

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

a. Deskripsi Dataset

Dataset yang digunakan berjumlah 541.909 data yang diunduh dengan format CSV. 3336 data unspecified, 89132 data null dan 73921 data dengan keterangan “C” pada atribut InvoiceNo yang artinya transaksi tersebut di dibatalkan. Dataset ini akan dijadikan bahan analisis untuk mengetahui tren pasar pada perusahaan ritel di Inggris. Dataset ini memiliki 8 atribut yaitu InvoiceNo, StockCode, Description, Quantity, Invoicedate, UnitPrice, CustomerID dan Country.

1	InvoiceNo,StockCode,Description,Quantity,InvoiceDate,UnitPrice,CustomerID,Country
541890	581585,23145,ZINC T-LIGHT HOLDER STAR LARGE,12,12/9/2011 12:31,0.95,15804,United Kingdom
541891	581585,22466,FAIRY TALE COTTAGE NIGHT LIGHT,12,12/9/2011 12:31,1.95,15804,United Kingdom
541892	581586,22061,LARGE CAKE STAND HANGING STRAWBERRY,8,12/9/2011 12:49,2.95,13113,United Kingdom
541893	581586,23275,SET OF 3 HANGING OWLS OLLIE BEAK,24,12/9/2011 12:49,1.25,13113,United Kingdom
541894	581586,21217,RED RETROSPOT ROUND CAKE TINS,24,12/9/2011 12:49,8.95,13113,United Kingdom
541895	581586,20685,DOORMAT RED RETROSPOT,10,12/9/2011 12:49,7.08,13113,United Kingdom
541896	581587,22631,CIRCUS PARADE LUNCH BOX ,12,12/9/2011 12:50,1.95,12680,France
541897	581587,22556,PLASTERS IN TIN CIRCUS PARADE ,12,12/9/2011 12:50,1.65,12680,France
541898	581587,22555,PLASTERS IN TIN STRONGMAN,12,12/9/2011 12:50,1.65,12680,France
541899	581587,22728,ALARM CLOCK BAKELIKE PINK,4,12/9/2011 12:50,3.75,12680,France
541900	581587,22727,ALARM CLOCK BAKELIKE RED ,4,12/9/2011 12:50,3.75,12680,France
541901	581587,22726,ALARM CLOCK BAKELIKE GREEN,4,12/9/2011 12:50,3.75,12680,France
541902	581587,22730,ALARM CLOCK BAKELIKE IVORY,4,12/9/2011 12:50,3.75,12680,France
541903	581587,22367,CHILDRENS APRON SPACEBOY DESIGN,8,12/9/2011 12:50,1.95,12680,France
541904	581587,22629,SPACEBOY LUNCH BOX ,12,12/9/2011 12:50,1.95,12680,France
541905	581587,23256,CHILDRENS CUTLERY SPACEBOY ,4,12/9/2011 12:50,4.15,12680,France
541906	581587,22613,PACK OF 20 SPACEBOY NAPKINS,12,12/9/2011 12:50,0.85,12680,France
541907	581587,22899,CHILDREN'S APRON DOLLY GIRL ,6,12/9/2011 12:50,2.1,12680,France
541908	581587,23254,CHILDRENS CUTLERY DOLLY GIRL ,4,12/9/2011 12:50,4.15,12680,France
541909	581587,23255,CHILDRENS CUTLERY CIRCUS PARADE,4,12/9/2011 12:50,4.15,12680,France
541910	581587,22138,BAKING SET 9 PIECE RETROSPOT ,3,12/9/2011 12:50,4.95,12680,France

Gambar 4.1 Dataset

b. Analisa Dataset

Analisa dataset memberikan gambaran umum sebelum data memasuki proses ETL. Sebelumnya, data tersebut diubah ke format xlsx guna

mengevaluasi kualitasnya. Setelah berhasil diekstraksi dalam format tersebut, teridentifikasi banyak data yang null pada kolom CustomerID. Kehadiran data kosong akan menghambat proses pengolahan data. Oleh karena itu, penulis memutuskan untuk menghapus entri yang kosong, tidak spesifik pada kolom Country, dan data dengan kode "C" pada kolom InvoiceNo.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
581492	23369	SET 36 COLOUR PENCILS LO	2	12/09/2011 10:03	2.46		United Kingdom
581492	23370	SET 36 COLOURING PENCILS	2	12/09/2011 10:03	2.46		United Kingdom
581492	23371	SET 36 COLOUR PENCILS SP	3	12/09/2011 10:03	2.46		United Kingdom
581492	23372	SET 36 COLOUR PENCILS DO	1	12/09/2011 10:03	2.46		United Kingdom
581492	23375	50'S CHRISTMAS PAPER GIF	1	12/09/2011 10:03	1.63		United Kingdom
581492	23376	PACK OF 12 VINTAGE CHRIS	3	12/09/2011 10:03	0.83		United Kingdom
581492	23377	PACK OF 12 DOLLY GIRL TISS	6	12/09/2011 10:03	0.83		United Kingdom
581492	23378	PACK OF 12 50'S CHRISTMA	9	12/09/2011 10:03	0.83		United Kingdom
581492	23379	PACK OF 12 RED APPLE TISS	4	12/09/2011 10:03	0.83		United Kingdom
581492	23380	PACK OF 12 VINTAGE DOILY	3	12/09/2011 10:03	0.83		United Kingdom
581492	23381	PACK OF 12 VINTAGE LEAF T	1	12/09/2011 10:03	0.83		United Kingdom
581492	23382	BOX OF 6 CHRISTMAS CAKE	4	12/09/2011 10:03	7.46		United Kingdom
581492	23391	I LOVE LONDON MINI BACK	2	12/09/2011 10:03	8.29		United Kingdom
581492	23392	SPACEBOY ROCKET LOLLY M	2	12/09/2011 10:03	4.13		United Kingdom
581492	23399	HOME SWEET HOME HANGI	4	12/09/2011 10:03	2.08		United Kingdom
581492	23405	HOME SWEET HOME 2 DRAV	3	12/09/2011 10:03	10.79		United Kingdom
581492	23418	LAVENDER TOILETTE BOTTL	1	12/09/2011 10:03	4.13		United Kingdom
581492	23426	METAL SIGN DROP YOUR PA	1	12/09/2011 10:03	5.79		United Kingdom
581492	23434	3 RAFFIA RIBBONS 50'S CHR	9	12/09/2011 10:03	0.83		United Kingdom
581492	23435	3 RAFFIA RIBBONS VINTAGE	3	12/09/2011 10:03	0.83		United Kingdom
581492	23437	50'S CHRISTMAS GIFT BAG L	1	12/09/2011 10:03	2.46		United Kingdom
581492	23439	HAND WARMER RED LOVE H	1	12/09/2011 10:03	4.13		United Kinedom

Gambar 4.2 Data Null

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
549687	20685	DOORMAT RED RETROSPOT	2	04/11/2011 13:29	7.95	12363	Unspecified
549687	22691	DOORMAT WELCOME SUNR	2	04/11/2011 13:29	7.95	12363	Unspecified
549687	48116	DOORMAT MULTICOLOUR S	2	04/11/2011 13:29	7.95	12363	Unspecified
549687	21213	PACK OF 72 SKULL CAKE CA	24	04/11/2011 13:29	0.55	12363	Unspecified
549687	21977	PACK OF 60 PINK PAISLEY C	24	04/11/2011 13:29	0.55	12363	Unspecified
549687	21976	PACK OF 60 MUSHROOM CA	24	04/11/2011 13:29	0.55	12363	Unspecified
549687	21212	PACK OF 72 RETROSPOT CA	24	04/11/2011 13:29	0.55	12363	Unspecified
549687	84992	72 SWEETHEART FAIRY CAK	24	04/11/2011 13:29	0.55	12363	Unspecified
549687	84991	60 TEATIME FAIRY CAKE CA	24	04/11/2011 13:29	0.55	12363	Unspecified
549687	21974	SET OF 36 PAISLEY FLOWER	12	04/11/2011 13:29	1.45	12363	Unspecified
549687	21973	SET OF 36 MUSHROOM PAP	12	04/11/2011 13:29	1.45	12363	Unspecified
549687	21211	SET OF 72 SKULL PAPER DO	12	04/11/2011 13:29	1.45	12363	Unspecified
549687	21124	SET/10 BLUE POLKADOT PA	24	04/11/2011 13:29	1.25	12363	Unspecified
549687	21122	SET/10 PINK POLKADOT PA	24	04/11/2011 13:29	1.25	12363	Unspecified
549687	21123	SET/10 IVORY POLKADOT PA	24	04/11/2011 13:29	1.25	12363	Unspecified
549687	21121	SET/10 RED POLKADOT PAR	24	04/11/2011 13:29	1.25	12363	Unspecified
552695	22953	BIRTHDAY PARTY CORDON F	24	05/10/2011 15:31	1.25	16320	Unspecified
552695	23078	ICE CREAM PEN LIP GLOSS	24	05/10/2011 15:31	1.25	16320	Unspecified
552695	22619	SET OF 6 SOLDIER SKITTLES	4	05/10/2011 15:31	3.75	16320	Unspecified
552695	21888	BINGO SET	8	05/10/2011 15:31	3.75	16320	Unspecified
552695	21890	S/6 WOODEN SKITTLES IN C	6	05/10/2011 15:31	3.25	16320	Unspecified
552695	22620	4 TRADITIONAL SPINNING T	12	05/10/2011 15:31	1.45	16320	Unspecified

