

PHÂN TÍCH BÁO CÁO VÀ DỰ ĐOÁN XU HƯỚNG BẢN HÀNG

GVHD: Thầy Nguyễn Minh Hiển

Nguyễn Thanh Duy Dương (Trưởng nhóm)

Huỳnh Sơn Thạch

Ngô Đạt Mẫn

Lâm Nguyễn

December 19,
2024

Presented by
BALL DATA

Tổng quan dự án

A

Giới thiệu dự án

Về các thành viên trong dự án, đề tài dự án và giới thiệu về nhóm



B

Các vấn đề cần giải quyết

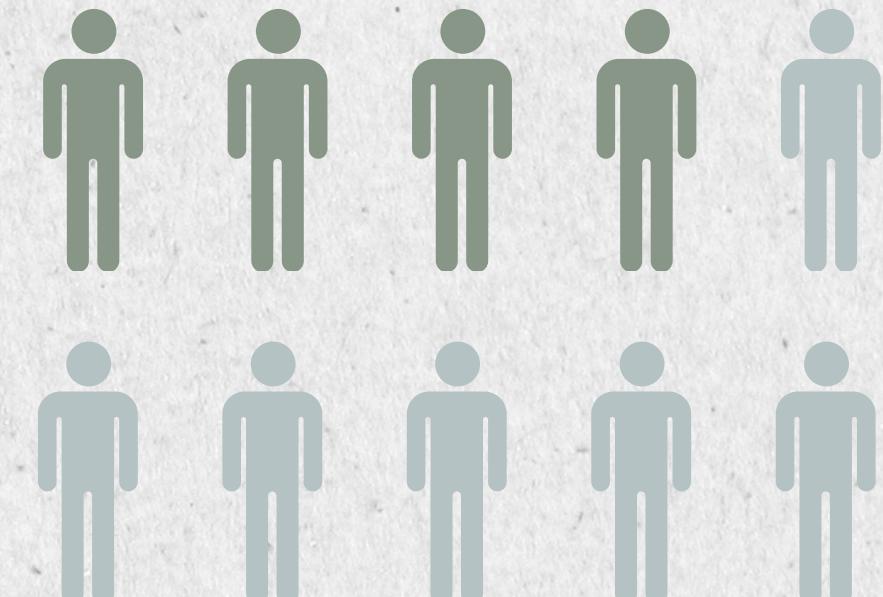
Những vấn đề cần phải giải quyết về tình hình kinh doanh của sàn thương mại



C

Hướng giải quyết và kết luận

Đưa ra các giải pháp cho từng vấn đề, kết luận lại về mục tiêu và kết quả đã đạt được của nhóm





**"Dữ liệu giống như quả bóng - khi ta hiểu rõ,
kiểm soát và linh hoạt sử dụng, nó sẽ đưa ta đến mục tiêu mong muôn."**

MỤC TIÊU DỰ ÁN



Analyze

Sử dụng các công cụ như Tableau, Python để xử lý, làm sạch và modeling dữ liệu, dùng cho việc phân tích và dự đoán.

Overview

Sau khi làm sạch, tiến hành phân tích và dự đoán sau đó đưa ra tổng quan về tình hình của Sàn Thương Mại.

Problem

Tìm hiểu và đưa ra các vấn đề gây khó khăn cho doanh nghiệp và khách hàng, sau đó tiến hành phân tích.

Problem solving

Đưa ra các giải pháp để khắc phục những khó khăn cản trở, sau đó thử nghiệm trong thời gian đưa ra và kiểm tra tính khả thi.



Analyze

NGUỒN GỐC BỘ DỮ LIỆU

Nguồn cung cấp: Được Olist công khai trên Kaggle để hỗ trợ các nhà phân tích và nghiên cứu.

Thời gian: Dữ liệu được thu thập trong 3 năm 2016-2018, phản ánh rõ về xu hướng và sự phát triển của thương mại điện tử tại Brazil.

Định dạng: Gồm nhiều tệp CSV với thông tin chi tiết từ giao dịch, người bán, khách hàng đến phản hồi đánh giá.



Brazilian E-Commerce Public Dataset by Olist

VỀ BỘ DỮ LIỆU



Tập dữ liệu này được cung cấp rộng rãi bởi Olist, cửa hàng bách hóa lớn nhất ở các khu chợ ở Brazil. Olist kết nối các doanh nghiệp nhỏ từ khắp Brazil với các kênh mà không gặp rắc rối và chỉ bằng một hợp đồng duy nhất.

VỀ BỘ DỮ LIỆU



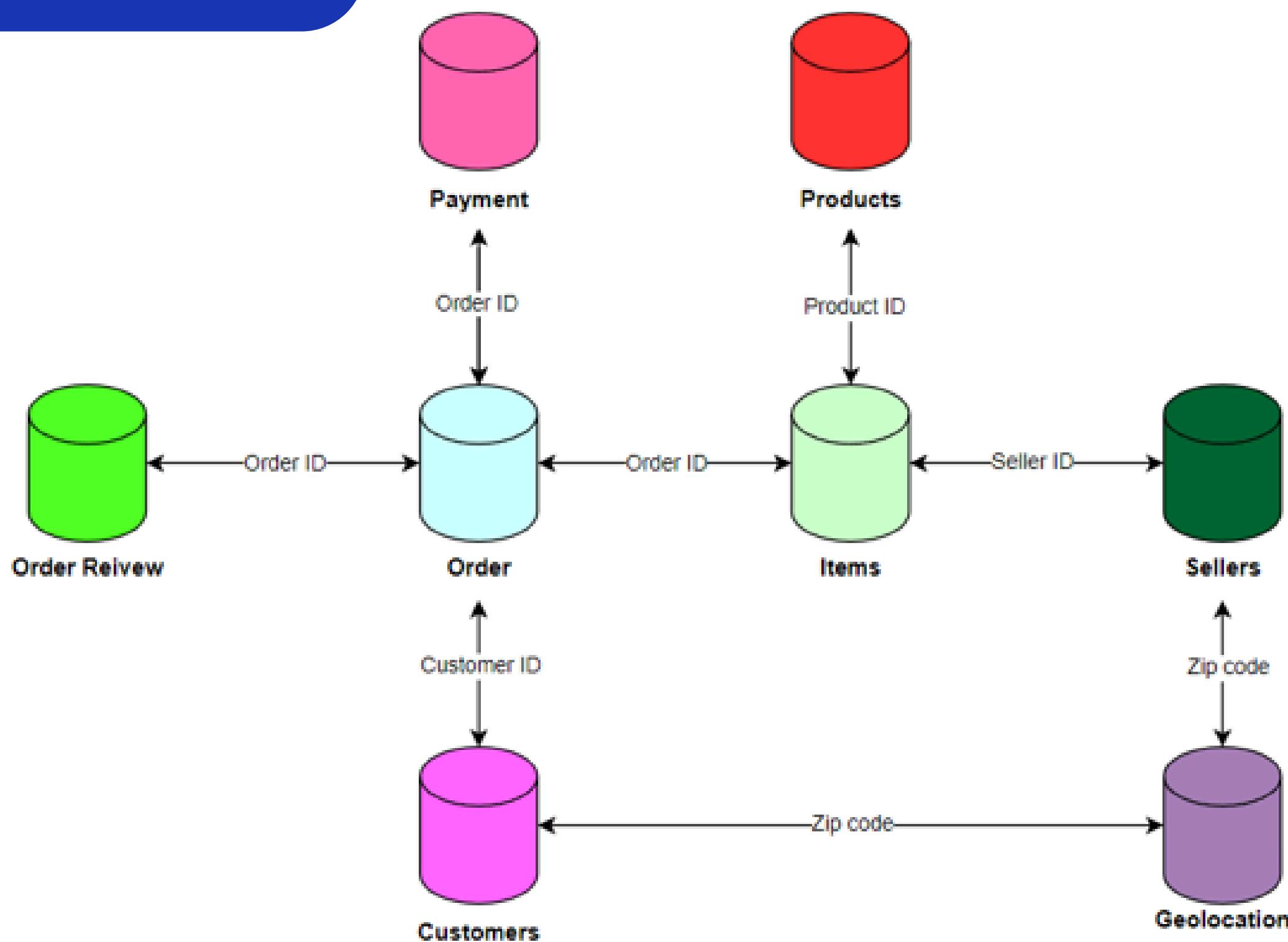
Những người bán này có thể bán sản phẩm của mình thông qua Cửa hàng Olist và vận chuyển trực tiếp đến khách hàng thông qua các đối tác hậu cần của Olist. Sau khi khách hàng mua sản phẩm từ Olist Store, người bán sẽ nhận được thông báo thực hiện đơn đặt hàng đó.

VỀ BỘ DỮ LIỆU



Sau khi khách hàng nhận được sản phẩm hoặc đến hạn giao hàng dự kiến, khách hàng sẽ nhận được một bản khảo sát về mức độ hài lòng qua email, nơi họ có thể đưa ra ghi chú về trải nghiệm mua hàng và viết ra một số nhận xét.

MÔ HÌNH DỮ LIỆU



Hình 1 Mô hình dữ liệu

MÔ TẢ TỪNG BẢNG: CUSTOMERS

	customer_id	customer_unique_id	customer_zip_code_prefix	customer_city	customer_state
0	06b8999e2fba1a1fbc88172c00ba8bc7	861eff4711a542e4b93843c6dd7febb0	14409	franca	SP
1	18955e83d337fd6b2def6b18a428ac77	290c77bc529b7ac935b93aa66c333dc3	9790	sao bernardo do campo	SP
2	4e7b3e00288586ebd08712fdd0374a03	060e732b5b29e8181a18229c7b0b2b5e	1151	sao paulo	SP
3	b2b6027bc5c5109e529d4dc6358b12c3	259dac757896d24d7702b9acbbff3f3c	8775	mogi das cruzes	SP
4	4f2d8ab171c80ec8364f7c12e35b23ad	345ecd01c38d18a9036ed96c73b8d066	13056	campinas	SP
5	879864dab9bc3047522c92c82e1212b8	4c93744516667ad3b8f1fb645a3116a4	89254	jaragua do sul	SC
6	fd826e7cf63160e536e0908c76c3f441	addec96d2e059c80c30fe6871d30d177	4534	sao paulo	SP
7	5e274e7a0c3809e14aba7ad5aae0d407	57b2a98a409812fe9618067b6b8ebe4f	35182	timoteo	MG
8	5adf08e34b2e993982a47070956c5c65	1175e95fb47ddff9de6b2b06188f7e0d	81560	curitiba	PR
9	4b7139f34592b3a31687243a302fa75b	9afe194fb833f79e300e37e580171f22	30575	belo horizonte	MG
10	9fb35e4ed6f0a14a4977cd9aea4042bb	2a7745e1ed516b289ed9b29c7d0539a5	39400	montes claros	MG
11	5aa9e4fdd4dfd20959cad2d772509598	2a46fb94aef5cbeeb850418118cee090	20231	rio de janeiro	RJ
12	b2d1536598b73a9abd18e0d75d92f0a3	918dc87cd72cd9f6ed4bd442ed785235	18682	lencois paulista	SP
13	eabebad39a88bb6f5b52376faec28612	295c05e81917928d76245e842748184d	5704	sao paulo	SP
14	1f1c7bf1c9b041b292af6c1c4470b753	3151a81801c8386361b62277d7fa5ecf	95110	caxias do sul	RS
15	206f3129c0e4d7d0b9550426023f0a08	21f748a16f4e1688a9014eb3ee6fa325	13412	piracicaba	SP
16	a7c125a0a07b75146167b7f04a7f8e98	5c2991dbd08bbf3cf410713c4de5a0b5	22750	rio de janeiro	RJ
17	c5c61596a3b6bd0cee5766992c48a9a1	b6e99561fe6f34a55b0b7da92f8ed775	7124	guarulhos	SP
18	9b8ce803689b3562defaad4613ef426f	7f3a72e8f988c6e735ba118d54f47458	5416	sao paulo	SP
19	49d0ea0986edde72da777f15456a0ee0	3e6fd6b2f0d499456a6a6820a40f2d79	68485	pacaja	PA

Bảng Customer mô tả thông tin khách hàng như ID khách hàng, ID đơn hàng, địa chỉ khách hàng.

Dùng để phân tích đặc điểm khách hàng, vị trí địa lý hoặc liên kết với đơn hàng để tìm hiểu hành vi mua sắm

MÔ TẢ TỪNG BẢNG: SELLERS

	seller_id	seller_zip_code_prefix	seller_city	seller_state
0	3442f8959a84dea7ee197c632cb2df15	13023	campinas	SP
1	d1b65fc7debc3361ea86b5f14c68d2e2	13844	mogi guacu	SP
2	ce3ad9de960102d0677a81f5d0bb7b2d	20031	rio de janeiro	RJ
3	c0f3eea2e14555b6faeee3dd58c1b1c3	4195	sao paulo	SP
4	51a04a8a6bdcb23deccc82b0b80742cf	12914	braganca paulista	SP
5	c240c4061717ac1806ae6ee72be3533b	20920	rio de janeiro	RJ
6	e49c26c3edfa46d227d5121a6b6e4d37	55325	brejao	PE
7	1b938a7ec6ac5061a66a3766e0e75f90	16304	penapolis	SP
8	768a86e36ad6aae3d03ee3c6433d61df	1529	sao paulo	SP
9	ccc4bbb5f32a6ab2b7066a4130f114e3	80310	curitiba	PR

Chứa thông tin về người bán, gồm seller_id, vị trí và mã bưu điện của họ

Dùng để phân tích hiệu suất của từng người bán hoặc so sánh doanh số giữa các khu vực.

MÔ TẢ TỪNG BẢNG: ORDERS

order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_carrier_date
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2017-10-02 10:56:33	2017-10-02 11:07:15	2017-10-04 19:55:00
53cdb2fc8bc7dce0b6741e2150273451	b0830fb4747a6c6d20dea0b8c802d7ef	delivered	2018-07-24 20:41:37	2018-07-26 03:24:27	2018-07-26 14:31:00
47770eb9100c2d0c44946d9cf07ec65d	41ce2a54c0b03bf3443c3d931a367089	delivered	2018-08-08 08:38:49	2018-08-08 08:55:23	2018-08-08 13:50:00
949d5b44dbf5de918fe9c16f97b45f8a	f88197465ea7920adcdbe7375364d82	delivered	2017-11-18 19:28:06	2017-11-18 19:45:59	2017-11-22 13:39:59
ad21c59c0840e6cb83a9ceb5573f8159	8ab97904e6daea8866dbdbc4fb7aad2c	delivered	2018-02-13 21:18:39	2018-02-13 22:20:29	2018-02-14 19:46:34
a4591c265e18cb1dcee52889e2d8acc3	503740e9ca751ccdda7ba28e9ab8f608	delivered	2017-07-09 21:57:05	2017-07-09 22:10:13	2017-07-11 14:58:04
136cce7faa42fdb2cefd53fdc79a6098	ed0271e0b7da060a393796590e7b737a	invoiced	2017-04-11 12:22:08	2017-04-13 13:25:17	Nan
6514b8ad8028c9f2cc2374ded245783f	9bdf08b4b3b52b5526ff42d37d47f222	delivered	2017-05-16 13:10:30	2017-05-16 13:22:11	2017-05-22 10:07:46
76c6e866289321a7c93b82b54852dc33	f54a9f0e6b351c431402b8461ea51999	delivered	2017-01-23 18:29:09	2017-01-25 02:50:47	2017-01-26 14:16:31
e69bfb5eb88e0ed6a785585b27e16dbf	31ad1d1b63eb9962463f764d4e6e0c9d	delivered	2017-07-29 11:55:02	2017-07-29 12:05:32	2017-08-10 19:45:24
e6ce16cb79ec1d90b1da9085a6118aeb	494dded5b201313c64ed7f100595b95c	delivered	2017-05-16 19:41:10	2017-05-16 19:50:18	2017-05-18 11:40:40
34513ce0c4fab462a55830c0989c7edb	7711cf624183d843aafe81855097bc37	delivered	2017-07-13 19:58:11	2017-07-13 20:10:08	2017-07-14 18:43:29
82566a660a982b15fb86e904c8d32918	d3e3b74c766bc6214e0c830b17ee2341	delivered	2018-06-07 10:06:19	2018-06-09 03:13:12	2018-06-11 13:29:00

Bảng Orders gồm ID đơn hàng, ID người mua, các mốc thời gian xác nhận đơn, vận chuyển và giao hàng, nhận hàng.

