

bi group

by Darren Choo

Submission date: 13-Dec-2019 04:57PM (UTC+0800)

Submission ID: 1233779209

File name: finalbigroup.docx (3.59M)

Word count: 1867

Character count: 10368

Table of Contents

Task1	2
Dataset	2
Task2	3
Business Models.....	3
Business Rules	4
Information Flow.....	5
Analytics spectrum questions	6
Data architecture	6
Data Preparation.....	6
Replace Missing Values with Rapid Miner	6
The Location intelligence technique	8
Task3	12
Finding Result /Valuable insight from the findings toward the organizational decision making	12

Task1

Dataset

Our chosen dataset is a dataset that we found online from Kaggle.com. The dataset is based on customer transactions for an e-commerce website. The chosen e-commerce dataset is suitable for us as it has location-based information such as the country whereby the customer made a transaction from. Our dataset has many different columns of information. There are 8 columns altogether which all consists of unique information. The columns are InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID & Country

Task2

Business Models

The Dataset of Ecommerce that comes from Kaggle is a data warehouse[2] (where it can access and store a lot of data, from centralized or decentralized).

The purpose choosing Ecommerce dataset is due to the amount of data produced to support decision making; this type of Business model is label as ³Online Analytical Processing Tools (OLAP). It is a computing method that enables users to easily and selectively extract and query data in order to analyse it from different points of view. Where it's often aid in trends analysis, financial reporting, budgeting, sales forecasting and other planning purposes.

²This type of analysis, data is collected from multiple data sources and stored in data warehouses and then organized into data cubes. Each OLAP cube contains data categorized by dimensions (such as customers, geographic sales region and time period) derived by dimensional tables in the data warehouses[1].

Dimensions are then populated by members (such as customer names, countries and months) that are organized hierarchically. OLAP cubes are often pre-summarized across dimensions to drastically improve query time over relational databases.

Business Rules

In a worldwide country, Ecommerce does business and treat their employees with respect from different countries of diverse cultural, social, and economic circumstances. We strive to work hard together to eliminate discriminatory practices. Ecommerce respects the principles of operating their business in a responsible and ethical manner, respecting the rights of individuals, and helping to protect the environment.

Here are the list of table that Ecommerce prohibited in table 1

List	Explanation
Child Labor	The economic exploitation of children under the age of 18 or their employment in work that is harmful to their health, safety or morals
Forced or Compulsory Labor	Work or service that is coerced or imposed with little or no freedom of choice and deprives a worker of a genuine possibility of terminating his or her employment without penalty or the threat of penalty, including deliberate withholding or non-payment of wages
Free Association	Free formation and joining of groups by workers and employers for the promotion and defense of occupational interests, and to influence matters that directly concern them
Collective Bargaining	A process through which employers or their organizations, and representatives designated by the workers, discuss and negotiate their relations, particularly the terms and conditions of work, to reach mutually acceptable collective agreements

Table 1

The OLAP process

How data is prepared for online analytical processing (OLAP)

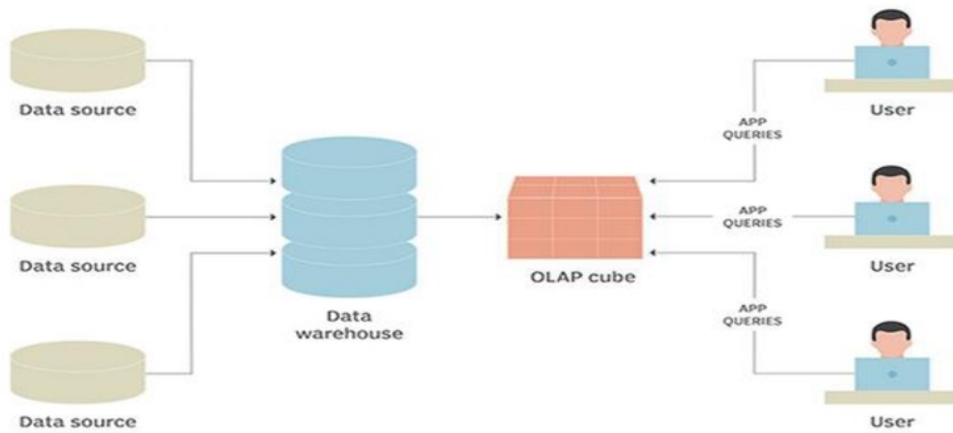


Figure 1

The above (Figure 1) shows the information flow process of Ecommerce dataset that comes from Kaggle in the form of data warehouse. Below show the OLAP cube (Figure 2).

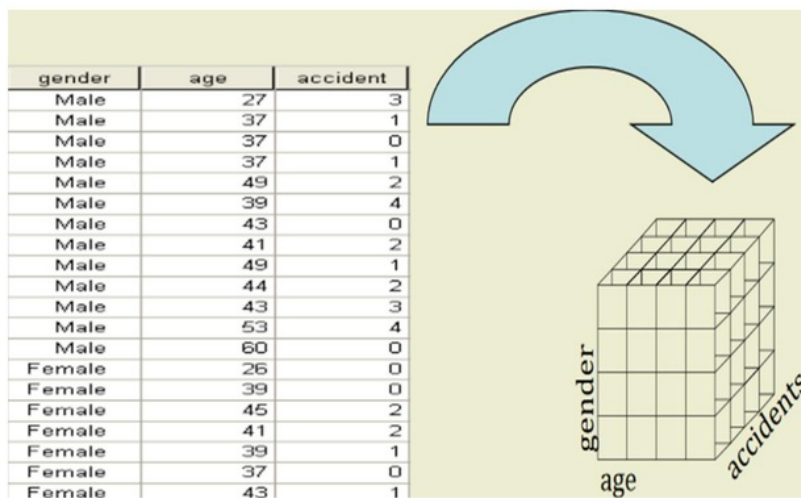


Figure 2

Analytics spectrum questions

- Which product are the highest selling number?
- The total amount of transaction in each country.
- What item that have in each invoice number?
- What are nationality of each customer?
- What are the number of customers from each country?
- Which are the highest selling item in each country?

Data architecture

-Data architecture is one of the technical very important to business decision. Businesses use data architecture to managing all the complex data and understanding business objectives and existing data infrastructure and assets. [3]

Data Preparation

-Data preparation able to cleaning data and improving the data quality. Because data is often created with missing values, inaccuracies or having other errors. In addition, data preparation able to fix errors quickly. In conclusion, using produced quality data will make better business decisions. Thus, we decided using rapid miner to replace the missing values.

Replace Missing Values with Rapid Miner

From Ecommerce Dataset there are a lot of missing values for attribute such as invoice No, Description and Customer ID.

Name	Type	Missing	Filter (8 / 8 attributes): <input type="text" value="Search for Attributes"/>	
✓ InvoiceNo	Integer	9291	Min 536365	Max 581587
✓ StockCode	Polynomial	0	Least m (1)	Most 85123A (
✓ Description	Polynomial	1454	Least wrongly sold sets (1)	Most WHITE H
✓ Quantity	Integer	0	Min -80995	Max 80995
✓ InvoiceDate	Date time	0	Earliest date Jan 12, 2010 8:26 AM	Latest date Jul 10, 20
✓ UnitPrice	Real	0	Min -11062.060	Max 38970
✓ CustomerID	Integer	135080	Min 12346	Max 18287

Diagram1.1(statistic)

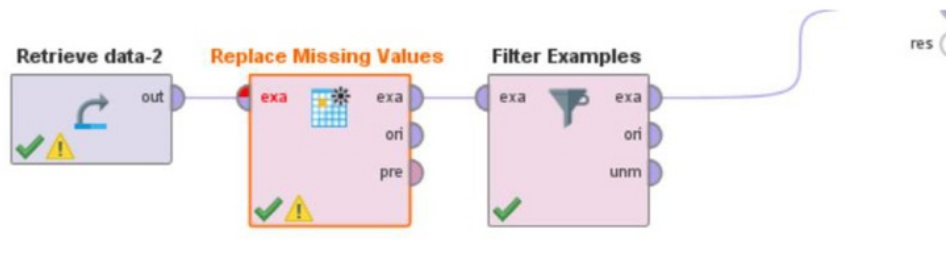


Diagram1.2(Flow)

Parameters

Filter Examples

condition class
no_missing_attri...

