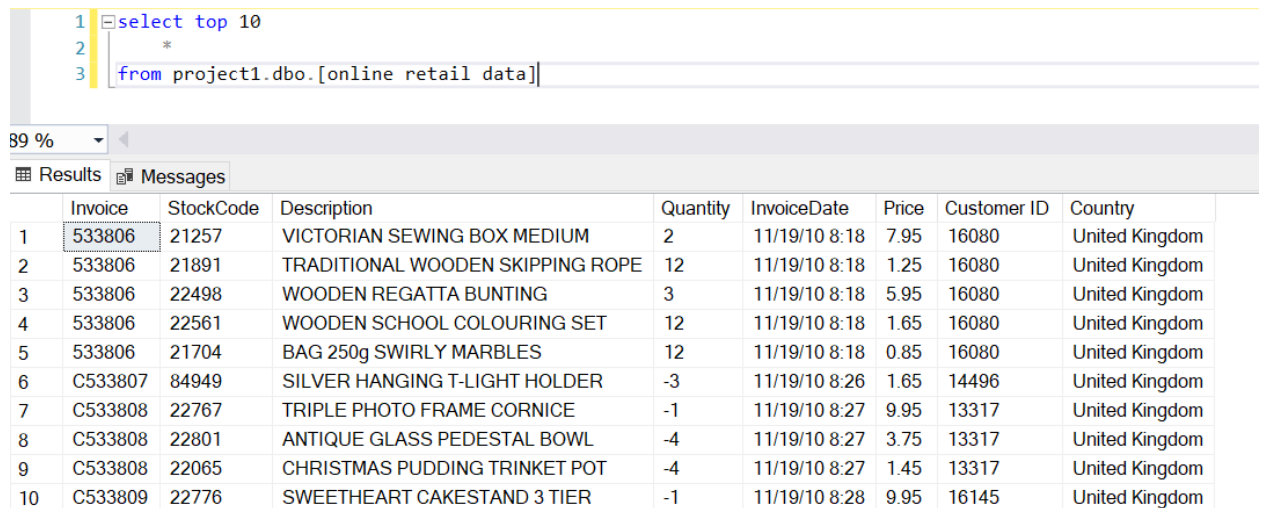


1. Data overview



The screenshot shows a SQL query window with the following text:

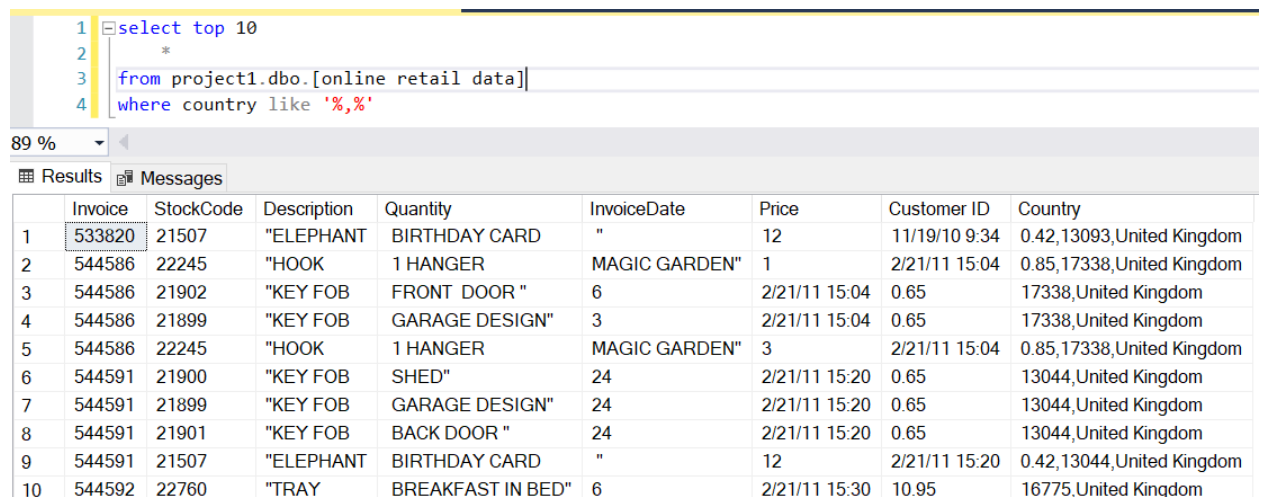
```
1 select top 10
2 *
3 from project1.dbo.[online retail data]
```

Below the query window, the 'Results' tab is active, displaying a table with 10 rows and 9 columns. The columns are Invoice, StockCode, Description, Quantity, InvoiceDate, Price, Customer ID, and Country. The data is as follows:

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
1	533806	21257	VICTORIAN SEWING BOX MEDIUM	2	11/19/10 8:18	7.95	16080	United Kingdom
2	533806	21891	TRADITIONAL WOODEN SKIPPING ROPE	12	11/19/10 8:18	1.25	16080	United Kingdom
3	533806	22498	WOODEN REGATTA BUNTING	3	11/19/10 8:18	5.95	16080	United Kingdom
4	533806	22561	WOODEN SCHOOL COLOURING SET	12	11/19/10 8:18	1.65	16080	United Kingdom
5	533806	21704	BAG 250g SWIRLY MARBLES	12	11/19/10 8:18	0.85	16080	United Kingdom
6	C533807	84949	SILVER HANGING T-LIGHT HOLDER	-3	11/19/10 8:26	1.65	14496	United Kingdom
7	C533808	22767	TRIPLE PHOTO FRAME CORNICE	-1	11/19/10 8:27	9.95	13317	United Kingdom
8	C533808	22801	ANTIQUE GLASS PEDESTAL BOWL	-4	11/19/10 8:27	3.75	13317	United Kingdom
9	C533808	22065	CHRISTMAS PUDDING TRINKET POT	-4	11/19/10 8:27	1.45	13317	United Kingdom
10	C533809	22776	SWEETHEART CAKESTAND 3 TIER	-1	11/19/10 8:28	9.95	16145	United Kingdom