Gambar 4.3 Data Unspecified

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	Customer	Country
CS81464	23458	DOLLY CABINET 3 DRAWERS	-1	12/08/2011 18:57	14.95	15951	United Kingdom
CS81464	71477	COLOURED GLASS STAR T-LI	-6	12/08/2011 18:57	3.95	15951	United Kingdom
CS81465	23660	HENRIETTA HEN MUG	-2	12/08/2011 18:59	1.65	15755	United Kingdom
CS81465	22171	3 HOOK PHOTO SHELF ANTI	-1	12/08/2011 18:59	8.5	15755	United Kingdom
CS81465	21876	POTTERING MUG	-4	12/08/2011 18:59	1.65	15755	United Kingdom
CS81465	20914	SET/5 RED RETROSPOT LID C	-3	12/08/2011 18:59	2.95	15755	United Kingdom
CS81466	22838	3 TIER CAKE TIN RED AND CI	-1	12/08/2011 19:20	14.95	13883	United Kingdom
CS81466	22720	SET OF 3 CAKE TINS PANTRY	-2	12/08/2011 19:20	4.95	13883	United Kingdom
CS81466	21216	SET 3 RETROSPOT TEA, COFF	-1	12/08/2011 19:20	4.95	13883	United Kingdom
CS81466	21535	RED RETROSPOT SMALL MIL	-2	12/08/2011 19:20	2.55	13883	United Kingdom
CS81466	21232	STRAWBERRY CERAMIC TRIF	-1	12/08/2011 19:20	1.25	13883	United Kingdom
CS81468	21314	SMALL GLASS HEART TRINKI	-10	12/08/2011 19:26	2.1	13599	United Kingdom
CS81468	22098	BOUDOIR SQUARE TISSUE B	-12	12/08/2011 19:26	0.39	13599	United Kingdom
CS81470	23084	RABBIT NIGHT LIGHT	-4	12/08/2011 19:28	2.08	17924	United Kingdom
CS81484	23843	PAPER CRAFT , LITTLE BIRDI	-80995	12/09/2011 09:27	2.08	16446	United Kingdom
CS81490	22178	VICTORIAN GLASS HANGING	-12	12/09/2011 09:57	1.95	14397	United Kingdom
CS81490	23144	ZINC T-LIGHT HOLDER STAR	-11	12/09/2011 09:57	0.83	14397	United Kingdom
CS81499	M	Manual	-1	12/09/2011 10:28	224.69	15498	United Kingdom
CS81568	21258	VICTORIAN SEWING BOX LA	-5	12/09/2011 11:57	10.95	15311	United Kingdom
CS81569	84978	HANGING HEART JAR T-LIGH	-1	12/09/2011 11:58	1.25	17315	United Kingdom
CS81569	20979	36 PENCILS TUBE RED RETRC	-5	12/09/2011 11:58	1.25	17315	United Kingdom

Gambar 4.4 Data Dengan Kode C

4.2 Proses ETL

ETL merupakan singkatan dari Extract, Transform, Load. Ini adalah suatu proses yang digunakan dalam pengolahan data untuk mengambil data dari berbagai sumber, mengubahnya agar sesuai dengan kebutuhan, dan memuatnya ke dalam penyimpanan atau sistem yang dituju

a. Extract

Tahap pertama dalam proses ETL adalah ekstraksi, yang melibatkan pengambilan data dari beberapa sumber, dalam penelitian ini data diekstraksi dari file CSV. Pada tahap ini dilakukan perubahan format dan tipe data pada atribut Invoicedate. Hal ini bertujuan agar PDI dapat membaca dataset yang di ekstrak. Selain itu pada tahap extract ini juga ditambahkan Length data pada beberapa atribut seperti StockCode dan Country. Hal ini juga bertujuan agar saat data disimpan dalam database, database dapat mengexcute dengan baik.

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim
1	InvoiceNo	Integer	#	30	0	IDR	.	,	none
2	StockCode	String		30		IDR	.	,	none
3	Description	String		35		IDR	.	,	none
4	Quantity	Integer	#	15	0	IDR	.	,	none
5	InvoiceDate	Date	MM/dd/yyyy			IDR	.	,	none
6	UnitPrice	Integer	#	5	2	IDR	.	,	none
7	CustomerID	Integer	#	15	0	IDR	.	,	none
8	Country	String		30		IDR	.	,	none

Gambar 4.5 Ekstraksi Data

Rows of step: CSV file input (1000 rows)

#	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/01/2010	2	17850	United Kingdom
2	536365	71053	WHITE METAL LANTERN	6	12/01/2010	3	17850	United Kingdom
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/01/2010	2	17850	United Kingdom
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/01/2010	3	17850	United Kingdom
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/01/2010	3	17850	United Kingdom
6	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/01/2010	7	17850	United Kingdom
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/01/2010	4	17850	United Kingdom
8	536366	22633	HAND WARMER UNION JACK	6	12/01/2010	1	17850	United Kingdom
9	536366	22632	HAND WARMER RED POLKA DOT	6	12/01/2010	1	17850	United Kingdom
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/01/2010	1	13047	United Kingdom
11	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/01/2010	2	13047	United Kingdom
12	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	12/01/2010	2	13047	United Kingdom
13	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	12/01/2010	3	13047	United Kingdom
14	536367	22310	IVORY KNITTED MUG COSY	6	12/01/2010	1	13047	United Kingdom
15	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	12/01/2010	4	13047	United Kingdom
16	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	12/01/2010	4	13047	United Kingdom
17	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	12/01/2010	9	13047	United Kingdom
18	536367	21754	HOME BUILDING BLOCK WORD	3	12/01/2010	5	13047	United Kingdom
19	536367	21755	LOVE BUILDING BLOCK WORD	3	12/01/2010	5	13047	United Kingdom
20	536367	21777	RECIPE BOX WITH METAL HEART	4	12/01/2010	7	13047	United Kingdom
21	536367	48187	DOORMAT NEW ENGLAND	4	12/01/2010	7	13047	United Kingdom
22	536368	22960	JAM MAKING SET WITH JARS	6	12/01/2010	4	13047	United Kingdom
23	536368	22913	RED COAT RACK PARIS FASHION	3	12/01/2010	4	13047	United Kingdom
24	536368	22912	YELLOW COAT RACK PARIS FASHION	3	12/01/2010	4	13047	United Kingdom
25	536368	22914	BLUE COAT RACK PARIS FASHION	3	12/01/2010	4	13047	United Kingdom
26	536369	21756	BATH BUILDING BLOCK WORD	3	12/01/2010	5	13047	United Kingdom
27	536370	22728	ALARM CLOCK BAKELIKE PINK	24	12/01/2010	3	12583	France
28	536370	22727	ALARM CLOCK BAKELIKE RED	24	12/01/2010	3	12583	France
29	536370	22726	ALARM CLOCK BAKELIKE GREEN	12	12/01/2010	3	12583	France
30	536370	21724	PANDA AND BUNNIES STICKER SHEET	12	12/01/2010	0	12583	France

Gambar 4.6 Preview Data

b. Transform

Setelah data diekstraksi, tahap berikutnya adalah transformasi, di mana data yang telah diambil akan diubah atau dimodifikasi agar sesuai dengan kebutuhan analisis yang digunakan. Transformasi ini bisa melibatkan berbagai tindakan, seperti pembersihan data (data cleansing), penggabungan data (data merging), normalisasi, agregasi, pemfilteran, dan

penambahan atau penghapusan kolom. Pada penelitian ini Transform data yang dilakukan dengan menambahkan satu kolom subtotal yang didapatkan dari hasil perkalian kolom Quantity dengan UnitPrice. Tujuan dari proses transform adalah menghasilkan data yang berkualitas tinggi, konsisten, dan sesuai dengan format yang diinginkan.

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping symbol	Currency symbol
1	subtotal	A * B	Quantity	UnitPrice		Integer			N				

Gambar 4.7 Transform Data

Rows of step: Calculator (1000 rows)

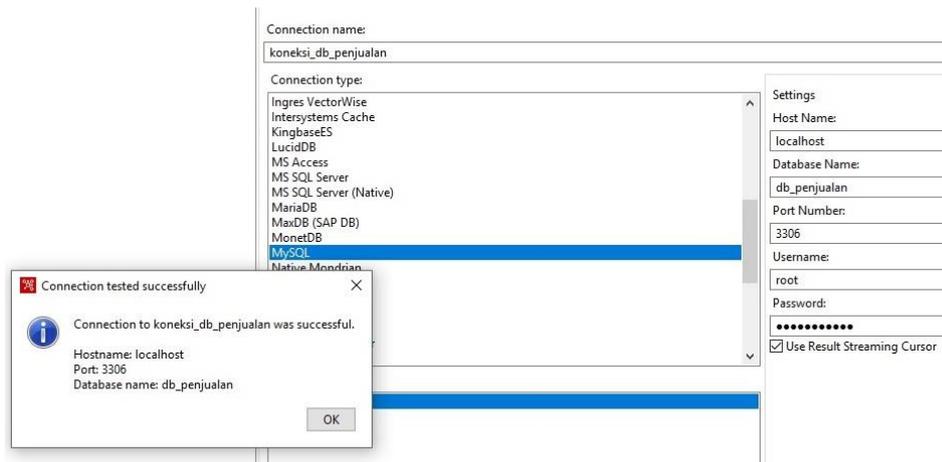
#	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	subtotal
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/01/2010	2	17850	United Kingdom	12
2	536365	71053	WHITE METAL LANTERN	6	12/01/2010	3	17850	United Kingdom	18
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/01/2010	2	17850	United Kingdom	16
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/01/2010	3	17850	United Kingdom	18
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/01/2010	3	17850	United Kingdom	18
6	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/01/2010	7	17850	United Kingdom	14
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/01/2010	4	17850	United Kingdom	24
8	536366	22633	HAND WARMER UNION JACK	6	12/01/2010	1	17850	United Kingdom	6
9	536366	22632	HAND WARMER RED POLKA DOT	6	12/01/2010	1	17850	United Kingdom	6
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/01/2010	1	13047	United Kingdom	32
11	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/01/2010	2	13047	United Kingdom	12
12	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	12/01/2010	2	13047	United Kingdom	12
13	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	12/01/2010	3	13047	United Kingdom	24
14	536367	22310	IVORY KNITTED MUG COSY	6	12/01/2010	1	13047	United Kingdom	6
15	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	12/01/2010	4	13047	United Kingdom	24
16	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	12/01/2010	4	13047	United Kingdom	12
17	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	12/01/2010	9	13047	United Kingdom	18
18	536367	21754	HOME BUILDING BLOCK WORD	3	12/01/2010	5	13047	United Kingdom	15
19	536367	21755	LOVE BUILDING BLOCK WORD	3	12/01/2010	5	13047	United Kingdom	15
20	536367	21777	RECIPE BOX WITH METAL HEART	4	12/01/2010	7	13047	United Kingdom	28
21	536367	48187	DOORMAT NEW ENGLAND	4	12/01/2010	7	13047	United Kingdom	28
22	536368	22960	JAM MAKING SET WITH JARS	6	12/01/2010	4	13047	United Kingdom	24
23	536368	22913	RED COAT RACK PARIS FASHION	3	12/01/2010	4	13047	United Kingdom	12
24	536368	22912	YELLOW COAT RACK PARIS FASHION	3	12/01/2010	4	13047	United Kingdom	12
25	536368	22914	BLUE COAT RACK PARIS FASHION	3	12/01/2010	4	13047	United Kingdom	12
26	536369	21756	BATH BUILDING BLOCK WORD	3	12/01/2010	5	13047	United Kingdom	15
27	536370	22728	ALARM CLOCK BAKELIKE PINK	24	12/01/2010	3	12583	France	72
28	536370	22727	ALARM CLOCK BAKELIKE RED	24	12/01/2010	3	12583	France	72
29	536370	22726	ALARM CLOCK BAKELIKE GREEN	12	12/01/2010	3	12583	France	36
30	536370	21724	PANDA AND BUNNIES STICKER SHEET	12	12/01/2010	0	12583	France	0