☐ invert filter

Diagram1.3(Parameter for the filter)

✓ InvoiceNo	Integer	0	536365	581587
✓ StockCode	Polynomial	0	Least m (1)	Most 85123A
✓ Description	Polynomial	0	Least wrongly sold sets (1)	Most WHITE
✓ Quantity	Integer	0	Min -80995	Max 80995
✓ InvoiceDate	Date time	0	Earliest date Jan 12, 2010 8:26 AM	Latest date Jul 10, 2
✓ UnitPrice	Real	0	Min -11062.060	Max 38970
✓ CustomerID	Integer	0	Min 12346	Max 18287

Diagram 1.4(Result)

The Location intelligence technique

Based on why customer analytics matters in a business is a better understanding to see the customer's buying habits or pattern to develop a more accurate predictive behaviors from it to satisfy customer journey. From the database that we choose which is the "E-commerce Dataset" it can classify that the items that are sold are mostly stuff or game for kids, house decoration and also kitchen items. From these details, we can analyze that the business target audience are someone who are does or interested in performing online shopping for their house's need. Consumer behavior can be analyses by the purchases based on the description of the product done by the customer. Below Figure 1 are few of the example of item that have been purchase by customer found from the dataset. Consumer are also to be found in different location like United Kingdom, Canada, Australia and many more country. Figure 2 are all the countries included in this "Ecommerce Dataset".

Description	Search
WHITE HANGING HEART T-LIGHT HOLDER	<input checked="" type="checkbox"/> Australia
WHITE METAL LANTERN	<input checked="" type="checkbox"/> Austria
CREAM CUPID HEARTS COAT HANGER	<input checked="" type="checkbox"/> Bahrain
KNITTED UNION FLAG HOT WATER BOTTLE	<input checked="" type="checkbox"/> Belgium
RED WOOLLY HOTTIE WHITE HEART.	<input checked="" type="checkbox"/> Brazil
SET 7 BABUSHKA NESTING BOXES	<input checked="" type="checkbox"/> Canada
GLASS STAR FROSTED T-LIGHT HOLDER	<input checked="" type="checkbox"/> Channel Islands
HAND WARMER UNION JACK	<input checked="" type="checkbox"/> Cyprus
HAND WARMER RED POLKA DOT	<input checked="" type="checkbox"/> Czech Republic
ASSORTED COLOUR BIRD ORNAMENT	<input checked="" type="checkbox"/> Denmark
POPPY'S PLAYHOUSE BEDROOM	<input checked="" type="checkbox"/> EIRE
POPPY'S PLAYHOUSE KITCHEN	<input checked="" type="checkbox"/> European Community
FELTCRAFT PRINCESS CHARLOTTE DOLL	<input checked="" type="checkbox"/> Finland
IVORY KNITTED MUG COSY	<input checked="" type="checkbox"/> France
BOX OF 6 ASSORTED COLOUR TEASPOONS	<input checked="" type="checkbox"/> Germany
BOX OF VINTAGE JIGSAW BLOCKS	<input checked="" type="checkbox"/> Greece
BOX OF VINTAGE ALPHABET BLOCKS	<input checked="" type="checkbox"/> Hong Kong
HOME BUILDING BLOCK WORD	<input checked="" type="checkbox"/> Iceland
LOVE BUILDING BLOCK WORD	<input checked="" type="checkbox"/> Israel
RECIPE BOX WITH METAL HEART	<input checked="" type="checkbox"/> Italy
DOORMAT NEW ENGLAND	<input checked="" type="checkbox"/> Japan
JAM MAKING SET WITH JARS	<input checked="" type="checkbox"/> Lebanon
RED COAT RACK PARIS FASHION	<input checked="" type="checkbox"/> Lithuania
YELLOW COAT RACK PARIS FASHION	<input checked="" type="checkbox"/> Malta
BLUE COAT RACK PARIS FASHION	<input checked="" type="checkbox"/> Netherlands
BATH BUILDING BLOCK WORD	<input checked="" type="checkbox"/> Norway
ALARM CLOCK BAKELIKE PINK	<input checked="" type="checkbox"/> Poland
ALARM CLOCK BAKELIKE RED	<input checked="" type="checkbox"/> Portugal
ALARM CLOCK BAKELIKE GREEN	<input checked="" type="checkbox"/> RSA
PANDA AND BUNNIES STICKER SHEET	<input checked="" type="checkbox"/> Saudi Arabia
STARS GIFT TAPE	<input checked="" type="checkbox"/> Singapore
INFLATABLE POLITICAL GLOBE	<input checked="" type="checkbox"/> Spain
	<input checked="" type="checkbox"/> Sweden
	<input checked="" type="checkbox"/> Switzerland
	<input checked="" type="checkbox"/> United Arab Emirates
	<input checked="" type="checkbox"/> United Kingdom
	<input checked="" type="checkbox"/> Unspecified
	<input checked="" type="checkbox"/> USA

Figure 1: Sold Items

The screenshot shows the 'Filter By Date Range' dialog box in Microsoft Access. The 'Criteria' tab is selected. The 'Criteria' section shows a date range from 12/23/2010 to 1/4/2011. The 'Show All Dates' checkbox is checked. The 'Filter by' dropdown is set to 'Date Filter'. The 'Filter by' dropdown is set to 'Date Filter'. The 'Filter by' dropdown is set to 'Date Filter'.

Figure 3: Year included in the datasets

Figure 3 are the year included in this “Ecommerce Dataset” which is the year 2010 ends on the 23rd December and starts on 4th January on the year 2011. Other than that, by the transaction we could also identify the characteristic of customer by pattern spend and this are the most valuable customer to the business. The type of customer analytics is “Customer

Spend Analysis". Analyses by the transaction behavior and purchase behavior to understand spend patterns and potential. From the "E-commerce Dataset" we could analyses or divide to know the amount of transaction or purchase that are done by a specific consumer based on the invoice no. From Figure 4 we can see three transaction done by customer where 536389,537676 and C538723 is the path where it divides the transaction.

