Mini Project Data Engineering Report

Louis Mayco Dillon Wijaya
Github: *https://github.com/louis-not/DSLS_MiniProject*

Contents

Case Study

Northwind database merupakan database sampel milik Microsoft yang merupakan perusahaan fiktif yang bergerak di **bidang ekspor import makanan**. Sebagai perusahaan yang bergerak di bidang ekspor impor makanan, mempertahankan adanya alur transaksi antara **supplier** dan **customers** yang baik, akan menjadi model bisnis utama perusahaan.

Maka, bagaimana Northwind dapat **menghubungkan customer dan supplier dengan untuk membuat kesepakatan dalam peningkatan volume transaksi ekspor-impor produk.**

Alir Pembahasan Case Study

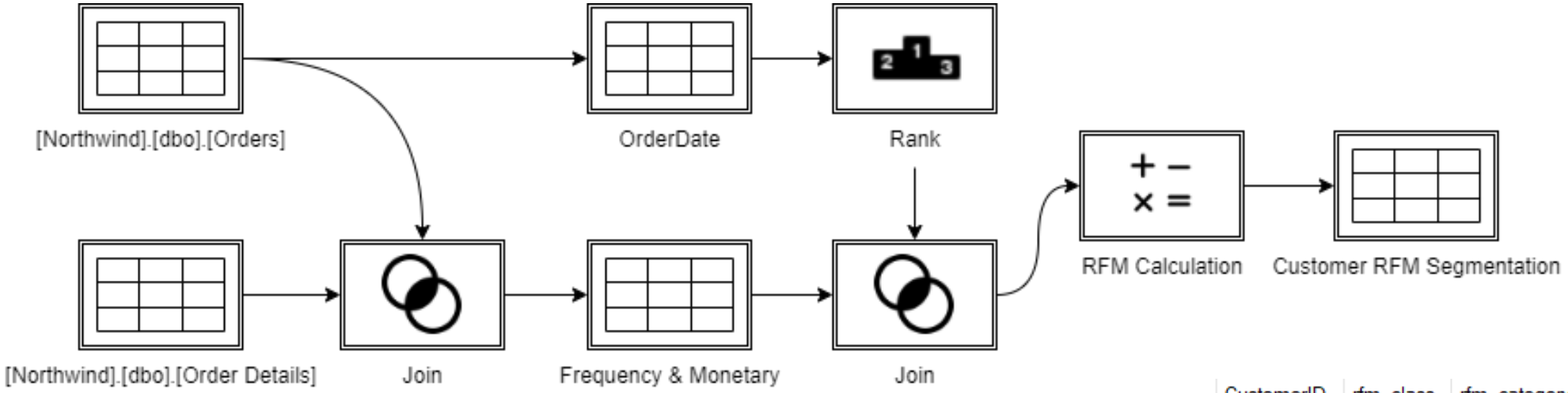
Customer Analysis

Objective: Melakukan kategorisasi customer yang memiliki potensi dalam melakukan hubungan kerja sama berdasarkan log transaksi.

Methodology: RFM analysis dapat digunakan karena cara kerja RFM analysis menekankan pada 3 faktor *(Recency, Frequency, Monetary)* yang dapat mengidentifikasi potensi suatu customer untuk melakukan transaksi dalam volume yang besar melalui kesepakatan dengan supplier.

Customer Analysis (cont.)

Flowchart:



	CustomerID	rfm_class	rfm_category
1	QUICK	311	NULL
2	QUICK	411	Churned Best Customers
3	QUICK	411	Churned Best Customers
4	SAVEA	311	NULL
5	SAVEA	311	NULL
6	SAVEA	411	Churned Best Customers
7	ERNSH	411	Churned Best Customers
8	ERNSH	411	Churned Best Customers
9	ERNSH	411	Churned Best Customers

