

BI Final Project Pakistan's Largest Ecommerce Dataset Analysis on Power BI

BI EXECUTION STRATEGY

SYED MUHAMMAD NAQI RAZA RIZVI - 18611

Data: [Pakistan's Largest E-Commerce Dataset | Kaggle](#)

Data wrangling:

Pakistan Largest Ecommerce Dataset.csv

File Origin: 1252: Western European (Windows) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

item_id	status	created_at	sku	price	qty_ordered	grand_total	increment_id	category_name_1	sales_commission_code	discount_amount	payment_method
211131	complete	7/1/2016	kreations_YI 06-L	1950	1	1950	100147443	Women's Fashion	\N	0	cod
211133	canceled	7/1/2016	kcc_Buy 2 Frey Air Freshener & Get 1 Kasual Body Spra...	240	1	240	100147444	Beauty & Grooming	\N	0	cod
211134	canceled	7/1/2016	Ego_UP0017-999-MR0	2450	1	2450	100147445	Women's Fashion	\N	0	cod
211135	complete	7/1/2016	kcc_krone deal	360	1	60	100147446	Beauty & Grooming	R-FSD-52352	300	cod
211136	order_refunded	7/1/2016	BK7010400AG	555	2	1110	100147447	Soghaat	\N	0	cod
211137	canceled	7/1/2016	UK_Namkino All In One 200 Gms	80	1	80	100147448	Soghaat	\N	0	cod
211138	complete	7/1/2016	kcc_krone deal	360	1	60	100147449	Beauty & Grooming	\N	300	cod
211139	complete	7/1/2016	UK_Namkino Mix Nimco 400 Gms	170	1	170	100147450	Soghaat	\N	0	cod
211140	canceled	7/1/2016	Apple iPhone 6S 64GB	96499	1	96499	100147451	Mobiles & Tablets	\N	0	ubilcreditcard
211141	canceled	7/1/2016	Apple iPhone 6S 64GB	96499	1	96499	100147452	Mobiles & Tablets	\N	0	mygateway
211142	complete	7/1/2016	GFC_Pedestal Myga Cross Base (Special Guard) 24"	5500	1	5500	100147453	Appliances	\N	0	cod
211143	received	7/1/2016	BK1070200PL	210	1	366	100147454	Soghaat	\N	0	cod
211144	received	7/1/2016	BK1130200CF	156	1	366	100147454	Soghaat	\N	0	cod
211145	complete	7/1/2016	kcc_Sultanat	120	1	120	100147455	Home & Living	105259	0	ubilcreditcard
211146	complete	7/1/2016	kcc_glamour deal	320	1	0	100147456	Beauty & Grooming	\N	0	customercredit
211147	canceled	7/1/2016	Assetmen_MD-346-M	1550	1	1550	100147457	Men's Fashion	105259	0	ubilcreditcard
211149	complete	7/1/2016	cr_DATES WITH CASHEW-400 GM	420	1	1270	100147458	Soghaat	R-KHW-104406	0	cod
211150	complete	7/1/2016	UK_Gift Box Mix Dry Fruit Sweets 500 Gms	360	1	1270	100147458	Soghaat	R-KHW-104406	0	cod
211151	complete	7/1/2016	itter_AB 1199	490	1	1270	100147458	Beauty & Grooming	R-KHW-104406	0	cod
211152	canceled	7/1/2016	RL_B005	899.25	1	2118.25	100147459	Home & Living	\N	0	cod

The data in the preview has been truncated due to size limits.

Load Transform Data Cancel

Go to excel > data > load data text/csv and then set it as above.

Then click on transform data

Table.TransformColumnTypes(#"Promoted Headers",{"item_id", Int64.Type}, {"status", type text}, {"created_at", ...

123 item_id	123 status	123 created_at	123 sku	123 price
1	211131	complete	7/1/2016 kreations_YI 06-L	
2	211133	canceled	7/1/2016 kcc_Buy 2 Frey Air Freshener & Get 1 Kasual Body Spray Free	
3	211134	canceled	7/1/2016 Ego_UP0017-999-MR0	
4	211135	complete	7/1/2016 kcc_krone deal	
5	211136	order_refunded	7/1/2016 BK7010400AG	
6	211137	canceled	7/1/2016 UK_Namkino All In One 200 Gms	
7	211138	complete	7/1/2016 kcc_krone deal	
8	211139	complete	7/1/2016 UK_Namkino Mix Nimco 400 Gms	
9	211140	canceled	7/1/2016 Apple iPhone 6S 64GB	
10	211141	canceled	7/1/2016 Apple iPhone 6S 64GB	
11	211142	complete	7/1/2016 GFC_Pedestal Myga Cross Base (Special Guard) 24"	
12	211143	received	7/1/2016 BK1070200PL	
13	211144	received	7/1/2016 BK1130200CF	
14	211145	complete	7/1/2016 kcc_Sultanat	
15	211146	complete	7/1/2016 kcc_glamour deal	
16	211147	canceled	7/1/2016 Assetmen_MD-346-M	
17	211149	complete	7/1/2016 cr_DATES WITH CASHEW-400 GM	
18	211150	complete	7/1/2016 UK_Gift Box Mix Dry Fruit Sweets 500 Gms	
19	211151	complete	7/1/2016 itter_AB 1199	
20	211152	canceled	7/1/2016 RL_B005	
21	211153	canceled	7/1/2016 bed&rest_S7	
22	211154	canceled	7/1/2016 L&L_LLHLE8224S	
23	211155	complete	7/1/2016 J&J_JJR-4	
24				

Query Settings

PROPERTIES

Name: Pakistan Largest Ecommerce Dataset

APPLIED STEPS

Source

Promoted Headers

Changed Type

For every column check its data type and change it accordingly

Now **reduce some rows**

By removing duplicates and removing errors

Query Editor

Choose Columns Remove Columns Manage Columns

Keep Rows Remove Rows Reduce Rows

Sort

Split Column Group By Transform

Data Type: Text Use First Row as Headers Replace Values

Merge Queries Append Queries Combine Files Combine

Manage Parameters Parameters

Data source settings Data Sources

New Source Recent Sources Enter Data New Query

Query Settings

PROPERTIES

Name

Pakistan Largest Ecommerce Dataset

APPLIED STEPS

Source

Promoted Headers

Changed Type

Removed Duplicates

Removed Errors

Table.TransformColumnTypes({"Promoted Headers",{"item_id", Int64.Type}, {"status", type text}, {"created_at",

item_id	status	created_at	sku	price
211131	complete	7/1/2016	kcreations_YI 06-L	
211133	canceled	7/1/2016	kcc_Buy 2 Frey Air Freshener & Get 1 Kasual Body Spray Free	
211134	canceled	7/1/2016	Ego_UP0017-999-MR0	
211135	complete	7/1/2016	kcc_krone deal	
211136	order_refunded	7/1/2016	BK7010400AG	
211137	canceled	7/1/2016	UK_Namkino All In One 200 Gms	
211138	complete	7/1/2016	kcc_krone deal	
211139	complete	7/1/2016	UK_Namkino Mix Nimco 400 Gms	
211140	canceled	7/1/2016	Apple iPhone 6S 64GB	
211141	canceled	7/1/2016	Apple iPhone 6S 64GB	
211142	complete	7/1/2016	GFC_Pedestal Myga Cross Base (Special Guard) 24"	
211143	received	7/1/2016	BK1070200PL	
211144	received	7/1/2016	BK1130200CF	
211145	complete	7/1/2016	kcc_Sultanat	
211146	complete	7/1/2016	kcc_glamour deal	
211147	canceled	7/1/2016	Assetmen_MD-346-M	
211149	complete	7/1/2016	cr_DATES WITH CASHEW-400 GM	
211150	complete	7/1/2016	UK_Gift Box Mix Dry Fruit Sweets 500 Gms	
211151	complete	7/1/2016	itter_AB 1199	
211152	canceled	7/1/2016	RL_B005	
211153	canceled	7/1/2016	bed&rest_S7	
211154	canceled	7/1/2016	L&L_LLHLE8224S	
211155	complete	7/1/2016	J&J_JJR-4	