Mục đích là theo dõi quá trình hoàn thành đơn hàng, phân tích hiệu suất giao hàng,
hoặc xác định xu hướng mua sắm theo thời gian.

MÔ TẢ TỪNG BẢNG: PRODUCTS

product_id	product_category_name	product_name lenght	product_description_lenght	product_photos_qty	product_weight_g
1e9e8ef04dbcff4541ed26657ea517e5	perfumaria	40.0	287.0	1.0	225.0
3aa071139cb16b67ca9e5dea641aaa2f	artes	44.0	276.0	1.0	1000.0
96bd76ec8810374ed1b65e291975717f	esporte_lazer	46.0	250.0	1.0	154.0
cef67bcfe19066a932b7673e239eb23d	bebés	27.0	261.0	1.0	371.0
9dc1a7de274444849c219cff195d0b71	utilidades_domesticas	37.0	402.0	4.0	625.0
41d3672d4792049fa1779bb35283ed13	instrumentos_musicais	60.0	745.0	1.0	200.0
732bd381ad09e530fe0a5f457d81becb	cool_stuff	56.0	1272.0	4.0	18350.0
2548af3e6e77a690cf3eb6368e9ab61e	moveis_decoracao	56.0	184.0	2.0	900.0
37cc742be07708b53a98702e77a21a02	eletrodomesticos	57.0	163.0	1.0	400.0
8c92109888e8cdf9d66dc7e463025574	brinquedos	36.0	1156.0	1.0	600.0
14aa47b7fe5c25522b47b4b29c98dc9	cama_mesa_banho	54.0	630.0	1.0	1100.0
03b63c5fc16691530586ae020c345514	bebés	49.0	728.0	4.0	7150.0
cf55509ea8edaac1d28fdb16e48fc22	instrumentos_musicais	43.0	1827.0	3.0	250.0
7bb6f29c2be57716194f96496660c7c2	moveis_decoracao	51.0	2083.0	2.0	600.0
eb31436580a610f202c859463d8c7415	construcao_ferramentas_seguranca	59.0	1602.0	4.0	200.0
3bb7f144022e6732727d8d838a7b13b3	esporte_lazer	22.0	3021.0	1.0	800.0
6a2fb4dd53d2cdb88e0432f1284a004c	perfumaria	39.0	346.0	2.0	400.0
a1b71017a84f92fd8da4aeeefba108a24	informatica_acessorios	59.0	636.0	1.0	900.0
a0736b92e52f6cead290e30b578413b6	moveis_decoracao	56.0	296.0	2.0	1700.0
f53103a77d9cf245e579ea37e5ec51f0	cama_mesa_banho	52.0	206.0	1.0	500.0
1c1890ba1779090cd54008a3c3302921	moveis_decoracao	27.0	158.0	4.0	2550.0
518ef5de2c2b3a255e326a4594ba15d9	cama_mesa_banho	27.0	329.0	2.0	800.0
e3e020af31d4d89d2602272b315c3f6e	beleza_saude	58.0	1987.0	3.0	75.0
8b41fbc2b984a12030090112324d1bc4	bebés	55.0	1627.0	1.0	500.0

Bao gồm thông tin về các sản phẩm như product_id, loại sản phẩm và trọng lượng. Dùng để tìm hiểu danh mục sản phẩm, phân tích mức độ phổ biến của từng loại sản phẩm.

MÔ TẢ TỪNG BẢNG: ITEMS

order_id	order_item_id	product_id	seller_id	shipping_limit_date	price	freight_value
0010242fe8c5a6d1ba2dd792cb16214	1	4244733e06e7ecb4970a6e2683c13e61	48436dade18ac8b2bce089ec2a041202	2017-09-19 09:45:35	58.90	13.29
00018f77f2f0320c557190d7a144bdd3	1	e5f2d52b802189ee658865ca93d83a8f	dd7ddc04e1b6c2c614352b383efe2d36	2017-05-03 11:05:13	239.90	19.93
000229ec398224ef6ca0657da4fc703e	1	c777355d18b72b67abbeef9df44fd0fd	5b51032eddd242adc84c38acab88f23d	2018-01-18 14:48:30	199.00	17.87
0024acbcdf0a6daa1e931b038114c75	1	7634da152a4610f1595efa32f14722fc	9d7a1d34a5052409006425275ba1c2b4	2018-08-15 10:10:18	12.99	12.79
0042b26cf59d7ce69dfabb4e55b4fd9	1	ac6c3623068f30de03045865e4e10089	df560393f3a51e74553ab94004ba5c87	2017-02-13 13:57:51	199.90	18.14
0048cc3ae777c65dbb7d2a0634bc1ea	1	ef92defde845ab8450f9d70c526ef70f	6426d21aca402a131fc0a5d0960a3c90	2017-05-23 03:55:27	21.90	12.69
0054e8431b9d7675808bcb819fb4a32	1	8d4f2bb7e93e6710a28f34fa83ee7d28	7040e82f899a04d1b434b795a43b4617	2017-12-14 12:10:31	19.90	11.85
00576fe39319847cbb9d288c5617fa6	1	557d850972a7d6f792fd18ae1400d9b6	5996cddab893a4652a15592fb58ab8db	2018-07-10 12:30:45	810.00	70.75
005a1a1728c9d785b8e2b08b904576c	1	310ae3c140ff94b03219ad0adc3c778f	a416b6a846a11724393025641d4edd5e	2018-03-26 18:31:29	145.95	11.65
005f50442cb953dcd1d21e1fb923495	1	4535b0e1091c278dfd193e5a1d63b39f	ba143b05f0110f0dc71ad71b4466ce92	2018-07-06 14:10:56	53.99	11.40
0061f2a7bc09da83e415a52dc8a4af1	1	d63c1011f49d98b976c352955b1c4bea	cc419e0650a3c5ba77189a1882b7556a	2018-03-29 22:28:09	59.99	8.88
063b381e2406b52ad429470734ebd5	1	f177554ea93259a5b282f24e33f65ab6	8602a61d680a10a82cceeeda0d99ea3d	2018-07-31 17:30:39	45.00	12.98
006ec9db01a64e59a68b2c340bf65a7	1	99a4788cb24856965c36a24e339b6058	4a3ca9315b744ce9f8e9374361493884	2018-07-26 17:24:20	74.00	23.32
008288aa423d2a3f00fcbb17cd7d8719	1	368c6c730842d78016ad823897a372db	1f50f920176fa81dab994f9023523100	2018-02-21 02:55:52	49.90	13.37
008288aa423d2a3f00fcbb17cd7d8719	2	368c6c730842d78016ad823897a372db	1f50f920176fa81dab994f9023523100	2018-02-21 02:55:52	49.90	13.37
009792311464db532ff765bf7b182ae	1	8cab8abac59158715e0d70a36c807415	530ec6109d11eaaf87999465c6afee01	2018-08-17 12:15:10	99.90	27.65
009c9a17f916a706d71784483a5d643	1	3f27ac8e699df3d300ec4a5d8c5cf0b2	fcb5ace8bcc92f75707dc0f01a27d269	2018-05-02 09:31:53	639.00	11.34
0aed2e25dbad2f9ddb70584c5a2ded	1	4fa33915031a8cde03dd0d3e8fb27f01	fe2032dab1a61af8794248c8196565c9	2018-05-16 20:57:03	144.00	8.77
00c3e6612759851cc3cbb4b83257986	1	b50c950aba0dcead2c48032a690ce817	218d46b86c1881d022bce9c68a7d4b15	2017-08-21 03:33:13	99.00	13.71
00e562887b1f2006d75e0be9558292e	1	5ed9eaf534f6936b51d0b6c5e4d5c2e9	8cbcac7e12637ed9cffa18c7875207478	2018-02-28 12:08:37	25.00	16.11
00e63d38ae8c00bbcb5a30573b99628	1	553e0e7590d3116a072507a3635d2877	1c129092bf23f28a5930387c980c0dfc	2018-03-29 20:07:49	47.90	8.88

Chứa các chi tiết từng sản phẩm trong mỗi đơn hàng, bao gồm product_id, giá và số lượng. Sử dụng để phân tích doanh thu, hiệu suất sản phẩm, hoặc chi tiết hóa từng đơn hàng.

MÔ TẢ TỪNG BẢNG: PAYMENTS

	order_id	payment_sequential	payment_type	payment_installments	payment_value
0	b81ef226f3fe1789b1e8b2acac839d17	1	credit_card	8	99.33
1	a9810da82917af2d9aefd1278f1dcfa0	1	credit_card	1	24.39
2	25e8ea4e93396b6fa0d3dd708e76c1bd	1	credit_card	1	65.71
3	ba78997921bbcdcc1373bb41e913ab953	1	credit_card	8	107.78
4	42fdf880ba16b47b59251dd489d4441a	1	credit_card	2	128.45
5	298fcdf1f73eb413e4d26d01b25bc1cd	1	credit_card	2	96.12
6	771ee386b001f06208a7419e4fc1bbd7	1	credit_card	1	81.16
7	3d7239c394a212faae122962df514ac7	1	credit_card	3	51.84
8	1f78449c87a54faf9e96e88ba1491fa9	1	credit_card	6	341.09
9	0573b5e23cbd798006520e1d5b4c6714	1	boleto	1	51.95

Lưu chi tiết các phương thức thanh toán, số lần thanh toán, và tổng giá trị thanh toán. Phân tích để hiểu xu hướng thanh toán, phổ biến phương thức thanh toán hoặc giá trị trung bình mỗi đơn hàng.

MÔ TẢ TỪNG BẢNG: GEOLOCATION

	geolocation_zip_code_prefix	geolocation_lat	geolocation_lng	geolocation_city	geolocation_state
0	1037	-23.545621	-46.639292	sao paulo	SP
1	1046	-23.546081	-46.644820	sao paulo	SP
2	1046	-23.546129	-46.642951	sao paulo	SP
3	1041	-23.544392	-46.639499	sao paulo	SP
4	1035	-23.541578	-46.641607	sao paulo	SP
5	1012	-23.547762	-46.635361	são paulo	SP
6	1047	-23.546273	-46.641225	sao paulo	SP
7	1013	-23.546923	-46.634264	sao paulo	SP
8	1029	-23.543769	-46.634278	sao paulo	SP
9	1011	-23.547640	-46.636032	sao paulo	SP
10	1013	-23.547325	-46.634184	sao paulo	SP
11	1032	-23.538418	-46.634778	sao paulo	SP
12	1014	-23.546435	-46.633830	sao paulo	SP
13	1012	-23.548946	-46.634671	sao paulo	SP
14	1037	-23.545187	-46.637855	são paulo	SP
15	1046	-23.546081	-46.644820	sao paulo	SP
16	1039	-23.541883	-46.639919	sao paulo	SP
17	1024	-23.541390	-46.629899	são paulo	SP
18	1009	-23.546935	-46.636588	sao paulo	SP
19	1046	-23.545884	-46.643163	sao paulo	SP
20	1046	-23.545166	-46.643786	sao paulo	SP
21	1020	-23.552235	-46.628441	são paulo	SP
22	1011	-23.546690	-46.635447	são paulo	SP
23	1043	-23.546657	-46.640788	são paulo	SP

Lưu trữ thông tin địa lý với tọa độ (latitude và longitude) dựa trên mã bưu điện. Dùng để phân tích vị trí giao hàng, tối ưu hóa lộ trình hoặc kết hợp với bản đồ để trực quan hóa.

MÔ TẢ TỪNG BẢNG: REVIEW

review_id	order_id	review_score	review_comment_title	review_comment_message	review_creation_date
7bc2406110b926393aa56f80a40eba40	73fc7af87114b39712e6da79b0a377eb	4	NaN	NaN	2018-01-18 00:00:00
80e641a11e56f04c1ad469d5645fdfde	a548910a1c6147796b98fdf73dbeba33	5	NaN	NaN	2018-03-10 00:00:00
228ce5500dc1d8e020d8d1322874b6f0	f9e4b658b201a9f2ecdecbb34bed034b	5	NaN	NaN	2018-02-17 00:00:00
e64fb393e7b32834bb789ff8bb30750e	658677c97b385a9be170737859d3511b	5	NaN	Recebi bem antes do prazo estipulado.	2017-04-21 00:00:00
f7c4243c7fe1938f181bec41a392bdeb	8e6bfb81e283fa7e4f11123a3fb894f1	5	NaN	Parabéns lojas lannister adorei comprar pela I...	2018-03-01 00:00:00
15197aa66ff4d0650b5434f1b46cda19	b18dcdf73be66366873cd26c5724d1dc	1	NaN	NaN	2018-04-13 00:00:00
07f9bee5d1b850860defd761afa7ff16	e48aa0d2dcec3a2e87348811bcfdf22b	5	NaN	NaN	2017-07-16 00:00:00
7c6400515c67679fbeef952a7525281ef	c31a859e34e3adac22f376954e19b39d	5	NaN	NaN	2018-08-14 00:00:00
a3f6f7f6f433de0aefbb97da197c554c	9c214ac970e84273583ab523dfaf09b	5	NaN	NaN	2017-05-17 00:00:00
8670d52e15e00043ae7de4c01cc2fe06	b9bf720beb4ab3728760088589c62129	4	recomendo	aparelho eficiente. no site a marca do aparelh...	2018-05-22 00:00:00
c9cf2d5ab5911836ababae136c3a10c	cdf9aa68e72324eeb25c7de974696ee2	5	NaN	NaN	2017-12-23 00:00:00
96052551d87e5f62e6c9f6974ec392e9	3d374c9e46530bb5ed4a7648915306a6	5	NaN	NaN	2017-12-19 00:00:00
4b49719c8a200003f700d3d986ea1a19	9d6f15f95d01e79bd1349cc208361f09	4	NaN	,travando... pelo valor ta Boa.\r\n	2018-02-16 00:00:00
23f75a37effc35d9a915b4e1ad483793	2eaf8e099d871cd5c22b83b5ea8f6e0e	4	NaN	NaN	2018-03-28 00:00:00
9a0abbb668bafb95a6d2b05db43284c4	d7bd0e4afdf94846eb73642b4e3e75c3	3	NaN	NaN	2017-04-30 00:00:00
3948b09f7c818e2d86c9a546758b2335	e51478e7e277a83743b6f9991dbfa3fb	5	Super recomendo	Vendedor confiável, produto ok e entrega antes...	2018-05-23 00:00:00
9314d6f9799f5bfa510cc7bcd468c01	0dacf04c5ad59fd5a0cc1faa07c34e39	2	NaN	GOSTARIA DE SABER O QUE HOUVE, SEMPRE RECEBI E...	2018-01-18 00:00:00

Bao gồm đánh giá từ khách hàng, với điểm đánh giá và thời gian nhận xét. Dùng để phân tích trải nghiệm khách hàng, xác định mức độ hài lòng hoặc lý do đánh giá thấp.

