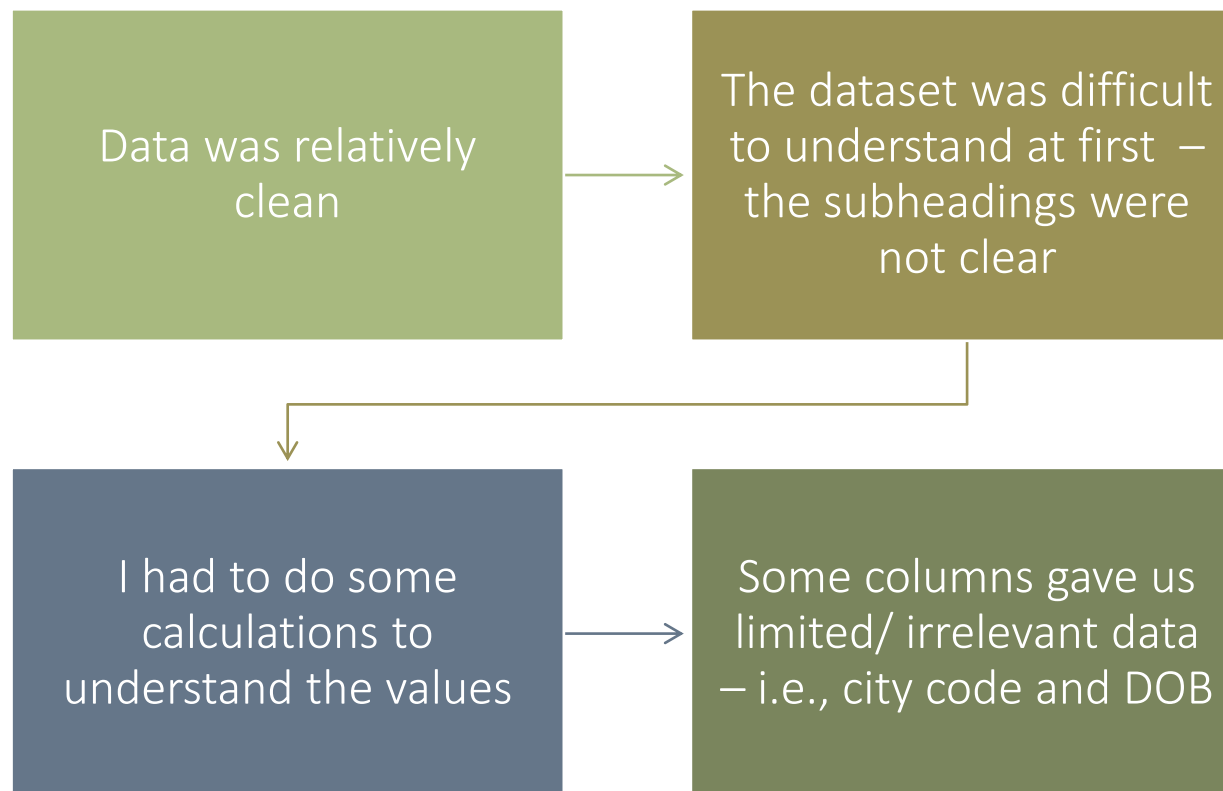# Retail Analysis: Customer Transactions

By Simisola Adebanjo

A little bit about me…

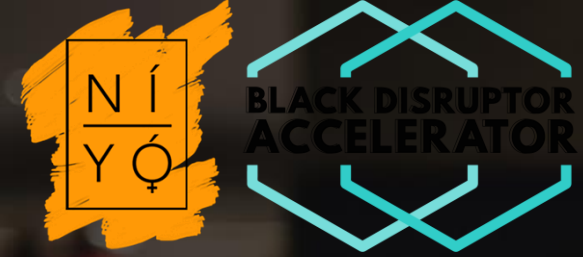❑I am Geography & Planning final year undergraduate

❑My interest in Data came when I launched my interior design business in March 2020

❑Many of my job roles have been in retail, and I have over 5 years of customer service experience

❑Part of my degree entails developing regeneration strategies for urban areas and there is a big push for retail regeneration in high streets espcially in Birmingham

❑Most recently I have taken on leadership roles as I love collaboration and sharing ideas

❑Fun fact: I enter the Jack Petchey Speak Up Competition in year 10 to work on my confidence and came runner-up!

