Group Assignment

Course: Business Intelligence for Data Science (880682-M-6)
Semester: Spring 2024/Block 4

Business Topic

Analyzing customer purchasing behavior and selling performance at SportOrder Inc. to optimize marketing strategy and operational efficiency.

Contents

1 Motivation & Goals 2
2 Source OLTP schema 2
3 Transferring OLTP schema to Tables in MySQL 3
4 DW schema 4
5 Dimensions & Facts of DW 5
6 ETL transformations in Pentaho Data Integration 6
7 View in MySQL 9
8 Report in Pentaho Report Designer 11

1 MOTIVATION & GOALS

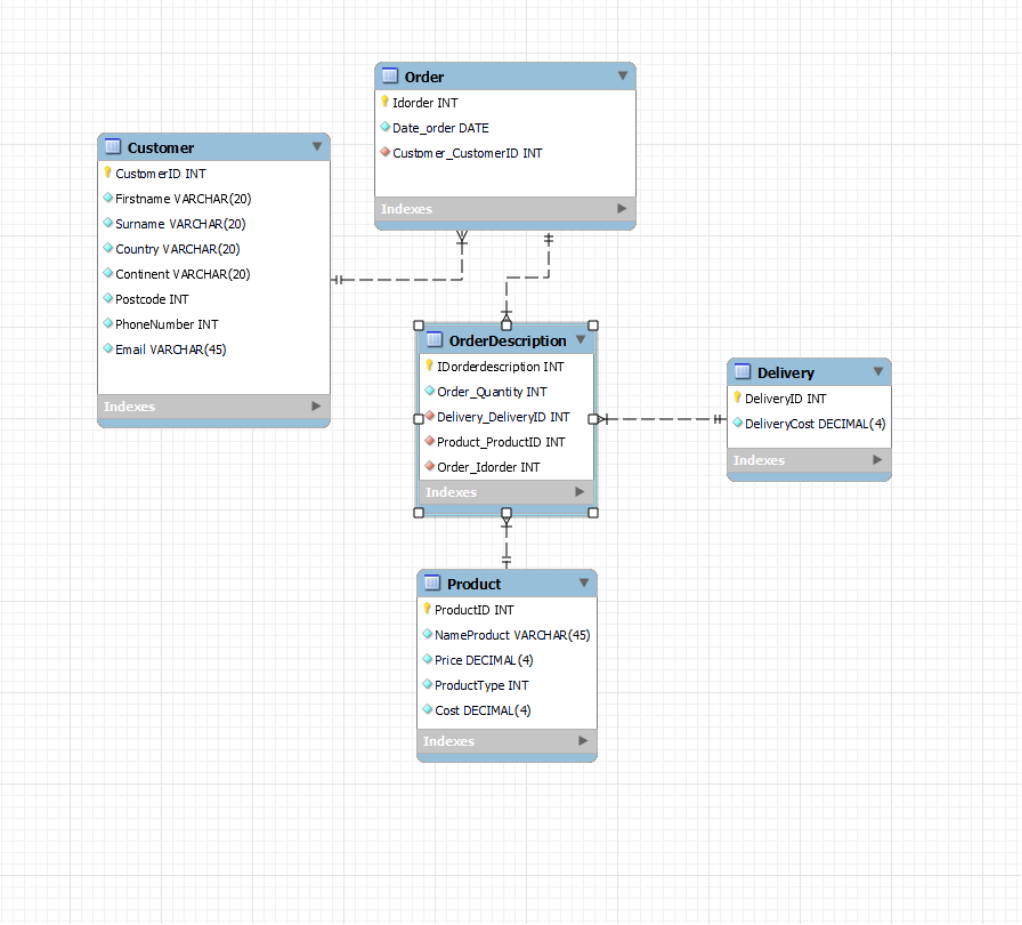
The main motivation for building a Data Warehouse (DW) within our chosen enterprise, SportOrder Inc., is to create a centralized repository that can consolidate data from various sources such as customer information, product details, orders, and deliveries. This centralization aims to improve data consistency and accuracy, reducing redundancy and inconsistency issues.

The goals include enhancing decision-making capabilities by providing comprehensive analytics and reports, optimizing inventory management through better tracking of sales and stock levels, and improving customer satisfaction by analyzing purchasing patterns and preferences. Additionally, the DW will help streamline operations by providing insights into marketing performance and quarterly profit. Ultimately, the DW will support the enterprise in making data-driven decisions, improving operational efficiency, and increasing overall business performance.

2 SOURCE OLTP SCHEMA

.

Figure 1: OLTP schema for the enterprise.