MÔ TẢ TỪNG BẢNG: PRODUCT_CATEGORY_TRANSLATE

	product_category_name	product_category_name_english
0	beleza_saude	health_beauty
1	informatica_acessorios	computers_accessories
2	automotivo	auto
3	cama_mesa_banho	bed_bath_table
4	moveis_decoracao	furniture_decor
5	esporte_lazer	sports_leisure
6	perfumaria	perfumery
7	utilidades_domesticas	housewares
8	telefonia	telephony
9	relogios_presentes	watches_gifts
10	alimentos_bebidas	food_drink
11	bebés	baby
12	papelaria	stationery
13	tablets_impressao_imagem	tablets_printing_image
14	brinquedos	toys
15	telefonia_fixa	fixed_telephony
16	ferramentas_jardim	garden_tools
17	fashion_bolsas_e_acessorios	fashion_bags_accessories
18	eletroportateis	small_appliances
19	consoles_games	consoles_games
20	audio	audio
21	fashion_calcados	fashion_shoes

Cung cấp bảng dịch các tên loại sản phẩm từ tiếng Bồ Đào Nha sang tiếng Anh.

Dùng để hiểu rõ ý nghĩa các loại sản phẩm và hỗ trợ phân tích nếu ngôn ngữ chính là tiếng Anh.

LÀM SẠCH VÀ XỬ LÝ DỮ LIỆU

```
raw_path = r"D:\data_du_an\\"

olist_customer = pd.read_csv(raw_path + 'olist_customers_dataset.csv')
olist_geolocation = pd.read_csv(raw_path + 'olist_geolocation_dataset.csv')
olist_orders = pd.read_csv(raw_path + 'olist_orders_dataset.csv')
olist_order_items = pd.read_csv(raw_path + 'olist_order_items_dataset.csv')
olist_order_payments = pd.read_csv(raw_path + 'olist_order_payments_dataset.csv')
olist_order_reviews = pd.read_csv(raw_path + 'olist_order_reviews_dataset.csv')
olist_products = pd.read_csv(raw_path + 'olist_products_dataset.csv')
olist_sellers = pd.read_csv(raw_path + 'olist_sellers_dataset.csv')
product_category_transaltion = pd.read_csv(raw_path + 'product_category_name_translation.csv')
```

Đọc tất cả table vào python

LÀM SẠCH VÀ XỬ LÝ DỮ LIỆU

```
datasets = {
    'olist_customer': olist_customer,
    'olist_geolocation': olist_geolocation,
    'olist_orders': olist_orders,
    'olist_order_items': olist_order_items,
    'olist_order_payments': olist_order_payments,
    'olist_order_reviews': olist_order_reviews,
    'olist_products': olist_products,
    'olist_sellers': olist_sellers,
    'product_category_translation' : product_category_tr
}
```

```
summary = []
##Điểm số dòng, số giá trị null và số cột của các bảng trên
for name, data in datasets.items():

    n_rows, n_cols = data.shape

    null_amount = data.isnull().sum().sum()

    qty_null_columns = (data.isnull().sum() > 0).sum()

    null_columns = data.columns[data.isnull().any()].tolist()

    summary.append({
        'dataset': name,
        'n_rows': n_rows,
        'n_cols': n_cols,
        'null_amount': null_amount,
        'qty_null_columns': qty_null_columns,
        'null_columns': ', '.join(null_columns)
    })

#tạo một dataframe để chứa các thông tin trên
summary_df = pd.DataFrame(summary)

summary_df
```

Kiểm tra số hàng, số cột và tình trạng các cột của các bảng

LÀM SẠCH VÀ XỬ LÝ DỮ LIỆU

	dataset	n_rows	n_cols	null_amount	qty_null_columns	null_colum
0	olist_customer	99441	5	0	0	
1	olist_geolocation	1000163	5	0	0	
2	olist_orders	99441	8	4908	3	order_approved_at, order_delivered_carrier_da
3	olist_order_items	112650	7	0	0	
4	olist_order_payments	103886	5	0	0	
5	olist_order_reviews	99224	7	145903	2	review_comment_title, review_comment_messa
6	olist_products	32951	9	2448	8	product_category_name, product_name_lenght, p
7	olist_sellers	3095	4	0	0	
8	product_category_translation	71	2	0	0	

Kết quả khi chạy các dòng code ở trang trước

LÀM SẠCH VÀ XỬ LÝ DỮ LIỆU

```
#chuyển sang kiểu dữ liệu datetime
olist_orders['order_purchase_timestamp'] = pd.to_datetime(olist_orders['order_purchase_timestamp'])
olist_orders['order_delivered_customer_date'] = pd.to_datetime(olist_orders['order_delivered_customer_date'])
olist_orders['order_estimated_delivery_date'] = pd.to_datetime(olist_orders['order_estimated_delivery_date'])
olist_orders['order_delivered_carrier_date'] = pd.to_datetime(olist_orders['order_delivered_carrier_date'])

#Tính thời gian vận chuyển
olist_orders['Shipping Days'] = olist_orders['order_delivered_customer_date'] - olist_orders['order_delivered_carrier_date']
olist_orders['Shipping Days'] = olist_orders['Shipping Days'].dt.days

#xác định trạng thái giao hàng dựa vào thời gian giao hàng dự kiến và thời gian giao hàng thực tế
def Delivery_status(row):
    day_diff = row['order_estimated_delivery_date'] - row['order_delivered_customer_date']
    if day_diff.days >= 0:
        return 'On time'
    else:
        return 'Late'

# Áp dụng hàm cho từng dòng
olist_orders['delivery_status'] = olist_orders.apply(Delivery_status, axis=1)

#Xử lý các hóa đơn từ bảng order mà không có tồn tại trong bảng order_item_id
olist_orders = olist_orders[olist_orders['order_id'].isin(olist_order_items['order_id'])]
```

Chuyển đổi kiểu dữ liệu các cột ngày đặt hàng, giao hàng , nhận hàng thành Datetime

Tính thời gian giao hàng, dựa vào ngày giao hàng dự kiến để xác định trạng thái giao hàng

Xóa bỏ các đơn hàng không tồn tại trong bảng order_item_id

LÀM SẠCH VÀ XỬ LÝ DỮ LIỆU

Kiểm tra lại kiểu dữ liệu

```

<class 'pandas.core.frame.DataFrame'>
Index: 98666 entries, 0 to 99440
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   order_id         98666 non-null   object  
 1   customer_id      98666 non-null   object  
 2   order_status      98666 non-null   object  
 3   order_purchase_timestamp 98666 non-null   datetime64[ns]
 4   order_approved_at 98652 non-null   object  
 5   order_delivered_carrier_date 97657 non-null   datetime64[ns]
 6   order_delivered_customer_date 96476 non-null   datetime64[ns]
 7   order_estimated_delivery_date 98666 non-null   datetime64[ns]
 8   Shipping Days     96475 non-null   float64 
 9   delivery_status    98666 non-null   object  
 10  dayofweek        98666 non-null   object  
 11  hour             98666 non-null   int32  
dtypes: datetime64[ns](4), float64(1), int32(1), object(6)
memory usage: 9.4+ MB

```

2 cột mới được thêm vào

	Shipping Days	delivery_status
0	6.0	On time
1	12.0	On time
2	9.0	On time
3	9.0	On time
4	1.0	On time
...
99436	7.0	On time
99437	20.0	On time
99438	23.0	On time
99439	13.0	On time
99440	6.0	On time

BRAZILIAN E-COMMERCE ANALYSIS

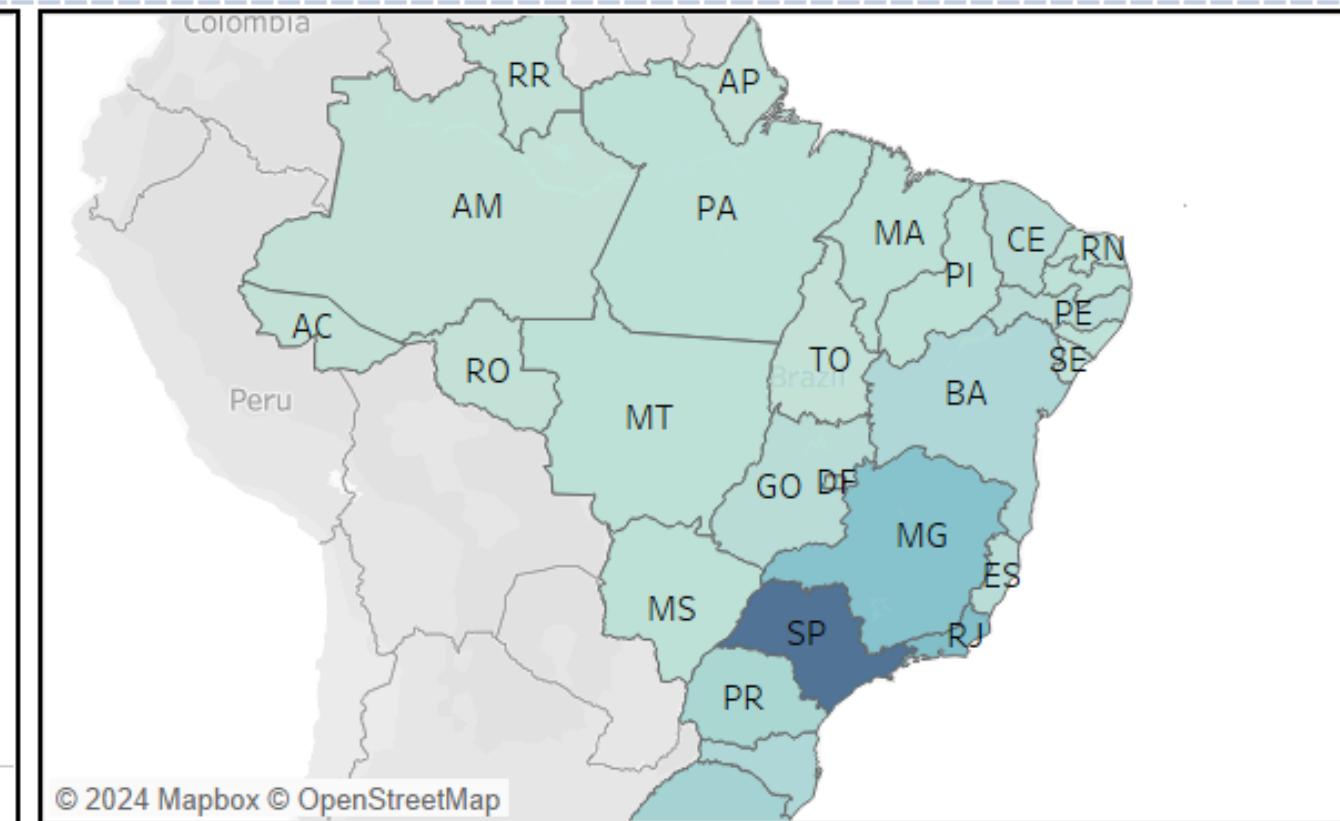
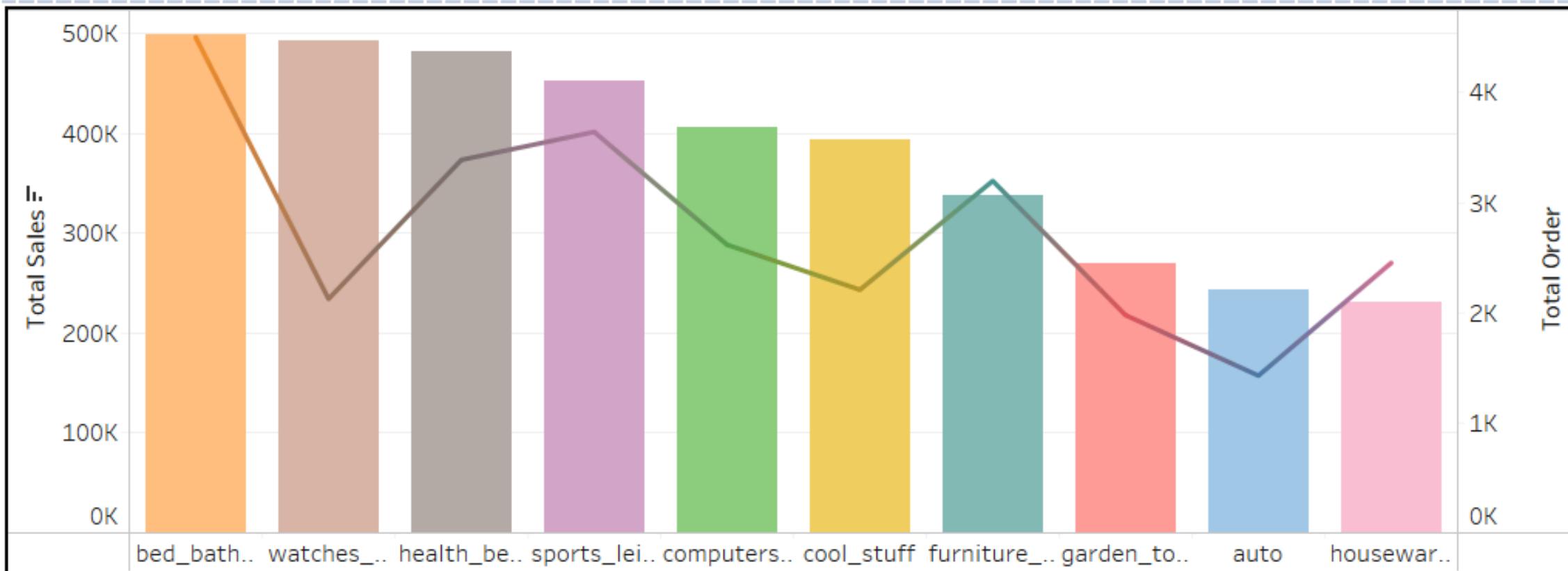
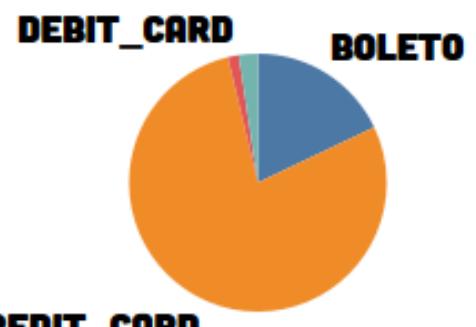
ORDERS
99,441