	A	B	C	D	E	F	G	H
	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536389	22941	CHRISTMAS LIGHTS 10 REINDEER	6	12/1/2010 10:03	8.5	12431	Australia
2	536389	23422	VINTAGE UNION JACK CUSHION COVER	8	12/1/2010 10:03	4.95	12431	Australia
3	536389	21791	VINTAGE HEADS AND TAILS CARD GAME	12	12/1/2010 10:03	1.25	12431	Australia
4	536389	35004C	SET OF 3 COLOURED FLYING DUCKS	6	12/1/2010 10:03	3.45	12431	Australia
5	536389	35004G	SET OF 3 GOLD FLYING DUCKS	4	12/1/2010 10:03	6.35	12431	Australia
6	536389	85014B	RED RETROSPOT UMBRELLA	6	12/1/2010 10:03	5.95	12431	Australia
7	536389	85014A	BLACK/BLUE POLKA DOT UMBRELLA	3	12/1/2010 10:03	5.95	12431	Australia
8	536389	22193	RED DINER WALL CLOCK	2	12/1/2010 10:03	8.5	12431	Australia
9	536389	22726	ALARM CLOCK BAKELIKE GREEN	4	12/1/2010 10:03	3.75	12431	Australia
10	536389	22727	ALARM CLOCK BAKELIKE RED	4	12/1/2010 10:03	3.75	12431	Australia
11	536389	22192	BLUE DINER WALL CLOCK	2	12/1/2010 10:03	8.5	12431	Australia
12	536389	22191	IVORY DINER WALL CLOCK	2	12/1/2010 10:03	8.5	12431	Australia
13	536389	22195	LARGE HEART MEASURING SPOONS	24	12/1/2010 10:03	1.65	12431	Australia
14	536389	22196	SMALL HEART MEASURING SPOONS	24	12/1/2010 10:03	0.85	12431	Australia
15	537676	22915	20 DOLLY PEGS RETROSPOT	24	12/8/2010 9:53	1.25	12386	Australia
16	537676	22915	ASSORTED BOTTLE TOP MAGNETS	120	12/8/2010 9:53	0.36	12386	Australia
17	537676	22926	IVORY GIANT GARDEN THERMOMETER	12	12/8/2010 9:53	5.95	12386	Australia
18	537676	22913	BIRTHDAY PARTY CORDON BARRIER TAPE	24	12/8/2010 9:53	1.25	12386	Australia
19	537676	21906	PHARMACE FIRST AID TIN	4	12/8/2010 9:53	6.75	12386	Australia
20	537676	22495	SET OF 2 ROUND TINS CAMEMBER	6	12/8/2010 9:53	2.95	12386	Australia
21	537676	22555	PLASTERS IN TIN STRONGMAN	12	12/8/2010 9:53	1.65	12386	Australia
22	537676	22557	PLASTERS IN TIN VINTAGE PAULIE	12	12/8/2010 9:53	1.65	12386	Australia
23	C538723	22333	RETROSPOT PARTY BAG + STICKER SET	-7	12/14/2010 11:12	1.85	12434	Australia
24	C538723	22094	RED RETROSPOT TISSUE BOX	-5	12/14/2010 11:12	1.25	12434	Australia
25	C538723	21217	RED RETROSPOT ROUND CAKE TIN	-1	12/14/2010 11:12	9.95	12434	Australia

Figure 4: Purchase by Customers

From that we could get to compare one demographic or country with another, and this could analyze the most transaction of purchase by customer in a particular country against the other country. Regarding that, potential customer in different country could be analyzed based on the total amount of transaction done in period of time. Below is just one of the example of comparison of the number of transaction perform by customer in Brazil comparing with another country called Bahrain. As mention above invoice no is one of the way to detect the transaction done by specific customer and based on Figure 5 it only contain one transaction from Brazil is which under 550201 "InvoiceNo". Where else Figure 6 contain 4 different transaction from Bahrain country.

	A	B	C	D	E	F	G	H	I	J	K
	InvoiceNo.	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country		Total	
1750	590001	22821	REGENCY CAKESTAND 3 TIER	18	4/13/2011 10:25	10.95	12789	Brazil		197.2	
1751	590001	22888	ROSE REGENCY TEACUP AND SAUCER	28	4/13/2011 10:25	2.35	12789	Brazil		65.8	
1752	590001	22887	GREEN REGENCY TEACUP AND SAUCER	28	4/13/2011 10:25	2.35	12789	Brazil		65.8	
1753	590001	22886	PINK REGENCY TEACUP AND SAUCER	28	4/13/2011 10:25	2.35	12789	Brazil		65.8	
1754	590001	22386	DOORMAT FABRIC	10	4/13/2011 10:25	6.75	12789	Brazil		67.5	
1755	590001	22480	SET/1 RED GINGHAM ROSE STORAGE BOX	24	4/13/2011 10:25	3.50	12789	Brazil		84.00	
1756	590001	22480	DOLLY GIRL LUNCH BOX	24	4/13/2011 10:25	3.75	12789	Brazil		90.00	
1757	590001	22862	LUNCH BAG DOLLY GIRL DESIGN	10	4/13/2011 10:25	1.65	12789	Brazil		16.5	
1758	590001	22829	SPACEBOY LUNCH BOX	12	4/13/2011 10:25	1.95	12789	Brazil		23.4	
1759	590001	22382	LUNCH BAG SPACEBOY DESIGN	10	4/13/2011 10:25	1.65	12789	Brazil		16.5	
1760	590001	1205686	EDWARDIAN PARADEL BLACK	3	4/13/2011 10:25	5.95	12789	Brazil		17.85	
1761	590001	1205697	EDWARDIAN PARADEL PINK	3	4/13/2011 10:25	5.95	12789	Brazil		17.85	
1762	590001	1205698	EDWARDIAN PARADEL NATURAL	3	4/13/2011 10:25	5.95	12789	Brazil		17.85	
1763	590001	20679	EDWARDIAN PARADEL RED	3	4/13/2011 10:25	5.95	12789	Brazil		17.85	
1764	590001	21181	COOK WITH WINE METAL SIGN	12	4/13/2011 10:25	2.08	12789	Brazil		24.96	
1765	590001	21181	PLEASE ONE PERSON METAL SIGN	12	4/13/2011 10:25	2.1	12789	Brazil		25.2	
1766	590001	21770	OPEN CLOSED METAL SIGN	10	4/13/2011 10:25	4.95	12789	Brazil		49.5	
1767	590001	22893	SET OF 4 PANTRY FULL NOODLES	28	4/13/2011 10:25	1.25	12789	Brazil		35	
1768	590001	22732	SET OF 4 SPICE TRAY PANTRY DESIGN	24	4/13/2011 10:25	3.45	12789	Brazil		82.8	
1769	590001	23178	CLOCK MAGNET MOUNT KITCHEN	6	4/13/2011 10:25	2.89	12789	Brazil		17.34	
1770	590001	23178	JAM CLOCK MAGNET	6	4/13/2011 10:25	2.89	12789	Brazil		17.34	
1771	590001	22884	EMERGENCY FIRST AID TIN	12	4/13/2011 10:25	1.25	12789	Brazil		15	
1772	590001	21288	FIRST AID TIN	6	4/13/2011 10:25	2.25	12789	Brazil		13.5	
1773	590001	22906	PHARMACEY FIRST AID TIN	6	4/13/2011 10:25	6.75	12789	Brazil		40.5	
1774	590001	240715	SMALL HEART FLOWER JAR	28	4/13/2011 10:25	0.85	12789	Brazil		23.8	
1775	590001	23034	RECYCLED ACAPULCO MAT LAVENDER	2	4/13/2011 10:25	8.25	12789	Brazil		16.5	
1776	590001	23033	RECYCLED ACAPULCO MAT PINK	2	4/13/2011 10:25	8.25	12789	Brazil		16.5	
1777	590001	23030	RECYCLED ACAPULCO MAT GREEN	2	4/13/2011 10:25	8.25	12789	Brazil		16.5	
1778	590001	23031	RECYCLED ACAPULCO MAT BLUE	2	4/13/2011 10:25	8.25	12789	Brazil		16.5	
1779	590001	23049	RECYCLED ACAPULCO MAT RED	2	4/13/2011 10:25	8.25	12789	Brazil		16.5	
1780	590001	23052	RECYCLED ACAPULCO MAT TURQUOISE	2	4/13/2011 10:25	8.25	12789	Brazil		16.5	
1781	590001	22488	NATURAL SLATE RECTANGULAR CHALKBOARD	12	4/13/2011 10:25	1.65	12789	Brazil		19.8	
541912											
541913				358						1143.6	
541914											

