Assumptions:

As there is no business user guide with a dataset so I assume that removing any necessary information would decrease the data enrichment.

I have utilized Unknown and 0 to the blank rows of Description and CustomerID columns respectively.

Finding:

Dataset Finding

Total Rows = 5 lacs.

Total Customer = 4373

Total Product = Almost 4000

Total Quantity Sales = 5386173

Total Invoicing = 25900

Total Country = 38

Unique Invoices = 25900

Unique Stockcode = 3958

Descriptive Analysis Finding

Highest Purchase (Item and Quantity) = WORLD WAR 2 GLIDERS ASSTD DESIGNS & 53847

Highest Purchase by Customer (ID and Quantity) = 14646 and 196719

Detailed Descriptive Analysis

Top 5 invoices contained different product items

	InvoiceNo	total_items_purchased	total_unique_item_purchased
1	573585	1114	1108
2	581219	749	748
3	581492	731	731
4	580729	721	720
5	558475	705	703

Let's find out those invoices that was cancelled or reduced in quantity after purchasing or may be due to machine or human error (top 10)

	InvoiceNo	total_purchase
1	573008	8050
2	556917	6331
3	538353	4571
4	546789	4180
5	562439	2989
6	563614	2559
7	559919	2164
8	576374	2094
9	541864	2044
10	562722	1833

Which stockcode items are purchased the most (top 10)

	StockCode	Description	total_purchased
1	84077	WORLD WAR 2 GLIDERS ASSTD DESI…	53847
2	85099B	JUMBO BAG RED RETROSPOT	47363
3	84879	ASSORTED COLOUR BIRD ORNAMENT	36413
4	22197	POPCORN HOLDER	36334
5	21212	PACK OF 72 RETROSPOT CAKE CASES	36135
6	85123a	WHITE HANGING HEART T-LIGHT HO…	35317
7	23084	RABBIT NIGHT LIGHT	30680
8	22492	MINI PAINT SET VINTAGE	26437
9	22616	PACK OF 12 LONDON TISSUES	26315
10	21977	PACK OF 60 PINK PAISLEY CAKE C…	24753

which stockcode items are returned the most (top 10)

	StockCode	Description	total_returned
1	23843	PAPER CRAFT , LITTLE BIRDIE	-80995
2	23166	MEDIUM CERAMIC TOP STORAGE JAR	-74494
3	84347	ROTATING SILVER ANGELS T-LIGHT HLDR	-9367
4	21108	FAIRY CAKE FLANNEL ASSORTED COLOUR	-3150
5	20971	PINK BLUE FELT CRAFT TRINKET BOX	-2617
6	85123A	WHITE HANGING HEART T-LIGHT HOLDER	-2578
7	21175	GIN + TONIC DIET METAL SIGN	-2030
8	22920	HERB MARKER BASIL	-1527
9	22273	FELTCRAFT DOLL MOLLY	-1447
10	47566B	TEA TIME PARTY BUNTING	-1424

which stock item are purchased with which items the most

	d1	d2	total_times_purch…
1	JUMBO BAG PINK POLKADOT	JUMBO BAG RED RETROSPOT	1010
2	GREEN REGENCY TEACUP AND SAUC…	PINK REGENCY TEACUP AND SAU…	913
3	JUMBO BAG RED RETROSPOT	JUMBO STORAGE BAG SUKI	900
4	GREEN REGENCY TEACUP AND SAUC…	ROSES REGENCY TEACUP AND SA…	887
5	JUMBO BAG RED RETROSPOT	JUMBO SHOPPER VINTAGE RED P…	856
6	LUNCH BAG BLACK SKULL.	LUNCH BAG RED RETROSPOT	721
7	ALARM CLOCK BAKELIKE GREEN	ALARM CLOCK BAKELIKE RED	718
8	JUMBO BAG BAROQUE BLACK WHITE	JUMBO BAG RED RETROSPOT	706
9	PINK REGENCY TEACUP AND SAUCER	ROSES REGENCY TEACUP AND SA…	705
10	JUMBO SHOPPER VINTAGE RED PAI…	JUMBO STORAGE BAG SUKI	695

Hypothesis: We can say that the top pairs are the most important like we can place this product nearer to each other for unintentional purchasing.