Profile based on top 1000 rows

Go to remove rows window and click on remove duplicates and remove errors

Query Editor

Transpose Reverse Rows Count Rows

Data Type: Any Detect Data Type Rename

Replace Values Fill Pivot Column

Unpivot Columns Move Convert to List

Split Column Format Text Column

Merge Columns Extract Parse

Statistics Standard Scientific

Trigonometry Rounding Information

Date Time Duration Date & Time Column

Expand Aggregate Extract Values Create Data Type

Structured Column

Query Settings

PROPERTIES

Name

Pakistan Largest Ecommerce Dataset

APPLIED STEPS

Source

Promoted Headers

Changed Type

Removed Duplicates

Removed Errors

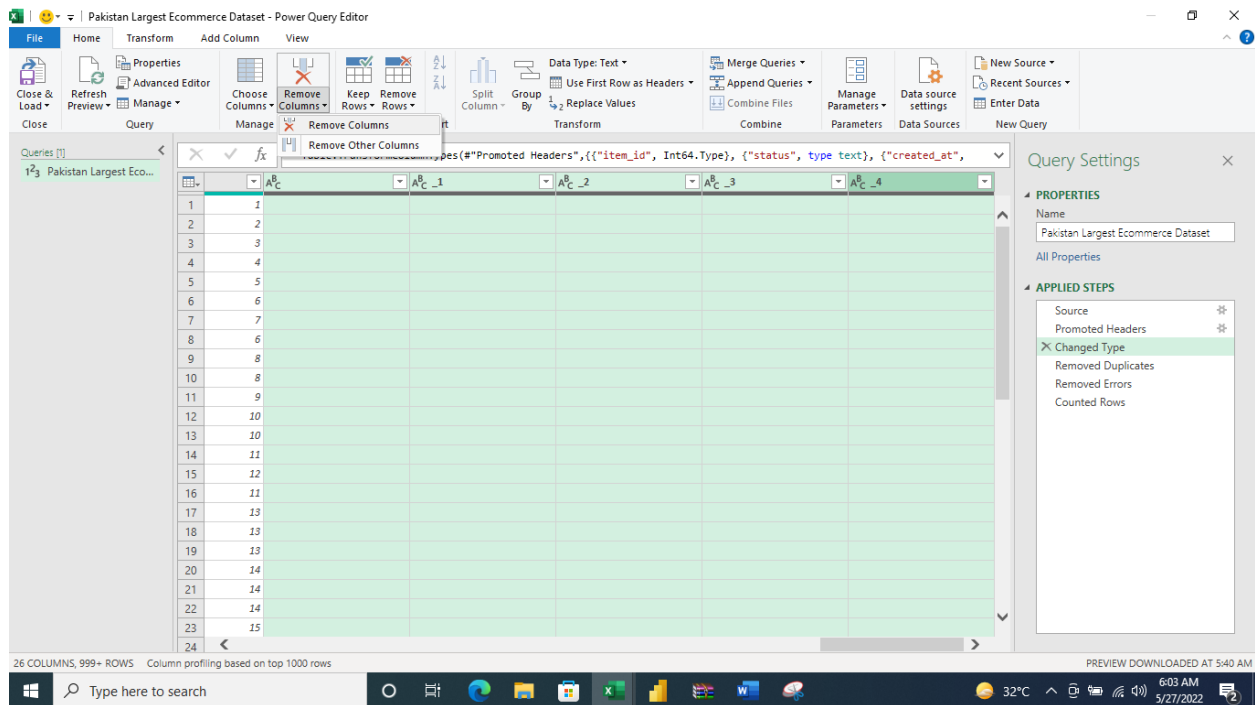
Counted Rows

Table.RowCount("#Removed Errors")

20

Go to transform > count rows and you see 20 rows have been removed.

Now remove some columns



all these selected columns have no values : all are null so remove them all

Now use group by to check null values in all columns and fill with mode values

`= Table.Group(#"Removed Columns", {"BI Status"}, {"Count", each Table.RowCount(_, Int64.Type)})`

	BI Status	Count
1	#REF!	1
2	Gross	201452
3	Net	234173
4	Valid	148887
5		464063
6	7	1
7	3	2
8	4	3
9	5	4
10	8	1

Remove #ref row and replace null value with mode

×

Replace Values

Replace one value with another in the selected columns.

Value To Find

Replace With

Net

▶ Advanced options

OK

Cancel

Now for status column so many empty values replace them with mode : completed garbage values like 2660,999,\N,2800 remove them all

×

✓

fx

= Table.Group(#"Filtered Rows", {"status"}, {"Count", each Table.RowCount(_), Int64.Type}))

	A ^B C status	1 ² 3 Count
1	canceled	201249
2	complete	233684
3	order_refunded	59529
4	received	77290
5	refund	8050
6	closed	494
7	2660	1
8	fraud	10
9	holded	31
10	exchange	4
11	pending_paypal	7
12	999	9
13	paid	1159
14	\N	4
15		464067
16	2800	1
17	cod	2859
18	pending	48
19	processing	33
20	payment_review	57

Filter garbage value:

✕ ✓ fx = Table.ReplaceValue("#Filtered Rows","", "complete", Replacer.

123 item_id	A ^B status	created_at	A ^B
		7/1/2016	kc
		7/1/2016	Eg
		7/1/2016	kc
		7/1/2016	Bh
		7/1/2016	UI
		7/1/2016	kc
		7/1/2016	UI
		7/1/2016	As
		7/1/2016	As
		7/1/2016	GI
		7/1/2016	Bh
		7/1/2016	Bh
		7/1/2016	kc
		7/1/2016	kc
		7/1/2016	As
		7/1/2016	cr
		7/1/2016	UI
		7/1/2016	itt
		7/1/2016	RL
		7/1/2016	be
		7/1/2016	L8
		7/1/2016	J8
		7/1/2016	J8

Sort Ascending

Sort Descending

Clear Sort

Clear Filter

Remove Empty

Text Filters

Search

☒ (Select All)

☐ 2660

☐ 2800

☐ 999

☒ canceled

☒ closed

☒ cod

☒ complete

☒ exchange

☒ fraud

☒ holded

☒ order_refunded

☒ paid

☒ payment_review

OK Cancel

Now replace null values with mode for sku:

created_at	sku	price	qty_ord
7/1/2016	cr_DATES WITH CASHEW-400 GM	420	
7/1/2016	UK_Gift Box Mix Dry Fruit Sweets 500 Gms	360	
7/1/2016	itter_AB 1199	490	
7/1/2016	RL_B005	899.25	
7/1/2016	bed&rest_S7	899	
7/1/2016	L&L_LLHLE8224S	320	
7/1/2016	J&J_JJR-4	149	
7/1/2016	J&J_JJR-20	149	

Replace Values

Replace one value with another in the selected columns.

Value To Find

Replace With

Advanced options

OK Cancel

Now replace null values of category name with mode Mobiles & Tablets of count 115710 (as seen in group by)

```
Table.ReplaceValue("#Filtered Rows1","", "MATSAM59DB75ADB2F80", Replacer.ReplaceValue, {"sku"})
```

1.2 grand_total	1.2 increment_id	A ^B C category_name_1	A ^B C sales_commission_code	1.2 discount
1270	100147458	Soghaat	R-KHW-104406	
1270	100147458	Soghaat	R-KHW-104406	
1270	100147458	Beauty & Grooming	R-KHW-104406	
2118.25	100147459	Home & Living	\N	
2118.25	100147459	Home & Living	\N	
2118.25	100147459	Home & Living	\N	
298	100147460	Kids & Baby	\N	
298	100147460	Kids & Baby	\N	

Replace null values in payment method with mode : cod which came 271960 times

Now for FY:

Now for FY:

Power Query Editor interface showing the 'Replace Values' dialog box.

Formulas Bar: `= Table.ReplaceValue("#Filtered Rows3",null,6348.8929917641672,Replacer.ReplaceValue,{"price"})`

Table Columns: 1.2 price, 1.2 Count

Table Data:

1.2 price	1.2 Count
6348.892992	464063

Replace Values Dialog:

Replace one value with another in the selected columns.

Value To Find:

Replace With:

Buttons: OK, Cancel

Power Query Editor interface showing the 'Replace Values' dialog box over a larger table.

Formulas Bar: `= Table.ReplaceValue("#Replaced Value3","", "cod",Replacer.ReplaceValue,{"payment_method"})`

Table Columns: 1.2 price, 1.2 qty_ordered, 1.2 grand_total, 1.2 increment_id

Table Data:

1.2 price	1.2 qty_ordered	1.2 grand_total	1.2 increment_id
240	1	240	100147444
2450	1	2450	100147445
360	1	60	100147446
555	2	1110	100147447
7448			7448
7449			7449
7450			7450
7451			7451
7452			7452
7453			7453
7454			7454
7455			7455
7456			7456
7457			7457
7458			7458
360	1	1270	100147458
490	1	1270	100147458
899.25	1	2118.25	100147459
899	1	2118.25	100147459
320	1	2118.25	100147459
149	1	298	100147460
149	1	298	100147460

Replace Values Dialog:

Replace one value with another in the selected columns.

