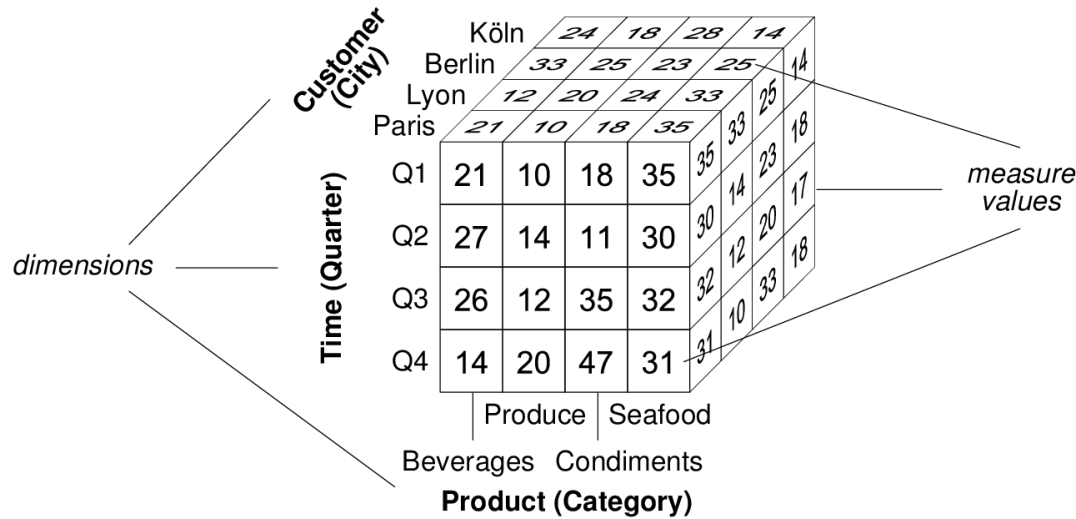# Data Analysis and Integration
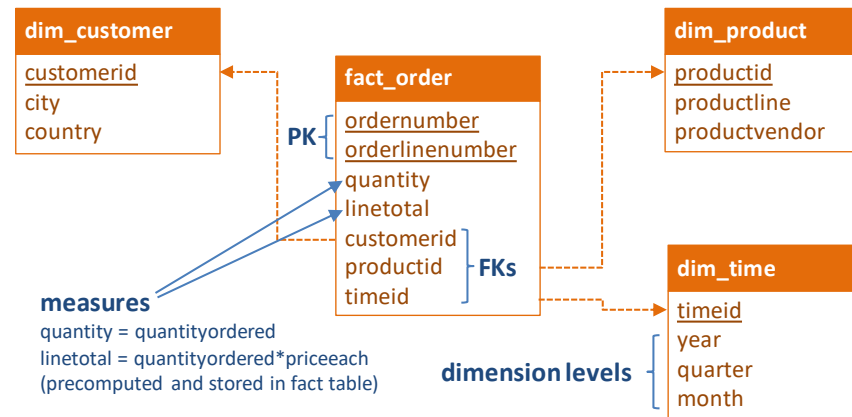
Defining the OLAP cube

Introduction

OLAP cube
OLAP server
MDX queries

DW tables
database system
SQL queries

OLAP cube
definition
(XML)

OLAP server

- ## What is an OLAP server?

  - serves data as a cube

    - data may or may not come from a star schema

  - allows a client to run queries over the data cube

    - we will talk about MDX in the next lectures

  - knows how to perform OLAP operations

    - drill-down, roll-up, slice, dice, pivot, etc.

  - automatically calculates and aggregates measures

    - according to the requested dimension levels

    - with a certain aggregation operation (sum, count, average, etc.)
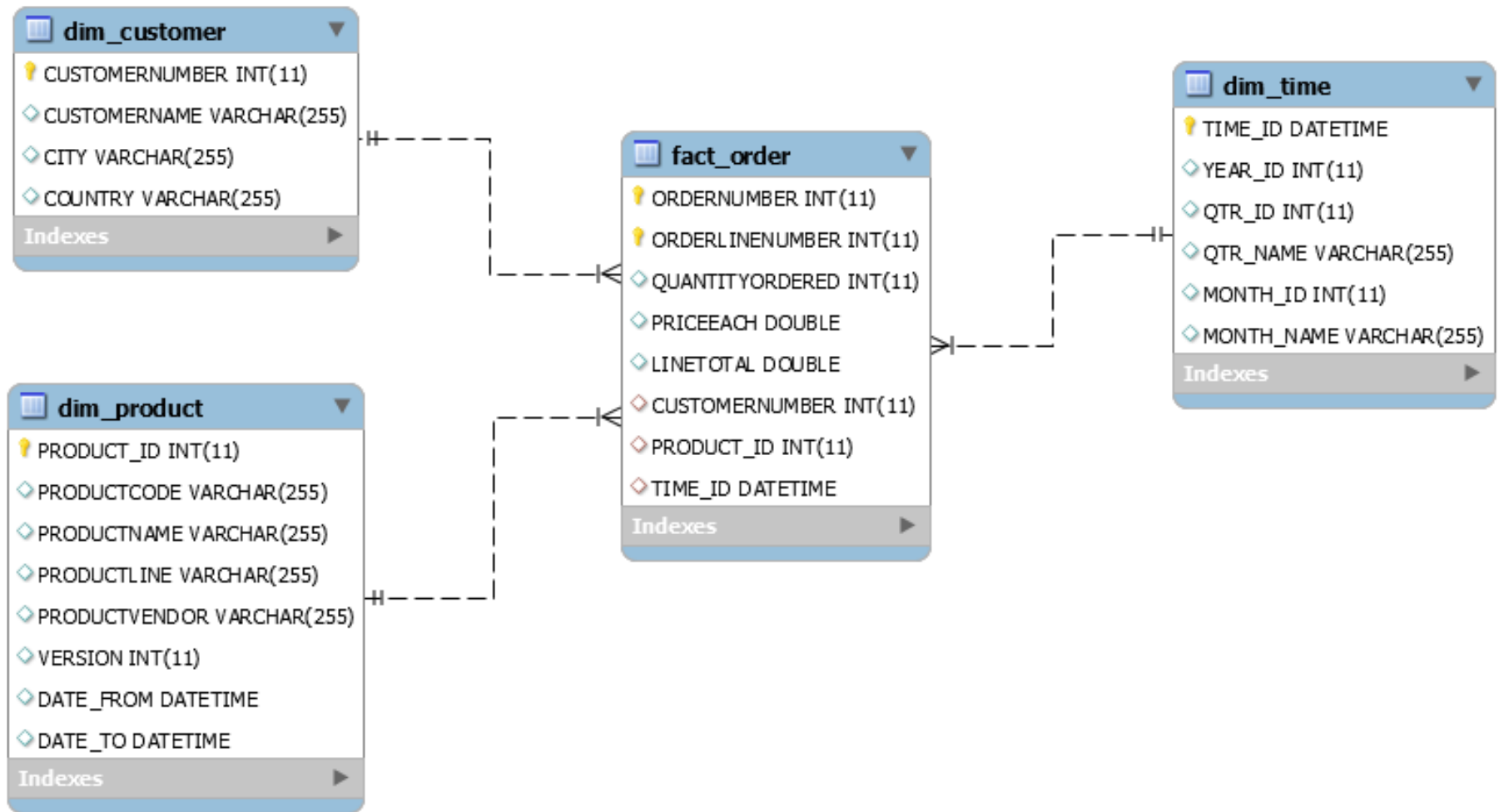
OLAP server