Figure 5: Transaction is Brazil

	A	B	C	D	E	F	G	H	I	J	K
1	InvoiceNo.	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country		Total	
1682	590000	22828	OCEAN SCENT CANDLE IN JEWELLED BOX	34	12/20/2010 11:02	3.81	12003	Bahrain		205.74	
1683	590000	22828	OCEAN SCENT CANDLE IN JEWELLED BOX	34	1/3/2011 11:31	3.81	12003	Bahrain		205.74	
1684	590000	22828	GROWN A FLATBAP OR SUNFLOWER IN TIN	98	5/9/2011 13:49	1.25	12003	Bahrain		122.5	
1685	590000	23076	ICE CREAM SUNDAE UP GLASS	98	5/9/2011 13:49	1.25	12003	Bahrain		122.5	
1686	590000	23077	DOORMOUNT UP GLASS	98	5/9/2011 13:49	1.25	12003	Bahrain		122.5	
1687	590000	22880	NOVELTY BISCUITS CAKE STAND 3 TIER	2	5/9/2011 13:49	9.95	12003	Bahrain		19.9	
1688	590000	22849	STRAWBERRY FAIRY CAKE TRAY	8	5/9/2011 13:49	4.95	12003	Bahrain		39.6	
1689	590000	22823	REGENCY CAKESTAND 3 TIER	2	5/9/2011 13:49	12.75	12003	Bahrain		25.5	
1690	590000	23048A	S/L PINK FLOWER CANDLES IN BOWL	12	5/9/2011 13:49	1.45	12003	Bahrain		17.4	
1691	590000	22889	ROSE REGENCY TEACUP AND SAUCER	6	5/9/2011 13:49	2.95	12003	Bahrain		17.7	
1692	590000	22887	GREEN REGENCY TEACUP AND SAUCER	6	5/9/2011 13:49	2.95	12003	Bahrain		17.7	
1693	590000	22886	PINK REGENCY TEACUP AND SAUCER	6	5/9/2011 13:49	2.95	12003	Bahrain		17.7	
1694	590000	22888A	ROSE SCENT CANDLE IN JEWELLED BOX	6	5/9/2011 13:49	4.25	12003	Bahrain		25.5	
1695	590000	22828B	OCEAN SCENT CANDLE IN JEWELLED BOX	6	5/9/2011 13:49	4.25	12003	Bahrain		25.5	
1696	590000	22828	VANILLA SCENT CANDLE IN JEWELLED BOX	6	5/9/2011 13:49	4.25	12003	Bahrain		25.5	
1697	590000	22849	CERAMIC CAKE STAND - HANGING CAGES	2	5/9/2011 17:47	9.95	12003	Bahrain		19.9	
1698	590000	22849	CAKE STAND WITH HANGING CAGES	2	5/9/2011 17:47	9.95	12003	Bahrain		19.9	
1699	590000	22880	NOVELTY BISCUITS CAKE STAND 3 TIER	4	5/9/2011 17:47	9.95	12003	Bahrain		39.8	
1700	590000	22820	CERAMIC CAKE BOWL - HANGING CAGES	6	5/9/2011 17:47	2.85	12003	Bahrain		17.1	
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919											
16919</											

Task3

Finding Result /Valuable insight from the findings toward the organizational decision making

1)Which product are the highest selling number

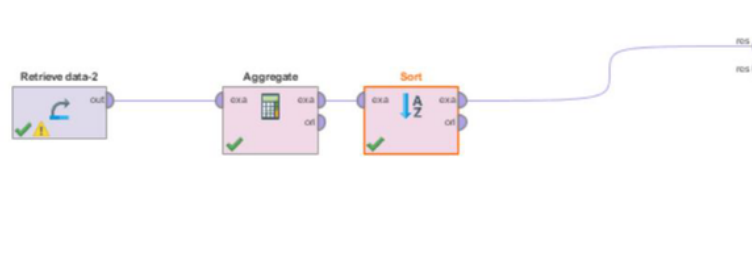


Diagram 1 (flow)

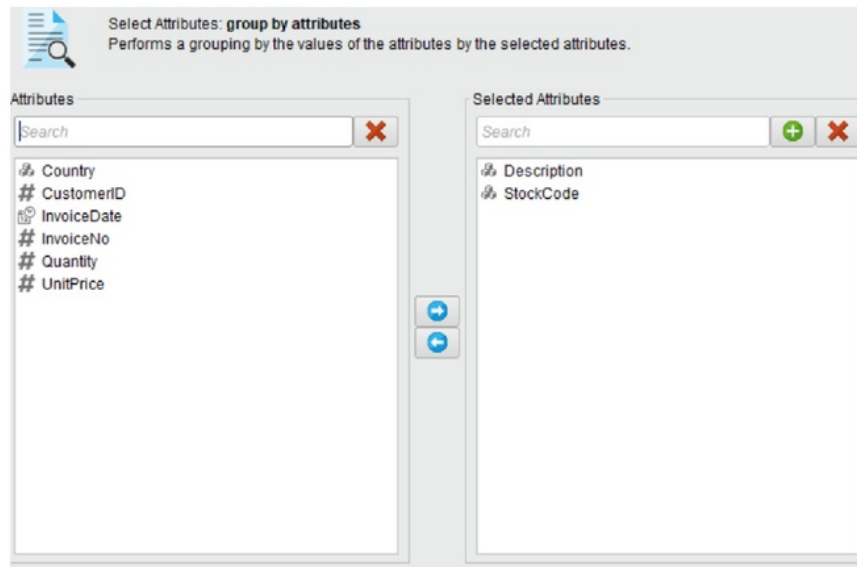


Diagram 1.1 (group attributes)

aggregation attribute	aggregation functions
Quantity	sum

Diagram 1.2 (aggregation attributes)

Parameters ✕

↓
2

Sort

attribute name

sum(Quantity) ▼ ⓘ

sorting direction

decreasing ▼ ⓘ

Diagram1.3(sort filter)

Row No.	StockCode	Description	sum(Quantity)
1	84077	WORLD WAR...	53847
2	85099B	JUMBO BAG ...	47363
3	84879	ASSORTED ...	36381
4	22197	POPCORN H...	36334
5	21212	PACK OF 72 ...	36039
6	85123A	WHITE HANG...	35025
7	23084	RABBIT NIGH...	30680
8	22492	MINI PAINT S...	26437

Diagram 1.4 (result)