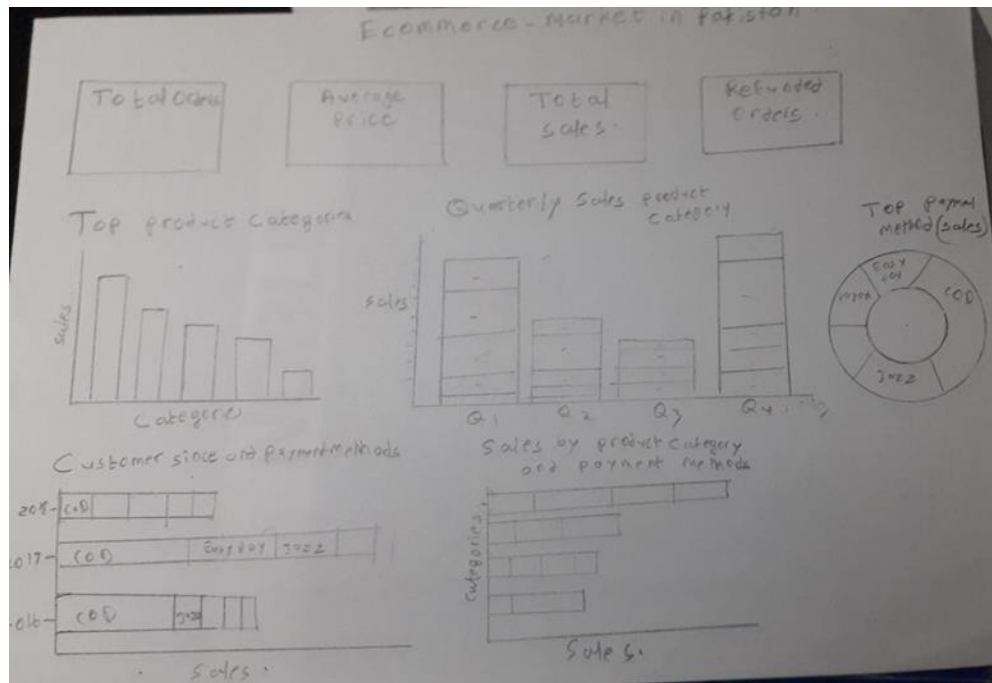
Value To Find:

Replace With:

Buttons: OK, Cancel

Step 2:

You are supposed to draw a rough diagram on paper in which one or more chart is answering each of your queries. Show images of these papers. Remember this is the first step and the bottom-up approach.



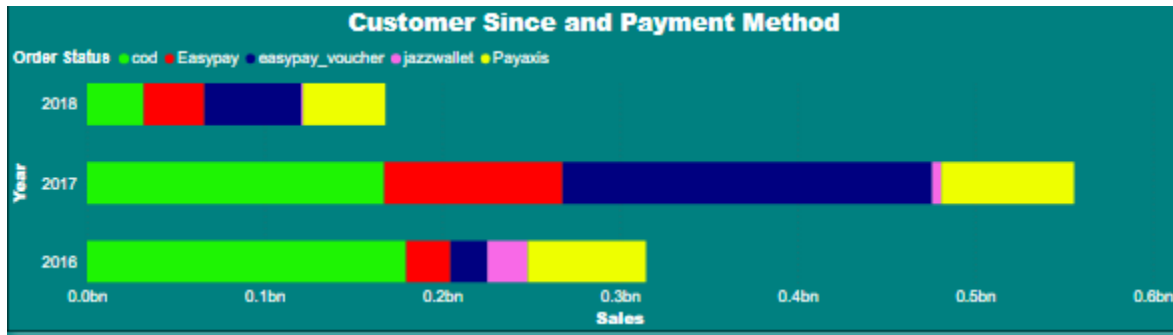
Step 3: Think about how you will present your story based on the charts in Step 2. Story can be presented across time typically (if you are able to find temporal patterns through analysis in Step 4). Right now, make a rough plan and show it. If you come up with any dashboards along the way, i.e., a dashboard for each story point as in Tableau, then pen down the contents of those dashboards as well.

Answer:

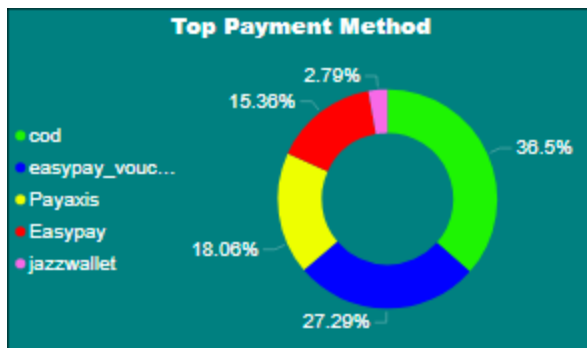
1. The bar chart tells us the top product categories according to sale
2. The score cards helps us understand the average price and informs us the refunded orders
3. We can see that the quarterly charts help us understand the sales of each product in each quarter to understand the trend of sales and where Owner needs to be prepared to keep inventory.
4. The factor of time and how long a customer is using the service shows us the impact of the payment methods, the longer the customer have stayed what payment method is preferred.
5. The sales by product category and payment method, help us identify which payment

method, a particular prefers in dealing.

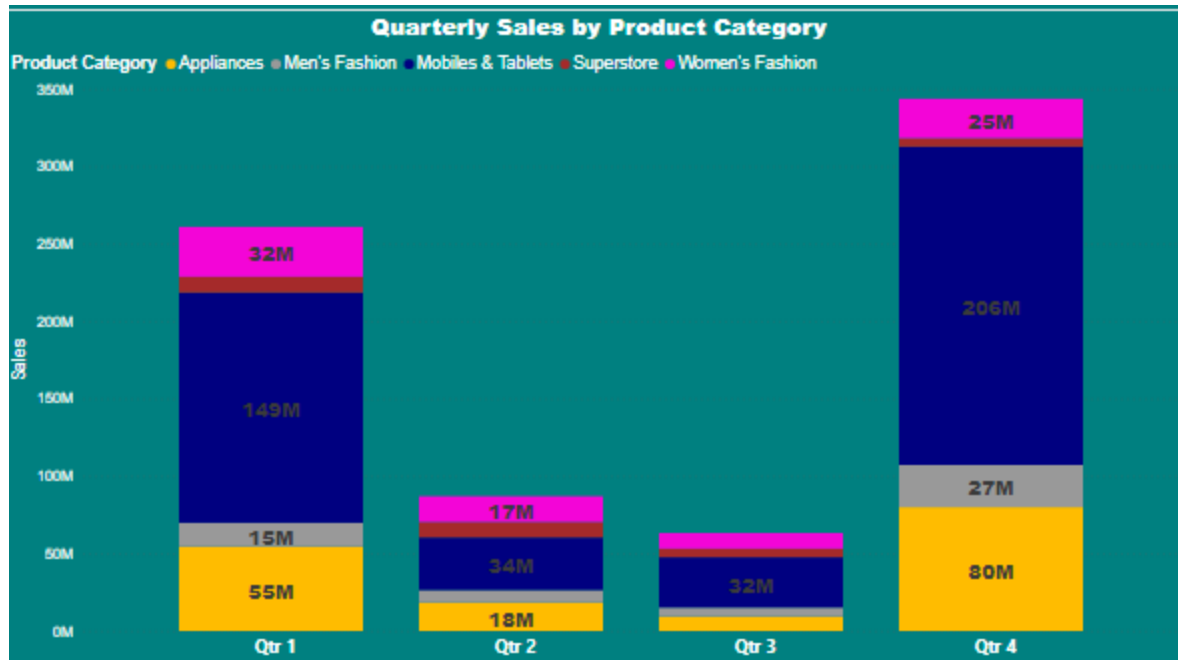
Step 4: Replicate Step 2 in the tool. Now, apply top-down approach to extract more information from these basic charts, e.g., by adding more dimensions in the chart attributes in Power BI, Qlik etc., or using the central pane in Tableau (which presents the dimensional analysis framework).