- An OLAP server needs a cube definition
    - where is the fact table?
    - where are the dimensions?
        - where are the hierarchies for each dimension?
        - where are the levels for each hierarchy?
        - where are the tables where those levels can be found?
        - is it a regular dimension or a time dimension?
    - where are the measures?
        - which columns should be aggregated?
        - which aggregation operations should be used?

OLAP server

- Such cube definition is usually provided in an XML file

```
<Schema ...>
    <Cube ...>
        <Table ... />
        <Dimension ...>
            <Hierarchy ...>
                <Table ... />
                <Level ... />
            </Hierarchy>
        </Dimension>
        <Measure ... />
    </Cube>
</Schema>
```

Data warehouse

OLAP cube

- Defining the cube and its fact table

```
<Schema name="steelwheels_dw">
    <Cube name="Orders" ...>
        <Table name="fact_order">
        </Table>
        ...
    </Cube>
</Schema>
```

OLAP cube

- Defining the customer dimension and its hierarchy

```
<Dimension name="Customer"
           foreignKey="CUSTOMERNUMBER"
           type="StandardDimension">
    <Hierarchy name="Customer Hierarchy"
               primaryKey="CUSTOMERNUMBER"
               allMemberName="All Customers">
        <Table name="dim_customer">
        </Table>
        ...
    </Hierarchy>
</Dimension>
```

OLAP cube

- Defining the levels of the customer hierarchy

```
<Level name="Country" column="COUNTRY"
      type="String" levelType="Regular">
</Level>

<Level name="City" column="CITY"
      type="String" levelType="Regular">
</Level>

<Level name="Customer Name" column="CUSTOMERNAME"
      type="String" levelType="Regular">
</Level>
```

OLAP cube

- Defining the product dimension and its hierarchy

```
<Dimension name="Product"
          foreignKey="PRODUCT_ID"
          type="StandardDimension">
    <Hierarchy name="Product Hierarchy"
              primaryKey="PRODUCT_ID"
              allMemberName="All Products">
        <Table name="dim_product">
        </Table>
        ...
    </Hierarchy>
</Dimension>
```

OLAP cube

- Defining the levels of the product hierarchy

```
<Level name="Product Line" column="PRODUCTLINE"
      type="String" levelType="Regular">
</Level>

<Level name="Product Vendor" column="PRODUCTVENDOR"
      type="String" levelType="Regular">
</Level>

<Level name="Product Name" column="PRODUCTNAME"
      type="String" levelType="Regular">
</Level>
```

OLAP cube

- Defining the time dimension and its hierarchy

```
<Dimension name="Time"
           foreignKey="TIME_ID"
           type="TimeDimension">
    <Hierarchy name="Time Hierarchy"
               primaryKey="TIME_ID"
               allMemberName="All Years">
        <Table name="dim_time">
        </Table>
        ...
    </Hierarchy>
</Dimension>
```

