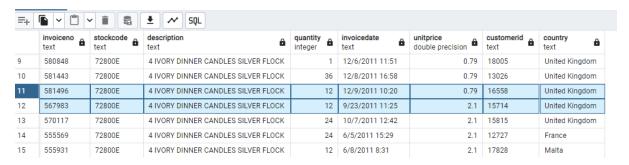# Exploratory Data Analysis (EDA)

1. Unit Price Changes Over Time

- Observation: Some stock codes show multiple prices even within the same month. This might be due to differences across countries.
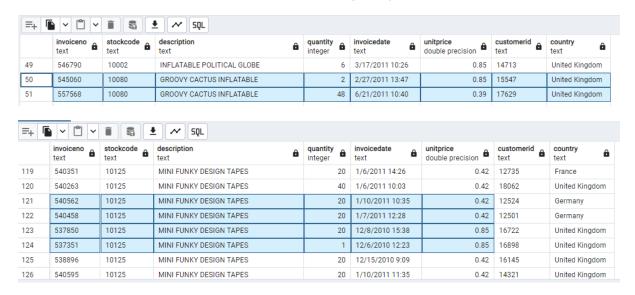
	invoiceno text	stockcode text	description text	quantity integer	invoicedate text	unitprice double precision	customerid text	country text
9	580848	72800E	4 IVORY DINNER CANDLES SILVER FLOCK	1	12/6/2011 11:51	0.79	18005	United Kingdom
10	581443	72800E	4 IVORY DINNER CANDLES SILVER FLOCK	36	12/8/2011 16:58	0.79	13026	United Kingdom
11	581496	72800E	4 IVORY DINNER CANDLES SILVER FLOCK	12	12/9/2011 10:20	0.79	16558	United Kingdom
12	567983	72800E	4 IVORY DINNER CANDLES SILVER FLOCK	12	9/23/2011 11:25	2.1	15714	United Kingdom
13	570117	72800E	4 IVORY DINNER CANDLES SILVER FLOCK	24	10/7/2011 12:42	2.1	15815	United Kingdom
14	555569	72800E	4 IVORY DINNER CANDLES SILVER FLOCK	24	6/5/2011 15:29	2.1	12727	France
15	555931	72800E	4 IVORY DINNER CANDLES SILVER FLOCK	12	6/8/2011 8:31	2.1	17828	Malta

2. Stock Codes Analysis:

- (a) Non-numeric Stock Codes:

  - Total: 35,809

  - Query: SELECT stockcode

    FROM Online_Retail

    WHERE stockcode !~ '^\d+$';
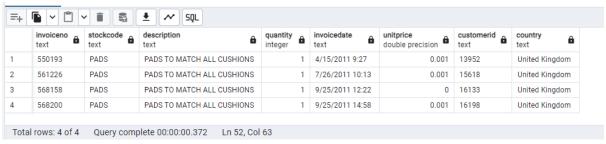
	invoiceno text	stockcode text	description text	quantity integer	invoicedate text	unitprice double precision	customerid text	country text
49	546790	10002	INFLATABLE POLITICAL GLOBE	6	3/17/2011 10:26	0.85	14713	United Kingdom
50	545060	10080	GROOVY CACTUS INFLATABLE	2	2/27/2011 13:47	0.85	15547	United Kingdom
51	557568	10080	GROOVY CACTUS INFLATABLE	48	6/21/2011 10:40	0.39	17629	United Kingdom

	invoiceno text	stockcode text	description text	quantity integer	invoicedate text	unitprice double precision	customerid text	country text
119	540351	10125	MINI FUNKY DESIGN TAPES	20	1/6/2011 14:26	0.42	12735	France
120	540263	10125	MINI FUNKY DESIGN TAPES	40	1/6/2011 10:03	0.42	18062	United Kingdom
121	540562	10125	MINI FUNKY DESIGN TAPES	20	1/10/2011 10:35	0.42	12524	Germany
122	540458	10125	MINI FUNKY DESIGN TAPES	20	1/7/2011 12:28	0.42	12501	Germany
123	537850	10125	MINI FUNKY DESIGN TAPES	20	12/8/2010 15:38	0.85	16722	United Kingdom
124	537351	10125	MINI FUNKY DESIGN TAPES	1	12/6/2010 12:23	0.85	16898	United Kingdom
125	538896	10125	MINI FUNKY DESIGN TAPES	20	12/15/2010 9:09	0.42	16145	United Kingdom
126	540595	10125	MINI FUNKY DESIGN TAPES	20	1/10/2011 11:35	0.42	14321	United Kingdom