Gambar 4.8 Preview Data

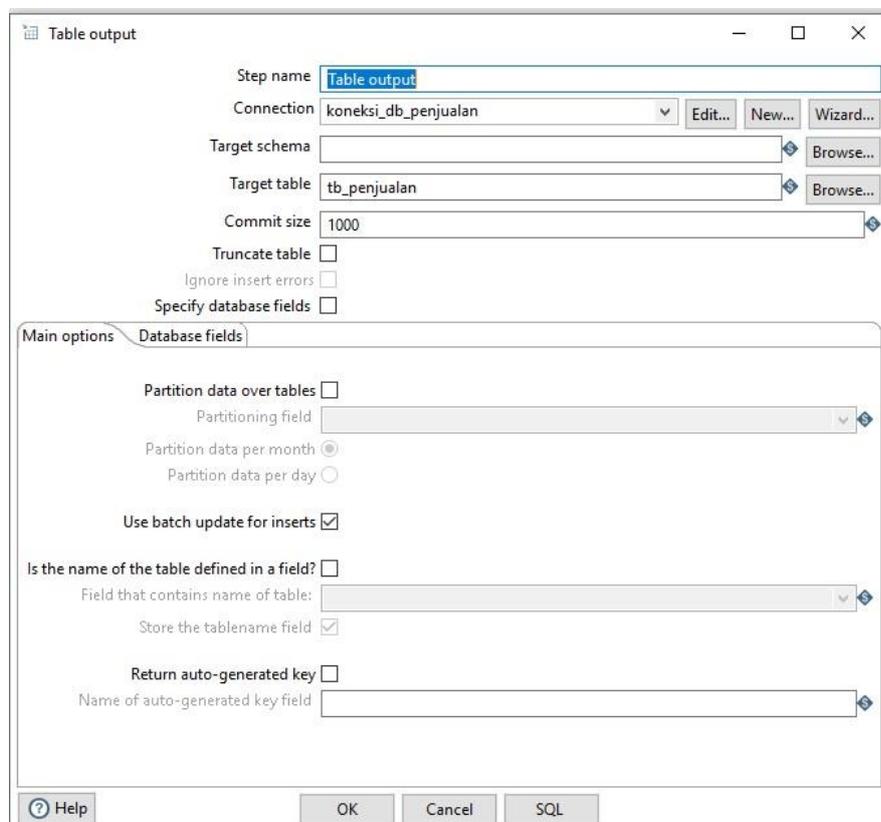
c. Load

Tahap terakhir dalam proses ETL adalah pemuatan, artinya data tersebut sudah melewati tahap ekstraksi dan transformasi sehingga data tersebut akan load ke dalam database yang dituju. Ini bisa berupa data warehouse, basis data operasional, atau sistem lainnya. Proses pemuatan dapat dilakukan secara langsung (direct loading), di mana data dimuat secara langsung ke dalam penyimpanan database. Sebelum data disimpan dalam

database, dilakukan tahap koneksi atau integrasi antara pentaho dengan database.



Gambar 4.9 Koneksi Database



Gambar 4.10 Pembuatan Tabel Penjualan



Gambar 4.11 Sukses ETL

4.3 Desain Data Warehouse

Data warehouse dapat diartikan sebagai wadah yang dibuat khusus untuk menyimpan dan mengelola data dari banyak sumber yang berbeda dalam suatu organisasi. Tujuan utamanya adalah untuk menyediakan akses yang mudah dan cepat ke data yang terintegrasi secara menyeluruh, sehingga memungkinkan pengguna untuk melakukan analisis data, pengambilan keputusan, dan pemahaman bisnis yang lebih baik. Data yang disimpan dalam data warehouse biasanya telah melalui proses transformasi, pembersihan, dan penggabungan agar sesuai dengan kebutuhan analisis. Hasil load data dari proses ETL yang sebelumnya akan ditampilkan dalam database sql yog yang sudah di integrasikan dengan pentaho.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	subtotal
536365	85123A	WHITE HARBING HEART T-LIGHT HOLDER	6	2010-12-01 00:00:00	2	17850	United Kingdom	12
536365	71053	WHITE METAL LANTERN	6	2010-12-01 00:00:00	3	17850	United Kingdom	18
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 00:00:00	2	17850	United Kingdom	16
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 00:00:00	3	17850	United Kingdom	18
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 00:00:00	3	17850	United Kingdom	18
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01 00:00:00	7	17850	United Kingdom	14
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010-12-01 00:00:00	4	17850	United Kingdom	24
536366	22633	HAND WARMER UNION JACK	6	2010-12-01 00:00:00	1	17850	United Kingdom	6
536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01 00:00:00	1	17850	United Kingdom	6
536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010-12-01 00:00:00	1	13047	United Kingdom	32
536367	22745	POPPY'S FLAYHOUSE BEDROOM	6	2010-12-01 00:00:00	2	13047	United Kingdom	12
536367	22748	POPPY'S FLAYHOUSE KITCHEN	6	2010-12-01 00:00:00	2	13047	United Kingdom	12
536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	2010-12-01 00:00:00	3	13047	United Kingdom	24
536367	22310	IVORY KNITTED MUG COSY	6	2010-12-01 00:00:00	1	13047	United Kingdom	6
536367	84869	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	2010-12-01 00:00:00	4	13047	United Kingdom	24
536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	2010-12-01 00:00:00	4	13047	United Kingdom	12
536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	2010-12-01 00:00:00	9	13047	United Kingdom	18
536367	21754	HOME BUILDING BLOCK WORD	3	2010-12-01 00:00:00	5	13047	United Kingdom	15
536367	21755	LOVE BUILDING BLOCK WORD	3	2010-12-01 00:00:00	5	13047	United Kingdom	15
536367	21777	RECIPE BOX WITH METAL HEART	4	2010-12-01 00:00:00	7	13047	United Kingdom	28
536367	48187	DOORMAT NEW ENGLAND	4	2010-12-01 00:00:00	7	13047	United Kingdom	28
536368	22960	JAM MAKING SET WITH JARS	6	2010-12-01 00:00:00	4	13047	United Kingdom	24

Gambar 4.12 Database Penjualan

4.4 Desain Data Multidimensi

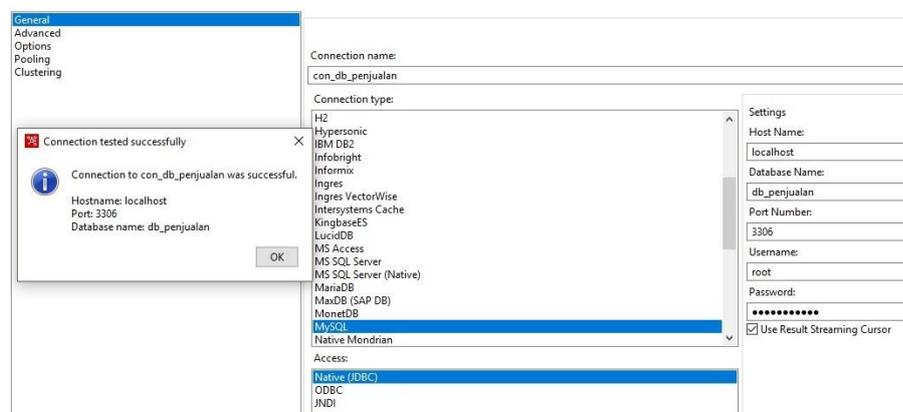
Desain data multidimensi mengacu pada pendekatan dalam merancang struktur data yang memungkinkan menyimpan dan mengelola data dalam berbagai format atau jenis. Data disimpan dalam format yang menggambarkan hubungan

antar-dimensi dan memungkinkan pengguna untuk melakukan analisis dengan cara yang intuitif dan efektif. model multidimensi mewakili sudut pandang yang berbeda terhadap data. Dimensi menyediakan kerangka atau konteks untuk data yang terkait, yang memungkinkan pengguna untuk mengelompokkan dan menyaring data berdasarkan kriteria tertentu.

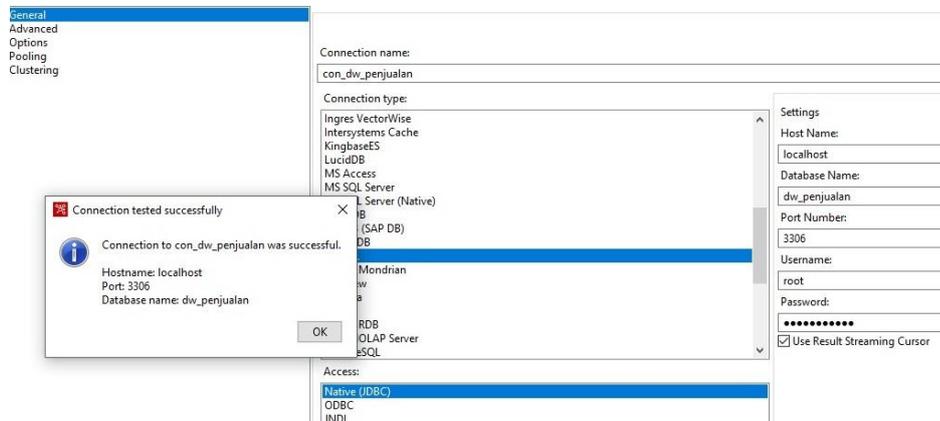
Istilah "multi-model" mengindikasikan bahwa struktur data ini dapat menangani beberapa model data, seperti model relasional, model dokumen, model grafik, dan sebagainya, tanpa harus membatasi diri pada satu model tertentu. Pada penelitian ini Data multidimensi akan dibuat menjadi 4 tabel yaitu tabel dimensi customer, dimensi produk, dimensi negara, dimensi waktu dan 1 tabel fakta. Fakta adalah data numerik yang terkait dengan dimensi tertentu. Masing-masing entitas fakta dalam model multidimensi merepresentasikan suatu peristiwa atau kejadian dalam domain bisnis, dan diukur atau dihitung dalam kaitannya dengan dimensi yang sesuai.

a. Pembuatan Tabel Dimensi Customer

Dalam Pembuatan tabel dimensi customer, langkah pertama yang harus dipersiapkan adalah mengintegrasikan database yang akan dijadikan sebagai sumber data dengan PDI. Pada penelitian ini database yang dijadikan sumber data adalah db_penjualan dan satu database lagi sebagai database output dengan nama dw_penjualan.