Sales made per year with legends of cash on delivery or online payment methods

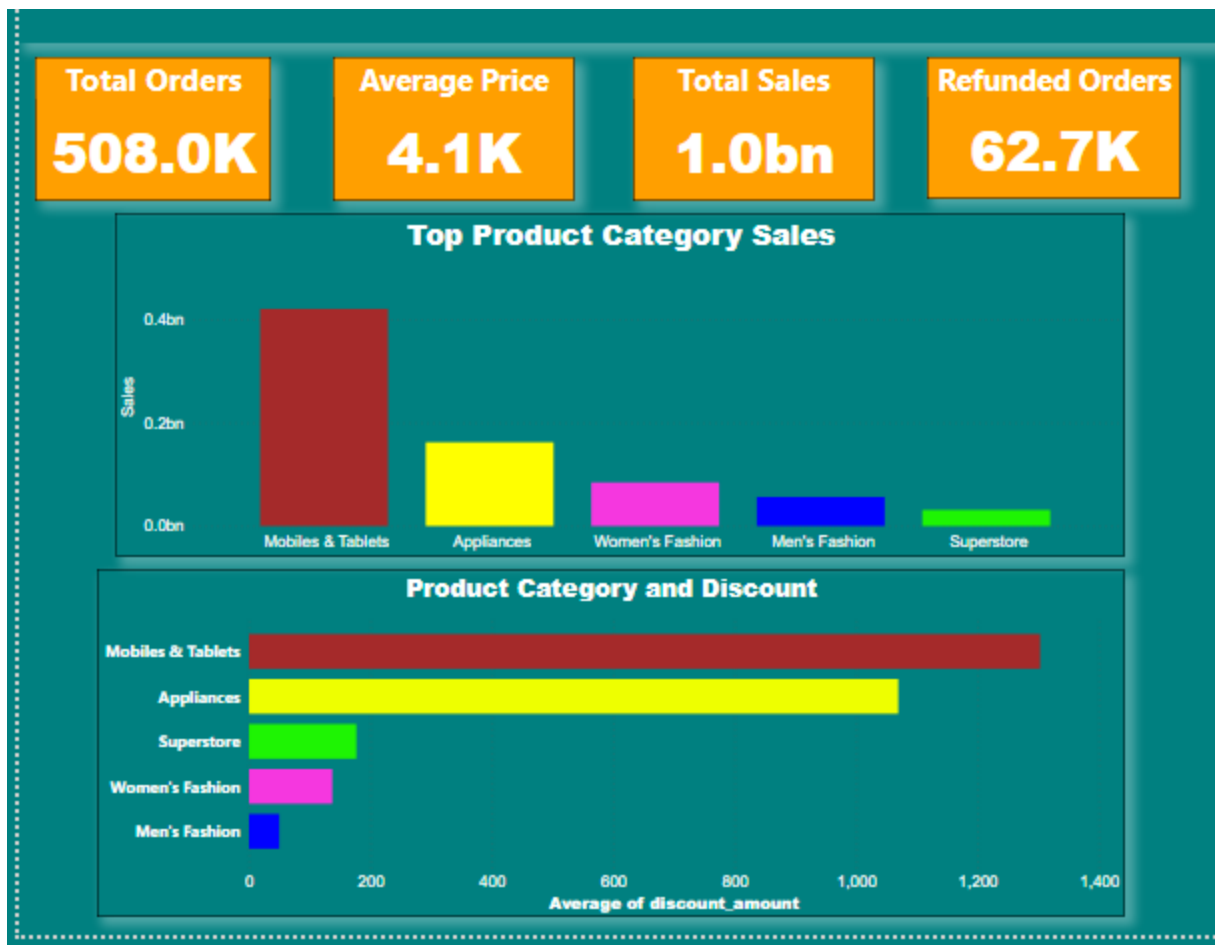


Sales made per payment methods



Sales made per quarter for each category

Step 5: Now develop the dashboards by combining different charts if needed, along with scorecards etc. (don't forget the bullet charts are better)



Total Orders

Overall Average of price is currently at 4120.33, and Average of price for category: Entertainment is at 17953.52 .

Overall Sales is currently at 1038136060.7, and Sales for category: Mobiles & Tablets is significantly higher than other segments at 420694791.26.

Refunded Orders

At 16,299, Count of refunds for Men's Fashion is unusually high.

Top Product Category Sales

The dashboard shows us that Mobiles and Tablets are the most sold category and then appliances.

At 420,694,791.26, Mobiles & Tablets had the highest Sales and was 1,249.06% higher than Superstore, which had the lowest Sales at 31,184,272.50

Mobiles & Tablets accounted for 55.72% of Sales.

Product Category and Discount

It can also be seen that the reason behind most of its sales could be the high discount rate on mobile phones and tablets with average amount of discount is approx. 1,400.

At 1,303.10, Mobiles & Tablets had the highest Average of discount_amount and was 2,520.28% higher than Men's Fashion, which had the lowest Average of discount_amount at 49.73.

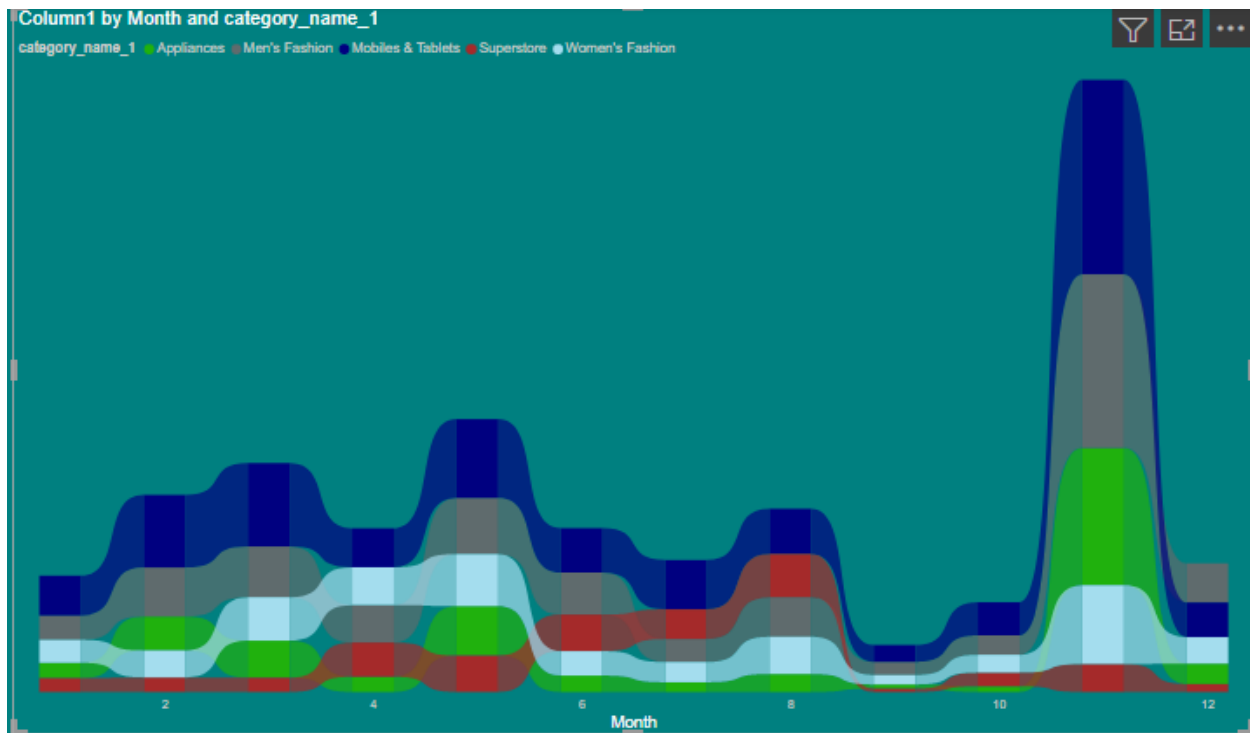
Across all 5 category_name_1, Average of discount_amount ranged from 49.73 to 1,303.10

Second is appliances which is about 1,200 average discount.

Least is on Fashion Products which is just in hundreds

Average Price

We also find out that the average price of mobile phone and tablet is 11k which is more than the given average among all categories



Ribbon chart shows in November orders shoot up reason is high demand of Mobiles and Tablets

Count of orders for category: Men's Fashion was unusually high (29916), which may have lifted Count of orders for November. Also appliances also came in demand in November after being low throughout.

