=reorder(Customer.Segment,n))) +
57   geom_bar(stat="identity", width = 1) +
58   coord_polar("y", start = 0) +
59   geom_text(aes(label = paste0(round((n/sum(n))*100), "%"), position = position_stack(vjust = 0.5)) +
60   scale_fill_manual(values=c("#550000", "#33658A", "#2F4858")) +
61   theme_classic() + theme(axis.line = element_blank(),
62     axis.text = element_blank(),
63     axis.ticks = element_blank(),
64     plot.title = element_text(hjust = 0.5, color = "black"))
65
66
67 ggplot(df1,aes(x=" ",y=n,fill=reorder(Customer.Segment,n))) +
68   geom_bar(stat="identity", width = 1) +
69   coord_polar("y", start = 0) +
70   geom_text(aes(label = paste0(n)), position = position_stack(vjust = 0.5)) +
71   scale_fill_manual(values=c("#550000", "#33658A", "#2F4858")) +
72   labs(x = NULL, y = NULL, fill = NULL, title = "Customer Segment Order Cancel") +
73   theme_classic() + theme(axis.line = element_blank(),
74     axis.text = element_blank(),
75     axis.ticks = element_blank(),
76     plot.title = element_text(hjust = 0.5, color = "black"))
77
78
79
80
```

Gambar 10. Source Code 2 Analisis Customer Segment

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Klustering_Final Project.R* Customer Segment.R* Tipe Pembayaran_Final Project.R* data6 Waktu_Final Project.R*
1 data1 <- read.csv("C:/Users/Lenovo/Documents/Vada Zafira/TUGAS/Semester 7/Business Analytics/Dataset 1.csv",sep=",")
2 head(data1)
3 library(dplyr)
4 library(ggplot2)
5 library(ggsci)
6 library(RColorBrewer)
7
8 View(data1)
9 Pembayaran <- select(data1, Status.terkirim, Tipe)
10
11 Pembayaran_sukses <- Pembayaran[!grep("canceled",Pembayaran$Status.terkirim),]
12 Pembayaran_cancel <- Pembayaran[grep("canceled",Pembayaran$Status.terkirim),]
13
14 View(Pembayaran_sukses)
15 View(Pembayaran_cancel)
16
17 Jumlah.pembayaran.sukses <- Pembayaran_sukses %>%
18 count(Tipe)
19
20 Jumlah.pembayaran.cancel <- Pembayaran_cancel %>%
21 count(Tipe)
22
23 Jumlah.pembayaran.sukses <- Jumlah.pembayaran.sukses[order(-Jumlah.pembayaran.sukses$`n`),]
24 Jumlah.pembayaran.cancel <- Jumlah.pembayaran.cancel[order(-Jumlah.pembayaran.cancel$`n`),]
25
26 ggplot(head(Jumlah.pembayaran.sukses,4),aes(x=reorder(Tipe,n),y=n,fill=reorder(Tipe,n))) +
27   geom_bar(stat="identity",
28     position=position_dodge(),
29     size=3) +
30   labs(y = "Jumlah Order",
31     x = "Tipe Pembayaran") +
32   ggtitle("Tipe Pembayaran Produk") +
33   theme_minimal() +
34   theme(plot.title=element_text(hjust=0.5, face='bold'))
35 View(Jumlah.pembayaran.cancel)
36
37 ggplot(head(Jumlah.pembayaran.cancel,3),aes(x=reorder(Tipe,n),y=n,fill=reorder(Tipe,n))) +
38   geom_bar(stat="identity",
39     position=position_dodge(),
40     size=3) +
41   labs(y = "Jumlah Order",
42     x = "Tipe Pembayaran Order Dibatalkan") +
43   ggtitle("Tipe Pembayaran Produk Cancel") +
44   theme_minimal() +
45   theme(plot.title=element_text(hjust=0.5, face='bold'))
46
47
48
```