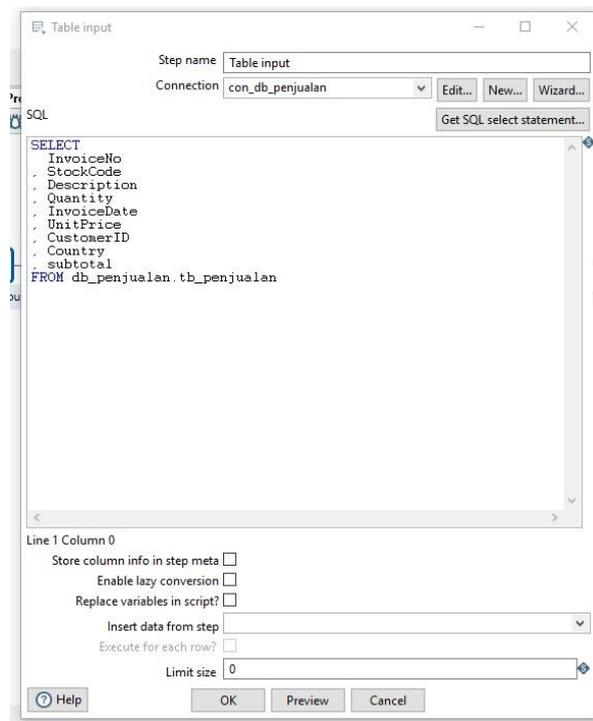


Gambar 4.13 Koneksi Database Penjualan



Gambar 4.14 Koneksi Database dw_penjualan

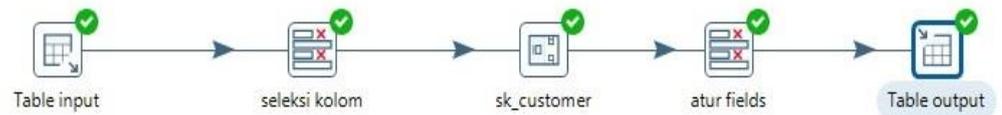
Setelah proses koneksi database berhasil proses selanjutnya adalah menampilkan atribut yang akan digunakan dalam pembuatan tabel dimensi customer.



Gambar 4.15 Tabel Input Penjualan

Setelah semua atribut tampil, selanjutnya adalah proses seleksi kolom atau seleksi atribut yang digunakan dalam pembuatan tabel dimensi customer. Pada tahap ini CustomerID akan digunakan sebagai atribut tabel dimensi

customer. Kemudian pembuatan sk_customer yang dijadikan sebagai primary key. Sebelum ke tahap akhir yaitu table output atribut akan disusun secara rapi dimana sk_customer akan diletakkan diatas atribut CustomerID.



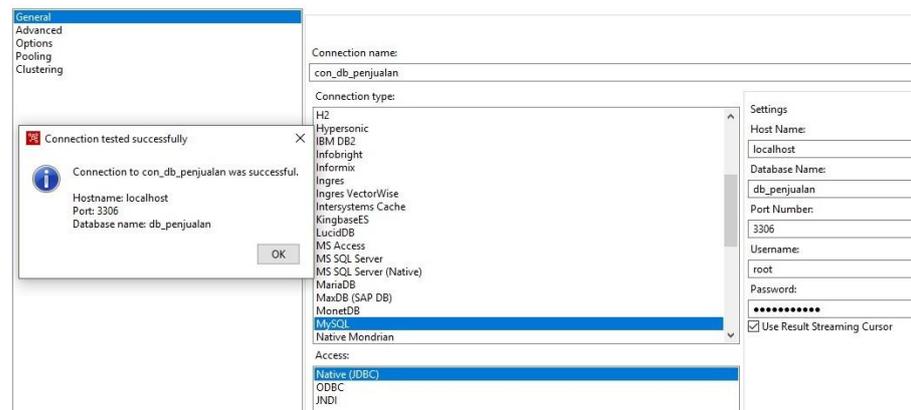
Gambar 4.16 Sukses dim_customer

sk_customer	CustomerID
1	17850
2	17850
3	17850
4	17850
5	17850
6	17850
7	17850
8	17850
9	17850
10	13047
11	13047
12	13047
13	13047
14	13047
15	13047
16	13047
17	13047
18	13047
19	13047
20	13047
21	13047
22	13047

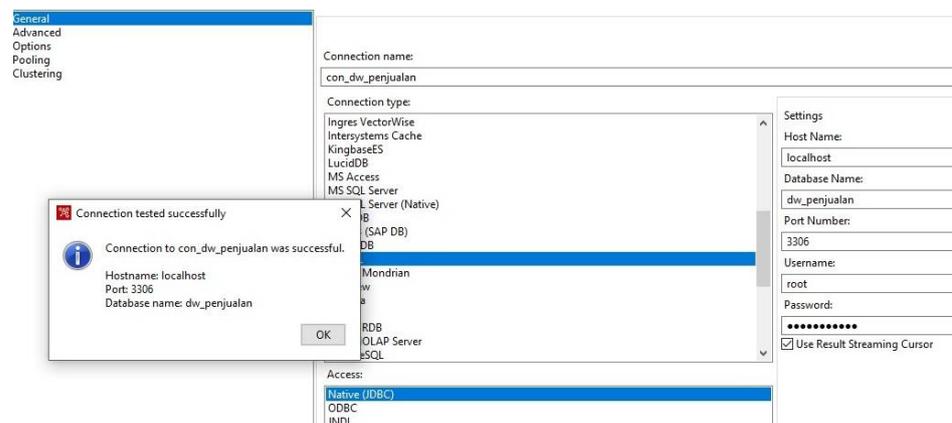
Gambar 4.17 Tabel dim_customer

b. Pembuatan Tabel Dimensi Produk

Dalam Pembuatan tabel dimensi produk, langkah pertama yang harus dipersiapkan adalah mengintegrasikan database yang akan dijadikan sebagai sumber data dengan PDI. Pada penelitian ini database yang dijadikan sumber data adalah db_penjualan dan satu database lagi sebagai database output dengan nama dw_penjualan

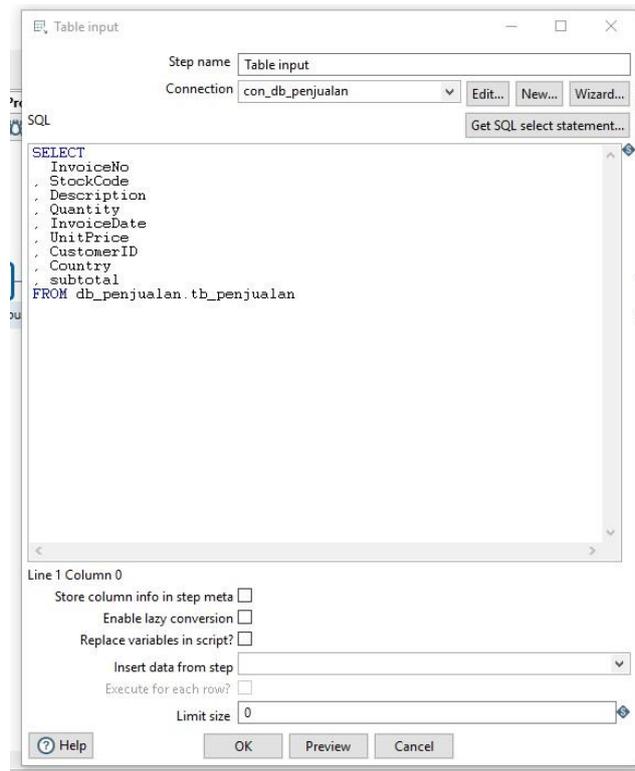


Gambar 4.18 Koneksi Database Penjualan



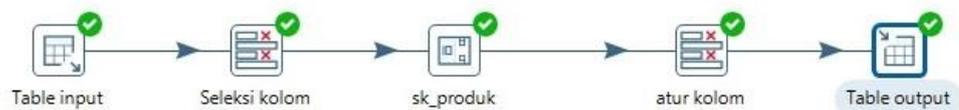
Gambar 4.19 Koneksi Database dw_penjualan

Setelah proses koneksi database berhasil proses selanjutnya adalah menampilkan atribut yang akan digunakan dalam pembuatan tabel dimensi produk.



Gambar 4.20 Tabel Input Penjualan

Setelah semua atribut tampil, selanjutnya adalah proses seleksi kolom atau seleksi atribut yang digunakan dalam pembuatan tabel dimensi produk. Pada tahap ini StockCode dan Description akan digunakan sebagai atribut tabel dimensi produk. Kemudian pembuatan sk_produk yang dijadikan sebagai primary key. Sebelum ke tahap akhir yaitu table output atribut akan disusun secara rapi dimana sk_produk diletakkan diatas atribut StockCode dan Description.