The OLTP schema consists of five tables: **Customer**, **Order**, **OrderDescription**, **Product**, and **Delivery**. Each customer can place multiple orders, as indicated by the foreign key *Customer_CustomerID* in the **Order** table. Each order can have multiple order descriptions, with the **OrderDescription** table containing the foreign key *Order_Idorder* linking back to the **Order** table. The **OrderDescription** table also references products and deliveries through the foreign keys *Product_ProductID* and *Delivery_DeliveryID*, respectively. This schema ensures that each order is associated with specific products and delivery details, maintaining the integrity and relational structure of the data.

3 TRANSFERRING OLTP SCHEMA TO TABLES IN MYSQL

Figure 2: Transferring the OLTP into Tables.

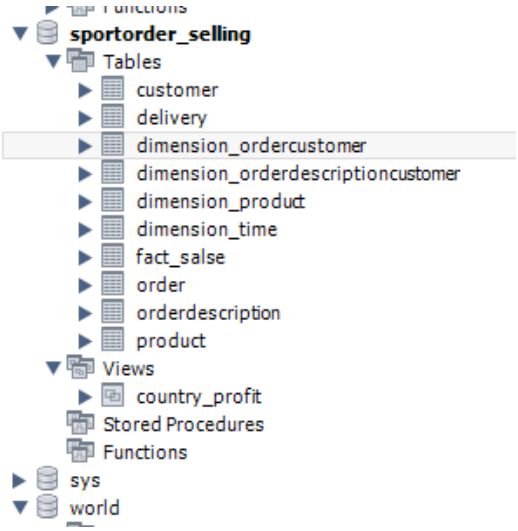


Figure 3: Imported data for the consumer table.

CustomerID	Firstname	Surname	Country	Continent	Postcode	PhoneNumber	Email
1	Ryan	Flynn	Spain	Europe	73091	9697354927	ryan.flynn@mail.es
2	Tonya	Jackson	Russia	Europe	8116	1958682839	tonya.jackson@rumail.com
3	Amy	Payne	Spain	Europe	72653	9963333984	amy.payne@mail.es
4	Seth	Phillips	UK	Europe	7437	1127978050	seth.phillips@example.co.uk
5	Stephen	Harris	Russia	Europe	66746	9703905708	stephen.harris@example.ru
6	Melissa	Phillips	Austria	Europe	2762	2801823865	melissa.phillips@mail.at
7	Joshua	Rhodes	Ireland	Europe	78816	4733616106	joshua.rhodes@example.ie
8	Tammy	Thomas	Norway	Europe	55775	6756332103	tammy.thomas@example.no
9	Ryan	Castaneda	France	Europe	49218	2445662552	ryan.castaneda@example.fr
10	Maria	Alexander	Sweden	Europe	71288	5710360937	maria.alexander@mail.se
11	Amanda	Hamilton	Egypt	Africa	17999	4466589547	amanda.hamilton@egmail.com
12	Carly	Johnson	Denmark	Europe	42262	3303082072	carly.johnson@mail.dk
13	Michelle	Zimmerman	Japan	Asia	72322	7665963680	michelle.zimmerman@jpmail....
14	Angela	Booker	Belgium	Europe	18399	3479708575	angela.booker@bemail.com
15	Steven	Warren	Japan	Asia	3860	9786748744	steven.warren@example.jp
16	Joseph	Church	Italy	Europe	10224	5231494181	joseph.church@example.it
17	Billy	Garrett	Spain	Europe	57536	6927596981	billy.garrett@mail.es
18	Laura	Johnson	Belgium	Europe	14604	5993561289	laura.johnson@mail.be

Figure 4: Imported data for the order table.

IDorderdescription	Order_Quantity	Delivery_DeliveryID	Product_ProductID	Order_Idorder
1	2	48	58	95822412
2	15	9	126	13356886
3	6	38	280	46913810
4	2	6	17	39958838
5	33	39	120	23756669
6	13	46	288	83197857
7	27	15	280	89254563
8	1	49	143	14265799
9	22	18	217	22575562
10	49	22	111	41227216
11	25	7	48	90801586
12	39	17	177	85329037
13	35	8	236	97226012
14	36	19	41	83140807
15	37	13	186	39587039
16	43	15	24	89089901
17	15	7	41	10872248
18	30	41	143	31429110
19	24	23	84	66722344
20	45	44	137	47295260

4 dw SCHEMA

Figure 5: DW schema for Sport Order Selling.