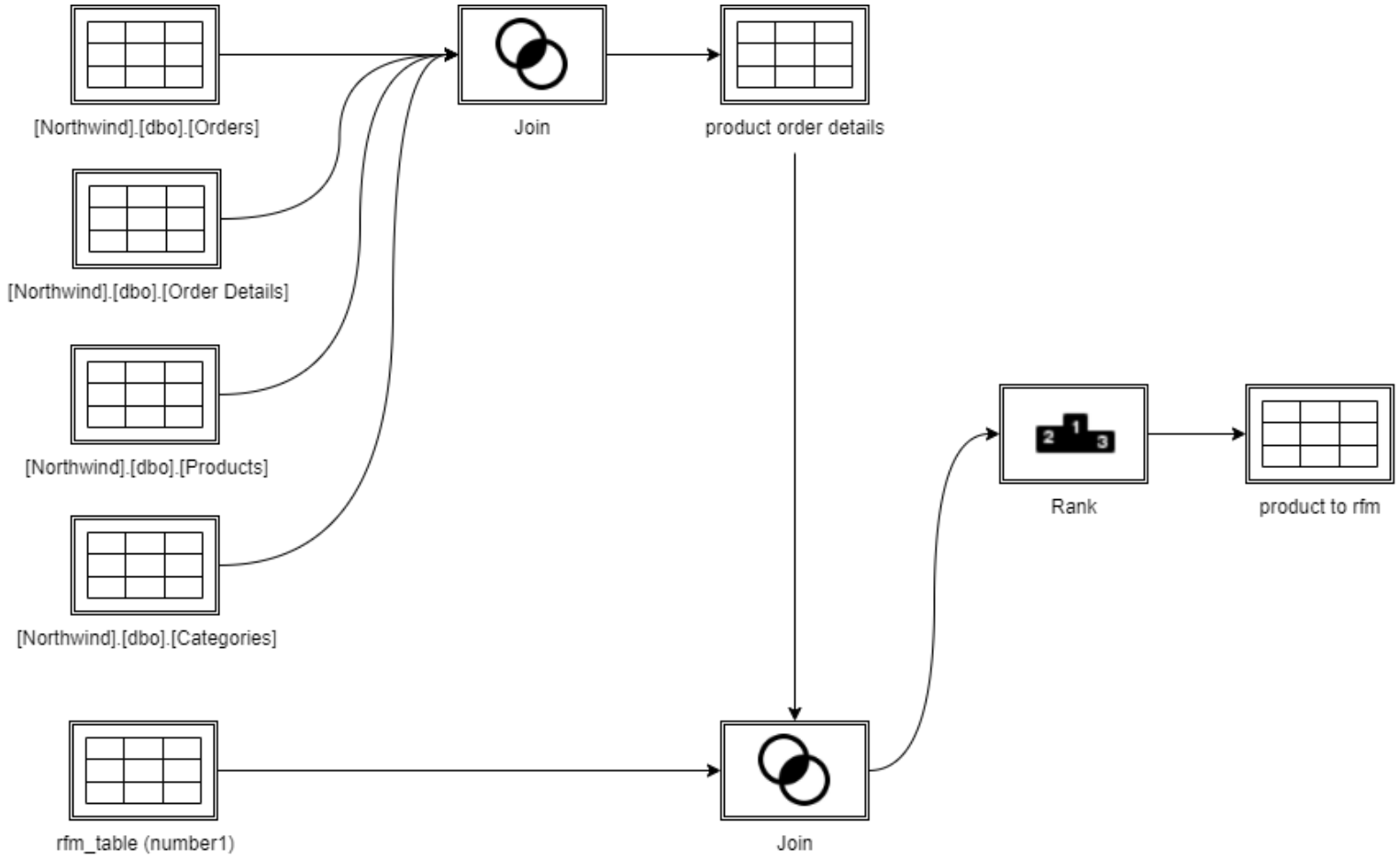
Contoh Hasil

Product Analysis

Objective: Mencari product category yang digemari oleh masing-masing segmen dari customer

	rfm_category	CategoryName	TotalQuantity	Rank_
1	Best Customers	Confections	5	1
2	Best Customers	Dairy Products	5	1
3	Best Customers	Beverages	4	2
4	Best Customers	Grains/Cereals	3	3
5	Best Customers	Seafood	3	3
6	Loyal Customers	Confections	20	1
7	Loyal Customers	Beverages	16	2
8	Loyal Customers	Dairy Products	15	3
9	Need Attention/Hibernating	Dairy Products	13	1
10	Need Attention/Hibernating	Seafood	12	2
11	Need Attention/Hibernating	Confections	9	3
12	Need Attention/Hibernating	Beverages	9	3
13	Potential Loyalist	Dairy Products	105	1
14	Potential Loyalist	Beverages	99	2
15	Potential Loyalist	Seafood	91	3