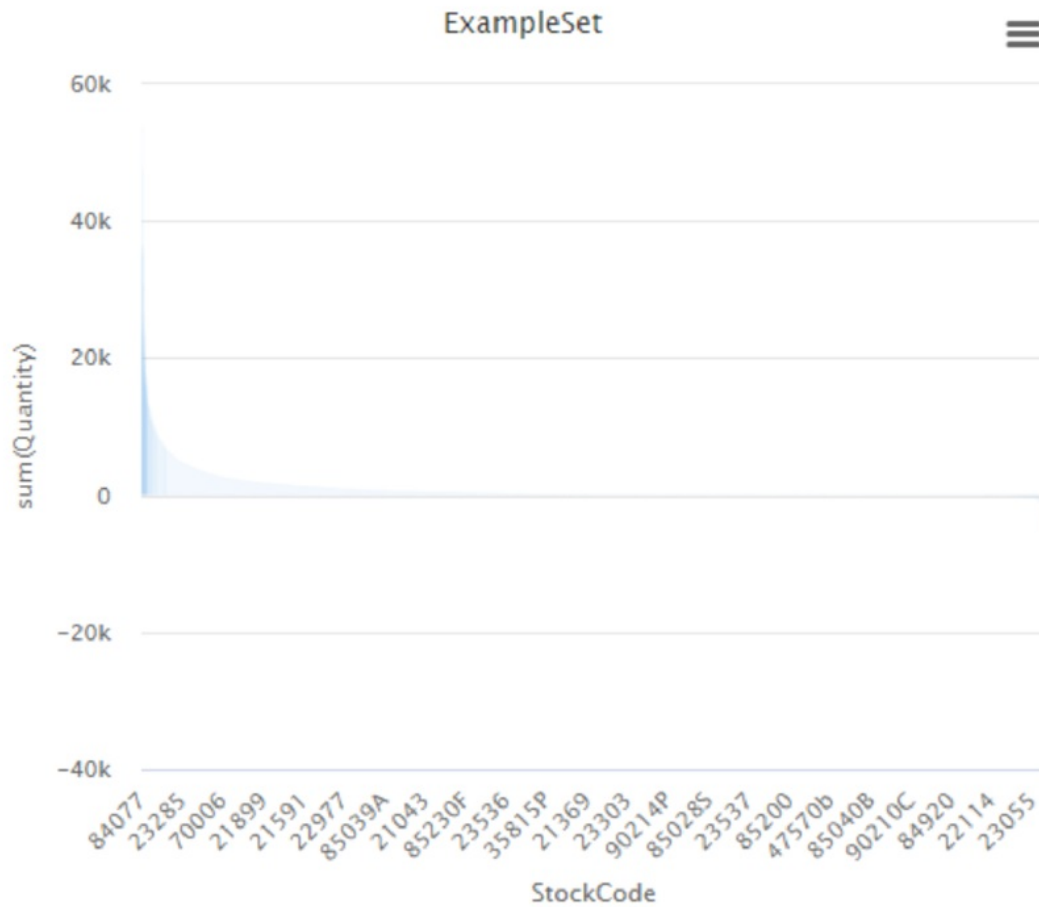


Diagram1.5(result with bar)

Explanation for Question 1

-The purpose of finding the highest number of sales is to help the e-commerce organization to be able to find out and to focus on which are their best-selling products. Once this information is available, the organization can then promote these products more to their customers since they are in such high demand.

2) The total amount of transaction in each country.

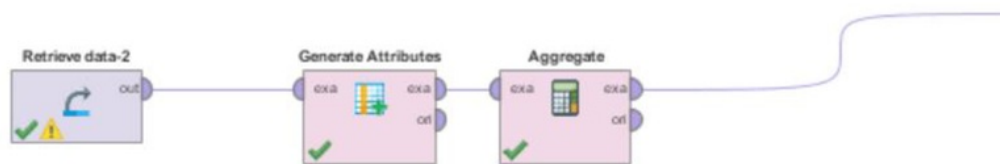


Diagram2.1(flow)

Edit Parameter List: function descriptions
 List of functions to generate.

attribute name	function expressions
Total unit price	UnitPrice*Quantity

Diagram2.1 (generate attribute)

Edit Parameter List: aggregation attributes
 The attributes which should be aggregated.

aggregation attribute	aggregation functions
Total unit price ▼	sum ▼

Diagram2.2 (aggregate attribute)

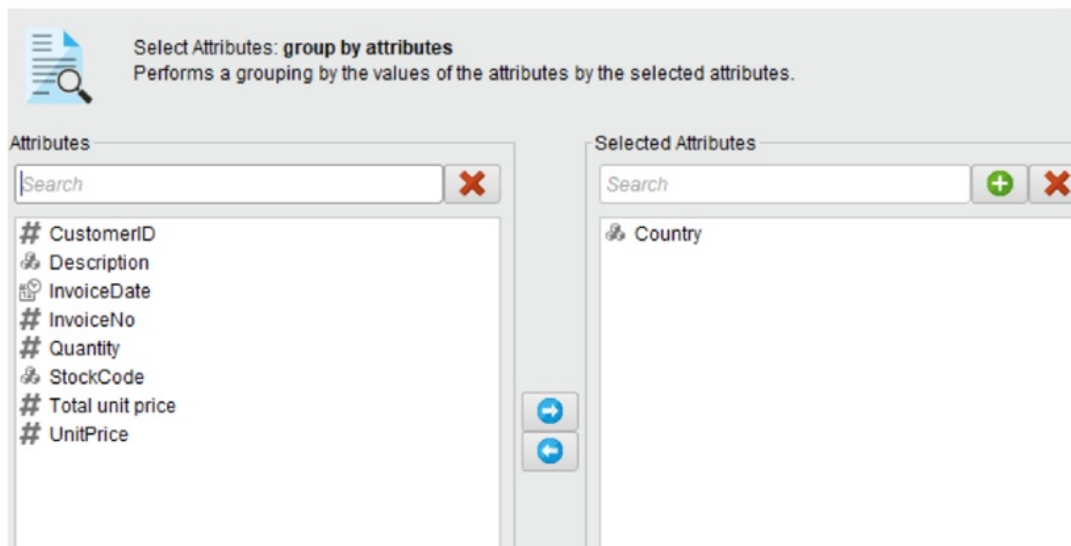


Diagram 2.3 (select attribute)

Row No.	Country	sum(Total u...
1	Australia	137077.270
2	Austria	10154.320
3	Bahrain	548.400
4	Belgium	40910.960
5	Brazil	1143.600
6	Canada	3666.380
7	Channel Isls	20086.200

Diagram 2.4 (result)