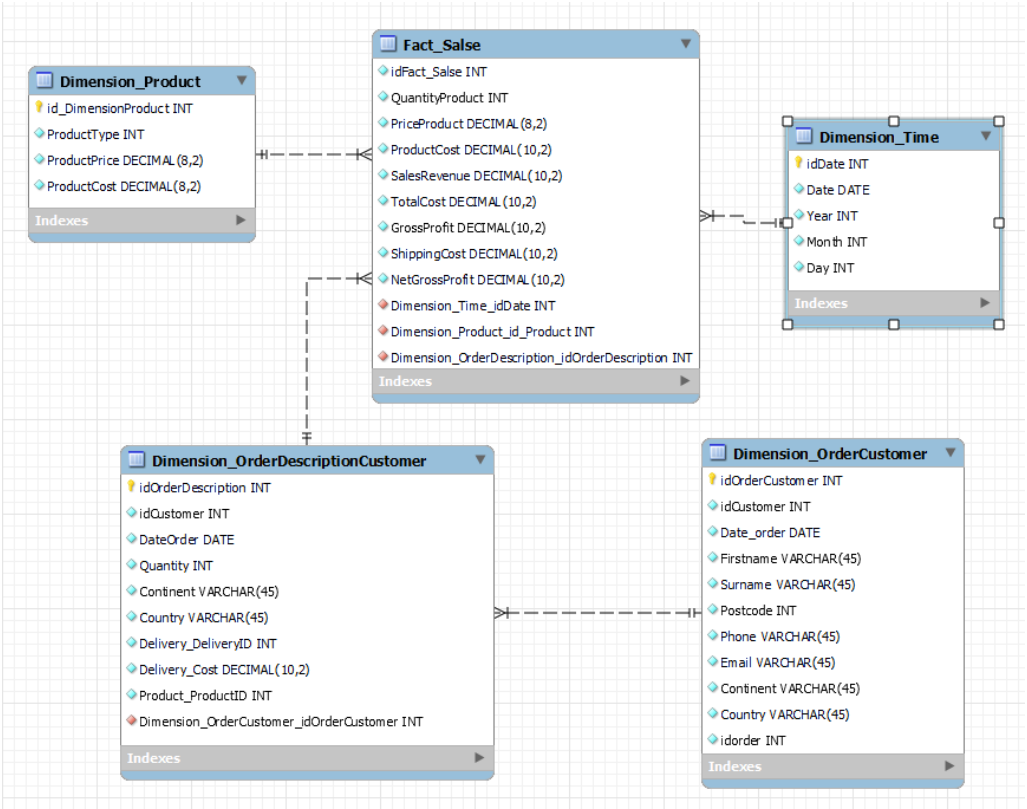
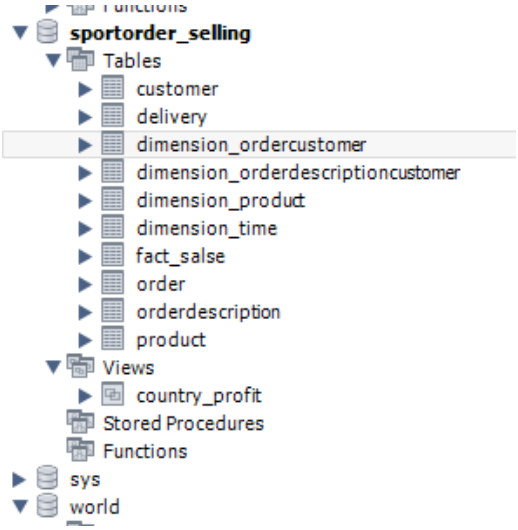


Figure 6: Transferring the DW into Tables.



5 dIMEnsIOns & fACTs of dw

The Data Warehouse (DW) schema consists of several dimension tables and one fact table.

Dimensions:

- **Dimension_Product**: Contains product details such as product ID, type, price, and cost.
- **Dimension_Time**: Holds time-related information including date, year, and month.
- **Dimension_OrderCustomer**: Stores customer order details such as customer ID, order date, and customer information.
- **Dimension_OrderDescriptionCustomer**: Includes order description details linked to customers, delivery, and product.

Fact Table:

- **Fact_Sales**: This table calculates and stores key performance metrics such as:
  - **QuantityProduct**: Number of products sold. This metric helps in understanding the sales volume and identifying the best-selling products.
  - **PriceProduct**: Sales price of the products. Tracking the sales price is crucial for revenue calculation and pricing strategy.
  - **ProductCost**: Cost of the products. Knowing the product cost is essential for calculating profit margins.
  - **SalesRevenue**: Total revenue generated from sales. This metric is vital for assessing the overall income from sales activities.
  - **TotalCost**: Overall cost generated from is obtained by multiplying the unit price of the product by the quantity selling of the product.
  - **GrossProfit**: Difference between sales revenue and product cost. Gross profit indicates the profitability before deducting operating expenses.