SELLERS
3,095

TOTAL SALES
13,591,644

Avg Freight
19.99

Avg Timeship
301.4



© 2024 Mapbox © OpenStreetMap

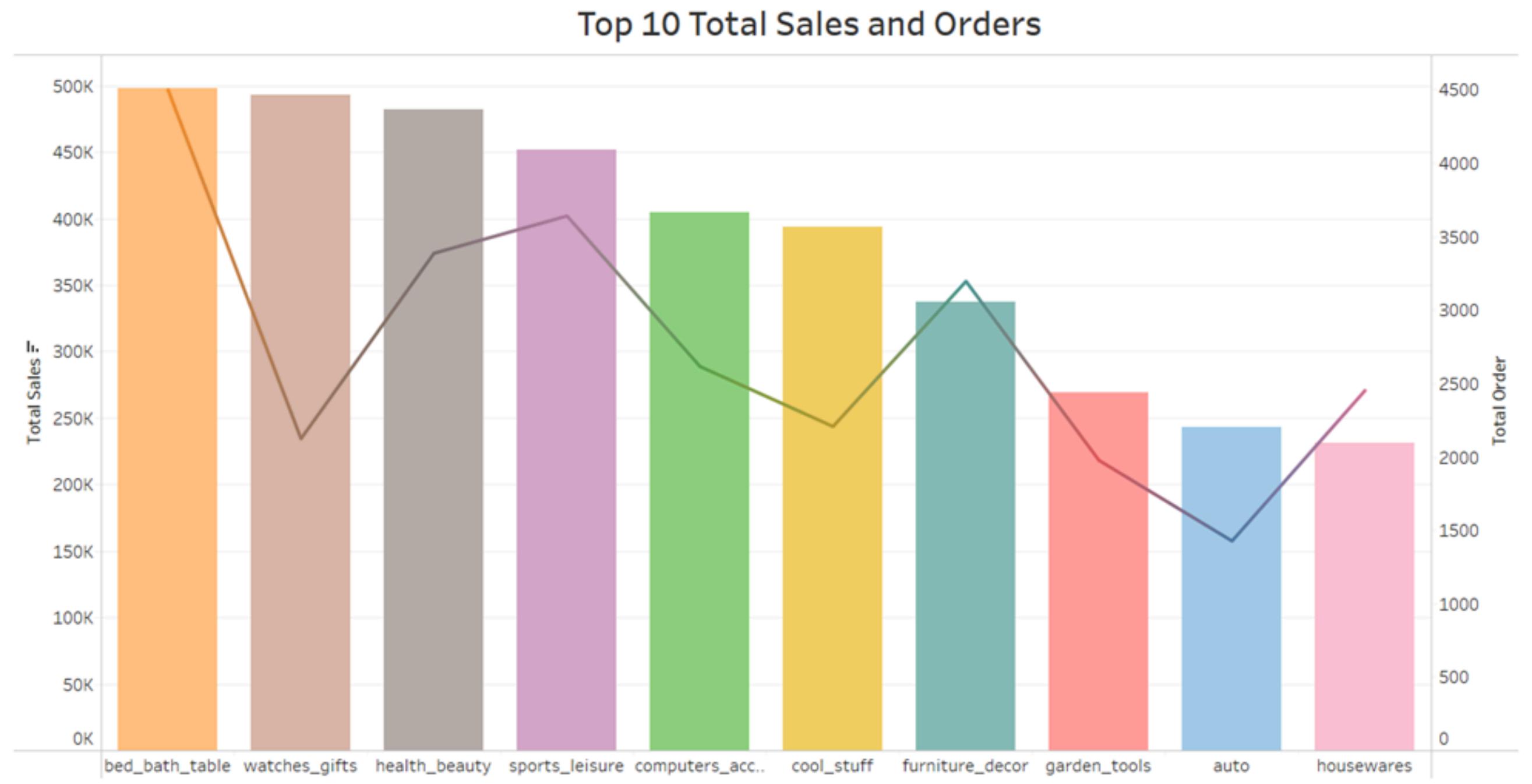
ĐỀ XUẤT KINH DOANH

Chương 1 Doanh thu của sàn thương mại

Về **sức khỏe** của Sàn Thương Mại và các Vấn
đề liên quan đến doanh thu

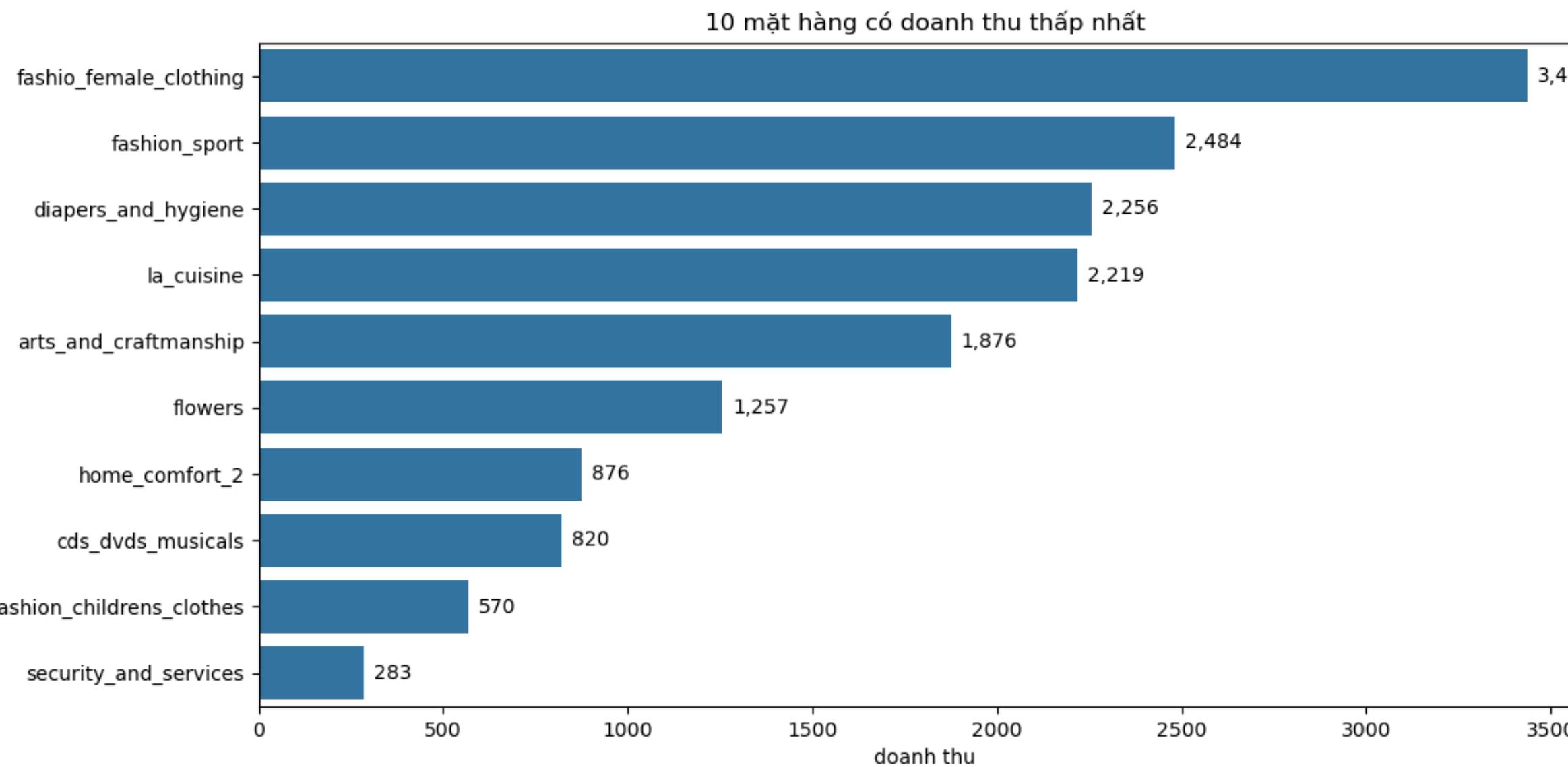


Chương 1: Doanh thu của sàn thương mại



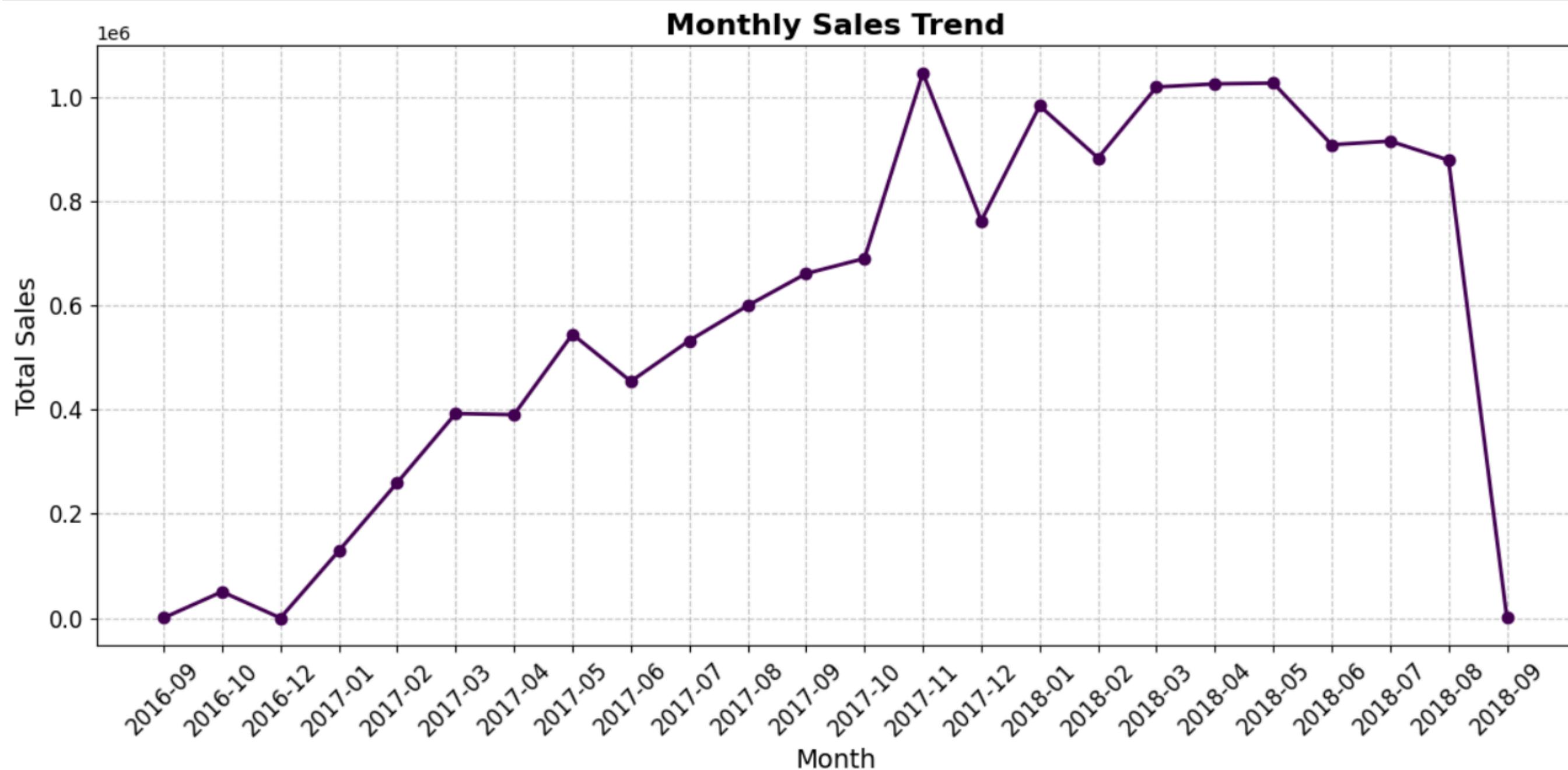
Top các sản phẩm bán chạy trong 3 năm, cùng với đường Line là tổng lượng đơn hàng trong từng danh mục

Chương 1: Doanh thu của sàn thương mại



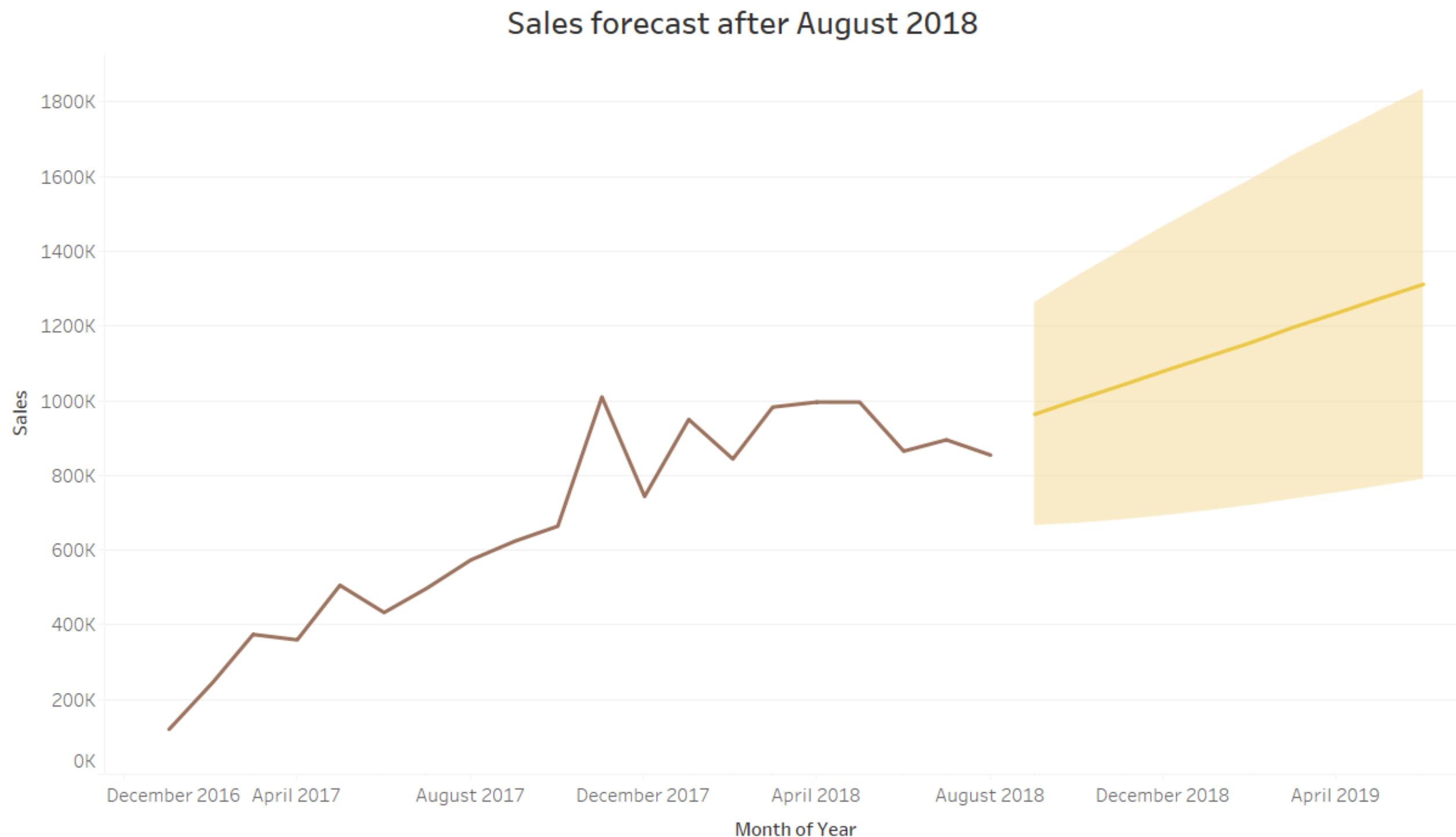
Top các sản phẩm bán không chạy trong 3 năm