- **(b) Numeric Stock Codes:**
  - Total: 371,020

- **(c) Special Codes:**
  - Codes like CRUK, D, DOT, C2, M, PADS, BANK CHARGES, and POST.
  - Query: SELECT DISTINCT(stockcode)[^3^][3]

    FROM Online_Retail

    WHERE stockcode ~ '^[A-Za-z]+$';
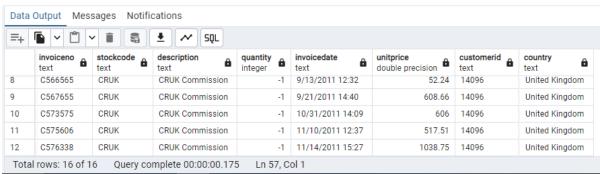
invoiceno text	stockcode text	description text	quantity integer	invoicedate text	unitprice double precision	customerid text	country text
536540	C2	CARRIAGE	1	12/1/2010 14:05	50	14911	EIRE
537368	C2	CARRIAGE	1	12/6/2010 12:40	50	14911	EIRE
537378	C2	CARRIAGE	1	12/6/2010 13:06	50	14911	EIRE
537963	C2	CARRIAGE	1	12/9/2010 11:30	50	13369	United Kingdom
538002	C2	CARRIAGE	1	12/9/2010 11:48	50	14932	Channel Islands
558908	C2	CARRIAGE	1	7/5/2011 9:06	50	13338	United Kingdom

	invoiceno text	stockcode text	description text	quantity integer	invoicedate text	unitprice double precision	customerid text	country text
17	537800	M	Manual	12	12/8/2010 13:08	0.19	12748	United Kingdom
18	537804	M	Manual	12	12/8/2010 13:17	2.95	12748	United Kingdom
19	537804	M	Manual	12	12/8/2010 13:17	0.19	12748	United Kingdom
20	546374	M	Manual	1	3/11/2011 12:05	10	17194	United Kingdom
21	C546907	M	Manual	-25	3/17/2011 18:46	0.67	16422	United Kingdom
22	C538124	M	Manual	-4	12/9/2010 15:43	0.5	15329	United Kingdom
23	538147	M	Manual	1	12/9/2010 16:11	15	13090	United Kingdom

	invoiceno text	stockcode text	description text	quantity integer	invoicedate text	unitprice double precision	customerid text	country text
1	550193	PADS	PADS TO MATCH ALL CUSHIONS	1	4/15/2011 9:27	0.001	13952	United Kingdom
2	561226	PADS	PADS TO MATCH ALL CUSHIONS	1	7/26/2011 10:13	0.001	15618	United Kingdom
3	568158	PADS	PADS TO MATCH ALL CUSHIONS	1	9/25/2011 12:22	0	16133	United Kingdom
4	568200	PADS	PADS TO MATCH ALL CUSHIONS	1	9/25/2011 14:58	0.001	16198	United Kingdom

Total rows: 4 of 4   Query complete 00:00:00.372   Ln 52, Col 63

	invoiceno text	stockcode text	description text	quantity integer	invoicedate text	unitprice double precision	customerid text	country text
1	536779	BANK CHARGES	Bank Charges	1	12/2/2010 15:08	15	15823	United Kingdom
2	541505	BANK CHARGES	Bank Charges	1	1/18/2011 15:58	15	15939	United Kingdom
3	549717	BANK CHARGES	Bank Charges	1	4/11/2011 14:56	15	14606	United Kingdom
4	551945	BANK CHARGES	Bank Charges	1	5/5/2011 11:09	15	16714	United Kingdom
5	565735	BANK CHARGES	Bank Charges	1	9/6/2011 12:25	15	16904	United Kingdom