OLAP cube

- Defining the levels of the time hierarchy

```
<Level name="Year" column="YEAR_ID"
      type="Integer" levelType="TimeYears">
</Level>


<Level name="Quarter" column="QTR_NAME"
      ordinalColumn="QTR_ID"
      type="String" levelType="TimeQuarters">
</Level>


<Level name="Month" column="MONTH_NAME"
      ordinalColumn="MONTH_ID"
      type="String" levelType="TimeMonths">
</Level>
```
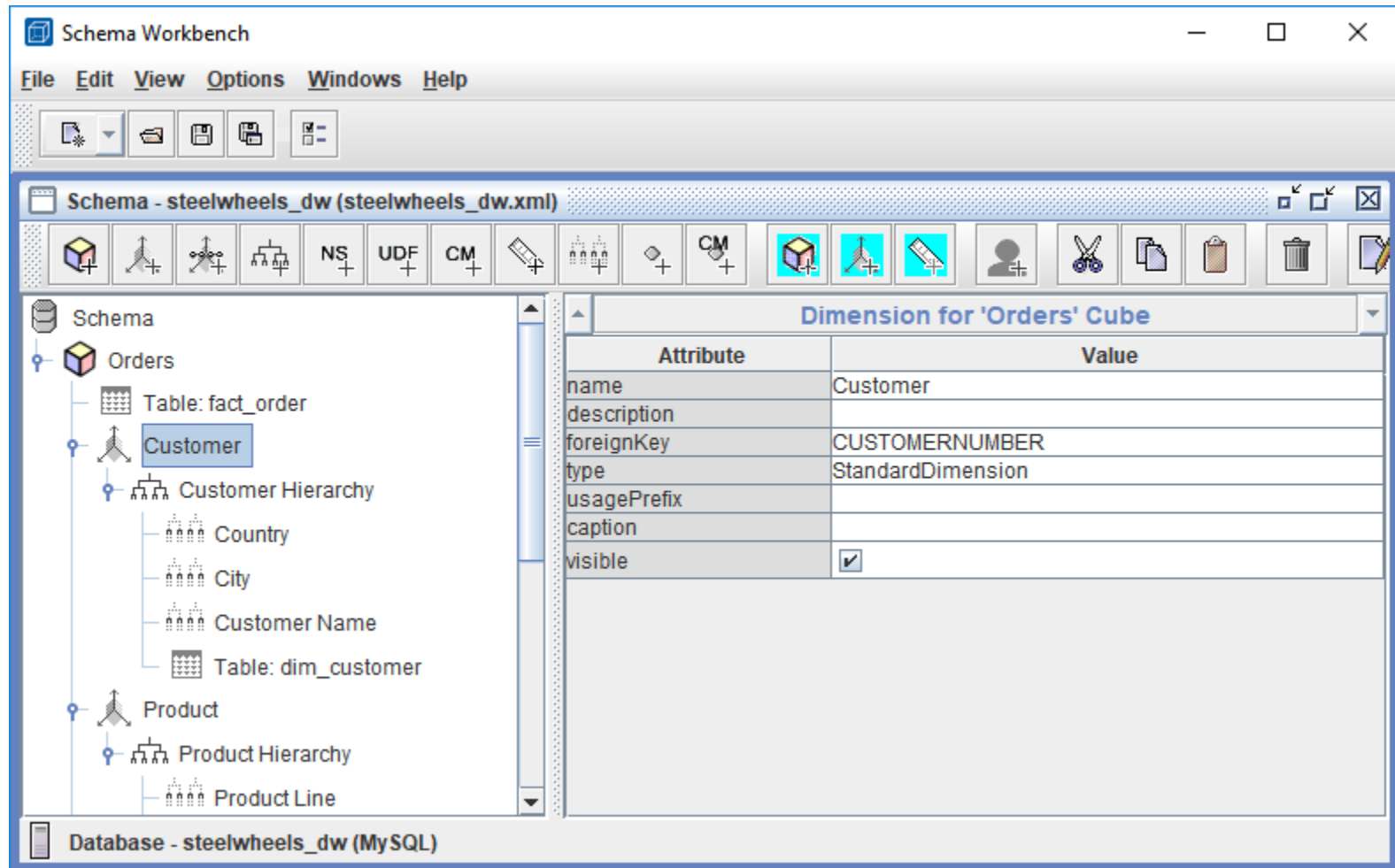
OLAP cube

- Defining the measures

```
<Measure name="Sales" column="LINETOTAL"
        aggregator="sum" datatype="Numeric"
        formatString="$ #,###.00">
</Measure>

<Measure name="Quantity" column="QUANTITYORDERED"
        aggregator="sum" datatype="Integer"
        formatString="#,###">
</Measure>
```

OLAP cube

- Pentaho Schema Workbench (PSW)

OLAP cube

- Pentaho Schema Workbench (PSW)

OLAP cube

- Cube definition is input to OLAP server
  - Mondrian is an OLAP server that can understand such XML cube definitions
    - Mondrian is a JAR file that can be loaded by Java apps

- Running queries over the data cube
  - requires a client tool (or front-end)
  - there are several front-ends for Mondrian
    - here we will use Saiku Analytics as a plug-in for Pentaho Server

OLAP front-end

OLAP front-end

- ## Manage data sources
  - database connection to the data warehouse
  - location of the XML file with the cube definition

OLAP front-end

- ## Manage data sources
  - database connection to the data warehouse
  - location of the XML file with the cube definition