At first glance…

Data was relatively clean

The dataset was difficult to understand at first – the subheadings were not clear

I had to do some calculations to understand the values

Some columns gave us limited/ irrelevant data – i.e., city code and DOB

Objective

"To query customer transaction data over a three-year trajectory across location and product category determining which store type performs the best and create dashboard visualizations of trends that support our findings"

Questions

1. "What is the best selling product?"

2. "Which store and product has the most returns?"

3. "What is the difference in the total amount spent by men and women?"

4. "What is the daily number of transactions per store and how much revenue is generated per transaction?"

Data Manipulation – Excel

USING SUMIF

`=SUMIF(J:J,L4,I:I)`

E	F	G	H	I	J	K	L	M
od_cat_ Qty		Rate	Tax	total_amt	Store_type		**Data Manipulation**	
1	-5	-772	£405.30	-£4,265.30	e-Shop		**USING SUMIF**	Total Amount sold/returned per store type
3	-5	-1497	£785.93	-£8,270.93	e-Shop		*Store_type*	*total_amt*
5	-2	-791	£166.11	-£1,748.11	TeleShop		e-shop	£19,824,828

USING COUNTIF

`=COUNTIF(E:F,L7)`

E	F	G	H	I	J	K	L	M
od_cat_ Qty		Rate	Tax	total_amt	Store_type		**Data Manipulation**	
1	-5	-772	£405.30	-£4,265.30	e-Shop		**USING SUMIF**	Total Amount sold/returned per store type
3	-5	-1497	£785.93	-£8,270.93	e-Shop		*Store_type*	*total_amt*
5	-2	-791	£166.11	-£1,748.11	TeleShop		e-shop	£19,824,828
6	-3	-1363	£429.35	-£4,518.35	e-Shop		**USING COUNTIF**	Total Qty per product_cat_code
5	-2	-791	£166.11	-£1,748.11	TeleShop		*product_cat_code*	*Qty*
3	-2	-824	£173.04	-£1,821.04	TeleShop		1	7174
6	-1	-1450	£152.25	-£1,602.25	e-Shop			

What is data validation?

is a feature that allows you to control the type of data entered into your worksheet.

USING SUMIF		Tota
Store_type		*tota*
e-shop	▼	
e-shop		
TeleShop		
MBR		
Flagship store		

Data Manipulation – Excel

USING DATEDIF

`=DATEDIF(B2,TODAY(), "y")`

Cell: C2

	A	B	C	D	E	F
1	customer_	DOB	Age	Gender	city_code	
2	268408	02/01/1970	52	M	4	
3	269696	07/01/1970	52	F	8	
4	268159	08/01/1970	52	F	8	
5	270181	10/01/1970	52	F	2	
6	268073	11/01/1970	52	M	1	

USING AVG

`=AVERAGE(C:C)`

D	E	F	G	H
ender	city_code			Data Manipulation
				Average Shopper Age
	4			
	8			41

Challenges and Troubleshooting

"Pound Signs and Blurred Lines"

-*One of my biggest challenges was that my aggregate functions were not working in SQL*

-*But why?*

-*After consulting Anna I learnt that...*

- *Having the '£' in the total_amt and tax column made the data type a text as opposed to an integer, so SQL wasn't able to recognise them as numbers to perform a calculation*

```
3      -- USING AGGREGATIONS TO SUMMARISE OUR DATASET
4  •   SELECT SUM(total_amt), prod_subcat_code
5      FROM data_analytics_project_niyo.transactions
6      GROUP BY prod_subcat_code
7      ORDER BY 1 DESC;
```

SUM(total_amt)	prod_subcat_code
0	1
0	5
0	6
0	11
0	8
0	12
0	3
0	7
0	4
0	9
0	10
0	2