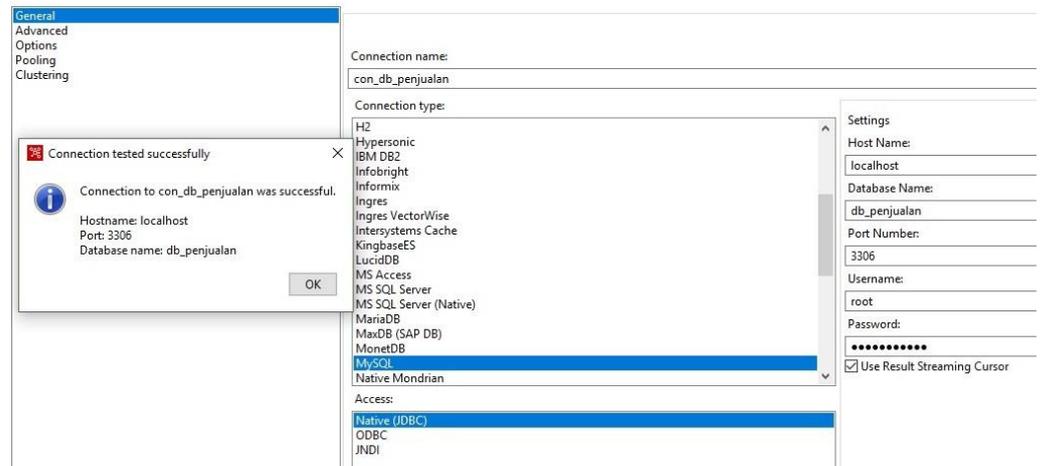
Gambar 4.21 Sukses dim_produk

<input type="checkbox"/> sk_produk	StockCode	Description
<input type="checkbox"/>	1 85123A	WHITE HANGING HEART T-LIGHT HOLDER
<input type="checkbox"/>	2 71053	WHITE METAL LANTERN
<input type="checkbox"/>	3 84406B	CREAM CUPID HEARTS COAT HANGER
<input type="checkbox"/>	4 84029G	KNITTED UNION FLAG HOT WATER BOTTLE
<input type="checkbox"/>	5 84029E	RED WOOLLY HOTTIE WHITE HEART.
<input type="checkbox"/>	6 22752	SET 7 BABUSHKA NESTING BOXES
<input type="checkbox"/>	7 21730	GLASS STAR FROSTED T-LIGHT HOLDER
<input type="checkbox"/>	8 22633	HAND WARMER UNION JACK
<input type="checkbox"/>	9 22632	HAND WARMER RED POLKA DOT
<input type="checkbox"/>	10 84879	ASSORTED COLOUR BIRD ORNAMENT
<input type="checkbox"/>	11 22745	POPPY'S PLAYHOUSE BEDROOM
<input type="checkbox"/>	12 22748	POPPY'S PLAYHOUSE KITCHEN
<input type="checkbox"/>	13 22749	FELTCRAFT PRINCESS CHARLOTTE DOLL
<input type="checkbox"/>	14 22310	IVORY KNITTED MUG COSY
<input type="checkbox"/>	15 84969	BOX OF 6 ASSORTED COLOUR TEASPOONS
<input type="checkbox"/>	16 22623	BOX OF VINTAGE JIGSAW BLOCKS
<input type="checkbox"/>	17 22622	BOX OF VINTAGE ALPHABET BLOCKS
<input type="checkbox"/>	18 21754	HOME BUILDING BLOCK WORD
<input type="checkbox"/>	19 21755	LOVE BUILDING BLOCK WORD
<input type="checkbox"/>	20 21777	RECIPE BOX WITH METAL HEART
<input type="checkbox"/>	21 48187	DOORMAT NEW ENGLAND
<input type="checkbox"/>	22 22960	JAM MAKING SET WITH JARS

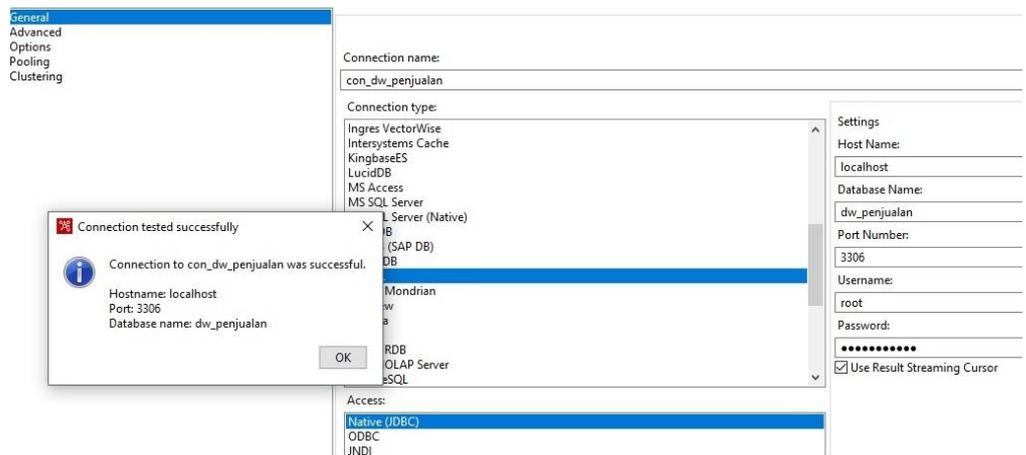
Gambar 4.22 Tabel dim_produk

c. Pembuatan Tabel Dimensi Negara

Dalam Pembuatan tabel dimensi negara, langkah pertama yang harus dipersiapkan adalah mengintegrasikan database yang akan dijadikan sebagai sumber data dengan PDI. Pada penelitian ini database yang dijadikan sumber data adalah db_penjualan dan satu database lagi sebagai database output dengan nama dw_penjualan

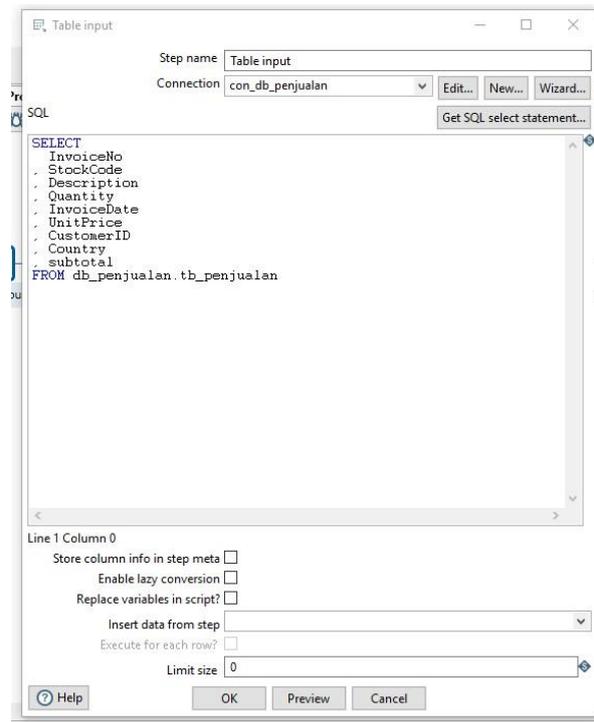


Gambar 4.23 Koneksi Database Penjualan



Gambar 4.24 Koneksi Database dw_penjualan

Setelah proses koneksi database berhasil proses selanjutnya adalah menampilkan atribut yang akan digunakan dalam pembuatan tabel dimensi negara.



Gambar 4.25 Tabel Input Penjualan

Setelah semua atribut tampil, selanjutnya adalah proses seleksi kolom atau seleksi atribut yang digunakan dalam pembuatan tabel dimensi negara. Pada tahap ini Country akan digunakan sebagai atribut tabel dimensi negara. Kemudian pembuatan sk_negara yang dijadikan sebagai primary key. Sebelum ke tahap akhir yaitu table output atribut akan disusun secara rapi dimana sk_negara diletakkan diatas atribut Country.



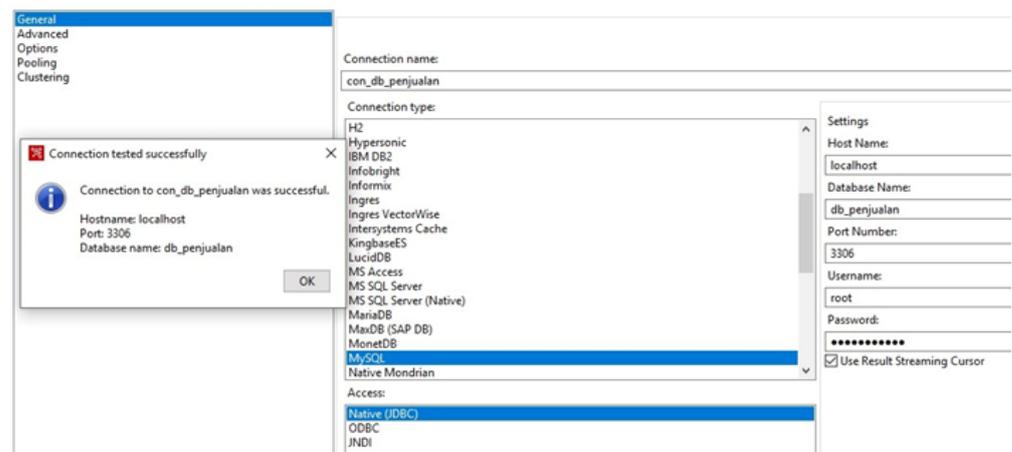
Gambar 4.26 Sukses dim_negara

sk_negara	Country
1	United Kingdom
2	United Kingdom
3	United Kingdom
4	United Kingdom
5	United Kingdom
6	United Kingdom
7	United Kingdom
8	United Kingdom
9	United Kingdom
10	United Kingdom
11	United Kingdom
12	United Kingdom
13	United Kingdom
14	United Kingdom
15	United Kingdom
16	United Kingdom
17	United Kingdom
18	United Kingdom
19	United Kingdom
20	United Kingdom
21	United Kingdom
22	United Kingdom