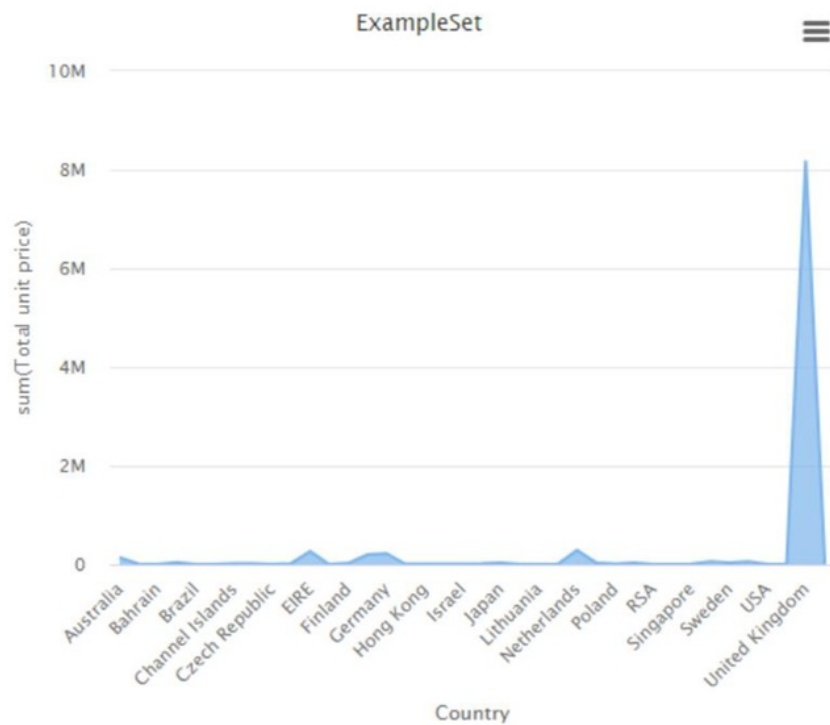
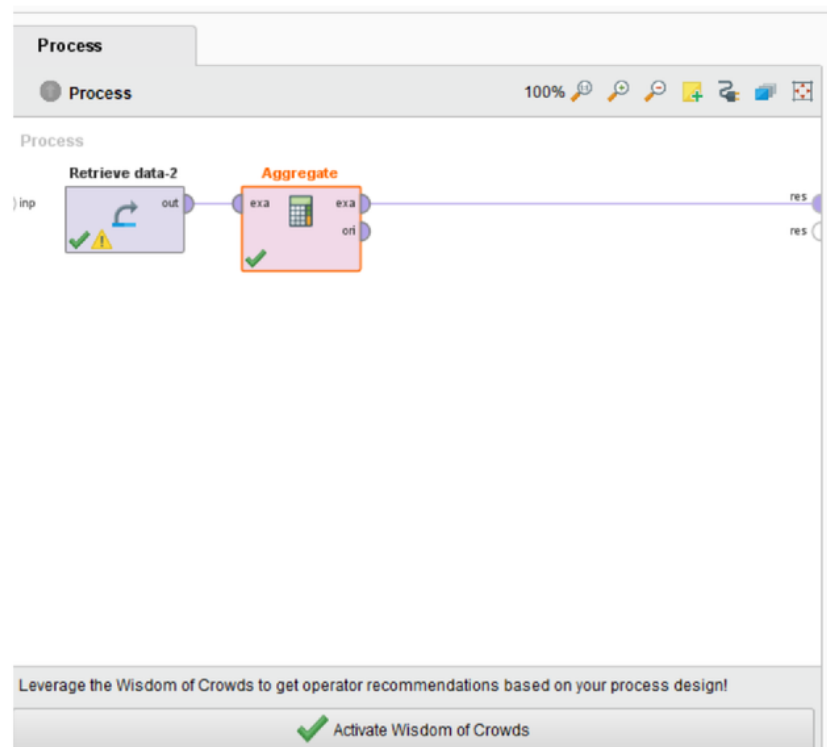


Diagram 2.5 (result with area)

Explanation for Question 2

- The purpose of finding the total number of transactions in each country is for the e-commerce organization to be able to find out which region/country produces the most sales and interests for consumers. The organization can then focus on promoting different products and items to different countries, based on which products are having the most transaction from that country.

3)What item that have in each invoice number?



Digram3(flow)

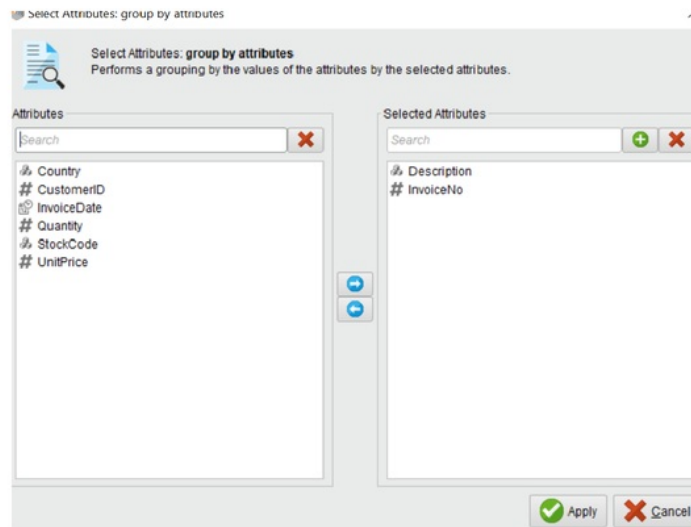


Diagram3.1 (group by attributes)

ExampleSet (Aggregate)

Open in Turbo Prep Auto Model

Filter (523,871 / 523,871 examples)

Row No.	InvoiceNo	Description
1	536365	CREAM CUPID HEARTS COATH...
2	536365	GLASS STAR FROSTED T-LIGHT...
3	536365	KNITTED UNION FLAG HOT BAT...
4	536365	RED WOOLLY HOTTIE WHITE H...
5	536365	SET 7 BABUSHKA NESTING BOX...
6	536365	WHITE HANGING HEART T-LIGH...
7	536365	WHITE METAL LANTERN
8	536366	HAND WARMER RED POLKA DOT
9	536366	HAND WARMER UNION JACK
10	536367	ASSORTED COLOUR BIRD ORN...
11	536367	BOX OF 6 ASSORTED COLOUR...
12	536367	BOX OF VINTAGE ALPHABET BL...

ExampleSet (523,871 examples, 0 special attributes, 2 regular attributes)

Repository

Import Data

Training Resources (connected)

Samples

- Deals (v1)
- Deals-TestSet (v1)
- Golf (v1)
- Golf-TestSet (v1)
- Iris (v1)
- Labor-Negotiations (v1)
- Market-Data (v1)
- Polynomial (v1)
- Products (v1)
- Purchases (v1)
- Ripley-Set (v1)
- Sonar (v1)
- Titanic (v1)
- Titanic-Training (v1)
- Titanic-Unlabeled (v1)
- Transactions (v1)

Diagram3.3(result)

Explanation for Question 3

- The purpose of finding out and compiling which invoice number is related to which item, based on the description of the product, is to identify which and how many products fall under the same invoice and it can help the organization to organize its information in a much more organized and cleaner structure.

4)What are nationality of each customer?

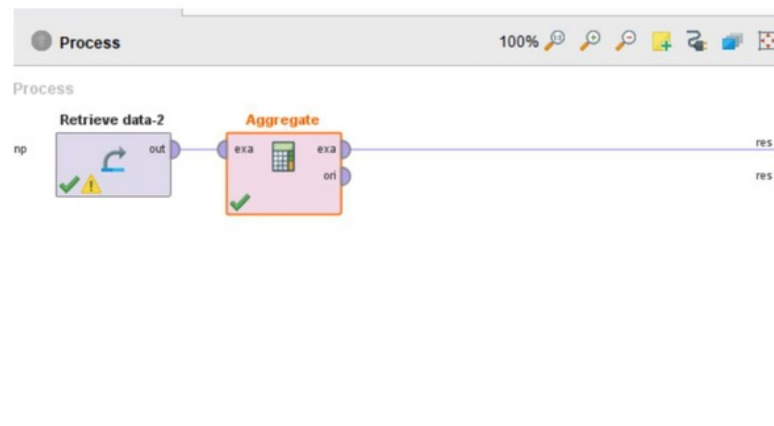


Diagram4.1 (flow)

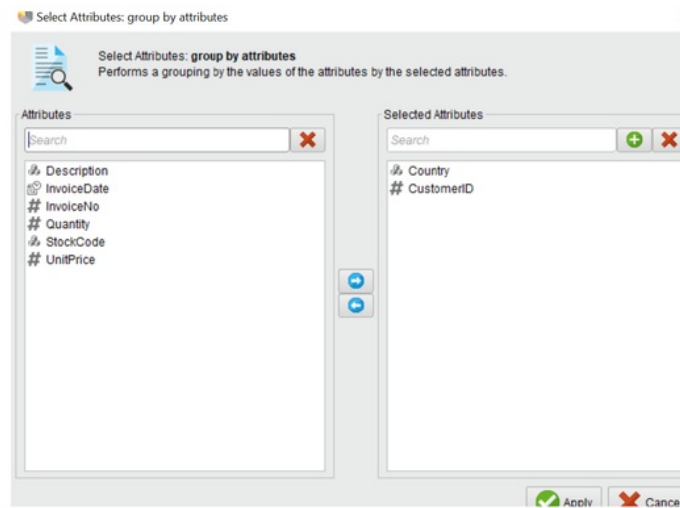


Diagram4.2(group by attributes)