- **ShippingCost**: Cost associated with delivering the products. This metric helps in analyzing delivery expenses.
- **NetGrossProfit**: Gross profit minus shipping cost. This metric offers a clearer picture of profitability by accounting for shipping costs.

These metrics are calculated to provide insights into sales performance, profitability, and cost management. They help in evaluating business performance, optimizing pricing strategies, managing costs effectively, and making informed decisions to enhance operational efficiency and financial health.

Schema Variant: The chosen schema variant is a **snowflake schema**, where dimension tables are normalized to reduce redundancy and improve data integrity. This design involves more complex queries due to additional tables and relationships, but it optimizes storage space and ensures consistent data.

6 ETl TRAnsfORMATIOns In pENTAHO DATA INTEGRATION

Figure 7: Dimension Time Transformation.

Figure 8: Dimension Time in MySQL.

idDate	Date	Year	Month	Day
1	2023-05-02	2023	5	2
2	2023-05-03	2023	5	3
3	2023-05-04	2023	5	4
4	2023-05-05	2023	5	5
5	2023-05-06	2023	5	6
6	2023-05-07	2023	5	7
7	2023-05-08	2023	5	8
8	2023-05-09	2023	5	9
9	2023-05-10	2023	5	10
10	2023-05-11	2023	5	11
11	2023-05-12	2023	5	12
12	2023-05-13	2023	5	13
13	2023-05-14	2023	5	14
14	2023-05-15	2023	5	15
15	2023-05-16	2023	5	16
16	2023-05-17	2023	5	17

Figure 9: Dimension Customer Order Transformation.

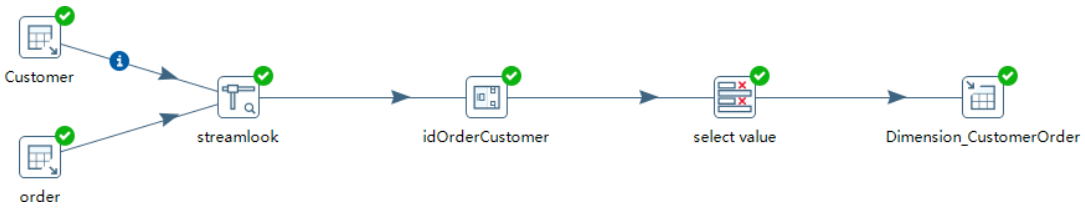


Figure 10: Dimension Customer Order in MySQL.

idOrderCustomer	idCustomer	Date_order	Firstname	Surname	Postcode	Phone	Email	Continent	Country	idorder
1	312	2024-04-21	Elizabeth	Santiago	12774	4014119601	elizabeth.santiago@nomail.com	Europe	Norway	10054484
2	614	2023-06-01	Jeffrey	Perez	45439	1993998799	jeffrey.perez@demail.com	Europe	Germany	10076758
3	419	2023-06-06	Robert	Hoffman	62340	3760786117	robert.hoffman@ukmail.com	Europe	UK	10359129
4	342	2023-12-31	Richard	Smith	94605	1754603393	richard.smith@mail.no	Europe	Norway	10475894
5	915	2024-03-21	Nicholas	Davis	17837	6102144866	nicholas.davis@example.jp	Asia	Japan	10744212
6	778	2023-08-16	Kathryn	Rose	89400	8761469086	kathryn.rose@mail.nl	Europe	Netherlands	10872248
7	684	2023-12-17	Christopher	Watkins	98090	4436189087	christopher.watkins@esmail.com	Europe	Spain	11049999
8	115	2023-08-26	Erin	White	86698	7049001459	erin.white@mail.es	Europe	Spain	11297845
9	530	2023-12-12	John	Maldonado	12568	3774110400	john.maldonado@example.it	Europe	Italy	11362459
10	697	2023-12-27	Grace	Graves	34604	9953941447	grace.graves@example.cn	Asia	China	11540956
11	588	2023-09-24	Matthew	White	46488	6582978524	matthew.white@example.de	Europe	Germany	11646234
12	870	2024-05-23	Jon	Copeland	33499	6663846004	jon.copeland@example.com	North A...	USA	11698142
13	726	2023-06-06	Michael	Bailey	2250	9914274467	michael.bailey@example.dk	Europe	Denmark	11898961
14	993	2024-01-13	Rebecca	Robertson	97109	4659306669	rebecca.robertson@usermail.com	North A...	USA	11980765
15	96	2024-03-03	Jason	Gray	8798	9155819156	jason.gray@example.co.uk	Europe	UK	12314351
16	768	2023-06-28	Anna	Palmer	76535	2331039876	anna.palmer@jpmail.com	Asia	Japan	12601580
17	115	2024-02-01	Erin	White	86698	7049001459	erin.white@mail.es	Europe	Spain	12614124
18	878	2023-06-22	Shannon	Hall	31830	5508149871	shannon.hall@esmail.com	Europe	Spain	12621534
19	637	2023-08-18	Derek	Fuentes	71303	4602080221	derek.fuentes@mail.ru	Europe	Russia	12735359
20	784	2023-12-10	Virginia	Lucas	23336	3185386224	virginia.lucas@example.es	Europe	Spain	12784407
21	676	2023-09-14	Ashley	Lawrence	29918	1455028354	ashley.lawrence@example.no	Europe	Norway	13176186