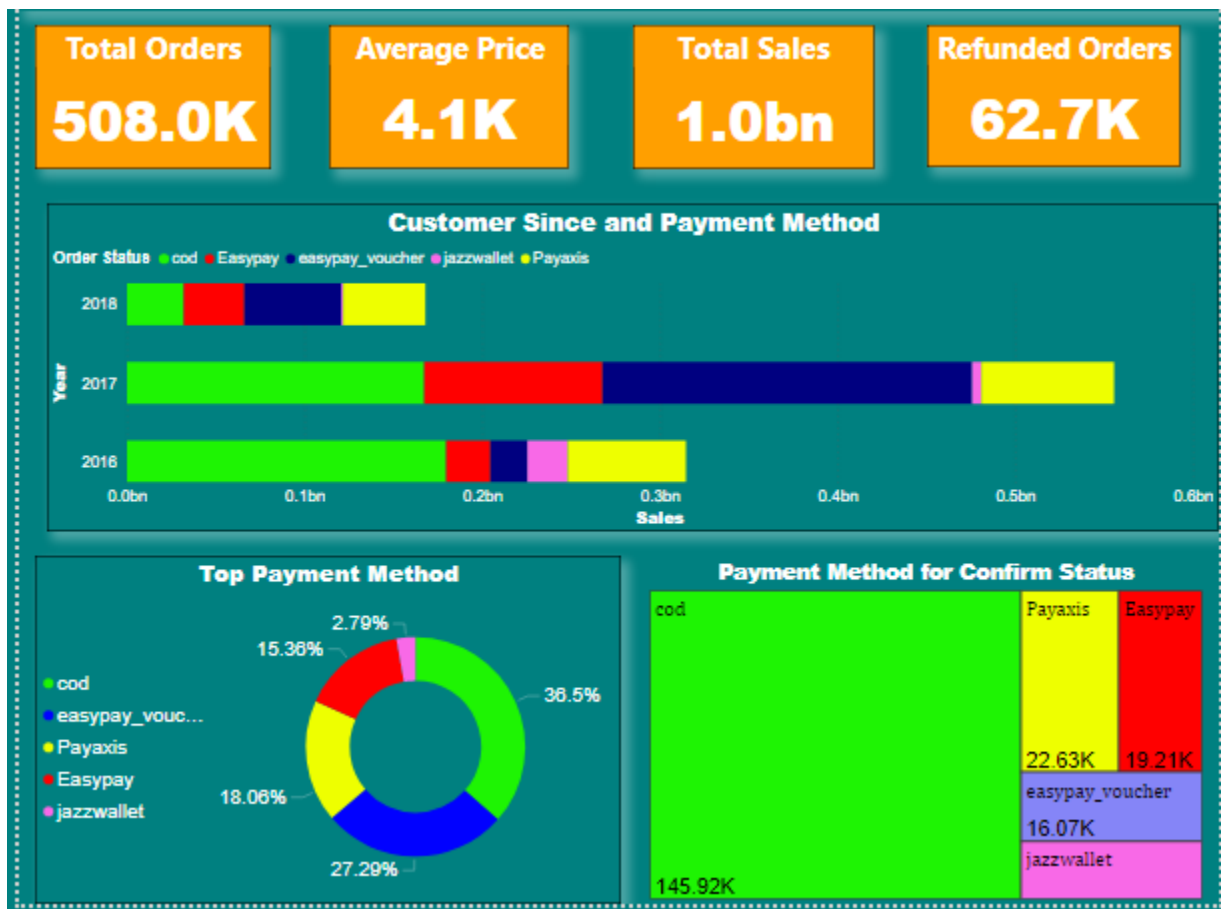
Women fashion increased in spring season then decreased in summer and beyond.

Mobile and tablet remain in high demand throughout the year

Superstore category increased in august only.

Note: Categories are filtered out on the basis on top 10 that contributes to sale.

This line and bar chart shows the diffe



The dashboard shows us that COD (cash on delivery) the most used payment method.

Customer since and payment method

Across payment_method, cod had the most interesting recent trend and started trending down on 2016, falling by 82.22% (147,693,772.89) in 2 years.

We can also observe from the bar chart that the longer the customer uses the service, it uses other payment options as well such as easypay_voucher.

Easypay_voucher emerges as most used in 2018 jumping from 20,270,560 to 207,968,417 in 2017 (the most used in 2017!) to 54,573,132 in 2018

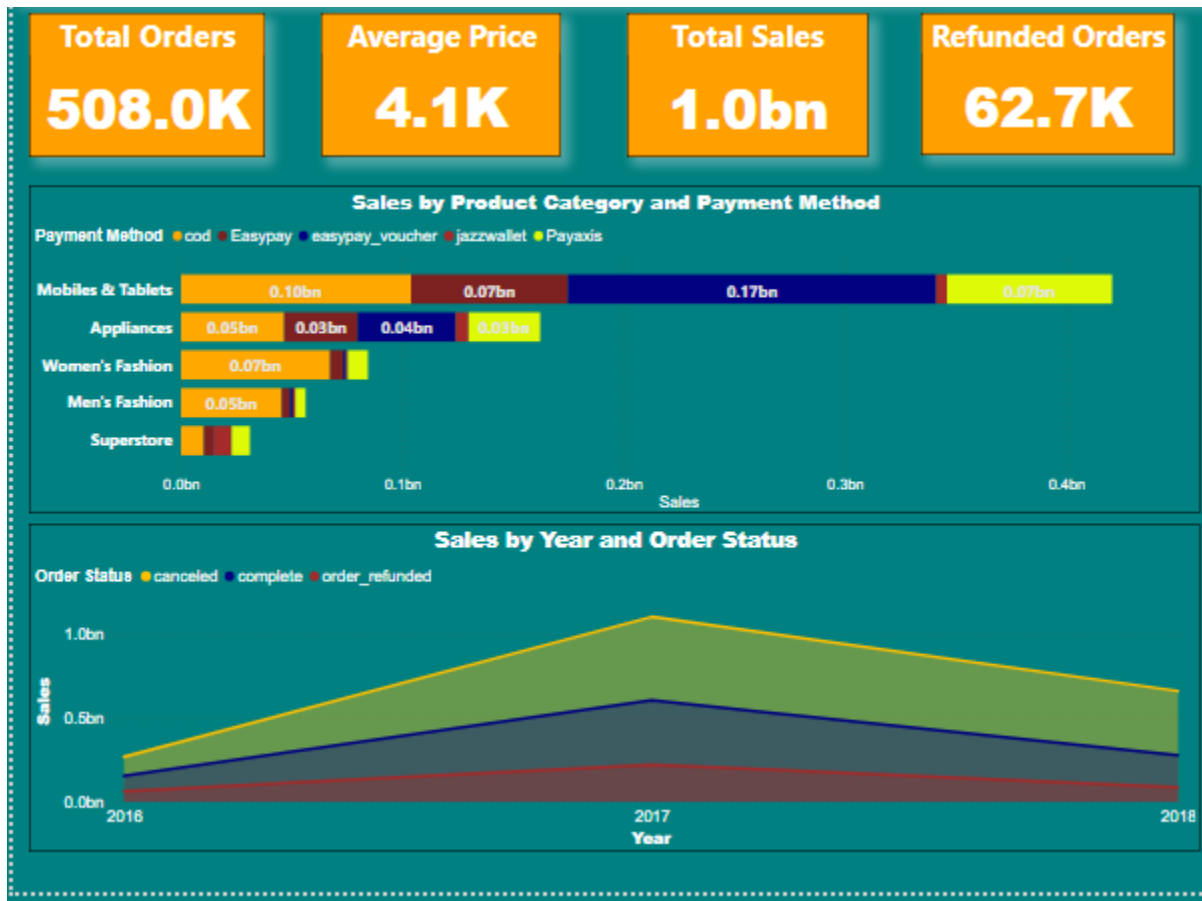
Pay axis goes from 66551k to 72699k back to 46191k in 2018.

Payment method for confirm Status

The tree chart also shows us that COD based payment consist of mostly confirm products.

Top payment Method

From donut chart we understand COD and Easypay_voucher are the most important payment method. cod accounted for 36.50% of Sales overall



Sales by product category and payment method

The bar chart shown in the dashboard inform us that most used product category mobile

and tablet uses the easypay_voucher as a preferred payment method.

Appliances customer prefer cash on delivery with 46,697k sales with it.

Women and Men Fashion also pay mostly as cod.