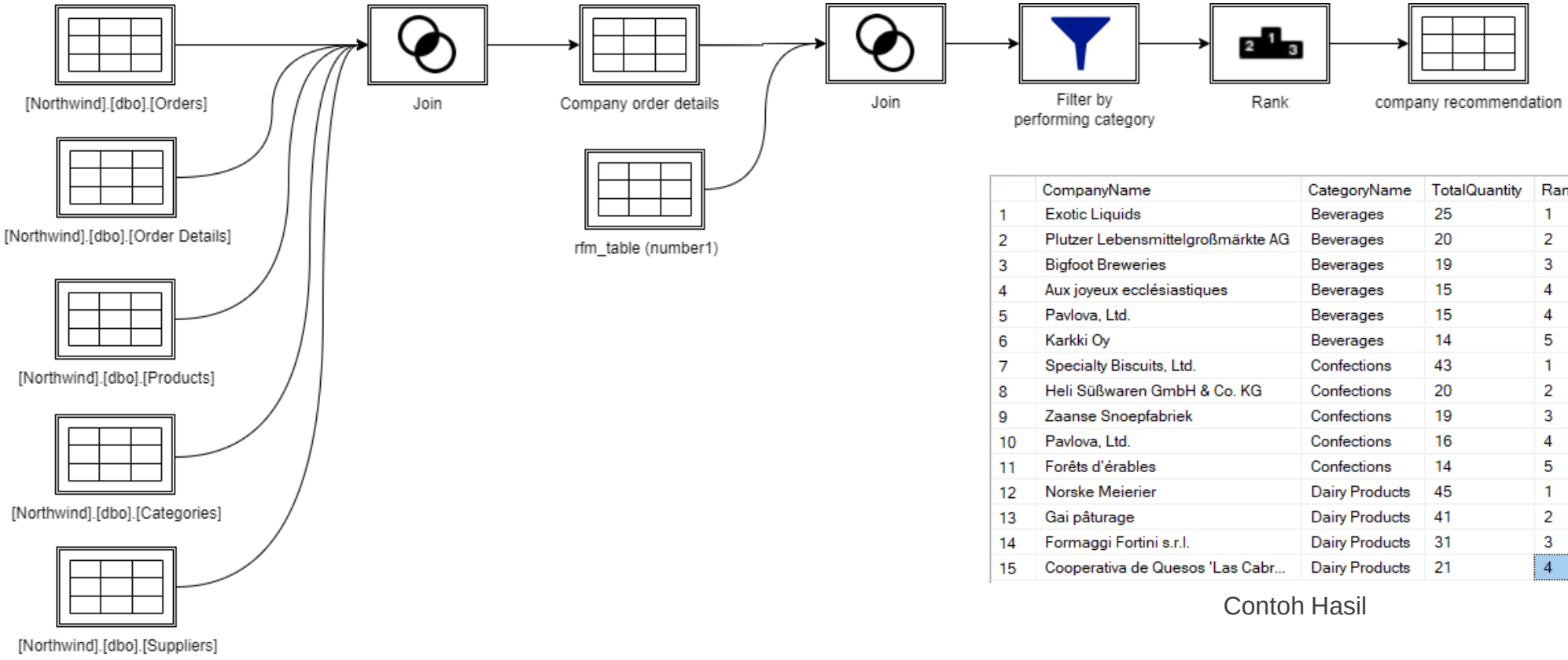
Contoh Hasil



Suppliers Analysis

Objective: Mencari Supplier dengan kategori yang sama dengan nilai harga termurah dan transaksi terbanyak