Figure 11: Dimension Order Description Customer Transformation.

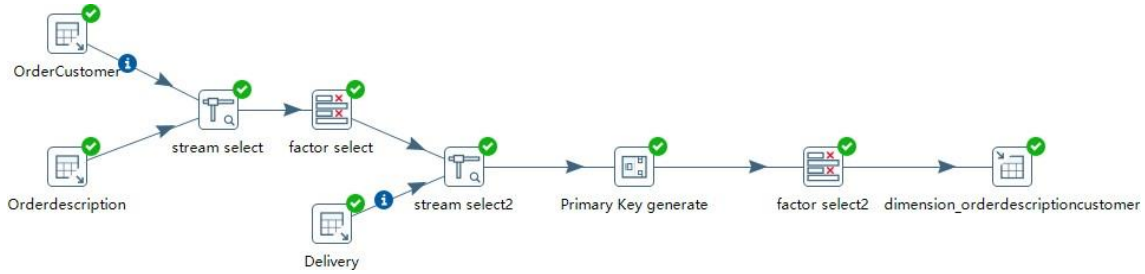


Figure 12: Dimension Order Description Customer in MySQL.

idOrderDescription	idCustomer	DateOrder	Quantity	Continent	Country	Delivery_DeliveryID	DeliveryCost	Dimension_OrderCustomer_idOrderCustomer	Product_ProductID
1	115	2024-01-17	2	Europe	Spain	48	11.00	568	58
2	760	2023-06-06	15	Asia	Japan	9	13.00	23	126
3	251	2023-09-05	6	Asia	Japan	38	17.00	249	280
4	143	2023-08-17	2	Europe	France	6	19.00	190	17
5	693	2024-02-21	33	Europe	Denmark	39	22.00	88	120
6	90	2024-01-30	13	Africa	Egypt	46	10.00	480	288
7	433	2024-04-18	27	Europe	Italy	15	18.00	528	280
8	31	2023-06-28	1	Europe	Denmark	49	18.00	30	143
9	224	2023-10-29	22	Europe	Sweden	18	17.00	81	217
10	518	2023-06-07	49	Europe	Spain	22	19.00	200	111
11	28	2023-08-16	25	Europe	Switzerl...	7	23.00	538	48
12	204	2023-11-28	39	Europe	Austria	17	9.00	498	177
13	719	2023-06-06	35	Europe	Austria	8	7.00	578	236
14	430	2023-08-08	36	Europe	Norway	19	21.00	479	41
15	460	2024-01-20	37	Europe	Switzerl...	13	6.00	188	186
16	285	2023-12-13	43	Europe	Norway	15	18.00	526	24
17	778	2023-08-16	15	Europe	Netherl...	7	23.00	6	41
18	715	2023-12-29	30	Africa	Egypt	41	19.00	139	143
19	349	2024-03-19	24	Asia	China	23	12.00	376	84
20	160	2023-05-30	45	Europe	Spain	44	11.00	252	137
21	981	2024-03-17	39	Europe	France	41	19.00	182	37

Figure 13: Dimension Product Transformation.

Figure 14: Dimension Product in MySQL.

idDimension_Product	ProductType	ProductPrice	ProductCost
1	6	1282.00	269.00
2	6	454.00	180.00
3	1	1095.00	123.00
4	5	445.00	49.00
5	4	406.00	76.00
6	2	1183.00	449.00
7	2	687.00	332.00
8	4	680.00	94.00
9	1	1697.00	665.00
10	1	1077.00	346.00
11	5	1661.00	294.00
12	3	148.00	73.00
13	4	1712.00	362.00
14	2	1275.00	316.00
15	6	427.00	203.00
16	5	1300.00	509.00
17	3	335.00	166.00
18	2	1284.00	480.00
19	2	1687.00	723.00
20	3	1612.00	204.00
21	6	1827.00	413.00

Figure 15: Fact Sale Transformation.

