Introduction

This project would primarily focus on using SQL, Power BI to analyze a dataset that was assumed to have been extracted from a data warehouse of a company after undergoing the ETL (Extract, Transform, Load) process. The dataset included five tables:

customers

customer_code	custmer_name	customer_type
Cus001	Surge Stores	Brick & Mortar
Cus002	Nomad Stores	Brick & Mortar
Cus003	Excel Stores	Brick & Mortar
Cus004	Surface Stores	Brick & Mortar
Cus005	Premium Stores	Brick & Mortar

transactions

product_code	customer_code	market_code	order_date	sales_qty	sales_amount	currency	profit_margin_percentage	profit_margin	cost_price
Prod279	Cus020	Mark011	2017-10-11	1	102	INR	0.39	39.78	62.22
Prod279	Cus020	Mark011	2017-10-18	1	102	INR	-0.12	-12.24	114.24
Prod279	Cus020	Mark011	2017-10-19	1	102	INR	0.29	29.58	72.42
Prod279	Cus020	Mark011	2017-11-08	1	102	INR	0.36	36.72	65.28
Prod279	Cus020	Mark011	2018-03-09	1	102	INR	-0.35	-35.7	137.7
Prod279	Cus020	Mark011	2018-03-20	1	102	INR	0.26	26.52	75.48

products

product_code	product_type
Prod001	Own Brand
Prod002	Own Brand
Prod003	Own Brand
Prod004	Own Brand
Prod005	Own Brand

markets

markets_code	markets_name	zone
Mark001	Chennai	South
Mark002	Mumbai	Central
Mark003	Ahmedabad	North
Mark004	Delhi NCR	North
Mark005	Kanpur	North

Date

date	cy_date	year	month_name	date_yy_mmm
2017-06-01	2017-06-01	2017	June	17-Jun
2017-06-02	2017-06-01	2017	June	17-Jun
2017-06-03	2017-06-01	2017	June	17-Jun
2017-06-04	2017-06-01	2017	June	17-Jun
2017-06-05	2017-06-01	2017	June	17-Jun

Analysis (product focused)

Top 5 customers by revenue

customer_name	contributed_revenue
Electricalsara Stores	64218400000
Premium Stores	31275100000
Surge Stores	16457100000
Leader	12467100000
Excel Stores	9740930000

Top 5 products by revenue

product_name	contributed_revenue
Prod090	52205700000
Prod318	13927800000
Prod316	9393830000
Prod232	8943010000
Prod239	5405690000

- **Which products were purchased the most by the top 5 customers?**

customer_name	most_sold_product	total_sales
Electricalsara Stores	Prod239	126906
Premium Stores	Prod090	79998
Surge Stores	Prod090	58306
Excel Stores	Prod090	47086
Leader	Prod260	16391
Leader	Prod318	16391

- **The top 5 products with the highest sales quantities.**

product_code	number_of_products_sold
Prod090	277959
Prod239	170458
Prod237	114170
Prod318	74195
Prod245	72935

- **The top 5 products with the lowest average price.**

product_code	avg_price_of_the_product
Prod111	65
Prod247	74.5
Prod115	106
Prod181	130
Prod154	134

- **The top 5 products with the highest average price.**

product_code	avg_price_of_the_product
Prod073	1477394
Prod163	644444
Prod107	316611
Prod083	263148
Prod077	253851.25

The top 5 markets by revenue.

market_name	contributed_revenue
Delhi NCR	81044700000
Mumbai	47316600000
Kochi	19240300000
Ahmedabad	13694400000
Bhopal	12467100000