Filter Rows: Export: Export: Wrap Cell Content: IA   Fetch rows:

at_code	prod_cat_code	Qty	Rate	Tax	total_amt	Store_type
1		-5	-772	£405.30	-£4,265.30	e-Shop
3		-5	-1497	£785.93	-£8,270.93	e-Shop
5		-2	-791	£166.11	-£1,748.11	TeleShop
6		-3	-1363	£429.35	-£4,518.35	e-Shop
5		-2	-791	£166.11	-£1,748.11	TeleShop
3		-2	-824	£173.04	-£1,821.04	TeleShop
6		-1	-1450	£152.25	-£1,602.25	e-Shop
6		-1	-1225	£128.63	-£1,353.63	TeleShop
1		-3	-908	£286.02	-£3,010.02	MBR
3		-4	-581	£244.02	-£2,568.02	e-Shop
3		5	1497	£785.93	£8,270.93	e-Shop
6		3	1360	£428.40	£4,508.40	e-Shop
5		4	587	£246.54	£2,594.54	e-Shop
2		-1	-611	£64.16	-£675.16	e-Shop

Data Analytics - SQL

USING GROUP BY TO FIND THE STORE WITH THE MOST RETURNS VS THE CATERGORY WITH THE MOST RETURNS

```sql
17 • SELECT SUM(Qty), Store_type
18   FROM data_analytics_project_niyo.transactions_v2
19   WHERE Qty < '0'
20   GROUP BY Store_type;
```

	SUM(Qty)	Store_type
▶	-2672	e-Shop
	-1241	TeleShop
	-1387	MBR
	-1327	Flagship store

Result Grid | Filter Rows: | Export: | Wrap Cell Cont

```sql
22 • SELECT prod_cat, SUM(Qty)
23   FROM data_analytics_project_niyo.transactions_v2 table1
24   INNER JOIN data_analytics_project_niyo.prod_cat_info table2
25   ON table1.prod_cat_code = table2.prod_cat_code
26   WHERE Qty < '0'
27   GROUP BY prod_cat
28   ORDER BY 2;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: IA

	prod_cat	SUM(Qty)
▶	Books	-10446
	Electronics	-6250
	Home and kitchen	-4980
	Clothing	-2682
	Footwear	-2565
	Bags	-1284

Data Analytics - SQL

USING WHERE TO COMPARE THE NUMBER OF PURCHASE MADE BY MEN VS WOMEN

```
32  SELECT COUNT(transaction_id)
33  FROM data_analytics_project_niyo.transactions_v2 table1
34  INNER JOIN data_analytics_project_niyo.customer table2
35  ON table1.cust_id = table2.customer_Id
36  WHERE Gender = 'M' AND Qty > '0';
```

COUNT(transaction_id)
10706

```
38  SELECT COUNT(transaction_id)
39  FROM data_analytics_project_niyo.transactions_v2 table1
40  INNER JOIN data_analytics_project_niyo.customer table2
41  ON table1.cust_id = table2.customer_Id
42  WHERE Gender = 'F' AND Qty > '0';
```

COUNT(transaction_id)
10154

Data Analytics - SQL

USING CASE WHEN TO ANALYSE THE DAILY TRANSACTION RESULTS ACROSS ALL STORES