Total rows: 12 of 12   Query complete 00:00:00.259   Ln 52, Col 1

	invoiceno text	stockcode text	description text	quantity integer	invoicedate text	unitprice double precision	customerid text	country text
77	540178	POST	POSTAGE	3	1/5/2011 12:42	18	12681	France
78	540239	POST	POSTAGE	4	1/5/2011 14:48	18	12682	France
79	540365	POST	POSTAGE	3	1/6/2011 16:12	18	12413	France
80	C540367	POST	POSTAGE	-3	1/6/2011 16:17	18	12413	France
81	540410	POST	POSTAGE	3	1/7/2011 10:04	18	12530	Germany

Total rows: 1000 of 1196   Query complete 00:00:00.319   Ln 52, Col 1

	invoiceno text	stockcode text	description text	quantity integer	invoicedate text	unitprice double precision	customerid text	country text
8	C566565	CRUK	CRUK Commission	-1	9/13/2011 12:32	52.24	14096	United Kingdom
9	C567655	CRUK	CRUK Commission	-1	9/21/2011 14:40	608.66	14096	United Kingdom
10	C573575	CRUK	CRUK Commission	-1	10/31/2011 14:09	606	14096	United Kingdom
11	C575606	CRUK	CRUK Commission	-1	11/10/2011 12:37	517.51	14096	United Kingdom
12	C576338	CRUK	CRUK Commission	-1	11/14/2011 15:27	1038.75	14096	United Kingdom

Total rows: 16 of 16   Query complete 00:00:00.175   Ln 57, Col 1

3. Description and Stock Code Relationships

- Stock Code with Multiple Descriptions:
    - Query: SELECT stockcode, COUNT(DISTINCT description)

        FROM Online_Retail

        GROUP BY stockcode

        HAVING COUNT(DISTINCT description) > 1;

	description text	count bigint
1	COLOURING PENCILS BROWN TUBE	2
2	COLUMBIAN CANDLE RECTANGLE	2
3	COLUMBIAN CANDLE ROUND	2
4	EAU DE NILE JEWELLED PHOTOFRAME	2
5	FRENCH FLORAL CUSHION COVER	2
6	FRENCH LATTICE CUSHION COVER	2
7	FROSTED WHITE BASE	2
8	METAL SIGN\,CUPCAKE SINGLE HOOK	3
9	PAPER LANTERN 9 POINT SNOW STAR	2
10	PINK FAIRY CAKE CUSHION COVER	2

Total rows: 19 of 19   Query complete 00:00:02.158

- Descriptions Linked to Multiple Stock Codes:

  - Query: SELECT description, COUNT(DISTINCT stockcode)

    FROM Online_Retail

    GROUP BY description

    HAVING COUNT(DISTINCT stockcode) > 1;

	stockcode text	count bigint
1	16156L	2
2	17107D	3
3	20622	2
4	20725	2
5	20914	2
6	21109	2
7	21112	2
8	21175	2
9	21232	2
10	21243	2

Total rows: 213 of 213   Query com

Data Cleaning Steps

1. Convert `invoicedate` to DateTime Format:

```
UPDATE Online_Retail
SET invoicedate = TO_TIMESTAMP(invoicedate, 'MM/DD/YYYY HH24:MI');
```

2. Handle Negative Quantities:

- Observation: A negative quantity indicates the order was canceled or returned

- Rows with quantity < 0 : {8,905} rows

3. Filter Out Records with `unitprice = 0 ':

- Observation: These are incorrect data entries.

- Rows: 40

4. Remove Customers Without IDs:

- Observation: There are 135,080 customers with no ID impacting on RFM customer segmentation.

5. Drop Special Stock Codes:

- Observation: Some stock codes have static unit prices and quantities, which might indicate they should be removed.

6. Clean Description Field:

- Observation: Some descriptions are stored with unnecessary spaces or special characters, such as '????'

- Distinct entries: 934 rows

- Query to clean descriptions: SELECT DISTINCT description

    FROM Online_Retail

    WHERE description LIKE ' %';

7. Handle Outliers in `CustomerID` for Total Sales:

- Observation: Some customer IDs are outliers in terms of their total spending.