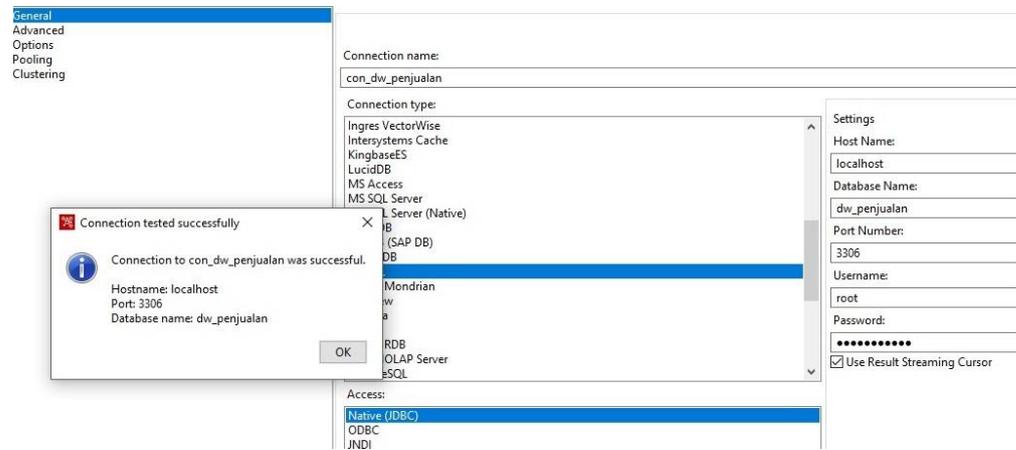
Gambar 4.27 Tabel dim_negara

d. Pembuatan Tabel Dimensi Waktu

Dalam Pembuatan tabel dimensi waktu, langkah pertama yang harus dipersiapkan adalah mengintegrasikan database yang akan dijadikan sebagai sumber data dengan PDI. Pada penelitian ini database yang dijadikan sumber data adalah db_penjualan dan satu database lagi sebagai database output dengan nama dw_penjualan

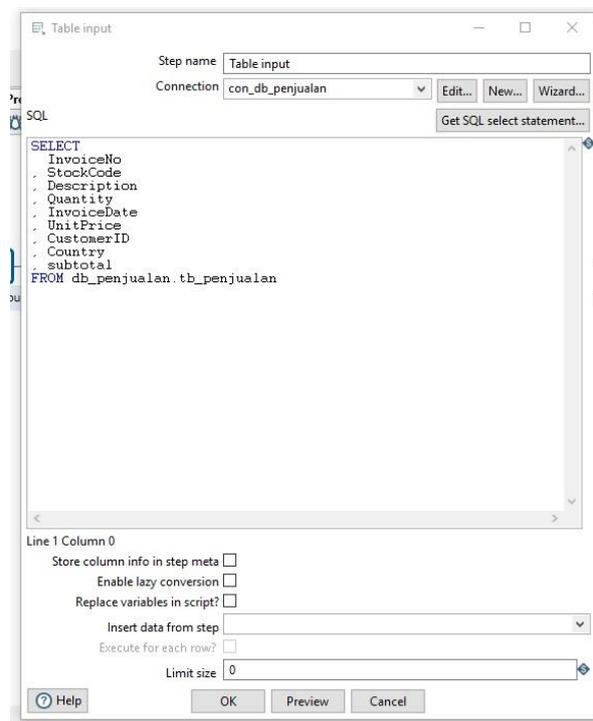


Gambar 4.28 Koneksi Database Penjualan



Gambar 4.29 Koneksi Database dw_penjualan

Setelah proses koneksi database berhasil proses selanjutnya adalah menampilkan atribut yang akan digunakan dalam pembuatan tabel dimensi waktu.



Gambar 4.30 Tabel Input Penjualan

Setelah semua atribut tampil, selanjutnya adalah proses seleksi kolom atau seleksi atribut yang digunakan dalam pembuatan tabel dimensi waktu.

Pada tahap ini InvoiceDate akan digunakan sebagai atribut tabel dimensi waktu. Kemudian pembuatan sk_waktu yang dijadikan sebagai primary key. Sebelum ke tahap akhir yaitu table output atribut akan disusun secara rapi dimana sk_waktu diletakkan diatas atribut InvoiceDate.



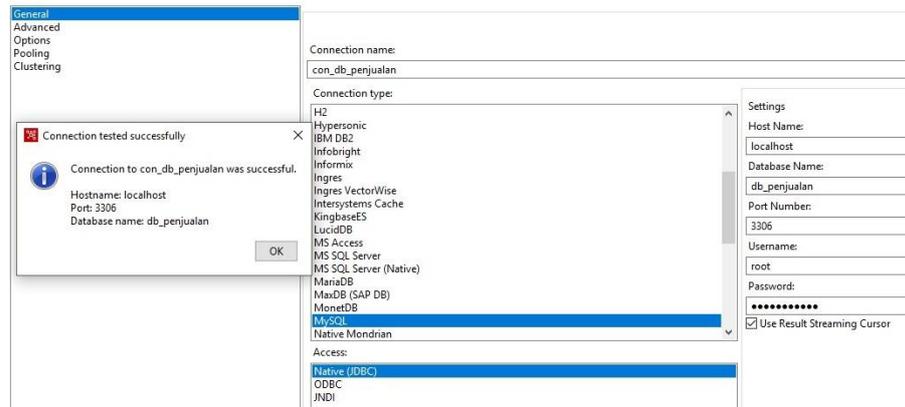
Gambar 4.31 Sukses dim_waktu

sk_waktu	InvoiceDate
1	2010-12-01 00:00:00
2	2010-12-01 00:00:00
3	2010-12-01 00:00:00
4	2010-12-01 00:00:00
5	2010-12-01 00:00:00
6	2010-12-01 00:00:00
7	2010-12-01 00:00:00
8	2010-12-01 00:00:00
9	2010-12-01 00:00:00
10	2010-12-01 00:00:00
11	2010-12-01 00:00:00
12	2010-12-01 00:00:00
13	2010-12-01 00:00:00
14	2010-12-01 00:00:00
15	2010-12-01 00:00:00
16	2010-12-01 00:00:00
17	2010-12-01 00:00:00
18	2010-12-01 00:00:00
19	2010-12-01 00:00:00
20	2010-12-01 00:00:00
21	2010-12-01 00:00:00
22	2010-12-01 00:00:00

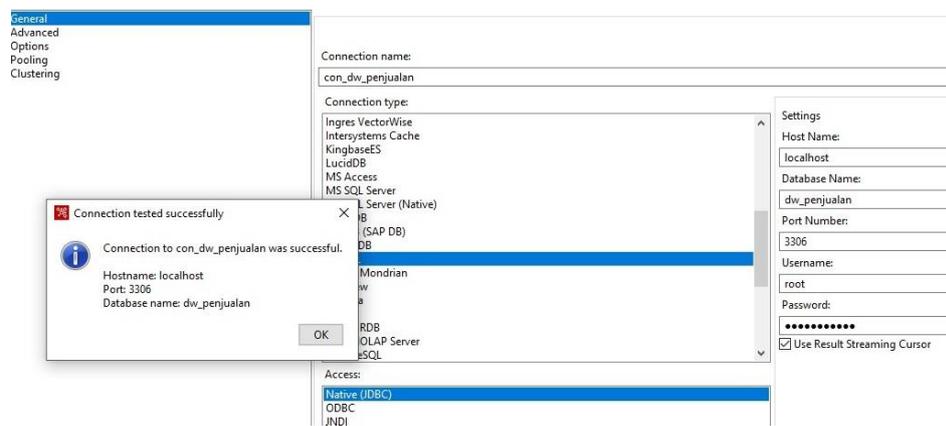
Gambar 4.32 Tabel dim_waktu

e. Pembuatan Tabel Fakta

Dalam Pembuatan tabel fakta, langkah pertama yang harus dipersiapkan adalah mengintegrasikan database yang akan dijadikan sebagai sumber data dengan PDI. Pada penelitian ini database yang dijadikan sumber data adalah db_penjualan dan satu database lagi sebagai database output dengan nama dw_penjualan

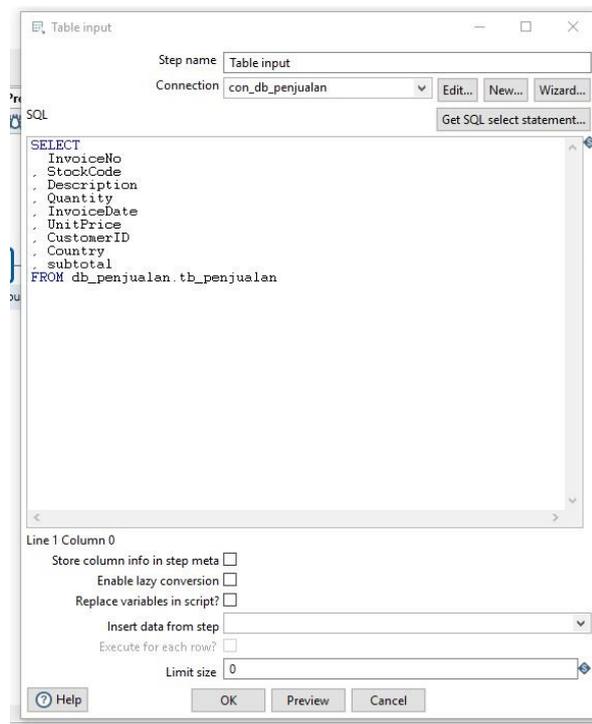


Gambar 4.33 Koneksi Database Penjualan



Gambar 4.34 Koneksi Database dw_penjualan

Setelah proses koneksi database berhasil proses selanjutnya adalah menampilkan atribut yang akan digunakan dalam pembuatan tabel fakta.



Gambar 4.35 tabel Input Penjualan

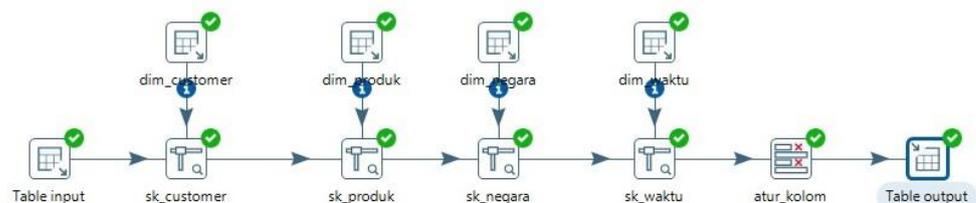
Setelah semua atribut tampil, selanjutnya adalah step transformation stream look up, step ini berfungsi untuk menggabungkan berbagai tabel dimensi menjadi 1 tabel fakta. Dalam step ini terjadi proses penggabungan berdasarkan nilai kunci atau *primary key* yang cocok. Jika tidak memiliki kunci yang cocok maka penggabungan data akan mengalami error. Pada step ini sumber data yang berasal dari tabel input (db_penjualan) memiliki atribut CustomerID kemudian sumber data kedua yang berasal dari tabel dimensi customer (dw_penjualan) juga memiliki CustomerID. Namun pada sumber data kedua tabel dimensi customer ada tambahan sk_customer yang berfungsi sebagai *surrogate key* atau kunci tiruan. Kunci tiruan ini digunakan dalam desain basis data untuk mengidentifikasi pelanggan dengan cara yang mungkin lebih efisien atau lebih sesuai dengan kebutuhan pengelolaan data. Jadi, sk_customer juga berperan sebagai kunci utama untuk mengidentifikasi pelanggan, tetapi mungkin dengan pendekatan yang berbeda atau lebih abstrak daripada CustomerID yang mungkin lebih langsung dan spesifik.