Chương 1: Doanh thu của sàn thương mại



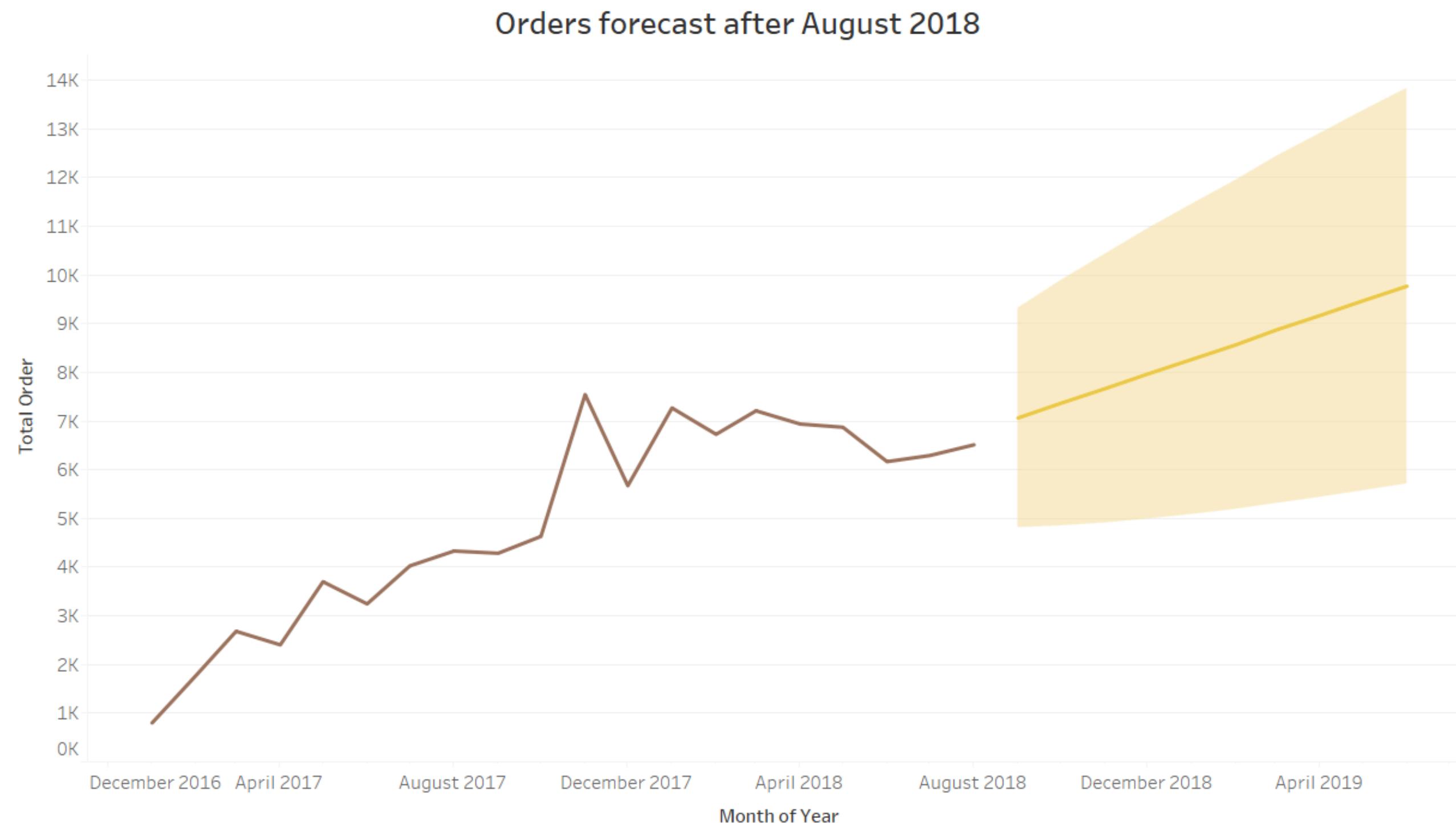
Biểu đồ thể hiện doanh thu bán hàng theo thời gian từ thời điểm 8/2016 đến 8/2018. Cho thấy doanh thu từ giữa đến cuối năm 2017 có sự tăng trưởng đáng kể.

Chương 1: Doanh thu của sàn thương mại



Dựa vào doanh thu từ của tháng trong năm 2016 đến tháng 8 năm 2018, có thể dự đoán doanh thu các tháng tiếp
sẽ diễn biến theo xu hướng như nào. Kết quả thu được là sự tăng trưởng ổn định đối với tình trạng doanh thu

Chương 1: Doanh thu của sàn thương mại



Dự đoán đơn hàng từ 8/2018 cho đến 4/2019

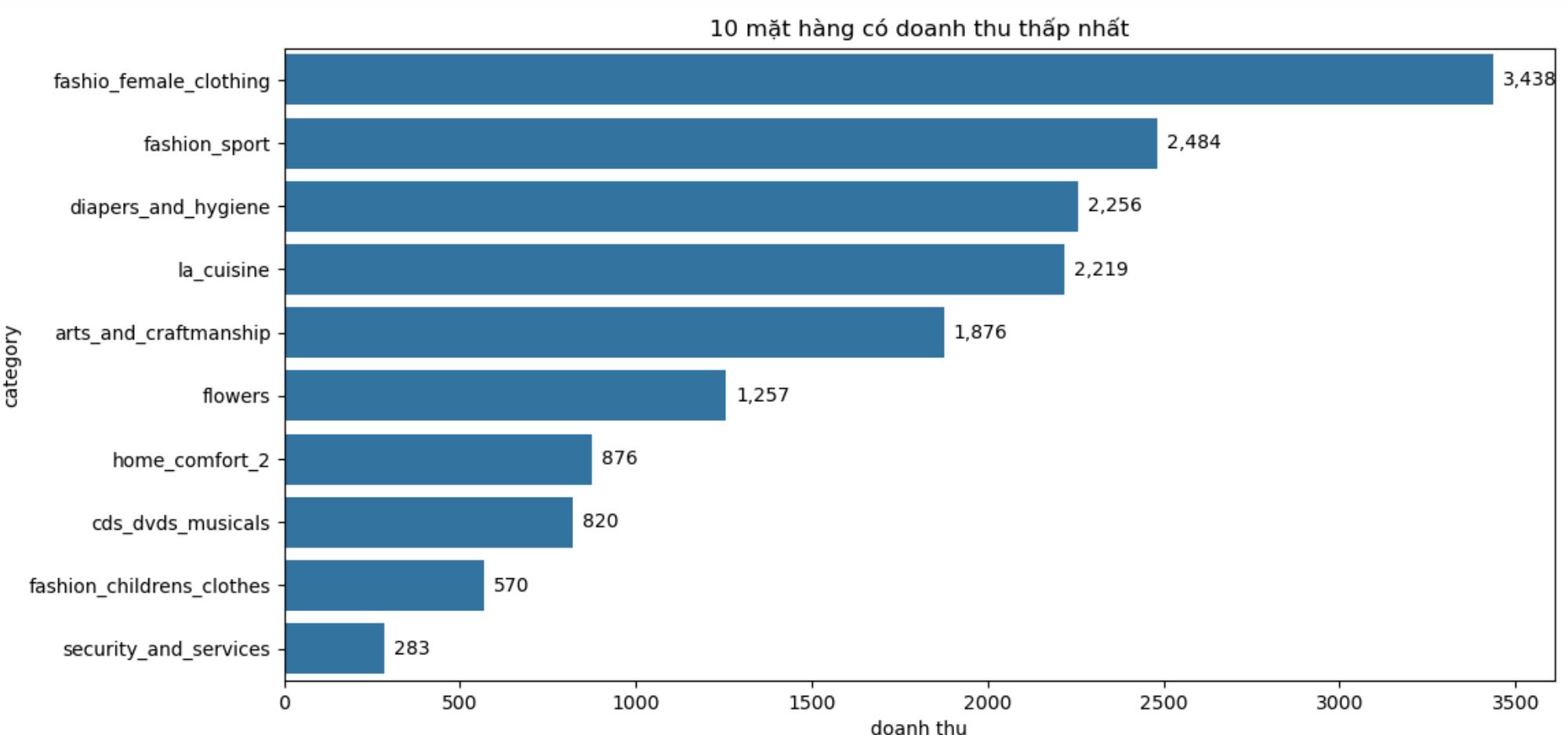
Vấn đề cần giải quyết:

Sản phẩm có doanh thu và lượt bán thấp:

- Sản phẩm không đáp ứng nhu cầu, chất lượng chưa nổi bật, hoặc thiếu chiến dịch quảng bá.
- Cạnh tranh cao từ đối thủ với sản phẩm tương tự

Tăng trưởng doanh thu trong các tháng thấp điểm:

- Thiếu chiến dịch kích cầu hoặc sự kiện hấp dẫn.
- Chưa khai thác hết đối tượng khách hàng mục tiêu, lãng phí thị trường tiềm năng.



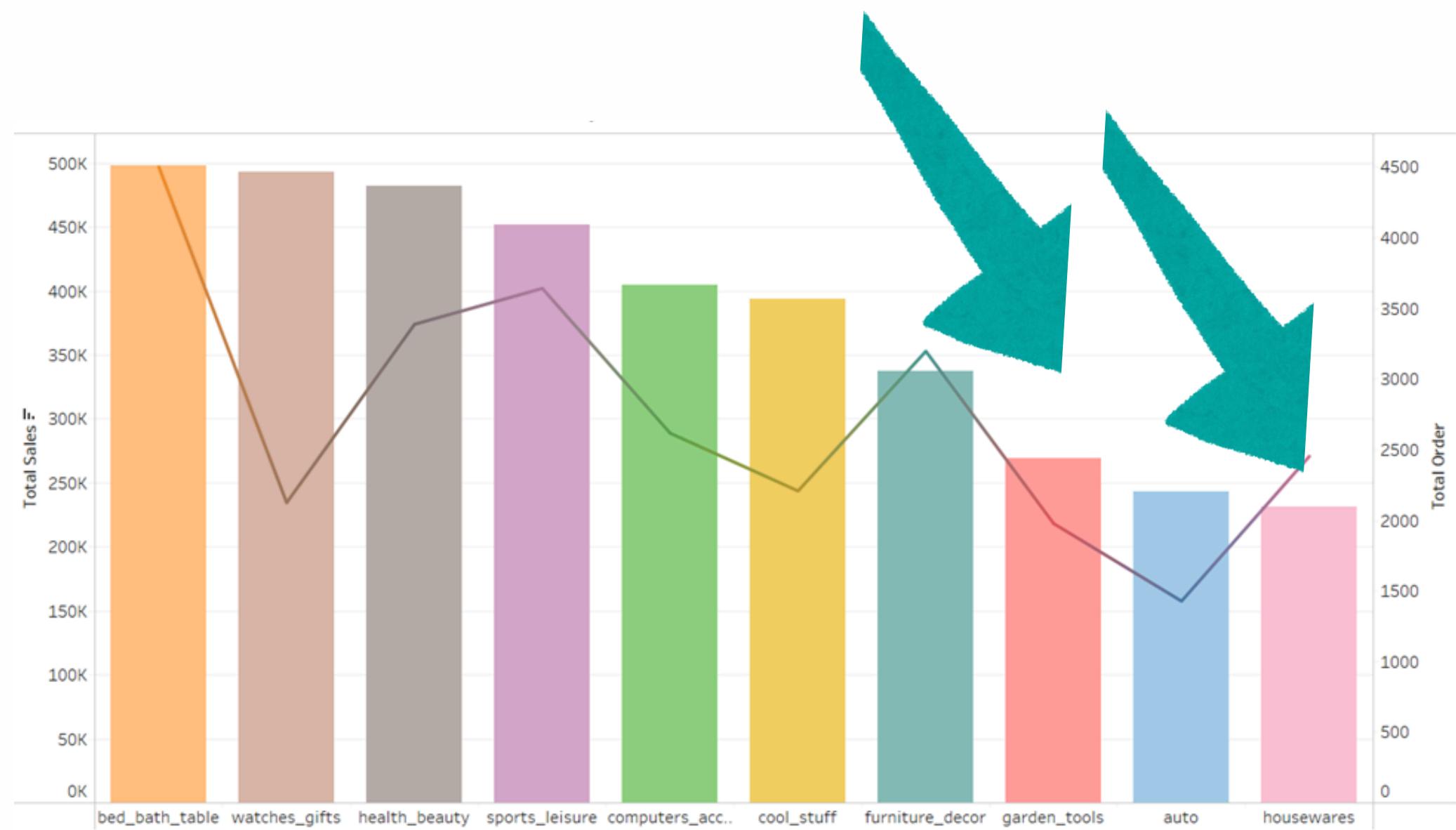
Vấn đề cần giải quyết:

Sản phẩm có doanh thu và lượt bán cao:

- Sản phẩm gần gũi, chất lượng tốt, được nhiều khách hàng biết tới.
- Có nhiều chương trình ưu đãi hấp dẫn từ người bán.