Manage Data Sources

☼˅   New Data Source

Datasource	Type
steelwheels_dw	Analysis
steelwheels_dw	JDBC

OLAP front-end

- Creating a new query
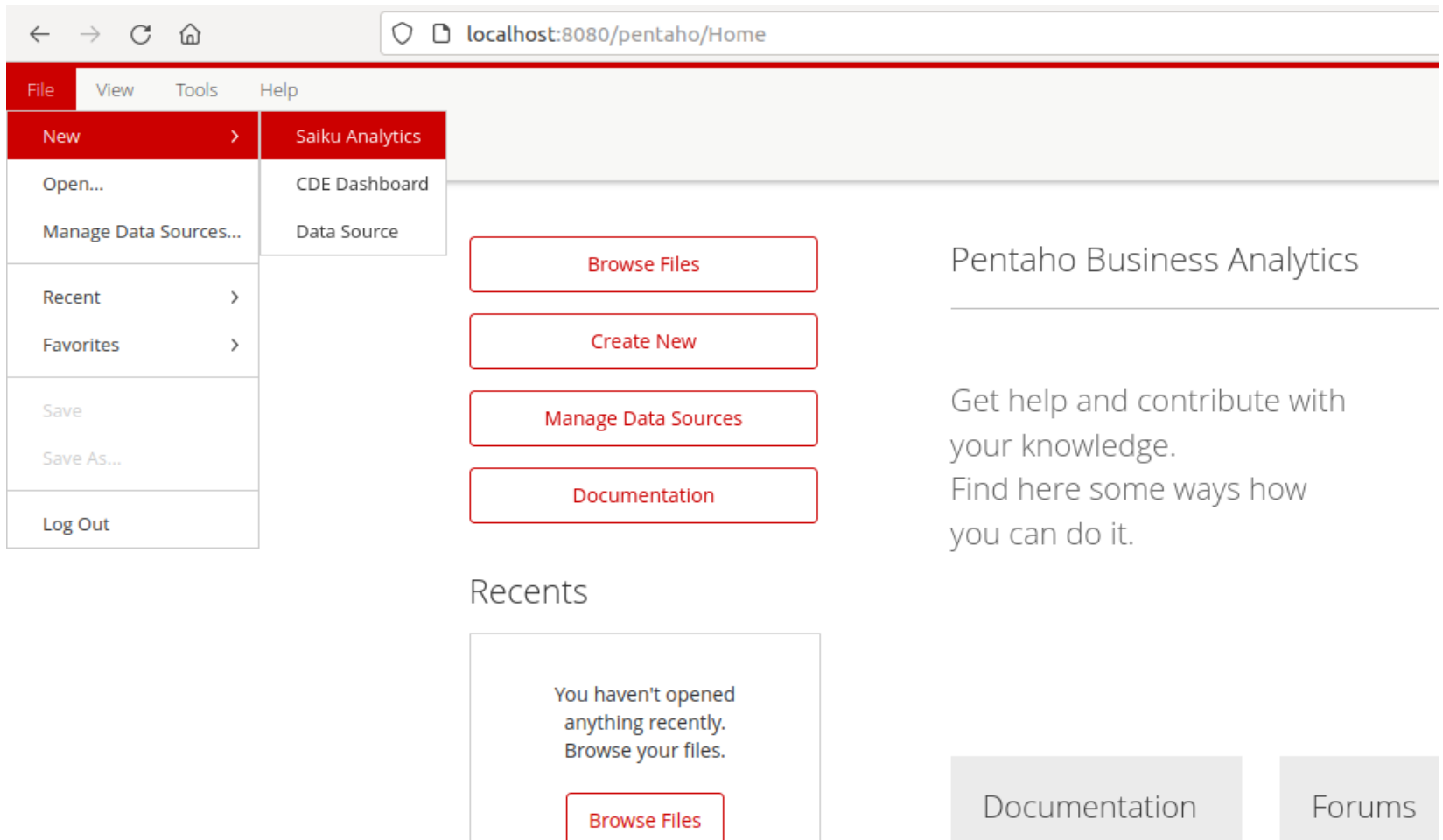
OLAP front-end

- Select the cube
  - if the cube does not show up, there is a problem in the XML
  - check for exceptions in the log file (pentaho.log)
  - go back to PSW and fix the problem
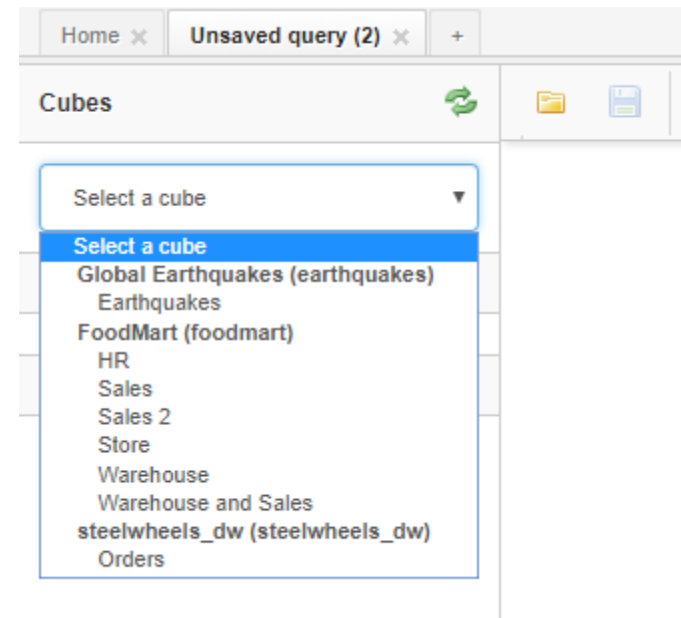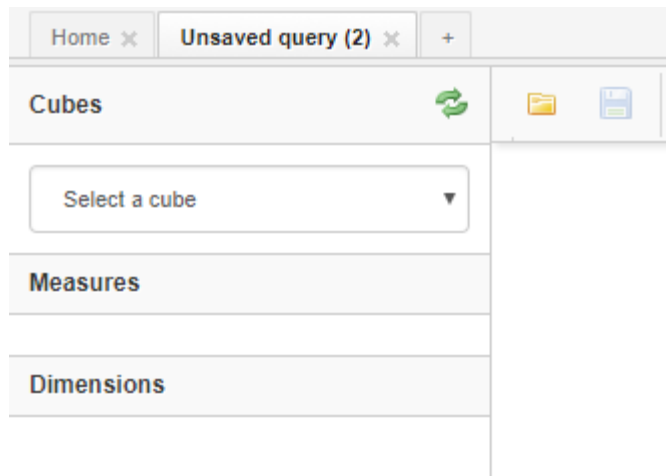
OLAP front-end