Selanjutnya pada tabel dimensi produk yang berasal dari dw_penjualan memiliki atribut sk_produk, StockCode dan Description. Sedangkan pada tabel input (db_prnjualan) juga memiliki StockCode. Jadi kedua sumber data ini dihubungkan dengan atribut StockCode dengan tambahan sk_produk sebagai kunci tiruan yang berguna juga untuk mengidentifikasi produk dengan cara yang mungkin lebih efisien atau lebih sesuai dengan kebutuhan pengelolaan.

Pada tabel dimensi negara yang berasal dari dw_penjualan memiliki atribut sk_negara dan Country. Sedangkan pada tabel input (db_prnjualan) juga memiliki Country. Jadi kedua sumber data ini dihubungkan dengan atribut Country dengan tambahan sk_negara sebagai kunci tiruan yang berguna juga untuk mengidentifikasi negara dengan cara yang mungkin lebih efisien atau lebih sesuai dengan kebutuhan pengelolaan.

Terakhir pada tabel dimensi waktu yang berasal dari dw_penjualan memiliki atribut sk_waktu dan InvoiceDate. Sedangkan pada tabel input (db_prnjualan) juga memiliki InvoiceDate. Jadi kedua sumber data ini dihubungkan dengan atribut InvoiceDate dengan tambahan sk_waktu sebagai kunci tiruan yang berguna juga untuk mengidentifikasi waktu dengan cara yang mungkin lebih efisien atau lebih sesuai dengan kebutuhan pengelolaan.

Sebelum ke tahap tabel output maka akan dilakukan seleksi kolom yang akan digunakan. Pada penelitian ini kolom atribut yang akan digunakan adalah sk_customer, sk_produk, sk_negara, sk_waktu, InvoiceNo, Quantity, UnitPrice dan subtotal.



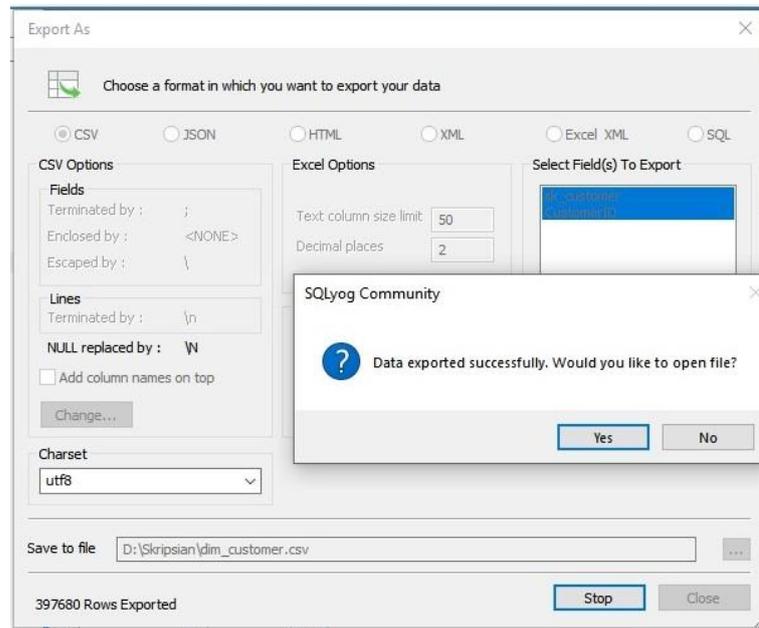
Gambar 4.36 Sukses Tabel Fakta

4.5 OLAP

OLAP (Online Analytical Processing) adalah metode pengolahan data yang memungkinkan pengguna untuk melakukan analisis multidimensional yang mendalam terhadap data. Dengan OLAP, pengguna dapat menjelajahi data dari berbagai sudut pandang, melakukan pengeboran (drill-down) ke level yang lebih dalam, dan melihat tren serta pola yang signifikan dalam data. Sebelum data dapat dimasukkan ke dalam sistem OLAP untuk analisis, data biasanya diekstrak dari sumber data operasional yang mendasarinya dan dimuat ke dalam sebuah penyimpanan data khusus yang disebut Data Warehouse. Pada penelitian ini data yang sudah di ekstrak disimpan ke dalam sebuah manajemen basis data SQL Yog.

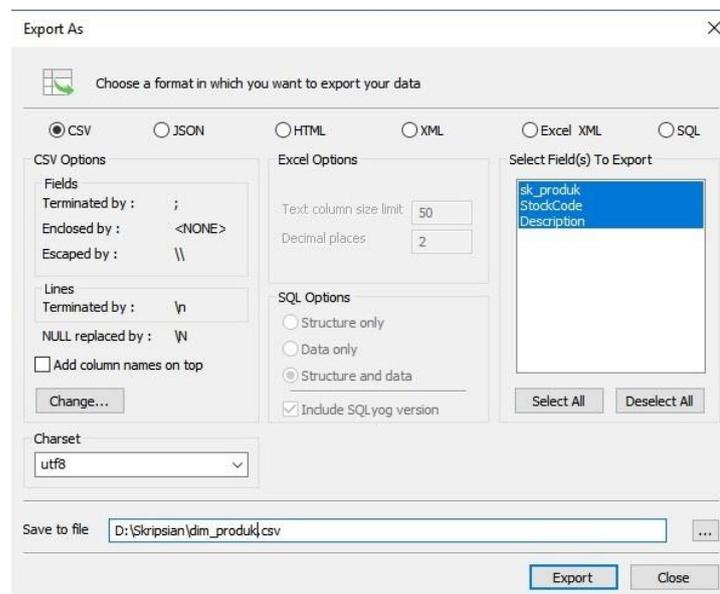
Dalam pembuatan data warehouse dan analisis OLAP, SQL Yog dapat digunakan untuk mengekspor data dari sumber data SQL ke dalam format file CSV. File CSV (Comma-Separated Values) adalah format file sederhana yang mengorganisir data dalam bentuk tabel, dengan setiap baris mewakili sebuah catatan dan setiap kolom dipisahkan oleh koma. Ekspor data ke dalam format CSV memiliki beberapa keuntungan, yaitu:

1. Portabilitas: Data dalam format CSV dapat dengan mudah dipindahkan antara berbagai sistem dan platform, karena hampir semua perangkat lunak dapat mengimpor dan membaca file CSV.
2. Kesederhanaan: Format CSV sederhana dan mudah dimengerti, yang membuatnya ideal untuk digunakan dalam proses pemuatan data dan integrasi.
3. Kompatibilitas dengan Alat Analisis: Banyak alat analisis data, termasuk alat OLAP, dapat dengan mudah mengimpor data dari format CSV, memungkinkan analisis yang lebih lanjut.



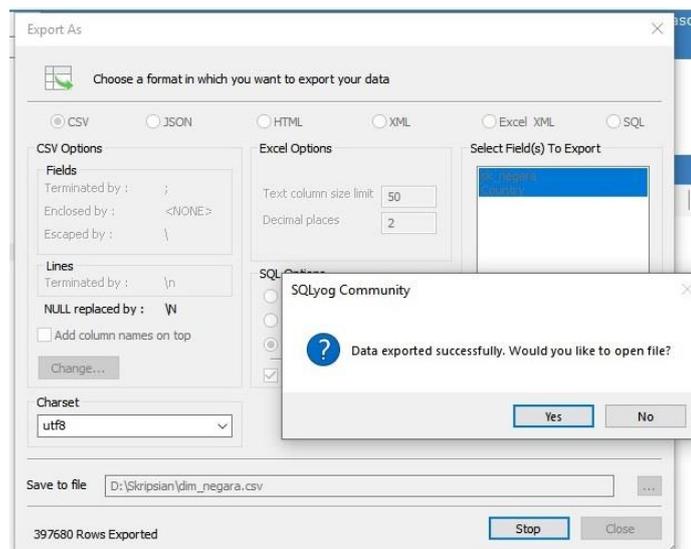
Gambar 4.37 Ekspor Tabel dim_customer

Gambar di atas merupakan proses ekspor tabel dimensi customer menjadi file csv yang diberi nama dim_customer. Fields yang di ekspor akan menyesuaikan fields yang ada dalam tabel dimensi customer. File CSV tersebut dipisahkan dengan tanda titik koma (;).



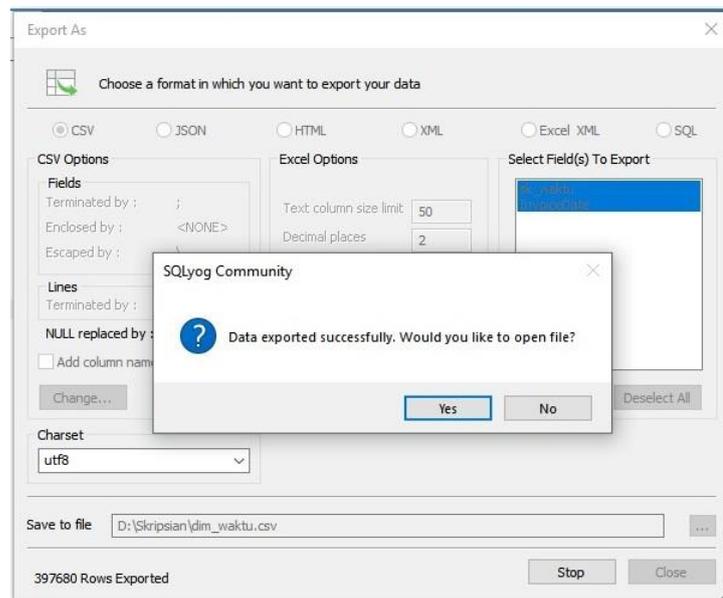
Gambar 4.38 Eksport Tabel Produk

Gambar di atas merupakan proses ekspor tabel dimensi customer menjadi file csv yang diberi nama dim_produk. Fields yang di ekspor akan menyesuaikan fields yang ada dalam tabel dimensi produk. File CSV tersebut dipisahkan dengan tanda titik koma (;).