Gambar 11. Source Code Tipe Pembayaran

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Klustering_Final Project.R* Customer Segment.R* Tipe Pembayaran_Final Project.R* data6 Waktu_Final Project.R*
42 group_by(waktu2)
43
44
45 View(Jumlah.order.sukses)
46
47 Jumlah.order.sukses$waktu2 <- ymd_hms(Jumlah.order.sukses$waktu2)
48 class(Jumlah.order.sukses$waktu2)
49
50 #Extraksi date info
51 just.hour <- hour(Jumlah.order.sukses$waktu2)
52 just.minut <- minute(Jumlah.order.sukses$waktu2)
53 just.day <- day(Jumlah.order.sukses$waktu2)
54 just.dayweek <- wday(Jumlah.order.sukses$waktu2)
55 just.dayweek2 <- wday(Jumlah.order.sukses$waktu2, label=TRUE)
56 just.month <- month(Jumlah.order.sukses$waktu2)
57 just.month2 <- month(Jumlah.order.sukses$waktu2, label = TRUE)
58 just.week <- week(Jumlah.order.sukses$waktu2)
59 just.year <- year(Jumlah.order.sukses$waktu2)
60 just.hour[1]; just.dayweek[1]; just.dayweek2[1]; just.month[1]; just.month2[1]; just.year[1]
61
62 Ex <- data.frame(original=Jumlah.order.sukses$waktu2,hour=just.hour,minute=just.minut,
63   day=just.day,dayofweek=just.dayweek,dayofweek2=just.dayweek2,month=just.month,
64   month2=just.month2,week=just.week,year=just.year)
65 head(Ex)
66
67 summary(month(Jumlah.order.sukses$waktu2,label=TRUE))
68 bulan <- table(month(Jumlah.order.sukses$waktu2,label=TRUE))
69
70 summary(year(Jumlah.order.sukses$waktu2))
71
72 plot.month <- Ex %>%
73   count(month)
74 view(plot.month)
75
76 table(hour(Jumlah.order.sukses$waktu2))
77 barplot(table(hour(Jumlah.order.sukses$waktu2)),
78   main = "Jumlah Order Perjam",
79   xlab = "Total Order",
80   ylab = "Bulan Order")
81
82 table(minute(Jumlah.order.sukses$waktu2))
83 barplot(table(minute(Jumlah.order.sukses$waktu2)),
84   main = "Jumlah Order Permenit",
85   xlab = "Total Order",
86   ylab = "Bulan Order")
87
88 table(month(Jumlah.order.sukses$waktu2))
89 barplot(bulan, main = "Jumlah Order Perbulan",
90   xlab = "Total Order",
91   ylab = "Bulan Order")
92
91:31 (Top Level) R Script
```

Gambar 12. Source Code 1 Analisis Waktu Order Produk


```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Klustering_Final Project.R* Customer Segment.R* Tipe Pembayaran_Final Project.R* data6 Waktu_Final Project.R*
Source on Save Run Source
43
44
45 View(Jumlah.order.sukses)
46
47 Jumlah.order.sukses$qwaktu2 <- ymd_hms(Jumlah.order.sukses$qwaktu2)
48 class(Jumlah.order.sukses$qwaktu2)
49
50 #Extraksi date info
51 just.hour <- hour(Jumlah.order.sukses$qwaktu2)
52 just.minut <- minute(Jumlah.order.sukses$qwaktu2)
53 just.day <- day(Jumlah.order.sukses$qwaktu2)
54 just.dayweek <- wday(Jumlah.order.sukses$qwaktu2)
55 just.dayweek2 <- wday(Jumlah.order.sukses$qwaktu2, label=TRUE)
56 just.month <- month(Jumlah.order.sukses$qwaktu2)
57 just.month2 <- month(Jumlah.order.sukses$qwaktu2, label=TRUE)
58 just.week <- week(Jumlah.order.sukses$qwaktu2)
59 just.year <- year(Jumlah.order.sukses$qwaktu2)
60 just.hour[1]; just.dayweek[1]; just.dayweek2[1]; just.month[1]; just.month2[1]; just.year[1]
61
62 Ex <- data.frame(original=Jumlah.order.sukses$qwaktu2, hour=just.hour, minute=just.minut,
63                 day=just.day, dayofweek=just.dayweek, dayofweek2=just.dayweek2, month=just.month,
64                 month2=just.month2, week=just.week, year=just.year)
65
66 head(Ex)
67
68 summary(month(Jumlah.order.sukses$qwaktu2, label=TRUE))
69
70 bulan <- table(month(Jumlah.order.sukses$qwaktu2, label=TRUE))
71
72 summary(year(Jumlah.order.sukses$qwaktu2))
73
74 plot.month <- Ex %>%
75   count(month)
76   view(plot.month)
77
78 table(hour(Jumlah.order.sukses$qwaktu2))
79 barplot(table(hour(Jumlah.order.sukses$qwaktu2)),
80         main = "Jumlah Order Perjam",
81         xlab = "Total Order",
82         ylab = "Bulan Order")
83
84 table(minute(Jumlah.order.sukses$qwaktu2))
85 barplot(table(minute(Jumlah.order.sukses$qwaktu2)),
86         main = "Jumlah Order Permenit",
87         xlab = "Total Order",
88         ylab = "Bulan Order")
89
90 table(month(Jumlah.order.sukses$qwaktu2))
91 barplot(bulan, main = "Jumlah Order Perbulan",
92         xlab = "Total Order",
93         ylab = "Bulan Order")LLI
94
91:32 (Top Level) R Script
```

Gambar 13. Source Code 2 Analisis Waktu Order Produk