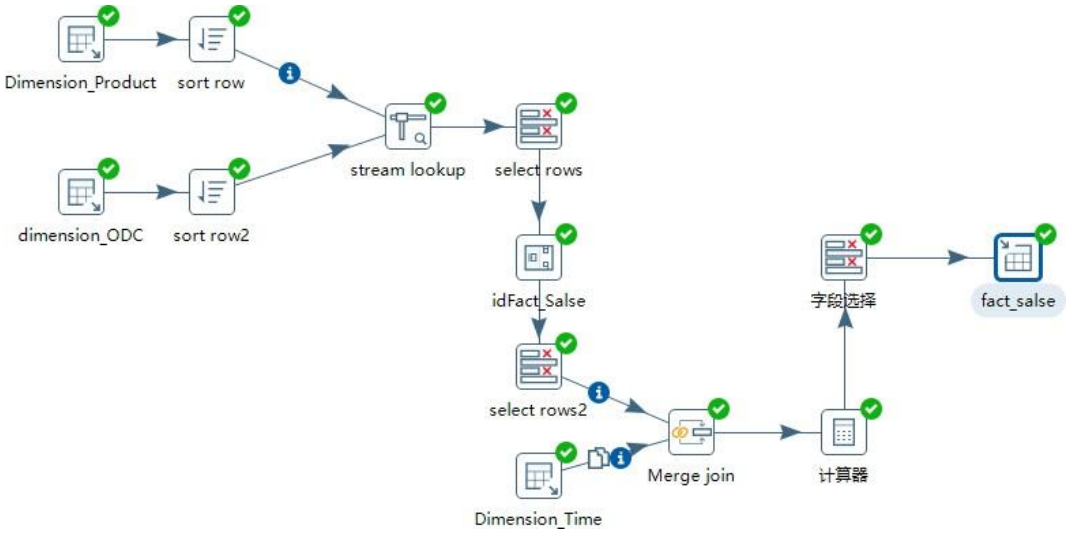


Figure 16: Fact Sale in MySQL.

idFact_Salse	QuantityProduct	PriceProduct	ProductCost	SalesRevenue	TotalCost	GrossProfit	ShippingCost	NetGrossProfit	Dimension_Time_idDate	Dimension_Product_id_Product	Dimension_OrderDescription_idOrderDescription
1	2	1086.00	511.00	2172.00	1022.00	1150.00	11.00	1139.00	1	58	1
2	15	1610.00	193.00	24150.00	2895.00	21255.00	13.00	21242.00	2	126	2
3	6	241.00	37.00	1446.00	222.00	1224.00	17.00	1207.00	3	280	3
4	2	335.00	166.00	670.00	332.00	338.00	19.00	319.00	4	17	4
5	33	1505.00	636.00	49665.00	20988.00	28677.00	22.00	28655.00	5	120	5
6	13	432.00	179.00	5616.00	2327.00	3289.00	10.00	3279.00	6	288	6
7	27	241.00	37.00	6507.00	999.00	5508.00	18.00	5490.00	7	280	7
8	1	877.00	206.00	877.00	206.00	671.00	18.00	653.00	8	143	8
9	22	814.00	296.00	17908.00	6512.00	11396.00	17.00	11379.00	9	217	9
10	49	1698.00	380.00	83202.00	18620.00	64582.00	19.00	64563.00	10	111	10
11	25	564.00	178.00	14100.00	4450.00	9650.00	23.00	9627.00	11	48	11
12	39	744.00	177.00	29016.00	6903.00	22113.00	9.00	22104.00	12	177	12
13	35	951.00	471.00	33285.00	16485.00	16800.00	7.00	16793.00	13	236	13
14	36	1746.00	823.00	62856.00	29628.00	33228.00	21.00	33207.00	14	41	14
15	37	1889.00	848.00	69893.00	31376.00	38517.00	6.00	38511.00	15	186	15
16	43	533.00	245.00	22919.00	10535.00	12384.00	18.00	12366.00	16	24	16
17	15	1746.00	823.00	26190.00	12345.00	13845.00	23.00	13822.00	17	41	17
18	30	877.00	206.00	26310.00	6180.00	20130.00	19.00	20111.00	18	143	18
19	24	1707.00	342.00	40968.00	8208.00	32760.00	12.00	32748.00	19	84	19
20	45	668.00	83.00	30060.00	3735.00	26325.00	11.00	26314.00	20	137	20
21	39	1665.00	205.00	64935.00	7995.00	56940.00	19.00	56921.00	21	37	21

7 VIEW IN MYSQL

Figure 17: View that contains only Europe from selling report in 2024 Q1.