```
46 •   SELECT tran_date, SUM(total_amt),
47 ⊖   CASE
48     WHEN SUM(total_amt) < 0 THEN 'LOSS'
49     ELSE 'PROFIT'
50     END AS 'Daily Performance Across All Stores'
51     FROM data_analytics_project_niyo.transactions_v2
52     GROUP BY tran_date;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tran_date	SUM(total_amt)	Daily Performance Across All Stores
28/08/2014	5695	PROFIT
27/05/2014	-4205	LOSS
24/04/2014	1603	PROFIT
24/11/2014	-860	LOSS
23/10/2014	-1244	LOSS
23/07/2014	-1217	LOSS
22/07/2014	1117	PROFIT
22/01/2014	-1759	LOSS
21/08/2014	2461	PROFIT
20/11/2014	106266	PROFIT
20/12/2014	87317	PROFIT
20/02/2014	115272	PROFIT
20/09/2014	113703	PROFIT

```
46 •   SELECT tran_date, SUM(total_amt),
47 ⊖   CASE
48     WHEN SUM(total_amt) < 0 THEN 'LOSS'
49     ELSE 'PROFIT'
50     END AS 'Daily Performance Across All Stores'
51     FROM data_analytics_project_niyo.transactions_v2
52     GROUP BY tran_date
53     ORDER BY 2 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell

tran_date	SUM(total_amt)	Daily Performance Across All Stores
20/07/2013	672288	PROFIT
20/10/2012	623933	PROFIT
20/05/2013	616231	PROFIT
20/11/2012	606334	PROFIT
20/09/2012	598933	PROFIT
20/04/2013	596400	PROFIT
20/10/2011	594715	PROFIT
20/06/2012	589854	PROFIT
20/08/2013	586797	PROFIT
20/12/2011	585571	PROFIT
20/08/2011	581649	PROFIT
20/08/2012	579599	PROFIT

Result 143

Data Analytics - SQL

USING INNER JOINS TO ANSWER *"WHAT IS THE BEST SELLING PRODUCT?"*

```sql
65    SELECT SUM(Qty), prod_subcat, prod_cat
66    FROM data_analytics_project_niyo.transactions_v2 table1
67    INNER JOIN data_analytics_project_niyo.prod_cat_info table
68    WHERE table1.prod_cat_code = table2.prod_cat_code
69    GROUP BY prod_sub_cat_code
70    ORDER BY 1 DESC;
71
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

SUM(Qty)	prod_subcat	prod_cat
36918	Audio and video	Electronics
31464	Mens	Clothing
29117	Kids	Clothing
24610	Children	Books
24610	Academic	Books
19156	Women	Clothing
14669	DIY	Books
14669	Fiction	Books
12308	Cameras	Electronics
12308	Personal Appliances	Electronics
12308	Computers	Electronics
9941	Furnishing	Home and...