There are 8 columns in total. There is no columns is Primary key but Invoice and StockCode combined make Primary key

2. Data problems



The screenshot shows a SQL query window with the following text:

```
1 select top 10
2 *
3 from project1.dbo.[online retail data]
4 where country like '%,%'
```

Below the query window, the 'Results' tab is active, displaying a table with 10 rows and 9 columns. The columns are Invoice, StockCode, Description, Quantity, InvoiceDate, Price, Customer ID, and Country. The data is as follows:

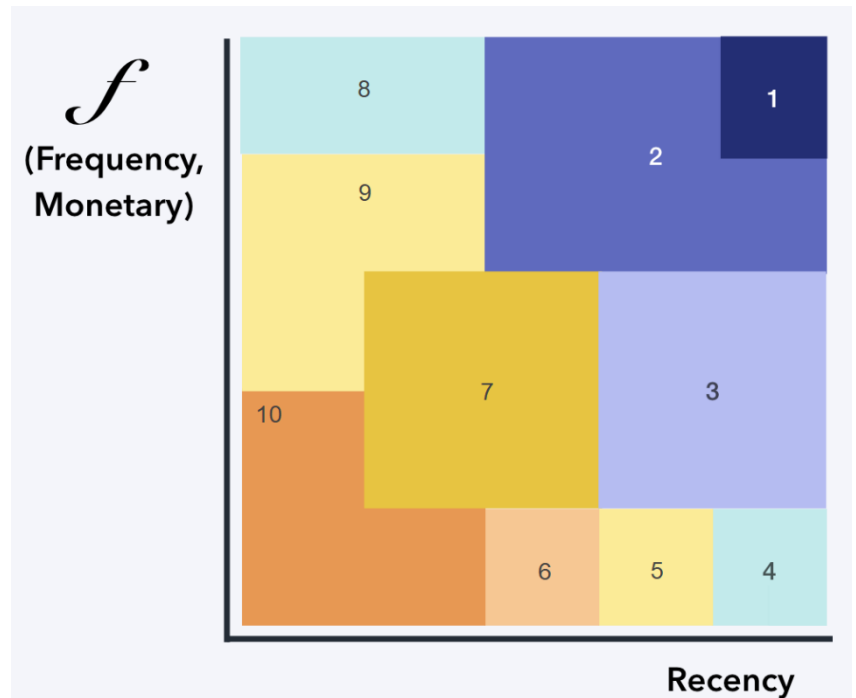
	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
1	533820	21507	"ELEPHANT BIRTHDAY CARD	"	12	11/19/10 9:34	0.42,13093,United Kingdom	
2	544586	22245	"HOOK 1 HANGER	MAGIC GARDEN"	1	2/21/11 15:04	0.85,17338,United Kingdom	
3	544586	21902	"KEY FOB FRONT DOOR "	6	2/21/11 15:04	0.65	17338,United Kingdom	
4	544586	21899	"KEY FOB GARAGE DESIGN"	3	2/21/11 15:04	0.65	17338,United Kingdom	
5	544586	22245	"HOOK 1 HANGER	MAGIC GARDEN"	3	2/21/11 15:04	0.85,17338,United Kingdom	
6	544591	21900	"KEY FOB SHED"	24	2/21/11 15:20	0.65	13044,United Kingdom	
7	544591	21899	"KEY FOB GARAGE DESIGN"	24	2/21/11 15:20	0.65	13044,United Kingdom	
8	544591	21901	"KEY FOB BACK DOOR "	24	2/21/11 15:20	0.65	13044,United Kingdom	
9	544591	21507	"ELEPHANT BIRTHDAY CARD	"	12	2/21/11 15:20	0.42,13044,United Kingdom	
10	544592	22760	"TRAY BREAKFAST IN BED"	6	2/21/11 15:30	10.95	16775,United Kingdom	

As we can see in the picture above, in some rows, parts of Description columns is in located in right columns, so the value of these columns also move to right, which make data come error.

3. Classification method

Customers which have purchase value ≤ 0 or product return count \geq purchase count is classify as Alert. Other customers are clasified as below:

1. Champion
2. Loyal Customer
3. Promising
4. New Customers
5. Warm Leads
6. Cold Leads
7. Need Attention
8. Should't Lose
9. Sleepers
10. Lost



4. Outcome

```

1 select top 10
2      *
3 from project1.dbo.Customer_classify

```

89 %

Results Messages

	Customer ID	Classify
1	18102	Champion
2	14646	Champion
3	14156	Champion
4	14911	Champion
5	17450	Champion
6	13694	Champion
7	17511	Champion
8	12415	Champion
9	16684	Champion
10	15061	Champion

Final result