- **Front-end**
  - measures
  - dimensions
    - hierarchies
    - levels
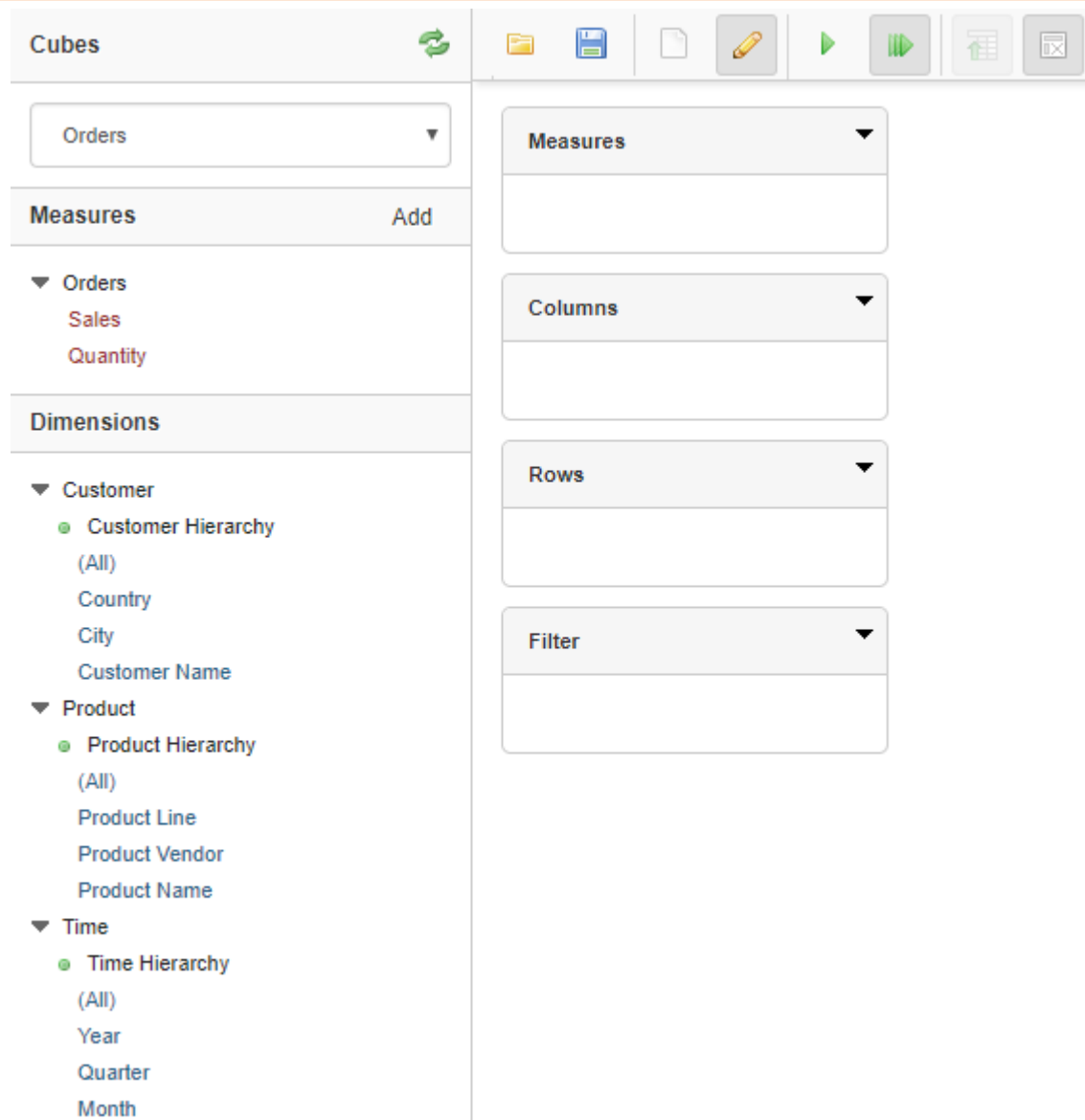
OLAP front-end

- Drag-and-drop
  - drag measures to
    - Measures
  - drag dimensions to
    - Columns
    - Rows
    - Filter