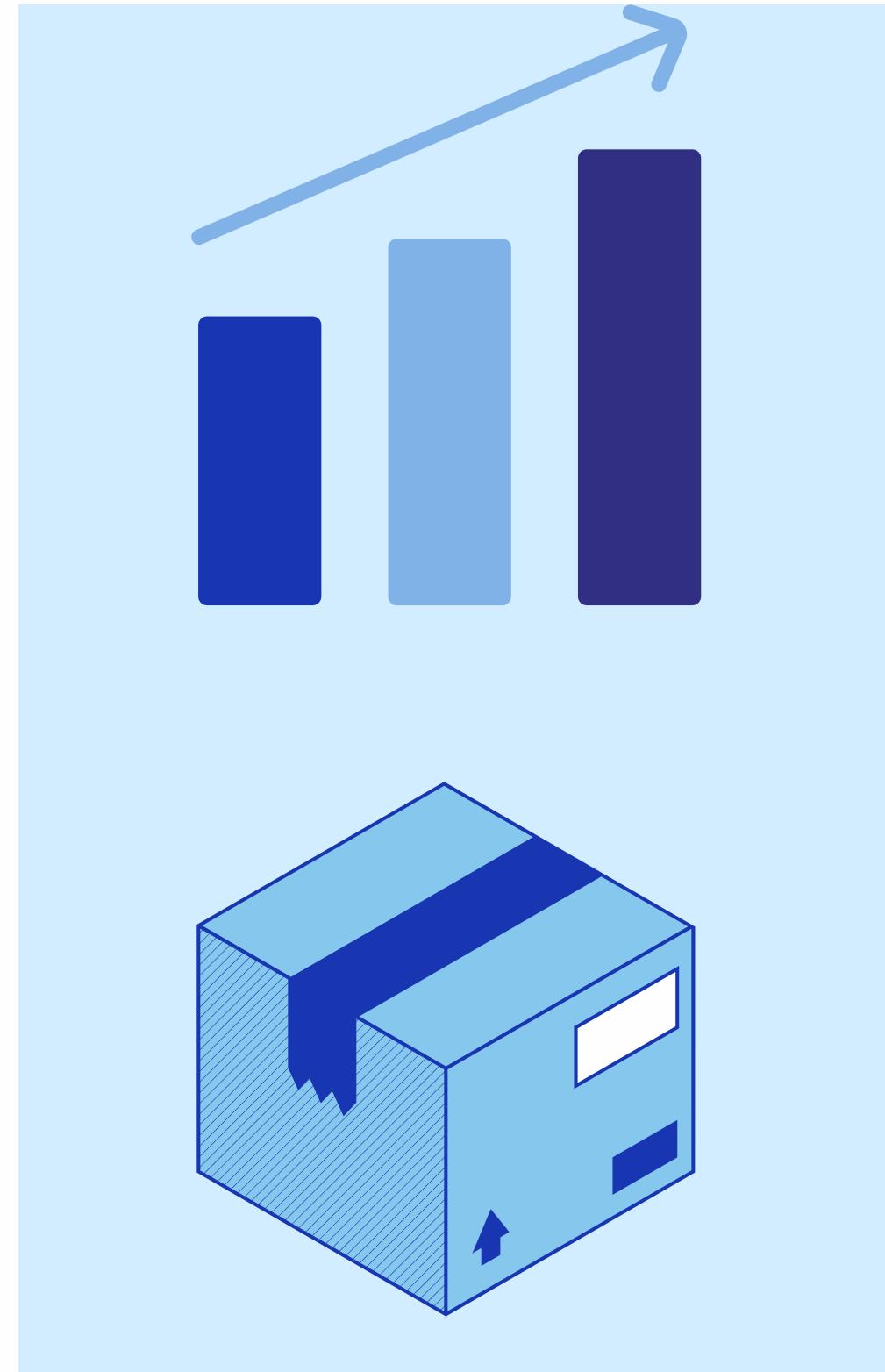
Tăng trưởng doanh thu hơn nữa của các sản phẩm bán chạy:

- Tiếp tục quản bá các sản phẩm bán chạy, dùng chính khách hàng sử dụng sản phẩm để Marketing cho chính thương hiệu.
- Khai thác thêm từ khách hàng cũ .



Hướng giải quyết

- **Nâng cấp sản phẩm:** Cải tiến chất lượng, dùng AI phân tích, quảng bá qua Ads và Influencers.
- **Nâng cao dịch vụ:** Tích hợp chatbot, đổi trả linh hoạt, giao hàng nhanh miễn phí.
- **Đa dạng hóa:** Bổ sung sản phẩm mới, mở rộng thị trường, thử nghiệm Amazon, eBay.
- **Tăng khuyến mãi:** Flash Sale, combo giảm giá, tháng khuyến mãi, tri ân khách hàng.



CHƯƠNG 2 Cơn đau đầu mang tên

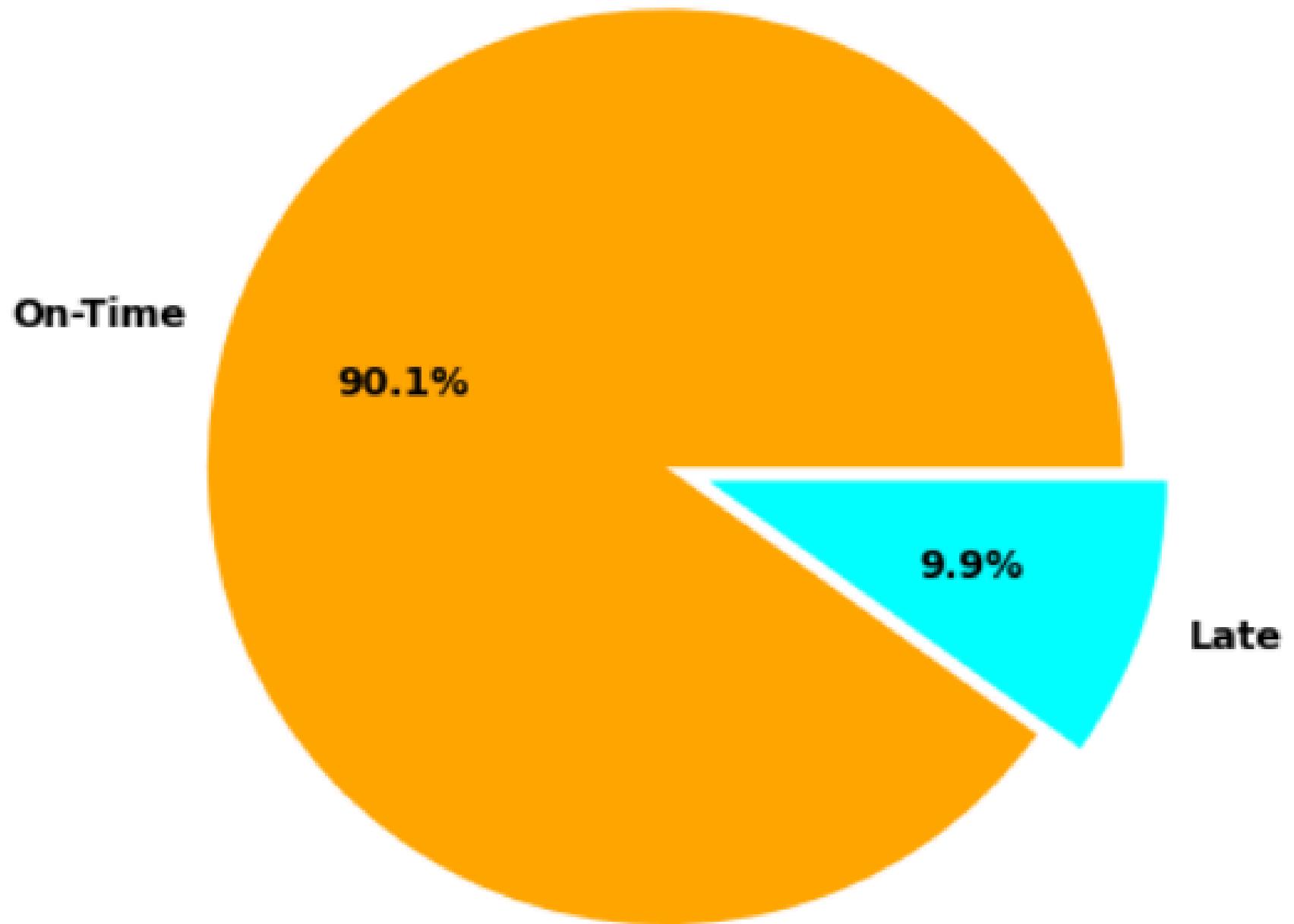
"GIAO HÀNG TRỄ HẠN"

Tỉ lệ giao hàng nhanh nhất và chậm nhất là bao nhiêu? Tại sao có sự chênh lệch này?



Chương 2: Cơn đau đầu mang tên "giao hàng trễ hạn"

Delivery Performance



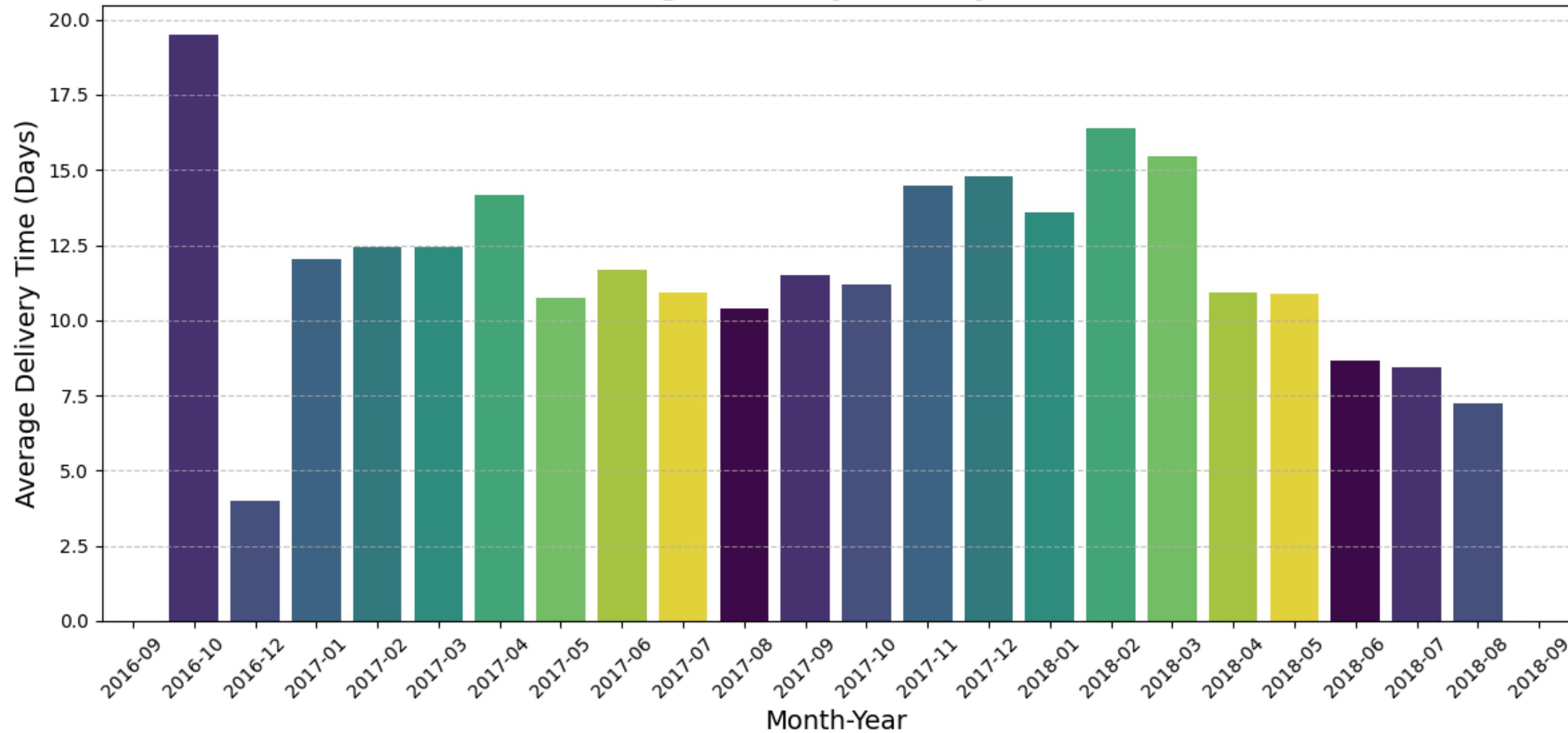
Tỷ lệ giao hàng đúng hạn so với giao hàng trễ

Thông tin chi tiết

Tổng số đơn hàng	99.441
Thời gian giao trung bình	216h
Tổng đơn giao muộn	8,949

Chương 2: Cơn đầu đầu mang tên "giao hàng trễ hạn"

Average Delivery Time by Month

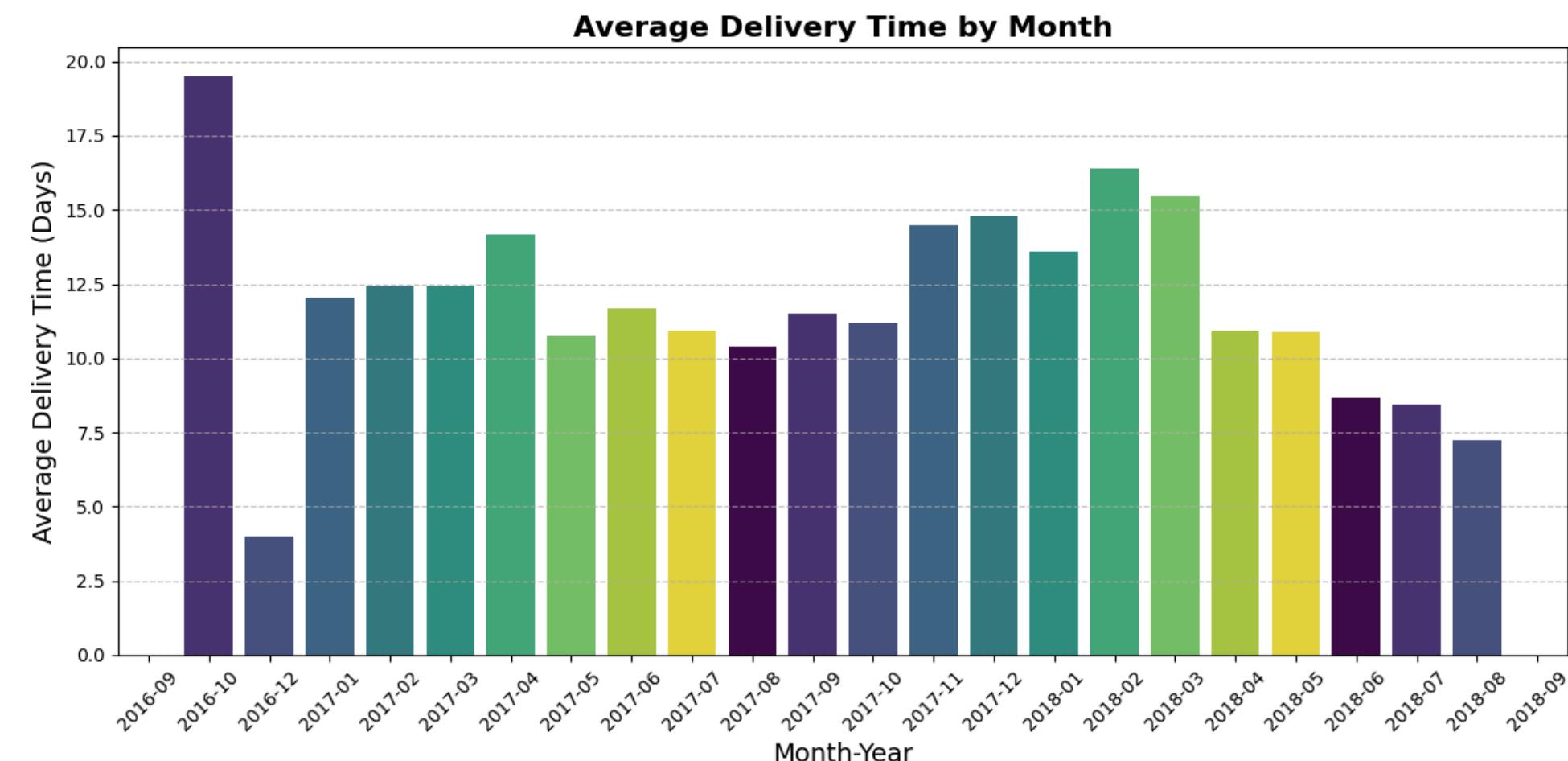


Biểu đồ thể hiện thời gian giao hàng trung bình trong từng thời điểm của năm

Vấn đề cần giải quyết:

Xu hướng tổng quan:

- Thời gian giao hàng trung bình đạt đỉnh vào tháng 9/2016 (hơn 50 ngày). Có thể do sự quá tải hoặc vấn đề vận hành nội bộ.
- Sau tháng 10/2016, thời gian giao hàng giảm và ổn định, thường từ 10–15 ngày.
- Đầu năm 2018 (tháng 2/2018) tăng nhẹ, có thể do cao điểm mua sắm lễ hội.