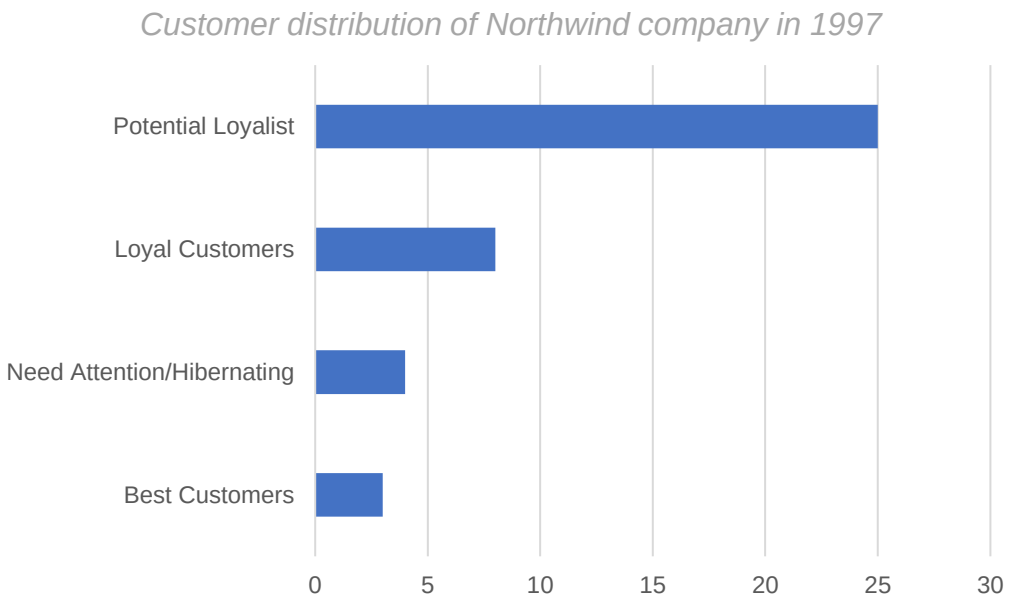
Flowchart:

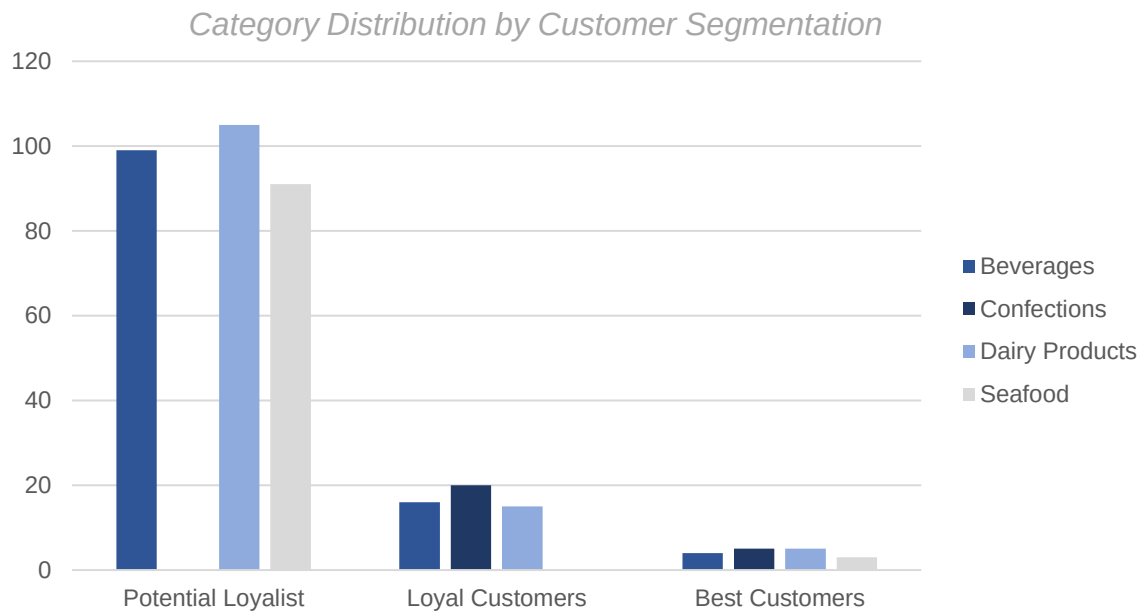


	CompanyName	CategoryName	TotalQuantity	Rank_
1	Exotic Liquids	Beverages	25	1
2	Plutzer Lebensmittelgroßmärkte AG	Beverages	20	2
3	Bigfoot Breweries	Beverages	19	3
4	Aux joyeux ecclésiastiques	Beverages	15	4
5	Pavlova, Ltd.	Beverages	15	4
6	Karkki Oy	Beverages	14	5
7	Specialty Biscuits, Ltd.	Confections	43	1
8	Heli Süßwaren GmbH & Co. KG	Confections	20	2
9	Zaanse Snoepfabriek	Confections	19	3
10	Pavlova, Ltd.	Confections	16	4
11	Forêts d'érables	Confections	14	5
12	Norske Meierier	Dairy Products	45	1
13	Gai pâturage	Dairy Products	41	2
14	Formaggi Fortini s.r.l.	Dairy Products	31	3
15	Cooperativa de Quesos 'Las Cabr...	Dairy Products	21	4

Contoh Hasil

Customer Segmentation





- *Berdasarkan data transasksi perusahaan **Northwind** di tahun 1997, pelanggan yang memiliki potensi untuk menjadi pelanggan memiliki jumlah terbanyak (41%), yang kemudian disusul dengan pelanggan setia (13%).*
- *Barang dengan kategori **Beverages, Dairy Products** dan **Seafood** memiliki potensi transaksi besar kedepannya.*
- *Berbeda dengan potensi pelanggan setia, kategori **Confections** selalu menempati peringkat 1 dalam segmen Loyal Customers dan Best Customers.*

Recommendation

- *Pihak **Northwind** direkomendasikan untuk memilih antara kategori **Confections, Beverages** dan **Dairy Products** untuk melakukan pengujian strategi pemasaran baru dengan melakukan hubungan Kerjasama pemesanan volume besar berfrekuensi dengan saran nama perusahaan sebagai berikut.*

- *Pengujian strategi pemasaran yang ditujukan pada segmen **Loyal Customers** dan juga **Best Customers** akan memiliki tingkat keberhasilan yang cukup tinggi, terutama di kategori **Confections**.*

- *Peluang keberhasilan tertinggi dapat dilakukan pada kategori **Dairy Products** pada segmen **Potential Customers** akibat target pemasaran yang paling besar.*

Beverages	Confections	Dairy Products
Exotic Liquids	Specialty Biscuits, Ltd.	Norske Meierier
Plutzer Lebensmittelgroßmärkte AG	Heli Süßwaren GmbH & Co. KG	Gai pâturage
Bigfoot Breweries	Zaanse Snoepfabriek	Formaggi Fortini s.r.l.
Aux joyeux ecclésiastiques	Pavlova, Ltd.	Cooperativa de Quesos 'Las Cabras'

Intermediate Queries

SQL Codes

Intermediate Queries

Nomor 1

	bulan	Jumlah Customer
1	1	27
2	2	21
3	3	24
4	4	27
5	5	26
6	6	25
7	7	25
8	8	30
9	9	27
10	10	28
11	11	27
12	12	37

Nomor 2

	LastName	FirstName	Title
1	Davolio	Nancy	Sales Representative
2	Leverling	Janet	Sales Representative
3	Peacock	Margaret	Sales Representative
4	Suyama	Michael	Sales Representative
5	King	Robert	Sales Representative
6	Dodsworth	Anne	Sales Representative

Intermediate Queries

Nomor 3

	OrderID	Quantity	ProductName	OrderDate
1	10401	70	Gnocchi di nonna Alice	1997-01-01 00:00:00.000
2	10403	70	Chocolade	1997-01-03 00:00:00.000
3	10430	70	Raclette Courdavault	1997-01-30 00:00:00.000
4	10402	65	Vegie-spread	1997-01-02 00:00:00.000
5	10401	60	Flotemysost	1997-01-01 00:00:00.000