Sales by country

Sales by country and product line

Sales by country, product line and year

Measures
Sales

Columns
Product Hierarchy
Product Line

Rows
Customer Hierarchy
Country
Time Hierarchy
Year

Filter

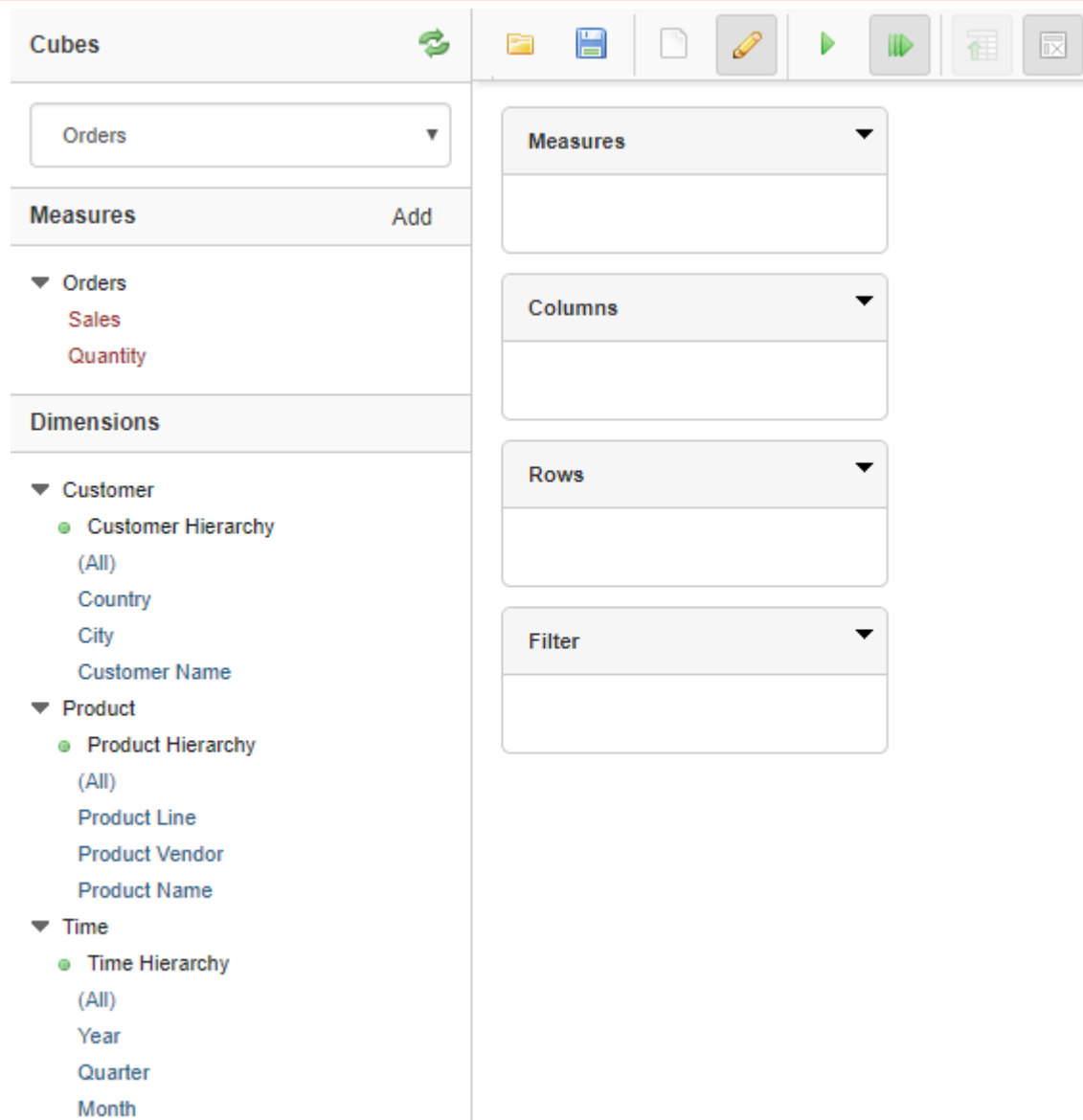| Product Line		Classic Cars	Motorcycles	Planes	Ships	Trains	Trucks and Buses	Vintage Cars
Country	Year	Sales	Sales	Sales	Sales	Sales	Sales	Sales
Australia	2003	$ 85,355.00	$ 42,337.00	$ 22,316.00	-	$ 1,683.00	$ 11,301.00	$ 90,133.00
	2004	$ 76,198.00	$ 33,077.00	$ 41,416.00	$ 1,080.00	-	$ 36,171.00	$ 44,325.00
	2005	$ 31,460.00	$ 14,478.00	$ 11,086.00	$ 3,072.00	-	$ 29,956.00	$ 55,194.00
Austria	2003	$ 26,647.00	-	$ 14,221.00	$ 9,028.00	-	$ 20,460.00	$ 11,775.00
	2004	$ 15,333.00	$ 26,060.00	$ 3,638.00	-	-	-	$ 6,672.00
	2005	$ 59,480.00	-	-	-	-	-	$ 8,775.00
Belgium	2003	-	-	-	-	$ 1,701.00	-	$ 1,640.00
	2004	$ 3,510.00	-	$ 5,617.00	$ 31,709.00	$ 7,315.00	-	$ 31,942.00
	2005	$ 16,631.00	-	-	-	-	-	$ 8,420.00
Canada	2003	$ 28,896.00	-	-	-	-	$ 10,162.00	$ 15,526.00
	2004	$ 20,600.00	$ 4,182.00	$ 25,529.00	$ 40,303.00	-	$ 22,351.00	$ 22,826.00
	2005	$ 12,138.00	-	-	-	-	$ 19,472.00	$ 2,100.00
Denmark	2003	$ 60,756.00	-	-	$ 20,444.00	$ 4,332.00	-	$ 13,673.00
	2004	$ 70,338.00	-	$ 7,570.00	$ 18,252.00	$ 7,150.00	$ 9,581.00	$ 7,512.00
	2005	$ 25,974.00	-	-	-	-	-	-
Finland	2003	$ 33,063.00	-	$ 23,111.00	$ 6,388.00	-	$ 40,469.00	$ 8,045.00
	2004	$ 54,255.00	$ 32,177.00	-	-	$ 5,136.00	-	-
	2005	$ 66,219.00	$ 15,662.00	$ 11,276.00	$ 23,362.00	-	-	$ 10,309.00
France	2003	$ 141,215.00	$ 89,175.00	$ 19,075.00	$ 6,660.00	-	$ 46,057.00	$ 10,731.00
	2004	$ 184,359.00	$ 95,359.00	$ 29,755.00	$ 38,747.00	$ 18,394.00	$ 40,077.00	$ 148,502.00
	2005	$ 63,314.00	$ 41,979.00	$ 59,354.00	$ 21,078.00	$ 8,955.00	$ 30,844.00	$ 17,392.00
Germany	2003	$ 55,476.00	$ 3,139.00	-	-	-	-	$ 11,424.00
	2004	$ 92,725.00	$ 4,362.00	$ 23,031.00	$ 5,484.00	$ 5,042.00	$ 10,182.00	$ 9,489.00