Hướng giải quyết

- Xác định các giai đoạn cao điểm
- Đảm bảo hàng hóa luôn sẵn có
- Xem xét việc ký hợp đồng với đơn vị giao hàng nhanh
- Sử dụng hệ thống định vị hoặc theo dõi đơn hàng
- Chủ động thông báo cho khách hàng về thời gian giao hàng
- Áp dụng các ưu đãi, giảm giá
- Theo dõi thời gian giao hàng trung bình



CHƯƠNG 3

Các phương thức thanh toán

Phương thức thanh toán nào

được sử dụng nhiều nhất ?

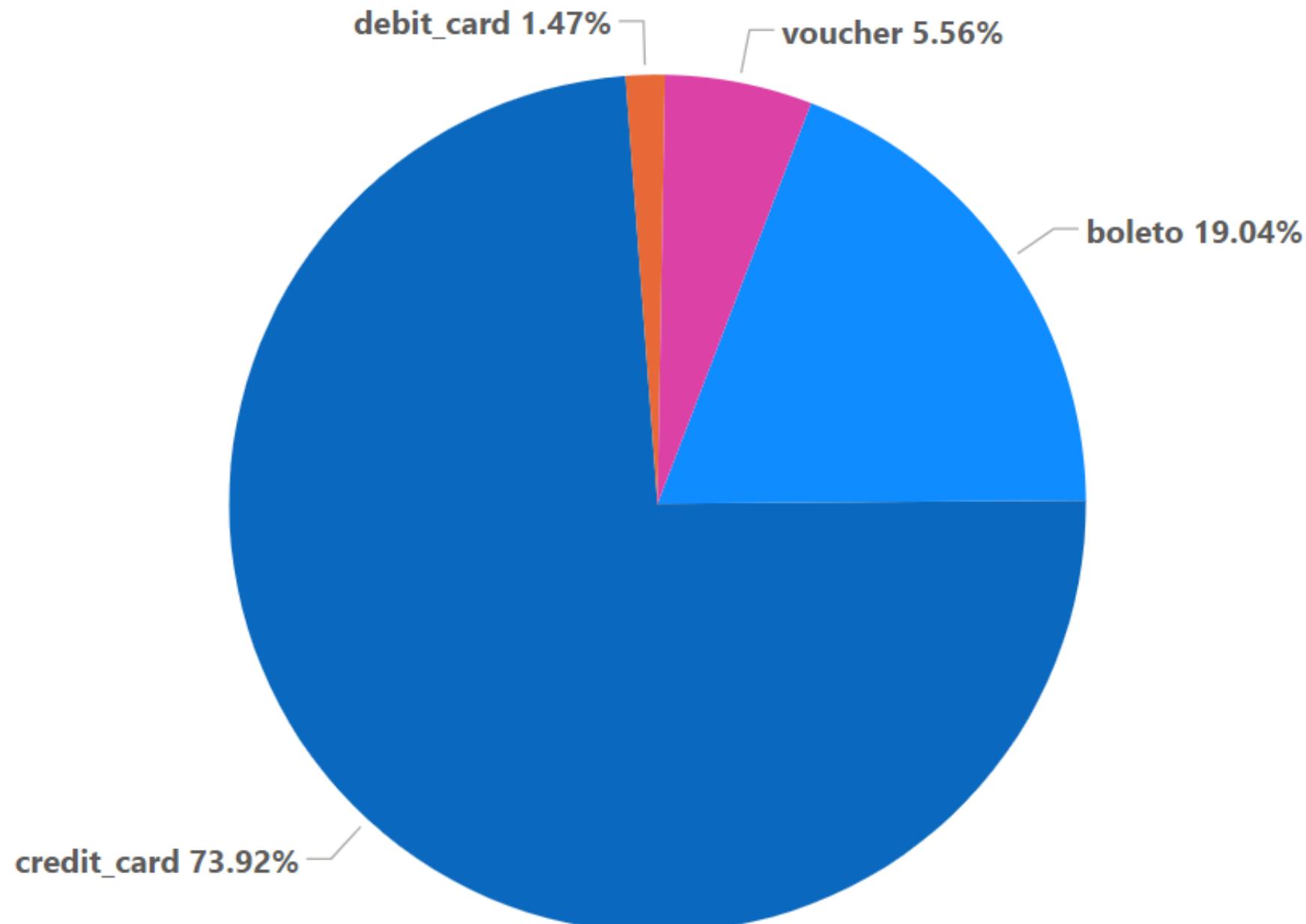
Độ an toàn của các phương

thức ra sao ?

TIỀM NĂNG



Chương 3: Về phương thức thanh toán

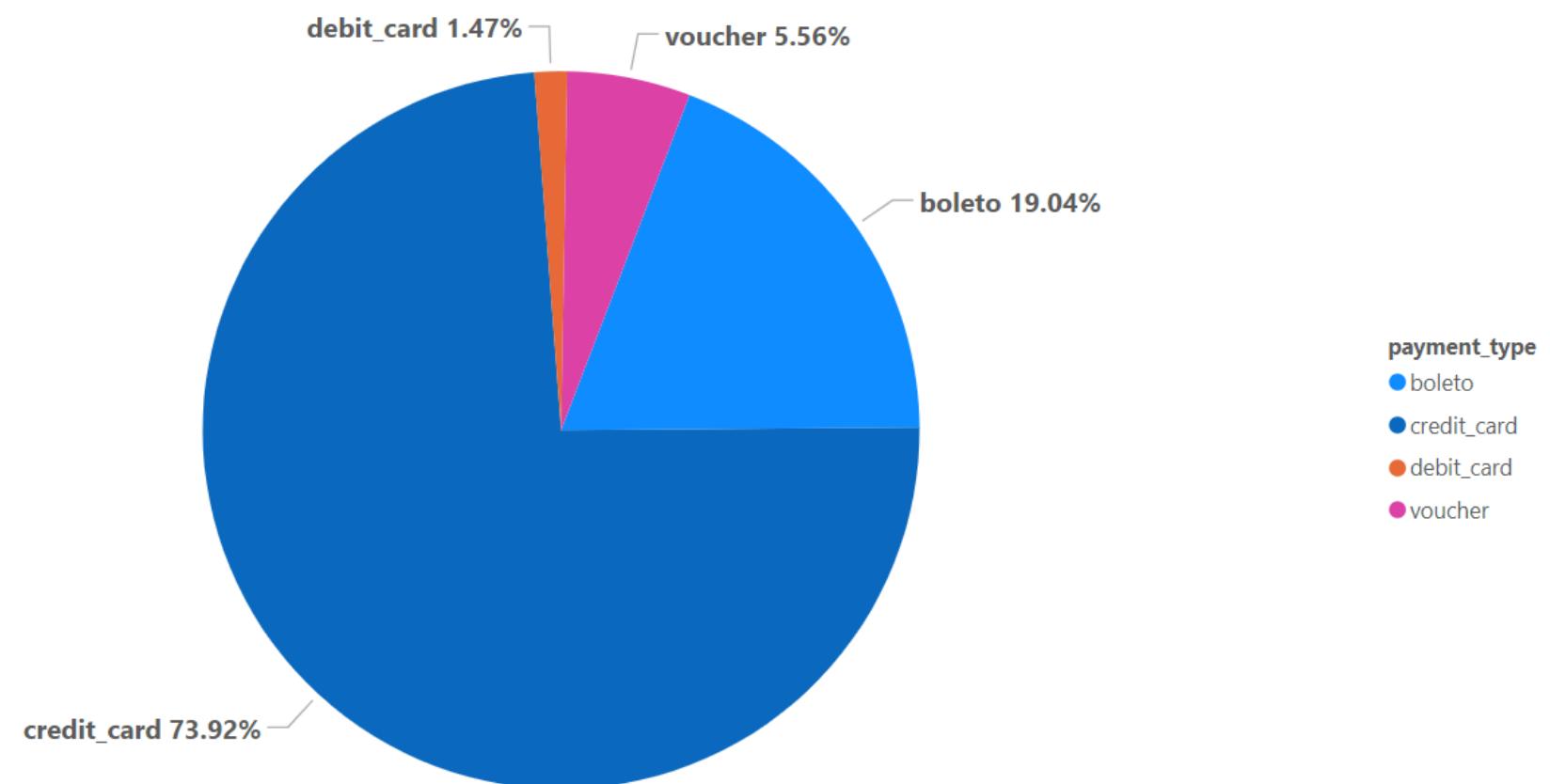


Payment_type_count

credit_card	76795
boleto	19784
voucher	5775
debit_card	1529

Name: count, dtype: int64

Biểu đồ này cho thấy sự phân phối các loại hình thanh toán của khách hàng



Vấn đề cần giải quyết:

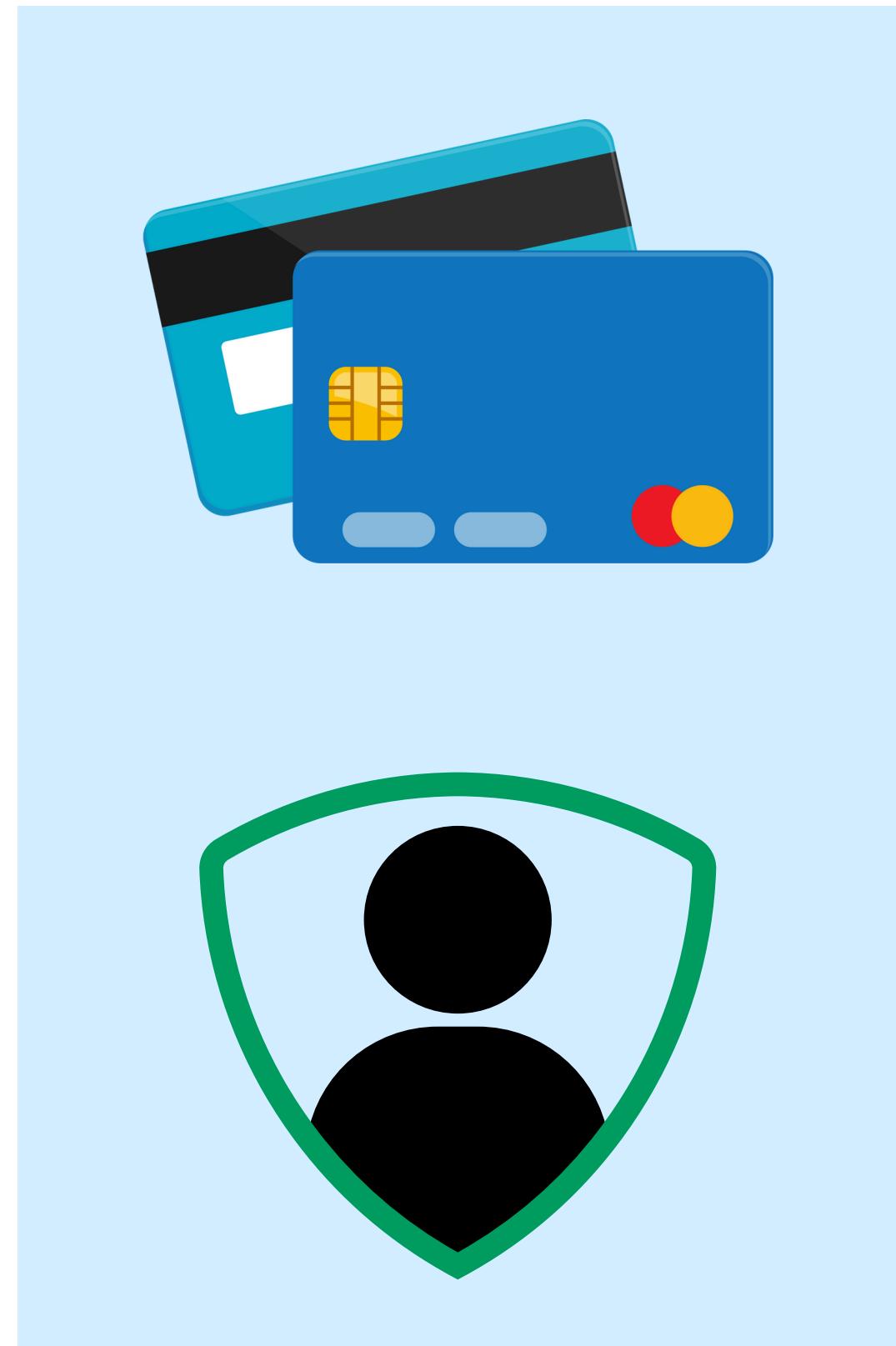
- Đa dạng hóa phương thức thanh toán (Châu Âu, Châu Á)
- Rủi ro an toàn của các phương thức thanh toán

Hướng giải quyết

- Đa dạng hóa phương thức thanh toán (Châu Âu và Châu Á)

Thanh toán qua ví điện tử, thanh toán địa phương

- Tăng cường an toàn
- Tối ưu trải nghiệm thanh toán
- Hướng dẫn khách hàng về an toàn thanh toán