Data Visualisation – Power BI
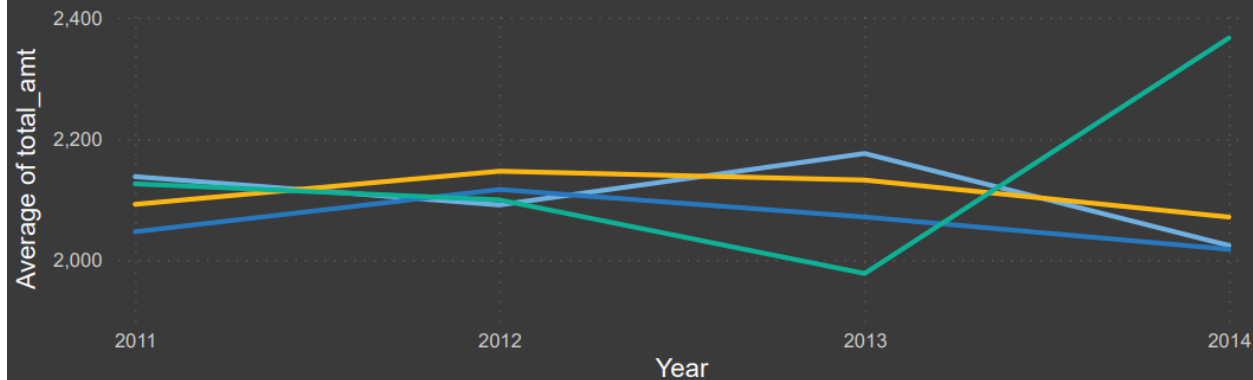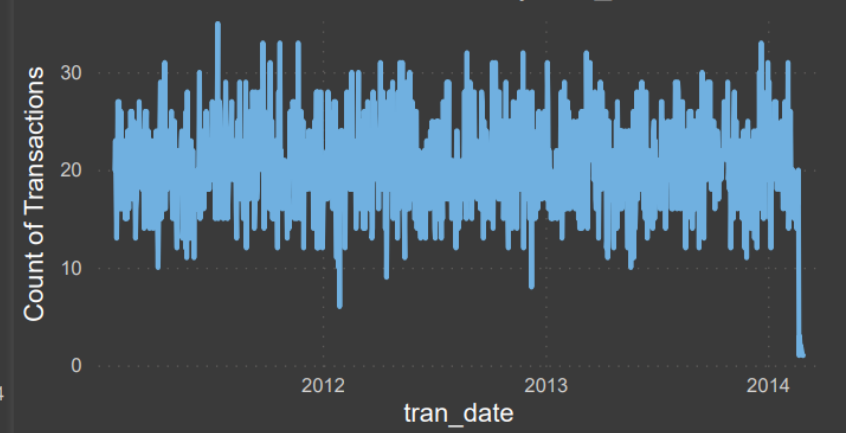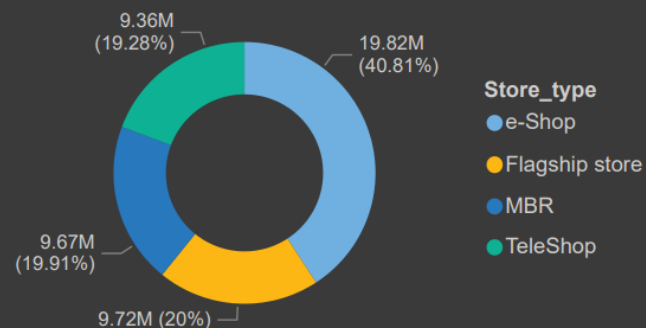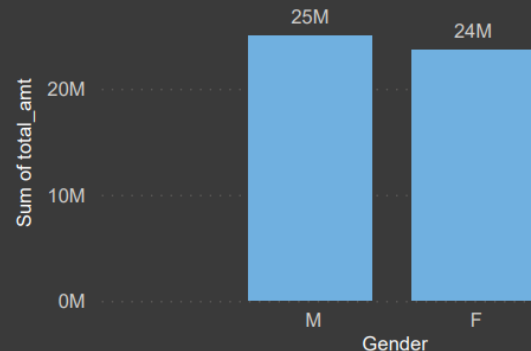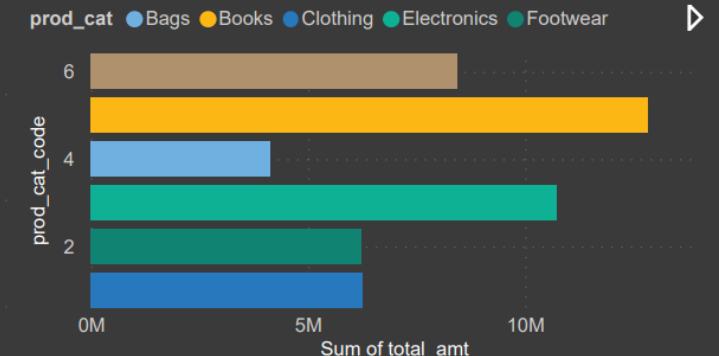
Business Strategy Recommendations

❑Loyalty rewards for customers who order frequently i.e. a discounted annual next day delivery promotion

❑Flash sale promotion for books

❑Till placement strategy on books

❑Continue to collect MORE data over a longer period time

❑Seasonal stocking of products

What I have learnt

Education

-I now understand a whole new language and can communicate with a computer and within the data world using SQL and Python!

-Learned that my interior design skills are transferable in creating appealing visualisations in Tableau and PowerBI

-Applied an agile approach to completing my final project constantly testing and reviewing my code

Personally

- That I am resilient and work well under pressure, for the past two months I have balanced bootcamp with my university deadlines and two part time jobs
- I have learnt the true value of sisterhood and great mentorships
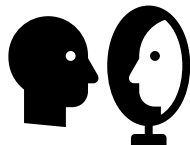
Career Wise

- I feel so much more confident in my skills to enter the tech industry as data analyst
- Faith is the key to success: I am currently receiving multiple interview offers as a data analyst in leading consultancy companies