Sales by country, product line and year

Measures
- Sales

Columns
- Product Hierarchy — Product Line
- Time Hierarchy — Year

Rows
- Customer Hierarchy — Country

Filter

Product Line	Classic Cars			Motorcycles			Planes	
Year	2003	2004	2005	2003	2004	2005	2003	2004
Country	Sales	Sales	Sales	Sales	Sales	Sales	Sales	Sales
Australia	$ 85,355.00	$ 76,198.00	$ 31,460.00	$ 42,337.00	$ 33,077.00	$ 14,478.00	$ 22,316.00	$ 41,416.00
Austria	$ 26,647.00	$ 15,333.00	$ 59,480.00	-	$ 26,060.00	-	$ 14,221.00	$ 3,638.00
Belgium	-	$ 3,510.00	$ 16,631.00	-	-	-	-	$ 5,617.00
Canada	$ 28,896.00	$ 20,600.00	$ 12,138.00	-	$ 4,182.00	-	-	$ 25,529.00
Denmark	$ 60,756.00	$ 70,338.00	$ 25,974.00	-	-	-	-	$ 7,570.00
Finland	$ 33,063.00	$ 54,255.00	$ 66,219.00	-	$ 32,177.00	$ 15,662.00	$ 23,111.00	-
France	$ 141,215.00	$ 184,359.00	$ 63,314.00	$ 89,175.00	$ 95,359.00	$ 41,979.00	$ 19,075.00	$ 29,755.00
Germany	$ 55,476.00	$ 92,725.00	-	$ 3,139.00	$ 4,362.00	-	-	$ 23,031.00
Hong Kong	-	-	-	$ 3,850.00	-	-	$ 39,617.00	-
Ireland	-	$ 31,665.00	-	-	$ 4,972.00	-	-	$ 11,785.00
Italy	$ 43,248.00	$ 89,937.00	-	$ 4,046.00	-	$ 7,586.00	$ 54,817.00	$ 28,563.00
Japan	-	$ 28,412.00	$ 18,838.00	-	$ 26,564.00	-	-	$ 49,202.00
New Zealand	$ 29,675.00	$ 71,433.00	$ 66,087.00	$ 18,450.00	$ 30,037.00	$ 51,357.00	$ 20,375.00	$ 26,258.00
Norway	$ 121,028.00	$ 13,659.00	-	$ 31,973.00	$ 19,773.00	-	$ 3,944.00	$ 25,569.00
Philippines	$ 42,572.00	$ 10,503.00	-	$ 12,630.00	$ 5,432.00	-	$ 20,926.00	-
Singapore	$ 78,033.00	$ 54,735.00	-	-	-	$ 4,180.00	-	-
Spain	$ 129,713.00	$ 209,484.00	$ 136,812.00	$ 13,647.00	$ 21,284.00	$ 39,734.00	$ 21,557.00	$ 41,496.00
Sweden	$ 13,161.00	$ 39,944.00	$ 16,029.00	-	-	$ 15,540.00	$ 8,912.00	-
Switzerland	-	$ 117,745.00	-	-	-	-	-	-
UK	$ 66,636.00	$ 92,659.00	-	-	-	$ 40,755.00	$ 8,990.00	$ 32,197.00
USA	$ 558,524.00	$ 560,410.00	$ 225,605.00	$ 178,145.00	$ 287,353.00	$ 55,056.00	$ 90,063.00	$ 177,503.00

Quantity by product line and year

Quantity and sales by product line and year

Measures ▼
Quantity
Sales

Columns ▼
Time Hierarchy
Year

Rows ▼
Product Hierarchy
Product Line

Filter ▼

Year	2003		2004		2005	
Product Line	Quantity	Sales	Quantity	Sales	Quantity	Sales
Classic Cars	12,762	$ 1,513,998.00	16,085	$ 1,837,904.00	6,705	$ 738,587.00
Motorcycles	4,031	$ 397,392.00	5,906	$ 590,632.00	2,771	$ 286,327.00
Planes	3,833	$ 347,924.00	5,820	$ 529,129.00	2,207	$ 200,077.00
Ships	2,844	$ 244,652.00	4,309	$ 375,498.00	1,346	$ 128,219.00
Trains	1,000	$ 72,857.00	1,409	$ 124,885.00	409	$ 36,920.00
Trucks and Buses	4,056	$ 420,523.00	5,024	$ 532,024.00	1,921	$ 201,903.00
Vintage Cars	10,003	$ 1,315,089.00	10,864	$ 997,708.00	4,116	$ 388,817.00