Nomor 4

	OrderID	ProductName	OrderDate	CompanyName
1	10576	Chai	1997-06-23 00:00:00.000	Tortuga Restaurante

Intermediate Queries

Nomor 5

	TotalPriceCategory	TOTAL_ORDERID
1	<=100	269
2	>500	502
3	100<x<=250	382
4	250<x<=500	425

Nomor 6

	CompanyName	Order_Quantity
1	Save-a-lot Markets	2574
2	QUICK-Stop	2172
3	Ernst Handel	1950
4	Hungry Owl All-Night Grocers	799
5	Mère Paillarde	750
6	HILARION-Abastos	650
7	Frankenversand	597
8	Wartian Herkku	550
9	Rattlesnake Canyon Grocery	540
10	Queen Cozinha	527
11	White Clover Markets	518

Intermediate Queries

Nomor 7

	Bulan	ProductName	Quantity
1	1	Gnocchi di nonna Alice	70
2	1	Chocolade	70
3	1	Raclette Courdavault	70
4	1	Vegie-spread	65
5	1	Flotemysost	60
6	2	Pâté chinois	120
7	2	Gula Malacca	100
8	2	Sirop d'érable	90
9	2	Tourtière	80
10	2	Louisiana Hot Spiced Okra	60
11	3	Guaraná Fantástica	80
12	3	Louisiana Hot Spiced Okra	60
13	3	Raclette Courdavault	60
14	3	Outback Lager	60
15	3	Raclette Courdavault	60
16	4	Schoggi Schokolade	120

Nomor 8

	OrderID	ProductID	ProductName	UnitPrice	Discount	DiscountPrice
1	10248	11	Queso Cabrales	14.00	0	14
2	10248	42	Singaporean Hokkien Fried Mee	9.80	0	9.8
3	10248	72	Mozzarella di Giovanni	34.80	0	34.8
4	10249	14	Tofu	18.60	0	18.6
5	10249	51	Manjimup Dried Apples	42.40	0	42.4
6	10250	41	Jack's New England Clam Chowder	7.70	0	7.7
7	10250	51	Manjimup Dried Apples	42.40	0.15	36.04
8	10250	65	Louisiana Fiery Hot Pepper Sauce	16.80	0.15	14.28
9	10251	22	Gustaf's Knäckebröd	16.80	0.05	15.96

Nomor 9

```
-- Pengujian Prosedur
EXEC print_invoice @CustomerID = 'HANAR';
```

100 %

Results | Messages

	CustomerID	CompanyName	OrderID	OrderDate	RequiredDate	ShippedDate
1	HANAR	Hanari Carnes	10250	1996-07-08 00:00:00.000	1996-08-05 00:00:00.000	1996-07-12 00:00:00.000
2	HANAR	Hanari Carnes	10253	1996-07-10 00:00:00.000	1996-07-24 00:00:00.000	1996-07-16 00:00:00.000
3	HANAR	Hanari Carnes	10541	1997-05-19 00:00:00.000	1997-06-16 00:00:00.000	1997-05-29 00:00:00.000
4	HANAR	Hanari Carnes	10645	1997-08-26 00:00:00.000	1997-09-23 00:00:00.000	1997-09-02 00:00:00.000
5	HANAR	Hanari Carnes	10690	1997-10-02 00:00:00.000	1997-10-30 00:00:00.000	1997-10-03 00:00:00.000