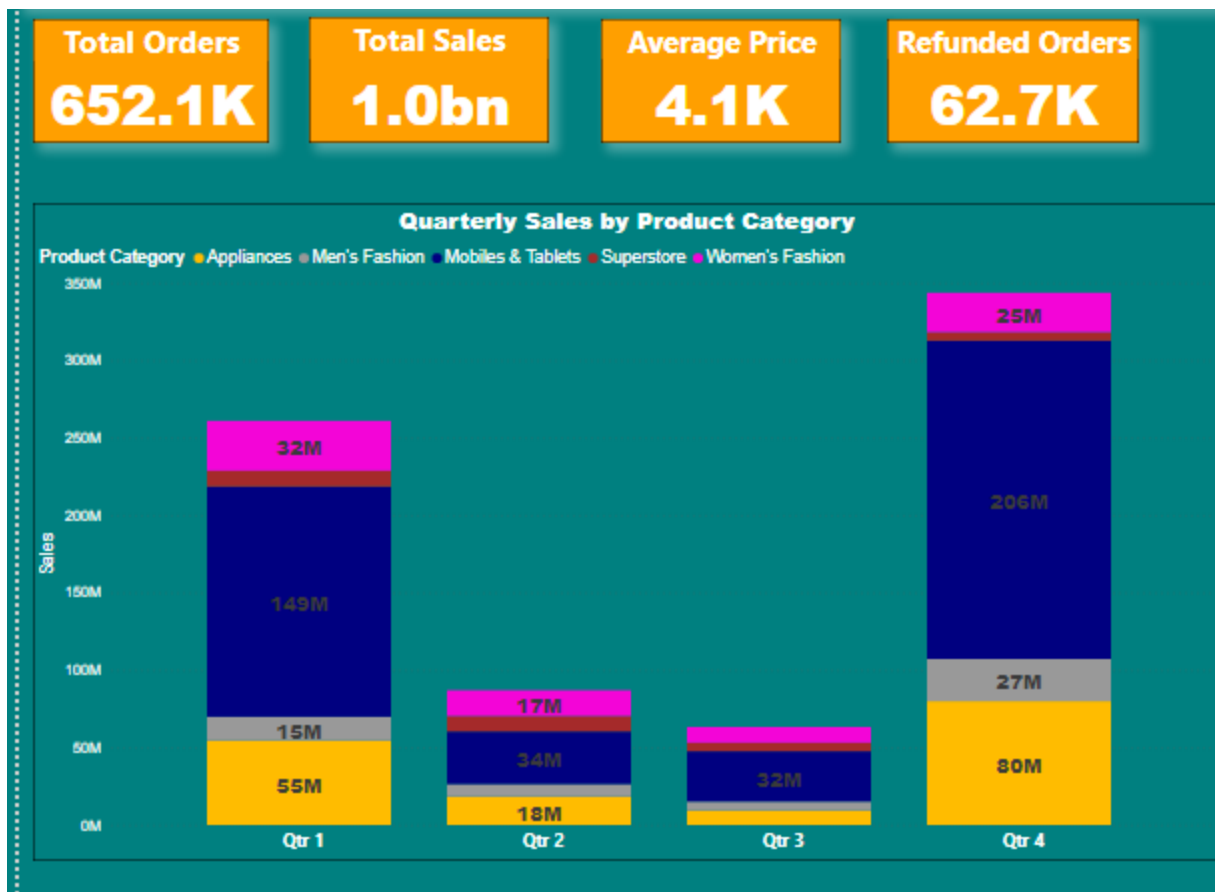
Mobiles & Tablets made up 22.02% of Sales.

Sales by year and Order status

The area chart shows us the loses that we see every year due to cancelled or refunded

order and in which category do we face the most refund.

Sales for canceled started trending up on 2016, rising by 145.91% (392,348,449.54) in 2 years.



Quarterly sales by Product Category

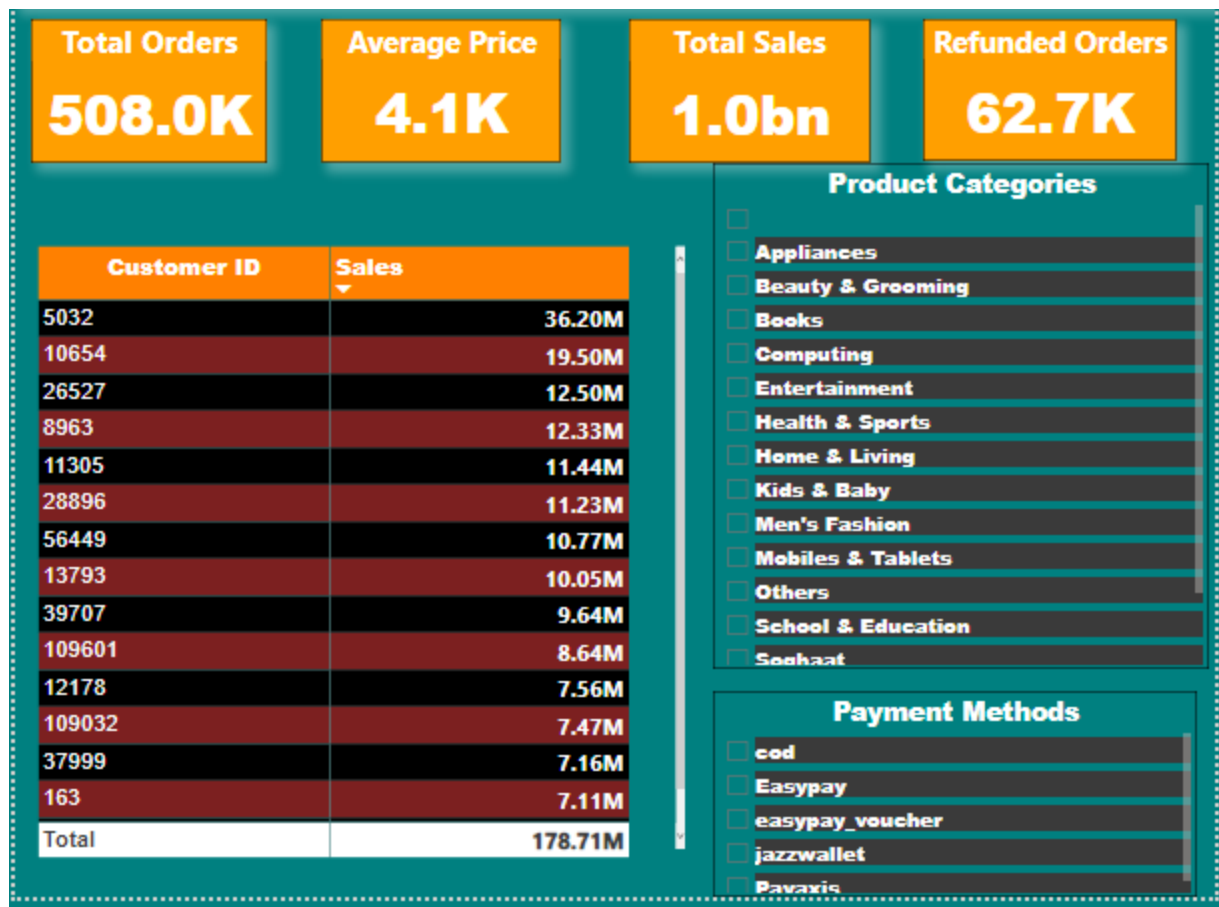
The dashboard shows us quarterly sales from each product category.

We see that mobile phones and tablet are mostly sold the highest in the first and last quarter

Appliances from 55 million in quarter 1 decreased to 18 Million then to 9 million in quarter 3 . Then it increased it sales to 80 million.

Mobile and tablet is sold the most in every quarter but their sales decreased from 149 million in first quarter to 32 million in quarter 3 but then jumped to 206 million in quarter 4.

Overall , Qtr 4 made up 27.22% of Sales.