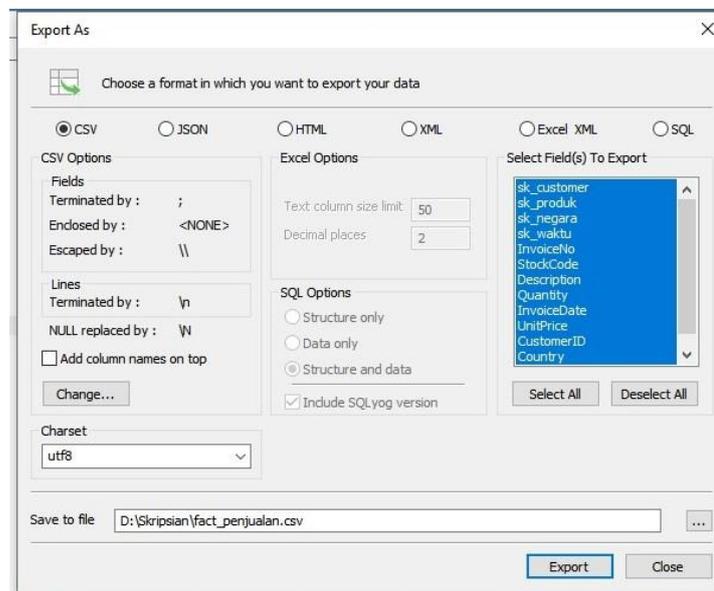
Gambar 4.39 Ekspor Tabel Negara

Gambar di atas merupakan proses ekspor tabel dimensi negara menjadi file csv yang diberi nama dim_negara. Fields yang di ekspor akan menyesuaikan fields yang ada dalam tabel dimensi negara. File CSV tersebut dipisahkan dengan tanda titik koma (;).



Gambar 4.40 Ekspor Tabel Waktu

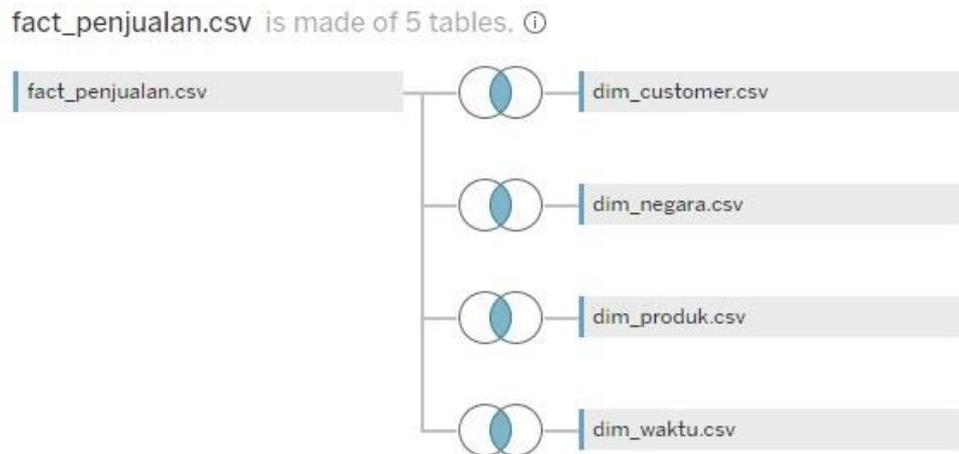
Gambar di atas merupakan proses ekspor tabel dimensi waktu menjadi file csv yang diberi nama dim_waktu. Fields yang di ekspor akan menyesuaikan fields yang ada dalam tabel dimensi waktu. File CSV tersebut dipisahkan dengan tanda titik koma (;).



Gambar 4.41 Ekspor Tabel Fakta

Gambar di atas merupakan proses ekspor tabel fakta menjadi file csv yang diberi nama fact_penjualan. Fields yang di ekspor akan menyesuaikan

fields yang ada dalam tabel fact penjualan. File CSV tersebut dipisahkan dengan tanda titik koma (;).



Gambar 4.42 Relasi Tabel

Setelah semua tabel dimensi dan tabel fakta di ubah menjadi file csv langkah selanjutnya adalah merelasikannya ke dalam tableau sebagai bahan analisis OLAP. Setiap file dapat direlasikan karena memiliki *primary key* dan *foreign key* yang sesuai. Relasi antar tabel dilakukan dalam antarmuka data source dengan fitur *drag and drop*.

The screenshot shows a Tableau OLAP view with the following dimensions and measures:

- Columns:** YEAR(Invoice Date), QUARTER(Invoice Date), MONTH(Invoice Date)
- Rows:** Customer ID, Country, Description
- Measures:** SUM(Quantity)

Customer ID	Count	Description	Invoice Date			
			2010 Q4	January	February	2011 March
12346	1	United King. MEDIUM CE.		74	215	
12347	1	Iceland 3D DOG PIC.	24	12		
		3D SHEET O.				
		60 TEATIM.		24		
		72 SWEETH.		24		
		AIRLINE BA.	4	4		
		AIRLINE BA.		4		
		AIRLINE BA.		4		
		AIRLINE BA.		4		
		ALARM CLO.	4	4		
		ALARM CLO.	4	4		
		ALARM CLO.	4	4		
		ALARM CLO.	4	4		
		BATHROOM.	12	4		
		BICYCLE PU.				

Gambar 4.43 OLAP

Analisis OLAP pada penelitian ini memanfaatkan berbagai dimensi seperti dimensi waktu, dimensi negara, dimensi produk dan dimensi

customer untuk memberikan pemahaman tentang kinerja penjualan perusahaan dari berbagai sudut pandang. Dalam visualisasi yang dibuat, InvoiceDate diposisikan sebagai kolom, sedangkan CustomerID, Country dan Description diposisikan sebagai baris. Hal ini memungkinkan untuk memperoleh pemahaman tentang penjualan berdasarkan tanggal pembelian, asal negara, nama produk yang dibeli dan pelanggan ID tertentu.



Gambar 4.44 Penggunaan OLAP

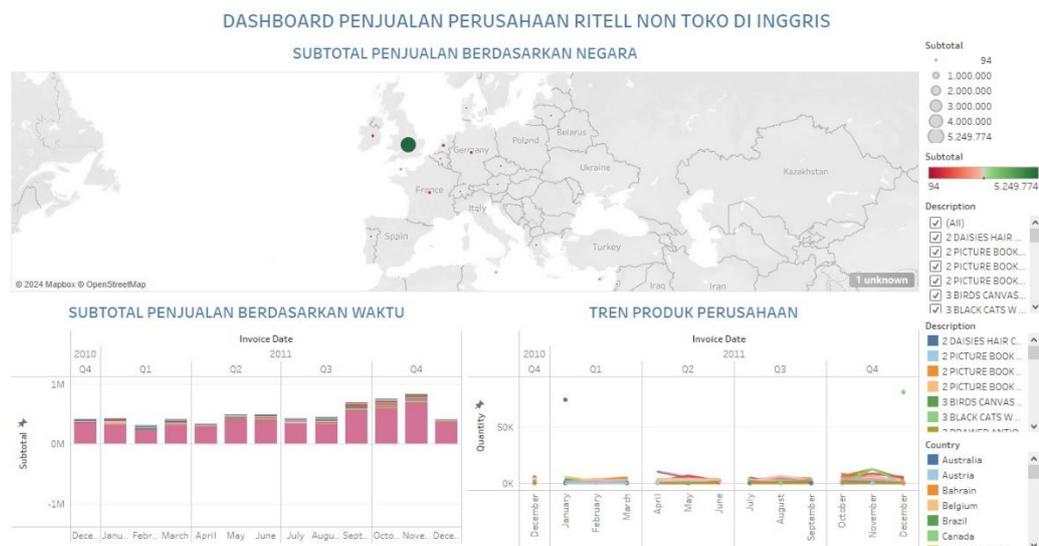
Untuk meningkatkan fleksibilitas analisis, penulis menggunakan fitur OLAP seperti drill down untuk melihat data secara detail, roll up untuk melihat data secara umum atau luas, slice untuk membatasi data dengan memilih subset tertentu dan dice untuk memilih dan membatasi data dengan menggunakan beberapa filter atau kriteria pada beberapa dimensi. Semua fitur yang digunakan memungkinkan untuk menyesuaikan tampilan data sesuai dengan preferensi dan kebutuhan analisis. Hal ini bertujuan agar dapat menampilkan data berdasarkan negara tertentu, produk spesifik, atau pelanggan tertentu dan periode tertentu untuk mendapatkan wawasan yang lebih terfokus.

4.6 Visualisasi

Visualisasi data memiliki penting dalam pengolahan informasi dan pengambilan keputusan. Dengan mengubah data menjadi representasi grafis yang menarik, visualisasi data membantu menyampaikan informasi kompleks dengan cara yang

lebih mudah dimengerti. Hal ini memberikan kesempatan bagi pengguna untuk mengidentifikasi pola, tren, dan anomali dalam data secara cepat. Dengan demikian, visualisasi data tidak hanya memfasilitasi pemahaman yang lebih baik tentang data, tetapi juga memungkinkan pengguna untuk membuat keputusan yang lebih tepat dan berbasis bukti.

Dalam penelitian ini berbagai jenis visualisasi data digunakan untuk menganalisis dan menyajikan informasi. Visualisasi dibuat menggunakan heatmap, diagram batang, dan grafik garis. Heatmap digunakan untuk menampilkan distribusi data subtotal penjualan berdasarkan skala warna, semakin gelap warna pada suatu negara menunjukkan semakin tinggi subtotal penjualan di negara tersebut, semakin terang warna pada suatu negara menunjukkan semakin kecil penjualan pada negara tersebut. Selanjutnya ada diagram batang yang menunjukkan subtotal penjualan dalam setiap bulan. Hal ini menunjukkan bahwa penjualan paling tinggi terjadi di bulan November 2011 di negara United Kingdom. Terakhir ada grafik garis yang menunjukkan tren produk perusahaan. Tren produk perusahaan didapatkan dengan cara melihat produk yang paling banyak dibeli dalam periode tertentu. Dalam dashboard tersebut bisa dilihat bahwa tren produk perusahaan yaitu word war 2 gliders asstd designs. Penjualan produk tersebut selalu ada dalam setiap bulannya.



Gambar 4.45 Dashboard