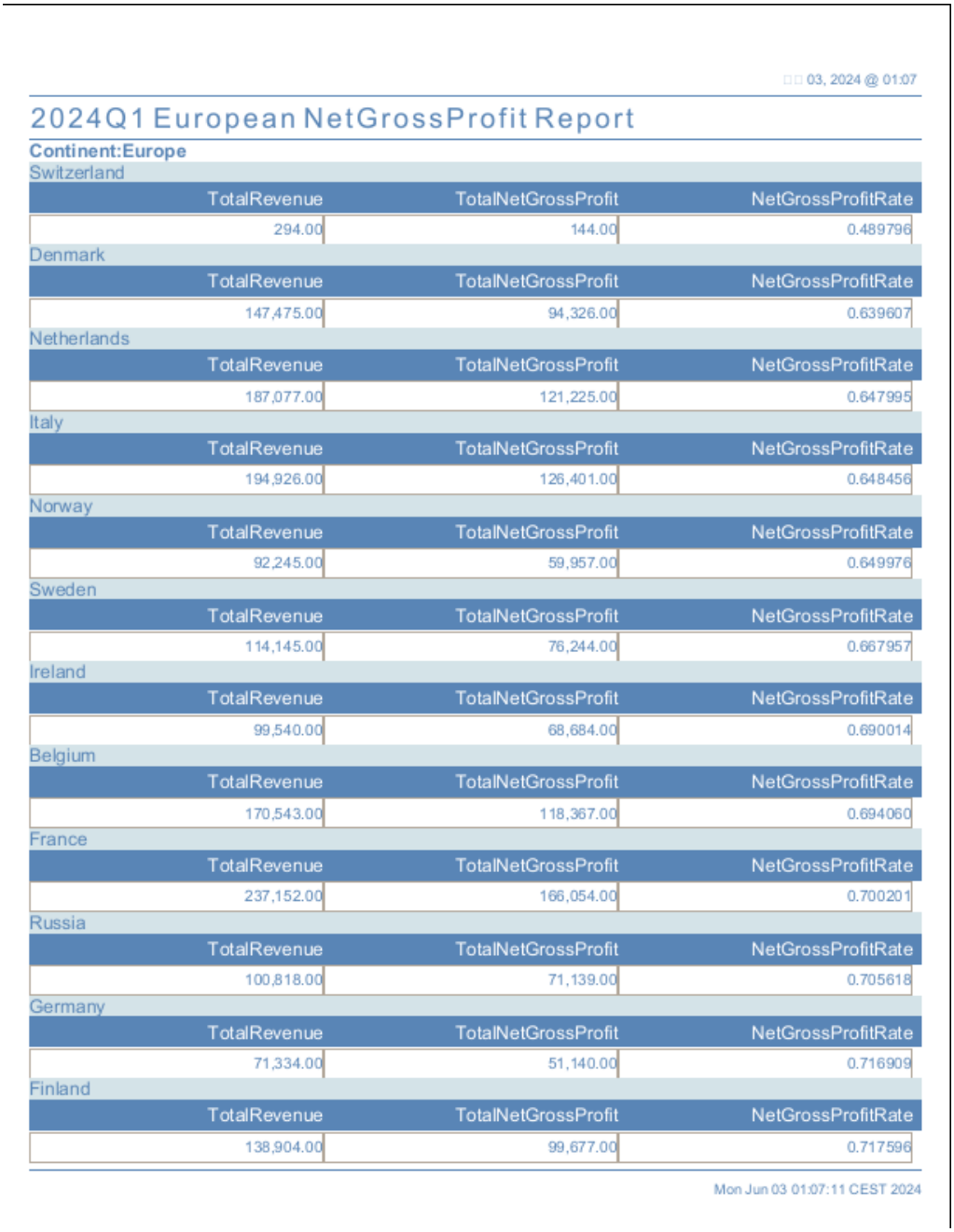
```sql
CREATE VIEW Country_Profit
AS
SELECT idFact_Salse,SalesRevenue,TotalCost,GrossProfit,ShippingCost,NetGrossProfit,ODC.Country,ODC.Continent,DT.Month,DT.Year
FROM fact_salse
INNER JOIN dimension_orderdescriptioncustomer AS ODC
ON fact_salse.Dimension_OrderDescription_idOrderDescription=ODC.idOrderDescription
INNER JOIN dimension_time AS DT
ON fact_salse.Dimension_Time_idDate=DT.idDate;

SELECT Continent, Country,
    SUM(SalesRevenue) AS TotalSalesRevenue,
    SUM(NetGrossProfit)  AS TotalNetGrossProfit,
    (SUM(NetGrossProfit)/SUM(SalesRevenue)) AS NetGrossProfit_Rate
FROM sportorder_selling.country_profit
WHERE Continent='Europe' AND Year='2024' AND Month IN (1,2,3)
GROUP BY Continent, Country
ORDER BY NetGrossProfit_Rate ASC;

```

Figure 18: Vieuw Results.