Lampiran

SQL Codes

RFM analysis

```sql
-- Recency Rank
CREATE VIEW view_r2 AS
SELECT X.CustomerID,
    DATEDIFF(day,X.OrderDate,'1997/12/31') AS Recency
    FROM (
        SELECT CustomerID,
        OrderDate,
        DENSE_RANK() OVER (PARTITION BY  CustomerID
            ORDER BY OrderDate DESC) AS Rank_
        FROM [Northwind].[dbo].[Orders]
        WHERE YEAR(OrderDate) = 1997
        ) AS X
    WHERE X.Rank_ =3 ;

-- Frequency & Monetary
CREATE VIEW view_fm2 AS
SELECT O.CustomerID, COUNT(O.OrderID) AS Frequency,
    SUM(OD.UnitPrice * Quantity) AS Amount
    FROM [Northwind].[dbo].[Orders] AS O
    JOIN [Northwind].[dbo].[Order Details] AS OD
    ON O.OrderID = OD.OrderID
    WHERE YEAR(O.OrderDate) = 1997
    GROUP BY CustomerID
```

```sql
-- RFM Value
CREATE VIEW view_rfm3 AS
SELECT view_fm2.CustomerID,
        view_r2.Recency,
        view_fm2.Frequency,
        view_fm2.Amount,
        NTILE(5) OVER (ORDER BY view_r2.Recency DESC) AS R,
        NTILE(5) OVER (ORDER BY view_fm2.Frequency DESC) AS F,
        NTILE(5) OVER (ORDER BY view_fm2.Amount DESC) AS M
    FROM view_fm2
    INNER JOIN view_r2
    ON view_fm2.CustomerID = view_r2.CustomerID;

CREATE VIEW view_rfm_class3 AS
SELECT view_rfm3.CustomerID,
    CONCAT(R, F, M) as rfm_class
    FROM view_rfm3;

CREATE VIEW view_rfm_result3 AS
SELECT *,
    CASE
        WHEN rfm_class LIKE '[4-5][4-5][4-5]' THEN 'Best Customers'
        WHEN rfm_class LIKE '[2-4][3-4][4-5]' THEN 'Loyal Customers'
        WHEN rfm_class LIKE  '[3-5][1-3][1-3]' THEN 'Potential Loyalist'
        WHEN rfm_class LIKE  '[4-5][1-3][1-3]' THEN 'New Customers'
        WHEN rfm_class LIKE  '[2-3][2-3][2-4]' THEN 'Need Attention/Hibernating'
        WHEN rfm_class LIKE  '111' THEN 'Lost'
    ELSE NULL
    END AS rfm_category
    FROM view_rfm_class3

SELECT *
    INTO table_rfm_result3
    FROM view_rfm_result3;
```

Product Analysis

```sql
-- view for product for easier queries
CREATE VIEW view_product_od2 AS
SELECT O.CustomerID,
    COUNT(O.OrderID) AS Quantity,
    P.ProductName,
    C.CategoryName
    FROM [Northwind].[dbo].[Orders] AS O
    JOIN [Northwind].[dbo].[Order Details] AS OD
    ON O.OrderID = OD.OrderID
    JOIN [Northwind].[dbo].[Products] AS P
    ON OD.ProductID=P.ProductID
    JOIN [Northwind].[dbo].[Categories] AS C
    ON P.CategoryID = C.CategoryID
    WHERE YEAR(O.OrderDate) = 1997
    GROUP BY O.CustomerID, P.ProductName, C.CategoryName;

-- Pivot on category on high performance customer
CREATE VIEW view_product_rfm AS
SELECT POD.CustomerID, POD.Quantity, POD.ProductName,
    POD.CategoryName, RFM.rfm_category
    FROM view_product_od2 AS POD
    LEFT JOIN table_rfm_result3 AS RFM
    ON RFM.CustomerID = POD.CustomerID
    WHERE RFM.rfm_category IS NOT NULL;
```

```sql
-- Most Popular Category Name by rfm category
CREATE VIEW view_product_to_rfm AS
SELECT *
    FROM (
        SELECT X.rfm_category, X.CategoryName, X.TotalQuantity,
        DENSE_RANK() OVER (PARTITION BY X.rfm_category
                ORDER BY X.TotalQuantity DESC) AS Rank_
        FROM (
            SELECT rfm_category, CategoryName, SUM(Quantity) AS TotalQuantity
                FROM view_product_rfm
                GROUP BY rfm_category, CategoryName
        ) AS X
    ) AS Y
    WHERE Y.Rank_ <= 3


SELECT *
    INTO table_product_to_rfm
    FROM view_product_to_rfm;
```

Supplier Analysis