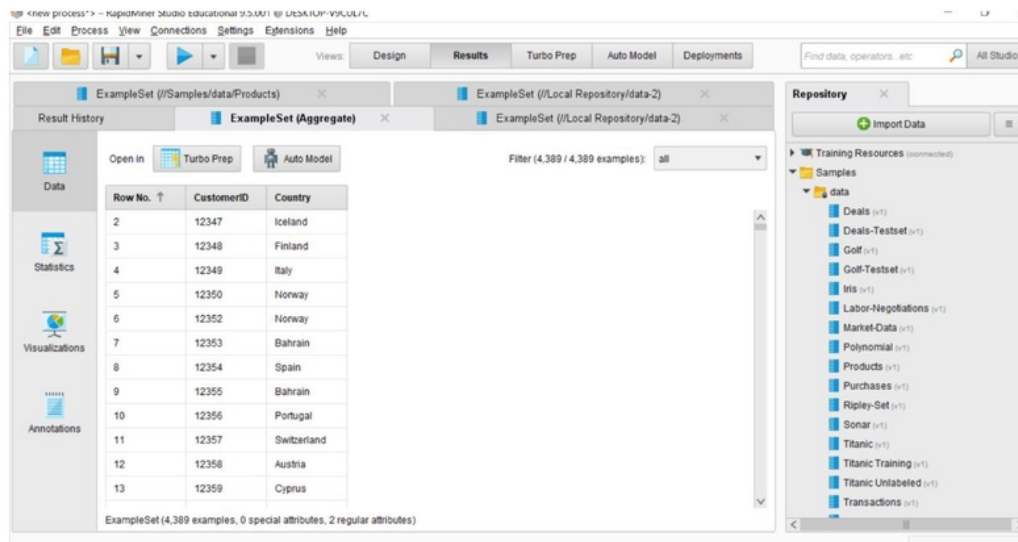


Diagram4.3 (Result)

Explanation for Question 4

- The purpose of finding out and gathering information about the customer and which country they are from is to help the organization to be able to identify which country the customer comes from so that suitable delivery information and different types of products can be promoted to them, according and suited to the country which they come from.

5)What are the number of customers from each country

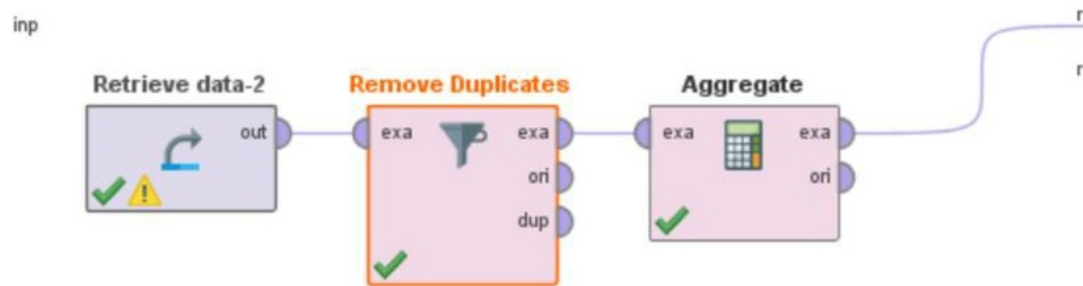


Diagram5.1(flow)

The screenshot shows the 'Parameters' window for the 'Remove Duplicates' tool. The window has a title bar with a close button. Below the title bar, the tool name 'Remove Duplicates' is displayed with a funnel icon. The parameters are as follows:

- attribute filter type:** A dropdown menu set to 'single'.
- attribute:** A dropdown menu set to 'CustomerID'.
- invert selection:** An unchecked checkbox.
- include special attributes:** An unchecked checkbox.
- treat missing values as duplicates:** An unchecked checkbox.

Diagram 5.2(Parameter)

aggregation attribute	aggregation functions
CustomerID	count

Diagram5.3(Aggregate)

 **Select Attributes: group by attributes**
Performs a grouping by the values of the attributes by the selected attributes.

Attributes

Search 

- # CustomerID
- # Description
- # InvoiceDate
- # InvoiceNo
- # Quantity
- # StockCode
- # UnitPrice

Selected Attributes

Search  

- Country





Diagram5.4(aggregate)

Row No. ↑	Country	count(Custo...
1	Australia	9
2	Austria	9
3	Bahrain	2
4	Belgium	24
5	Brazil	1

Diagram5.5(Result)

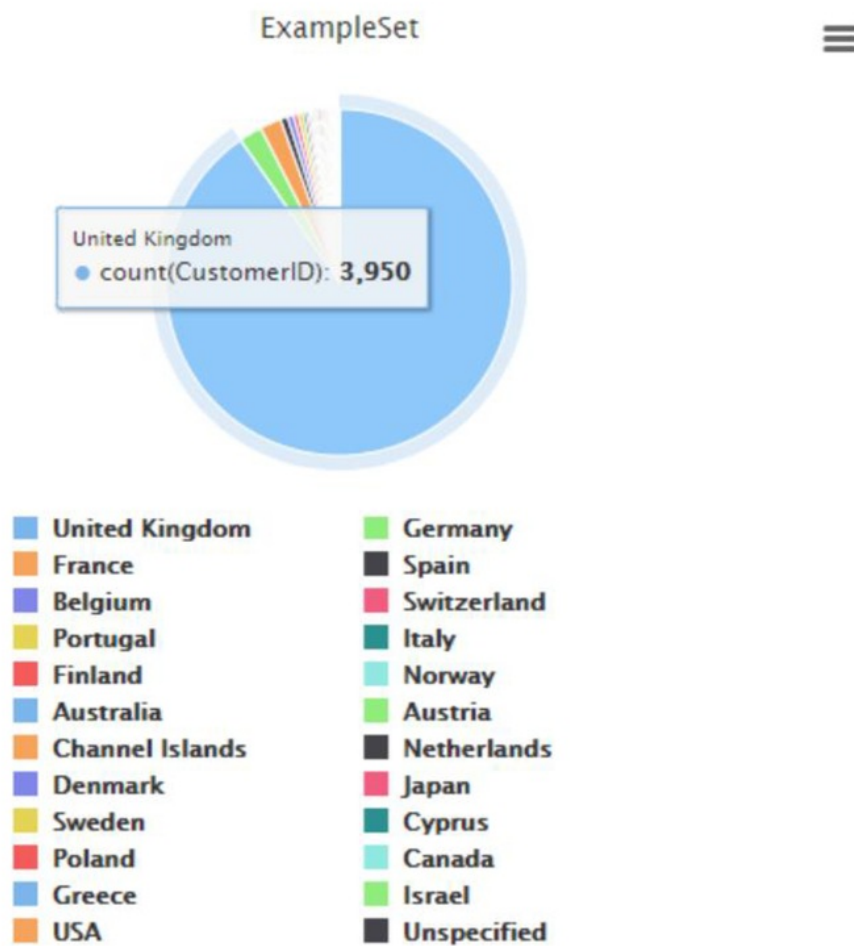


Diagram5.6(Result with pie chart)

Explanation for Question 5

The purpose of finding how many customers come from each country is to help the organization to realise how and where their greatest market is at. It can also help the organization to identify which countries and regions need better promotion and advertising to improve sales.