Monthly sales with running sale

	years	months	total_sales	running_sales
1	2010	12	346398	346398
2	2011	1	317948	664346
3	2011	2	283890	948236
4	2011	3	380418	1328654
5	2011	4	298536	1627190
6	2011	5	393592	2020782
7	2011	6	387560	2408342
8	2011	7	401525	2809867
9	2011	8	412690	3222557
10	2011	9	568013	3790570
11	2011	10	604503	4395073
12	2011	11	759625	5154698
13	2011	12	231475	5386173

which quarter was the best day in term of sales

	year	quarter	total_sales
1	2011	4	1595603
2	2011	3	1382228
3	2011	2	1079688
4	2011	1	982256
5	2010	4	346398

which month are concerning in terms of low sale

	years	months	total_sales	rank
1	2011	11	759625	1
2	2011	10	604503	2
3	2011	9	568013	3
4	2011	8	412690	4
5	2011	7	401525	5
6	2011	5	393592	6
7	2011	6	387560	7
8	2011	3	380418	8
9	2010	12	346398	9
10	2011	1	317948	10
11	2011	4	298536	11
12	2011	2	283890	12
13	2011	12	231475	13

which customer purchased the highest quantity

	CustomerID	total_purch…
1	14646	196719
2	12415	77242
3	14911	77180
4	17450	69029
5	18102	64122
6	17511	63012
7	13694	61803
8	14298	58021
9	14156	57025
10	16684	49390

Customer lifetime analysis

	Customer…	first_date	last_date	time_length
1	12433	2010-12-01 13:24:…	2011-12-09 10:02…	373
2	12662	2010-12-01 13:04:…	2011-12-09 11:59…	373
3	13777	2010-12-01 16:01:…	2011-12-09 12:25…	373
4	15311	2010-12-01 09:41:…	2011-12-09 12:00…	373
5	12748	2010-12-01 12:48:…	2011-12-09 12:20…	373
6	17924	2010-12-01 10:51:…	2011-12-08 19:28…	372
7	13408	2010-12-01 10:39:…	2011-12-08 09:05…	372
8	17841	2010-12-01 14:30:…	2011-12-08 12:07…	372
9	16210	2010-12-01 12:27:…	2011-12-08 12:36…	372
10	14911	2010-12-01 14:05:…	2011-12-08 15:54…	372

	CustomerID	total_quantity	total_invoice	quantity/invoice	time_length
1	14646	196719	77	2554	353
2	12415	77242	26	2970	313
3	14911	77180	248	311	372
4	17450	69029	55	1255	359
5	18102	64122	62	1034	367
6	17511	63012	46	1369	371
7	13694	61803	60	1030	370
8	14298	58021	45	1289	357
9	14156	57025	66	864	363
10	16684	49390	31	1593	354

Insights: it tells that which customer purchased the maximum quantity but we can also check with how many invoices, we can check quantity/invoice to not only focus on quantity like sometime high quantity purchased using single invoice that is sudden purchase, but those who have high no. of invoice with good quantity purchase, are important to consider as they are our regular customer.

Customer frequency among countries

	Country	total_customers
1	United Kingdom	3951
2	Germany	95
3	France	88
4	Spain	31
5	Belgium	25
6	Switzerland	22
7	Portugal	20
8	Italy	15
9	Finland	12
10	Austria	11

Hypothesis: This tells us that UK has broad customer frequency that allow to focus on culture specific marketing

Recommendation:

I think the data are very ambiguous in multiple ways like Stockcode(I believe it is the location of stock) but some products like discount, return, manual, etc) which are not normalized data values, should be there in a proper business context. Blank customerID tells that there are some machine errors or human errors that need to be worked on that side, training, and proper functional checkout machines.