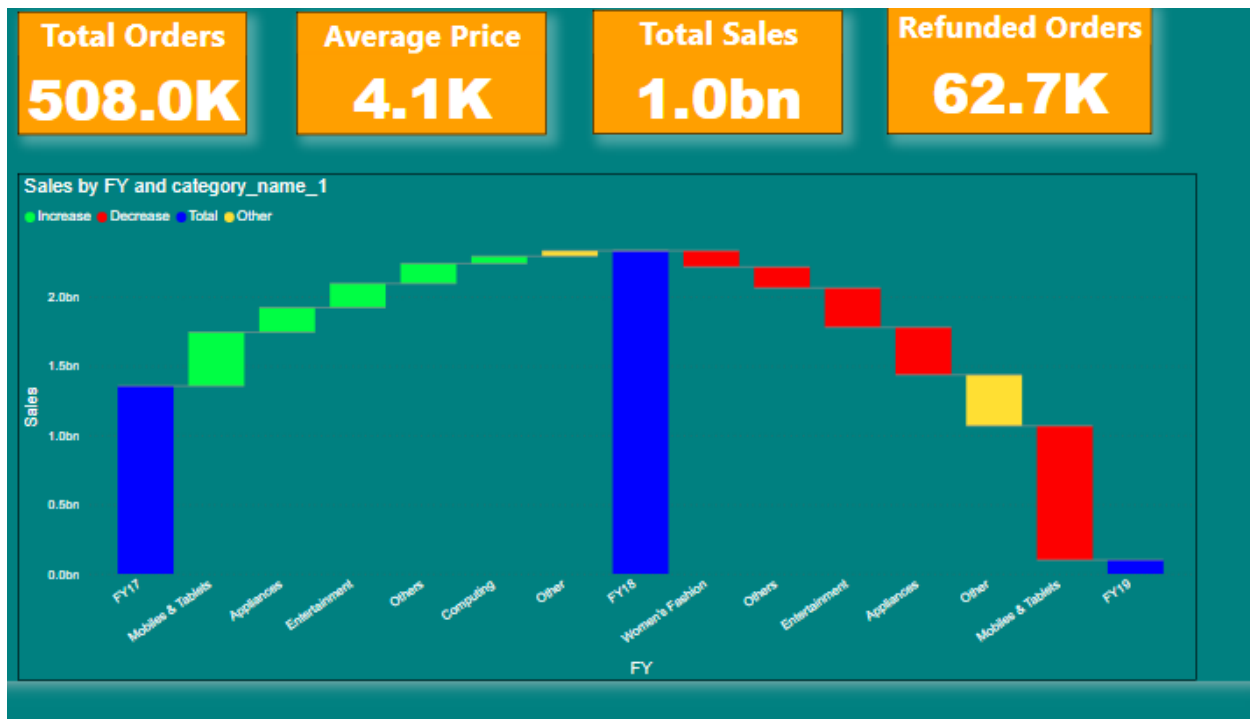
The purpose of this particular dashboard is to see use the slicer filter for product category and payment method.

This gives Owner, to forecast the possible future outcome of the product he chooses and the payment methods he offers for its customers from past history of customers

The table is filtered with top 15 customers with the most sales

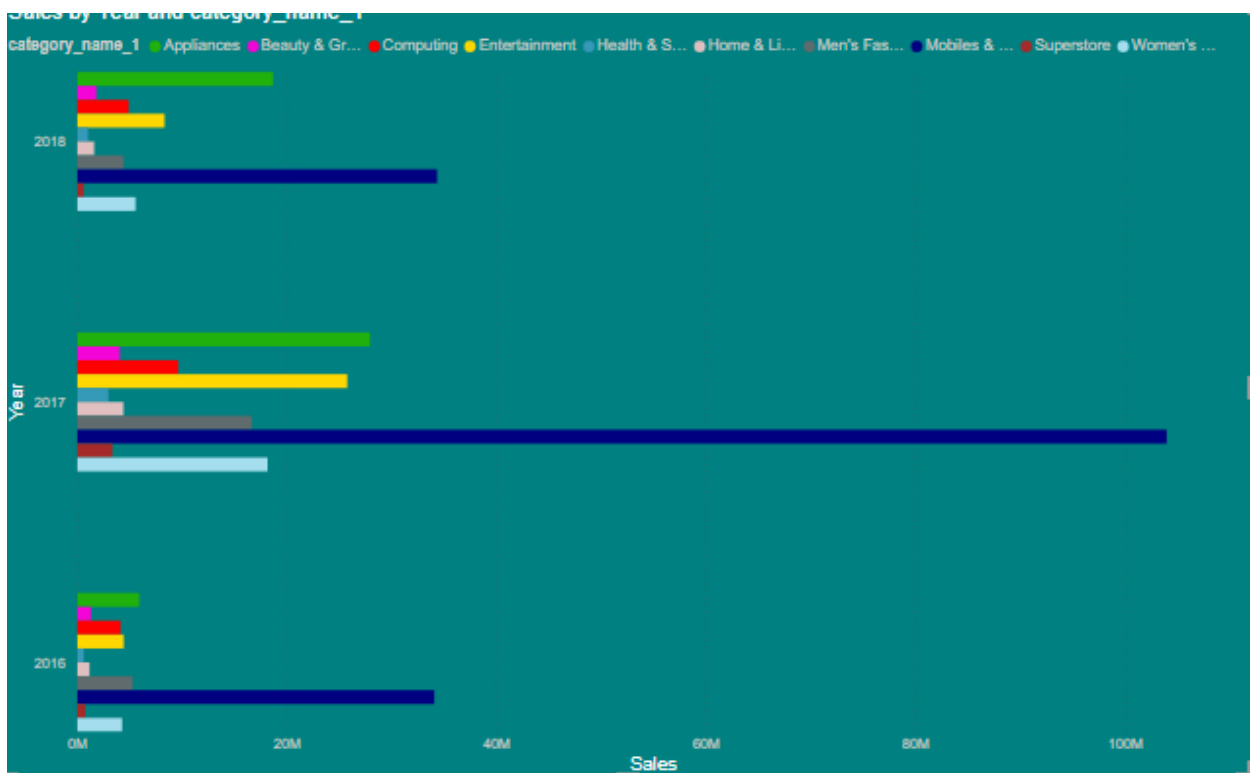
Customer 5032 did the most sales of 26.2 million

For mobile category top 2 customer are 5032 and 10664

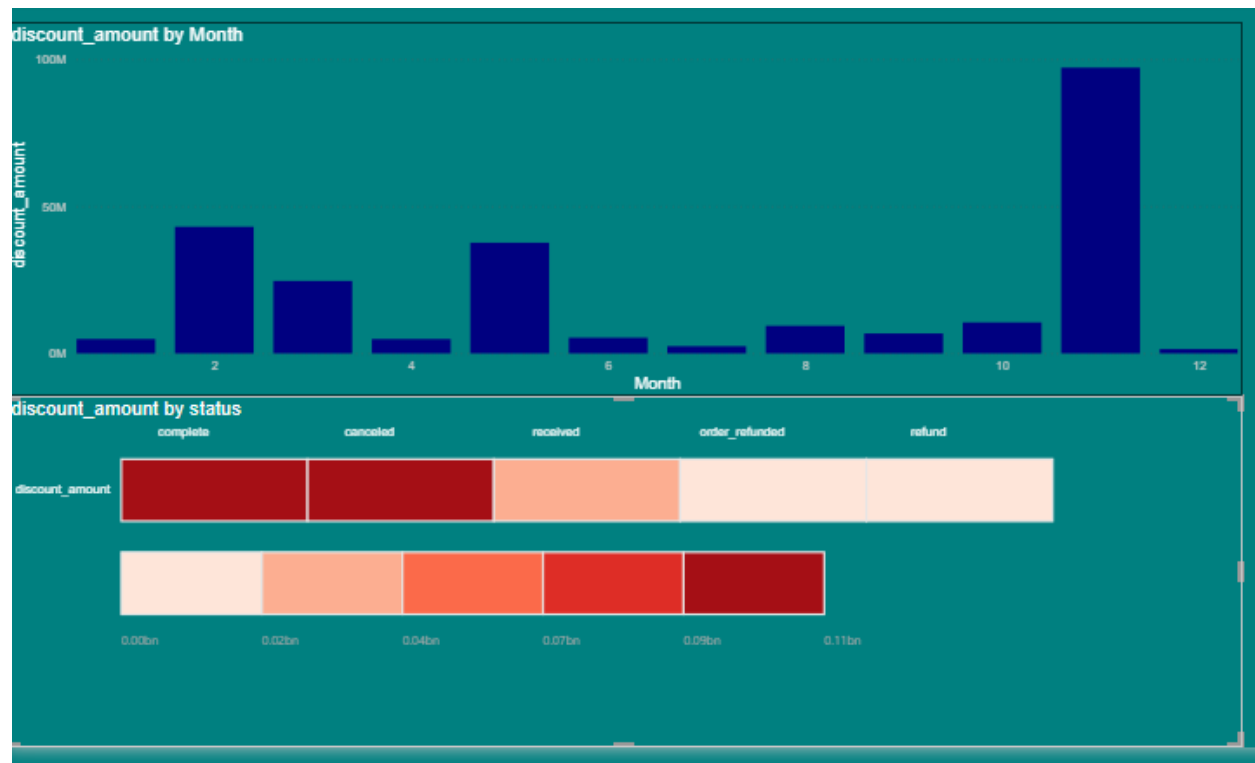


This waterfall charts shows from FY 17 to FY 18 sales increased mostly by mobiles and tablets category followed by appliances and entertainment.