Continent	Country	TotalSalesRevenue	TotalNetGrossProfit	NetGrossProfit_Rate
Europe	Switzerland	294.00	144.00	0.489796
Europe	Denmark	147475.00	94326.00	0.639607
Europe	Netherlands	187077.00	121225.00	0.647995
Europe	Italy	194926.00	126401.00	0.648456
Europe	Norway	92245.00	59957.00	0.649976
Europe	Sweden	114145.00	76244.00	0.667957
Europe	Ireland	99540.00	68684.00	0.690014
Europe	Belgium	170543.00	118367.00	0.694060
Europe	France	237152.00	166054.00	0.700201
Europe	Russia	100818.00	71139.00	0.705618
Europe	Germany	71334.00	51140.00	0.716909
Europe	Finland	138904.00	99677.00	0.717596
Europe	UK	278938.00	204630.00	0.733604
Europe	Spain	71984.00	53069.00	0.737233
Europe	Austria	83050.00	67894.00	0.817508

8 REPORT IN PENTAHO REPORT DESIGNER

Figure 19: Pentaho Report, NetGrossProfit Rate per country for Q1 2024 Europe.



□□ 03, 2024 @ 01:07

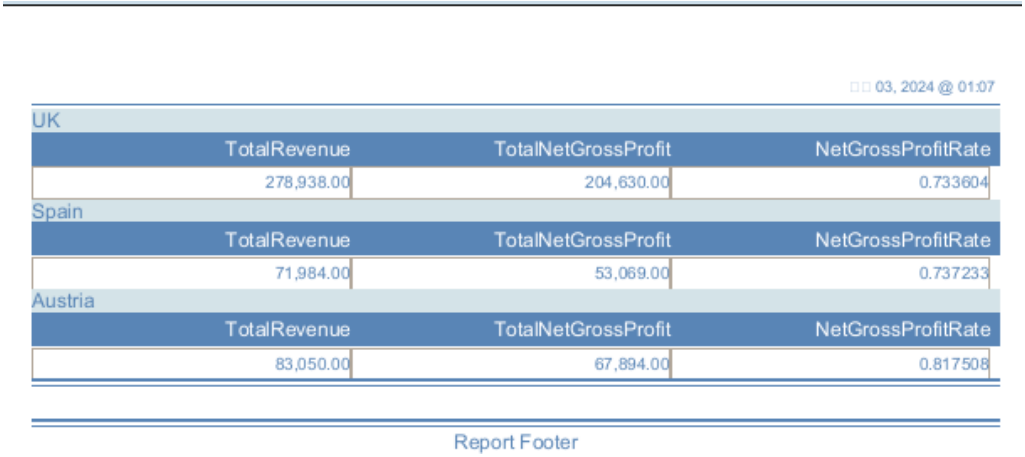
2024Q1 European NetGrossProfit Report

Continent:Europe

Switzerland		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
294.00	144.00	0.489796
Denmark		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
147,475.00	94,326.00	0.639607
Netherlands		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
187,077.00	121,225.00	0.647995
Italy		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
194,926.00	126,401.00	0.648456
Norway		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
92,245.00	59,957.00	0.649976
Sweden		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
114,145.00	76,244.00	0.667957
Ireland		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
99,540.00	68,684.00	0.690014
Belgium		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
170,543.00	118,367.00	0.694060
France		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
237,152.00	166,054.00	0.700201
Russia		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
100,818.00	71,139.00	0.705618
Germany		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
71,334.00	51,140.00	0.716909
Finland		
TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
138,904.00	99,677.00	0.717596

Mon Jun 03 01:07:11 CEST 2024

Figure 20: Continued.



□□ 03, 2024 @ 01:07

UK

TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
278,938.00	204,630.00	0.733604

Spain

TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
71,984.00	53,069.00	0.737233

Austria

TotalRevenue	TotalNetGrossProfit	NetGrossProfitRate
83,050.00	67,894.00	0.817508

Report Footer