6) which are the highest selling item in each country.

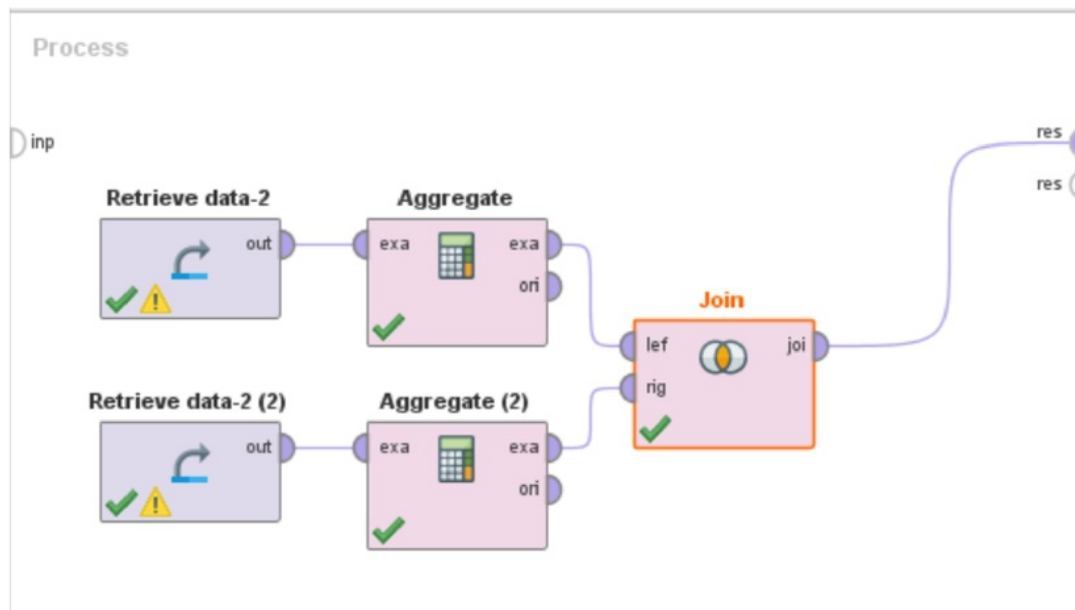


Diagram6.1(flow)

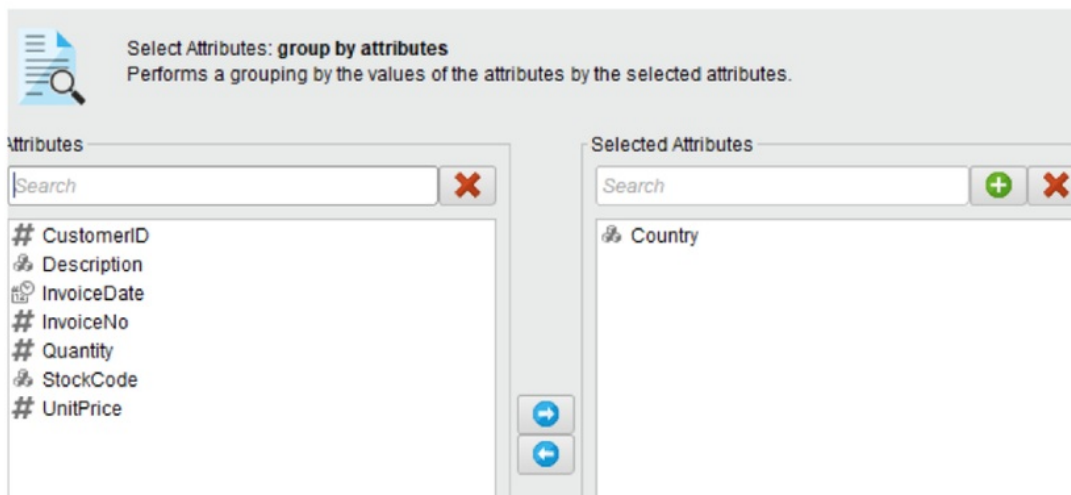




Diagram6.2 (aggregate1)


 Edit Parameter List: **aggregation attributes**
The attributes which should be aggregated.

aggregation attribute	aggregation functions
Quantity	maximum

Diagram6.3(aggregate1.2)



 Select Attributes: **group by attributes**
Performs a grouping by the values of the attributes by the selected attributes.

Attributes


Search 

- # CustomerID
- # InvoiceDate
- # InvoiceNo
- # Quantity
- # StockCode
- # UnitPrice

Selected Attributes

Search  

- Country
- Description








Diagram 6.4 (aggregate2.1)

 Edit Parameter List: **aggregation attributes**
The attributes which should be aggregated.

aggregation attribute	aggregation functions
Quantity	maximum

Diagram 6.3 (aggregate2.2)



Edit Parameter List: **key attributes**
The attributes which shall be used for join. Attributes which shall be matched must be of the same type.

left key attributes	right key attributes
Country ▼	Country ▼
maximum(Quantity) ▼	maximum(Quantity) ▼

Diagram 6.4 (join table)

Row No.	Country ↑	maximum(Q...	Description
37	Greece	48	4 LAVENDER...
38	Greece	48	4 PEAR BOT...
39	Hong Kong	144	WORLD WAR...
40	Iceland	240	ICE CREAM ...
41	Israel	100	RED RETRO...
42	Israel	100	WOODLAND ...
43	Italy	200	ICE CREAM ...
44	Japan	2040	RABBIT NIGH...
45	Lebanon	24	ASSTD FRUI...

Diagram 6.5(result)

Explanation for number 6

- The purpose of finding the highest selling item is very important to the e-commerce organisation because it is important for them to identify the highest selling item in each country. With this information, the organisation can then promote and identify which products to offer which countries, to increase the revenue and awareness of the organization

References

[1] M. Rouse, J. Biscobing, M. Rouse, and M. Rouse, "What is OLAP (online analytical processing)? - Definition from WhatIs.com," *SearchDataManagement*. [Online]. Available: <https://searchdatamanagement.techtarget.com/definition/OLAP>. [Accessed: 13-Dec-2019].

[2] data warehousing, Walker, J. (2018). *Top 5 data warehouses on the market today - Monitis Blog*. [online] Monitis Blog. Available at: <https://www.monitis.com/blog/top-5-data-warehouses-on-the-market-today/> [Accessed 10 Dec. 2019].

ORIGINALITY REPORT

19%

SIMILARITY INDEX

17%

INTERNET SOURCES

0%

PUBLICATIONS

19%

STUDENT PAPERS

PRIMARY SOURCES

1

www.starbucks.com

Internet Source

7%

2

searchdatamanagement.techtarget.com

Internet Source

4%

3

Submitted to Zambia Centre for Accountancy
Studies

Student Paper

2%

4

Submitted to Laureate Higher Education Group

Student Paper

2%

5

Submitted to Western International University

Student Paper

1%

6

Submitted to City University of Hong Kong

Student Paper

1%

7

www.hexucoc.com

Internet Source

1%

8

Submitted to Westminster International College -
Kuala Lumpur

Student Paper

<1%

9

Submitted to Coventry University

Student Paper

<1 %

10

idevji.com

Internet Source

<1 %

11

Submitted to University of Waikato

Student Paper

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off