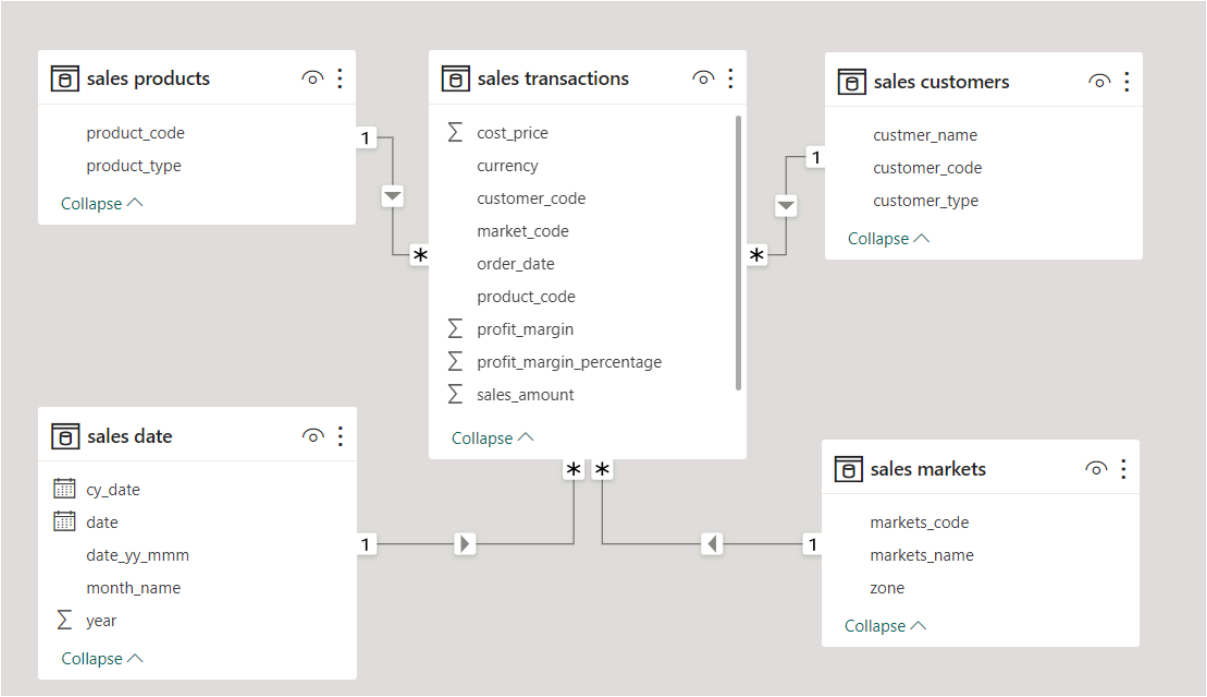
The top 5 months by revenue.

year	month	contributed_revenue
2018	January	16503800000
2018	February	13298700000
2019	November	11717900000
2018	July	10236700000
2019	January	9218410000