Quantity and sales by product line, country and year

Measures ▼
- Quantity
- Sales

Columns ▼
- Time Hierarchy
- Year

Rows ▼
- Product Hierarchy
- Product Line
- Customer Hierarchy
- Country

Filter ▼

Product Line	Country	Year 2003 Quantity	2003 Sales	2004 Quantity	2004 Sales	2005 Quantity	2005 Sales
Classic Cars	Australia	763	$ 85,355.00	644	$ 76,198.00	411	$ 31,460.00
	Austria	248	$ 26,647.00	180	$ 15,333.00	509	$ 59,480.00
	Belgium	-	-	30	$ 3,510.00	117	$ 16,631.00
	Canada	225	$ 28,896.00	147	$ 20,600.00	84	$ 12,138.00
	Denmark	478	$ 60,756.00	592	$ 70,338.00	174	$ 25,974.00
	Finland	280	$ 33,063.00	388	$ 54,255.00	616	$ 66,219.00
	France	1,210	$ 141,215.00	1,704	$ 184,359.00	626	$ 63,314.00
	Germany	525	$ 55,476.00	756	$ 92,725.00	-	-
	Ireland	-	-	202	$ 31,665.00	-	-
	Italy	304	$ 43,248.00	678	$ 89,937.00	-	-
	Japan	-	-	192	$ 28,412.00	122	$ 18,838.00
	New Zealand	289	$ 29,675.00	633	$ 71,433.00	604	$ 66,087.00
	Norway	1,001	$ 121,028.00	157	$ 13,659.00	-	-
	Philippines	391	$ 42,572.00	87	$ 10,503.00	-	-
	Singapore	507	$ 78,033.00	536	$ 54,735.00	-	-
	Spain	1,072	$ 129,713.00	2,070	$ 209,484.00	1,238	$ 136,812.00
	Sweden	73	$ 13,161.00	296	$ 39,944.00	183	$ 16,029.00
	Switzerland	-	-	1,078	$ 117,745.00	-	-
	UK	662	$ 66,636.00	845	$ 92,659.00	-	-
	USA	4,734	$ 558,524.00	4,870	$ 560,410.00	2,021	$ 225,605.00
Motorcycles	Australia	431	$ 42,337.00	324	$ 33,077.00	121	$ 14,478.00
	Austria	-	-	197	$ 26,060.00	-	-
	Canada	-	-	41	$ 4,182.00	-	-
	Finland	-	-	301	$ 32,177.00	146	$ 15,662.00

Quantity and sales by product line, country and quarter

Measures ▼
- Quantity
- Sales

Columns ▼
Time Hierarchy
- Year
- Quarter

Rows ▼
Product Hierarchy
- Product Line

Customer Hierarchy
- Country

Filter ▼

		Year							
		2003							
	Quarter	Q1		Q2		Q3		Q4	
Product Line	Country	Quantity	Sales	Quantity	Sales	Quantity	Sales	Quantity	Sales
Classic Cars	Australia	-	-	-	-	479	$ 52,520.00	284	$ 32,835.00
	Austria	-	-	46	$ 5,014.00	-	-	202	$ 21,633.00
	Belgium	-	-	-	-	-	-	-	-
	Canada	-	-	-	-	-	-	225	$ 28,896.00
	Denmark	120	$ 20,445.00	-	-	-	-	358	$ 40,311.00
	Finland	-	-	-	-	248	$ 28,551.00	32	$ 4,512.00
	France	-	-	780	$ 83,884.00	240	$ 32,409.00	190	$ 24,922.00
	Germany	-	-	-	-	-	-	525	$ 55,476.00
	Ireland	-	-	-	-	-	-	-	-
	Italy	-	-	-	-	-	-	304	$ 43,248.00
	Japan	-	-	-	-	-	-	-	-
	New Zealand	-	-	-	-	196	$ 19,434.00	93	$ 10,241.00
	Norway	110	$ 15,278.00	-	-	-	-	891	$ 105,750.00
	Philippines	391	$ 42,572.00	-	-	-	-	-	-
	Singapore	-	-	137	$ 20,630.00	85	$ 16,280.00	285	$ 41,123.00
	Spain	224	$ 25,915.00	105	$ 15,268.00	111	$ 17,332.00	632	$ 71,198.00
	Sweden	29	$ 7,221.00	-	-	-	-	44	$ 5,940.00
	Switzerland	-	-	-	-	-	-	-	-
	UK	297	$ 27,795.00	33	$ 4,389.00	-	-	332	$ 34,452.00
	USA	212	$ 27,419.00	636	$ 79,050.00	1,080	$ 132,984.00	2,806	$ 319,071.00
Motorcycles	Australia	-	-	212	$ 23,448.00	-	-	219	$ 18,889.00
	Austria	-	-	-	-	-	-	-	-
	Canada	-	-	-	-	-	-	-	-
	Finland	-	-	-	-	-	-	-	-

Quantity and sales by product line, country and month

Measures ▼

Quantity

Sales

⊗

Columns ▼

Time Hierarchy

Year

Quarter

Month

⊗

Rows ▼

Product Hierarchy

Product Line

Customer Hierarchy

Country

⊗

Filter ▼

	Year								
	Quarter			Q1					
	Month		Jan		Feb		Mar		Apr
Product Line	Country	Quantity	Sales	Quantity	Sales	Quantity	Sales	Quantity	Sales
Classic Cars	Australia	-	-	-	-	-	-	-	-
	Austria	-	-	-	-	-	-	46	$ 5,014.00
	Belgium	-	-	-	-	-	-	-	-
	Canada	-	-	-	-	-	-	-	-
	Denmark	-	-	120	$ 20,445.00	-	-	-	-
	Finland	-	-	-	-	-	-	-	-
	France	-	-	-	-	-	-	145	$ 17,983.00
	Germany	-	-	-	-	-	-	-	-
	Ireland	-	-	-	-	-	-	-	-
	Italy	-	-	-	-	-	-	-	-
	Japan	-	-	-	-	-	-	-	-
	New Zealand	-	-	-	-	-	-	-	-
	Norway	110	$ 15,278.00	-	-	-	-	-	-
	Philippines	-	-	-	-	391	$ 42,572.00	-	-
	Singapore	-	-	-	-	-	-	137	$ 20,630.00
	Spain	224	$ 25,915.00	-	-	-	-	-	-
	Sweden	-	-	-	-	29	$ 7,221.00	-	-
	Switzerland	-	-	-	-	-	-	-	-
	UK	-	-	-	-	297	$ 27,795.00	-	-
	USA	-	-	-	-	212	$ 27,419.00	137	$ 16,191.00
Motorcycles	Australia	-	-	-	-	-	-	212	$ 23,448.00
	Austria	-	-	-	-	-	-	-	-
	Canada	-	-	-	-	-	-	-	-