From FY-18 to FY-19 sales decreased and mostly it is because of mobiles and appliances



Per year refund per category it shows mobile and tablets although having most sales are also refunded the most. In 2017 Home and living products and appliances were also refunded but it decreased in 2018.



Discount_amount by month

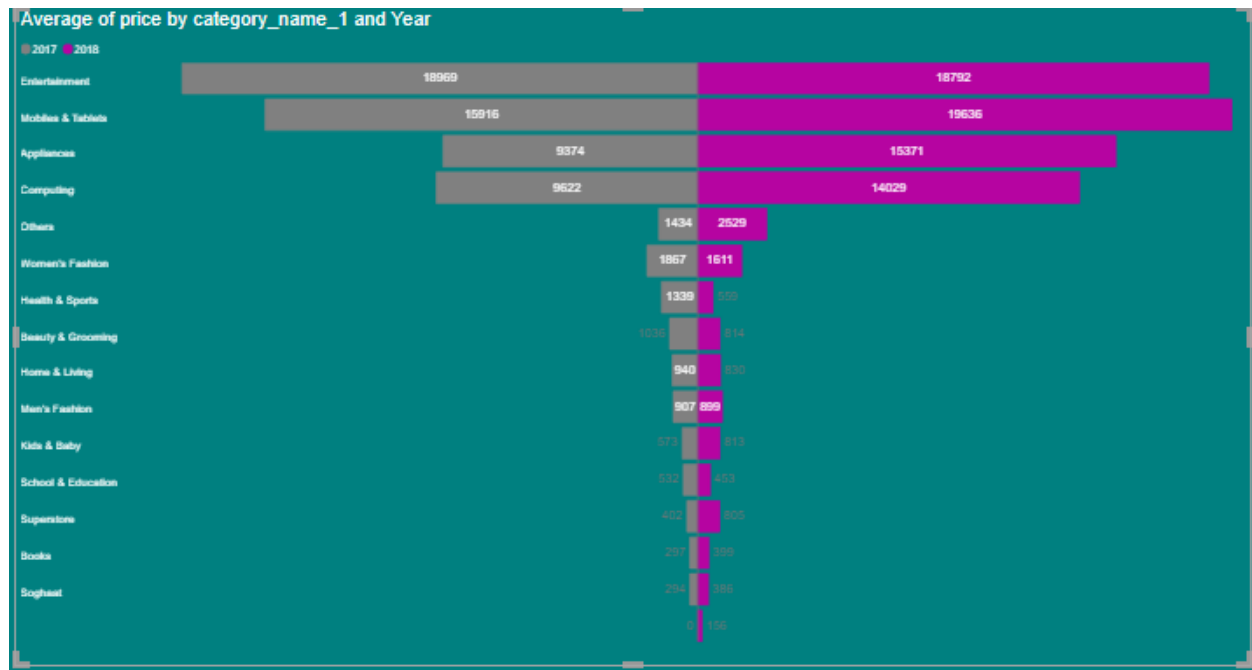
For every month we see the discounted amount. In November since sales are high so discount is high too. We also see a trend that whenever sales increase discounted amount also increases.

At 97,123,442.60, 11 had the highest discount_amount and was 6,655.95% higher than 12, which had the lowest discount_amount at 1,437,598.70.

11 accounted for 39.17% of discount_amount.

Discount_amount by status

Heatmap shows and answers a vital question that whether owner gave discounts to those orders which got cancelled? But that is not the case since more discount was given to orders which were completed in range of 0.09-0.11 billion. And the least discount was given to order which got refunded. Canceled order were also given discounts



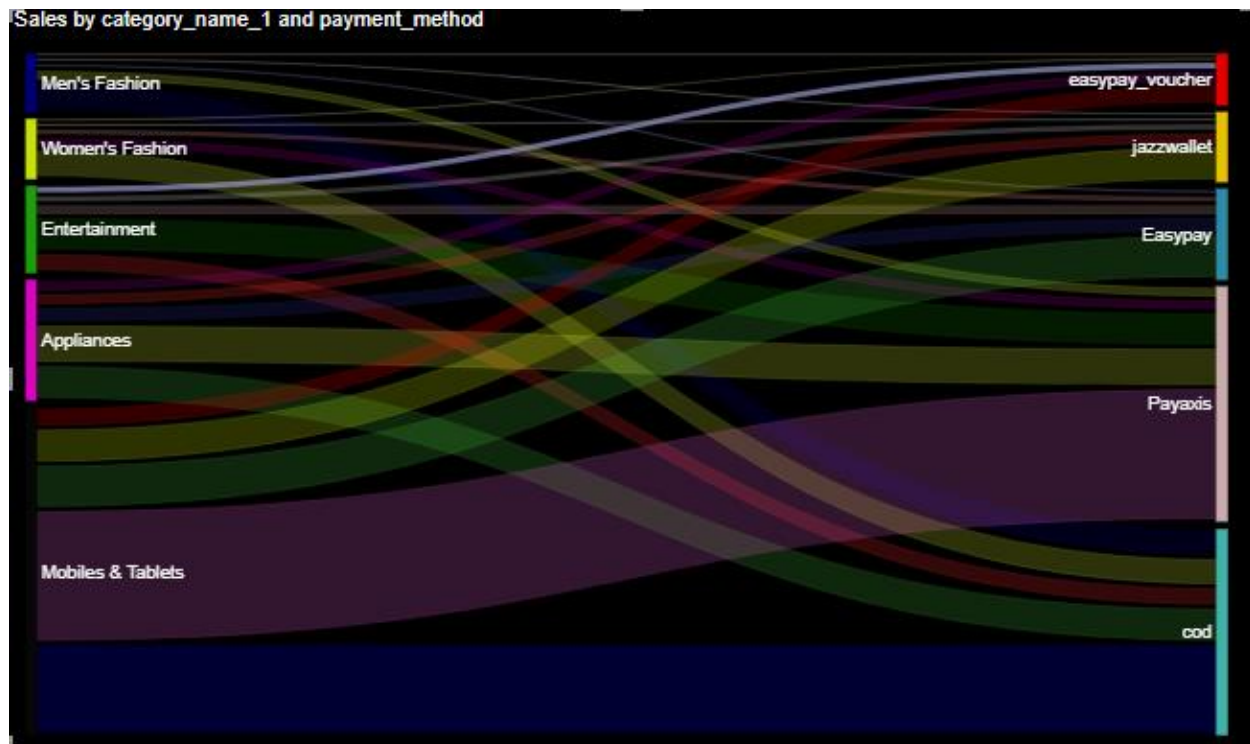
This tornado chart shows the average price change between 2018 and 2017 . Most cost of product entertainment products didn't changed much.

Mobile with large sale had increase in price change from 15915 to 19635.

Appliances prices also increased much more 9374 to 15371.

And so did computing

Surprisingly some prices decreased like health, beauty, women fashion products and home and living products.



Now we analyze the old customer habits. What do they buy? How do they pay? Since they are our loyal customers. We see they also buy mobile and tablets and pay on payaxis and second most they pay cod and then easy pay etc.

Highest payment method is payaxis.

Second highest product : appliances is also paid on payaxis the most and then on cod.

Step 6: Finally create the stories

Overall:

Setup: The owner gathered data of many products to find out hot products and market trends and in demand cash methods

Conflict: We see mobile and tablet sales chart go up significantly. But most refunds are also made of Mobile and tablet. And most discounts are also given in this category. Appliances also are getting refunded which is on contrary most sold item too and most discounted item . Overall , we see a decrease in sales from 2017 to 2018. We also see a shift in customer going towards easypay vouchers from cash on delivery in 2018. Men and Women Fashion are in demand product but their prices were decreased.

Resolution: blind discounts should be prevented. Owner should go for discounts based on seasons like in november Mobiles , appliances and men fashion was sold the most so put some offer. In spring women fashion rose to add some offer in that period. Increase prices for most in-demand products and add seasonal discounts and introduce loyal customer cards and voucher policy.

Step 7: Provide answer to each of your BI queries. Remember that multiple answers will

be available for each query.

Q) : What product is more profitable? And what payment method benefits the customer and the company

Mobile and Tablets and emerging payment method is pay axis after cod.

Q) What is causing customers to refund and why are they doing it?

Big discount make customer buy mobile and tablets later they refund in large amounts.