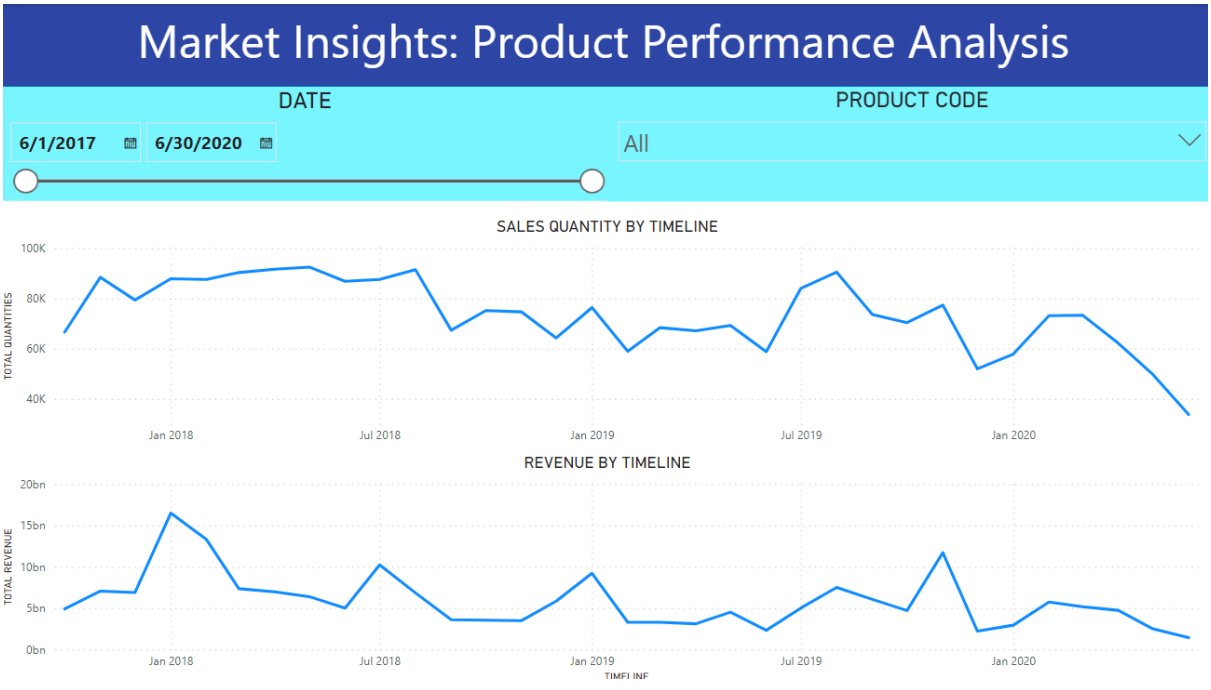
Visualization

Step 1

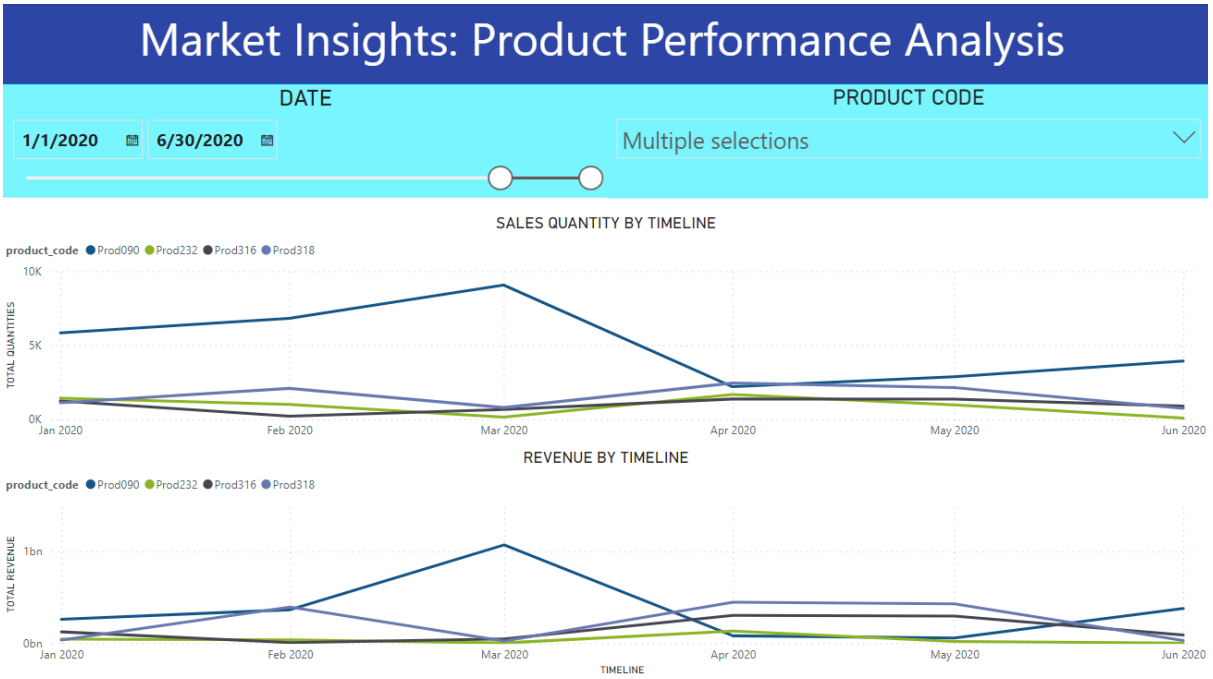
Made tables relate to transactions table.



I created a dashboard that provided insights to the audiences about sales quantities and revenue over a timeline. I also created two slicers, one to filter the timeline and one to filter product codes.



I filtered product codes by contributed revenue within the timeframe from the beginning of 2020 to the end of 2020 which prod90, prod318, prod 316, prod 232 were from the previous queries with SQL.



All of the above metrics were for practice purposes only and did not serve any business tasks.