Chương 4 Tận dụng tối đa

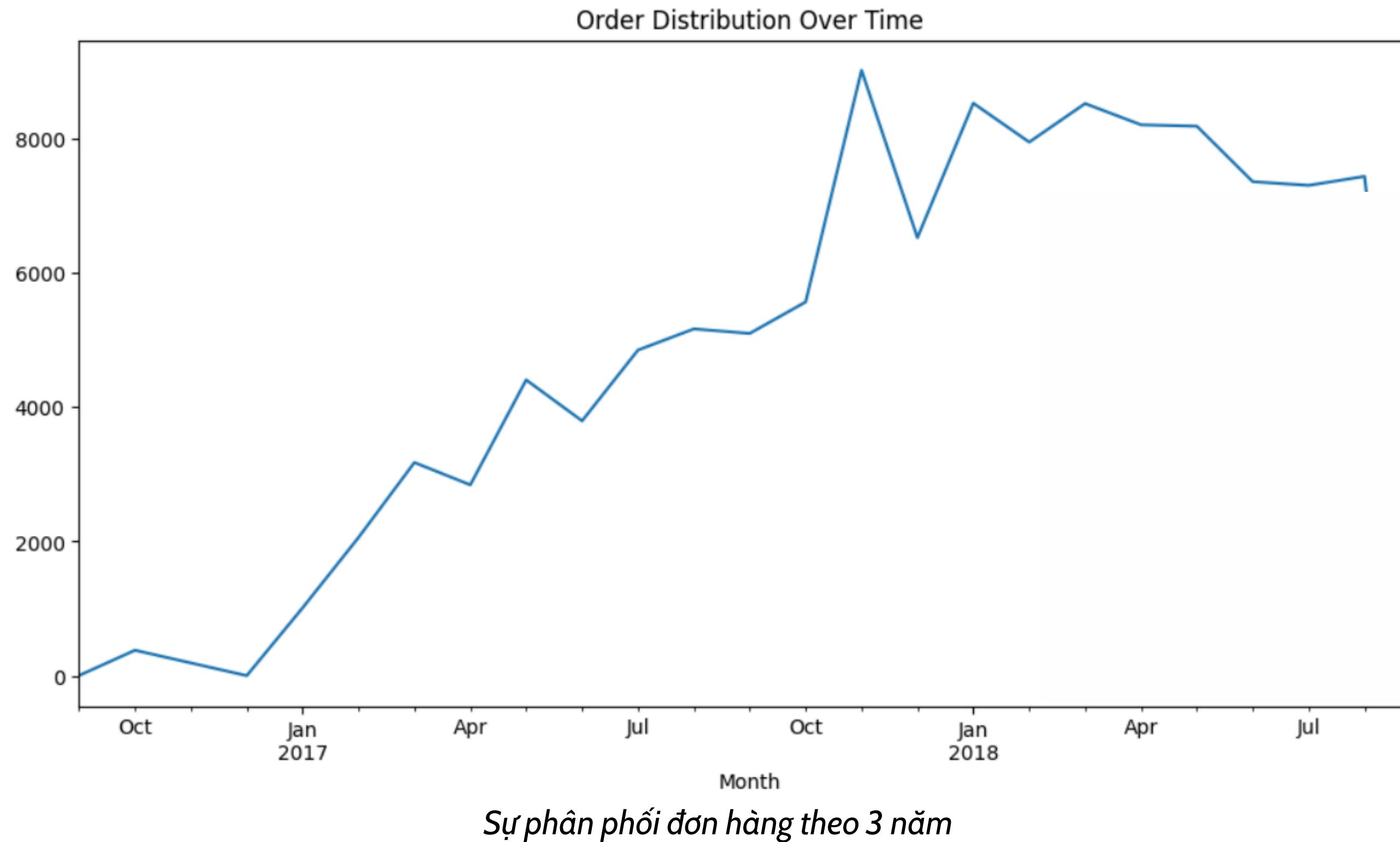
những “NGÀY VÀNG” trong
tháng

Thời gian nào thu hút nhiều khách hàng nhất

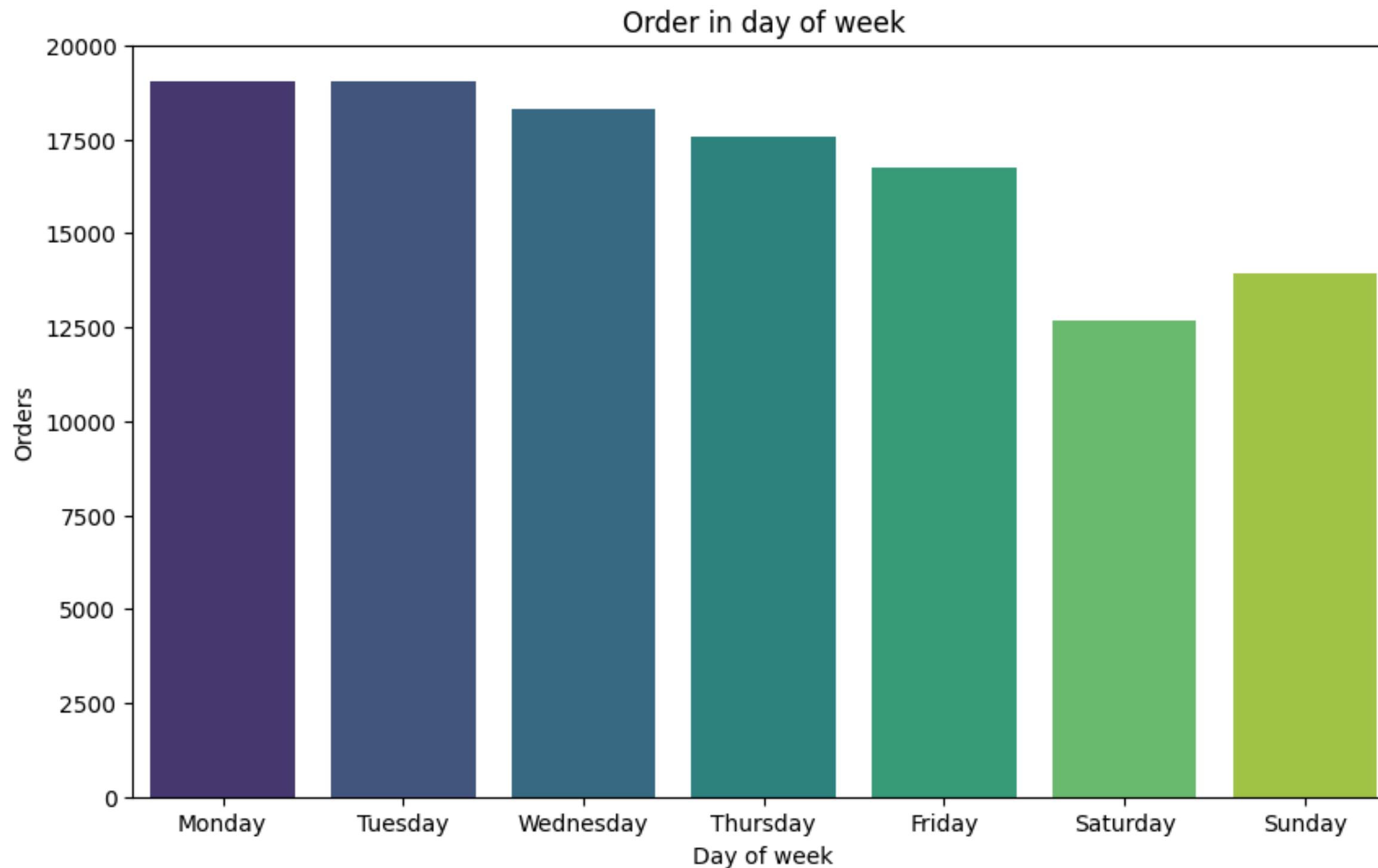
trong tháng? Và lý do đến từ yếu tố nào?



Chương 4: Những "ngày vàng" trong tháng

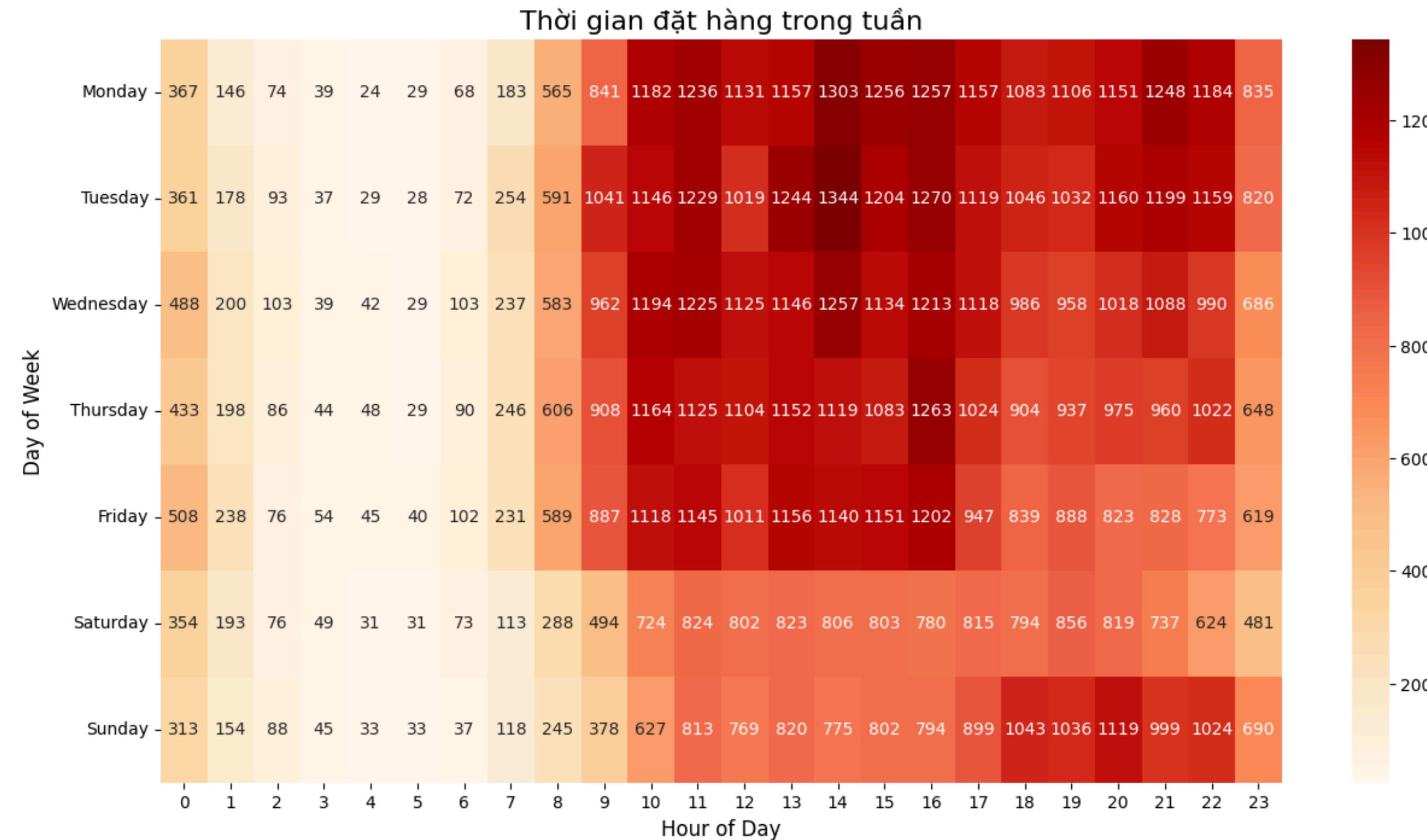


Chương 4: Những "ngày vàng" trong tháng và sự chuẩn bị cần thiết.



Biểu đồ phân phối đơn hàng theo từng ngày trong tuần

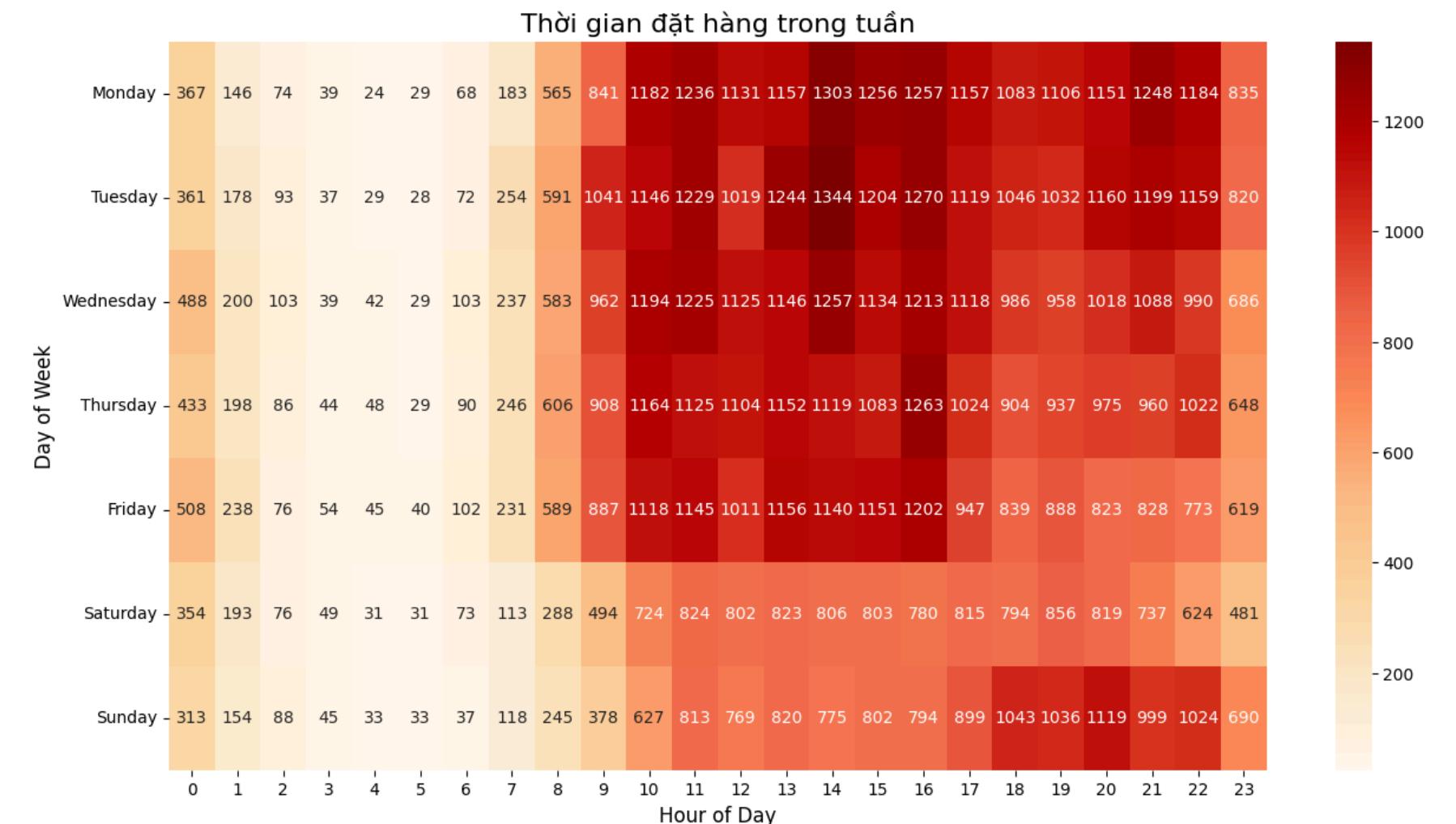
Chương 4: Những "ngày vàng" trong tháng.



Biểu đồ nhiệt hiển thị số đơn hàng theo giờ và ngày trong tuần. Màu sắc đậm biểu thị thời gian cao điểm.

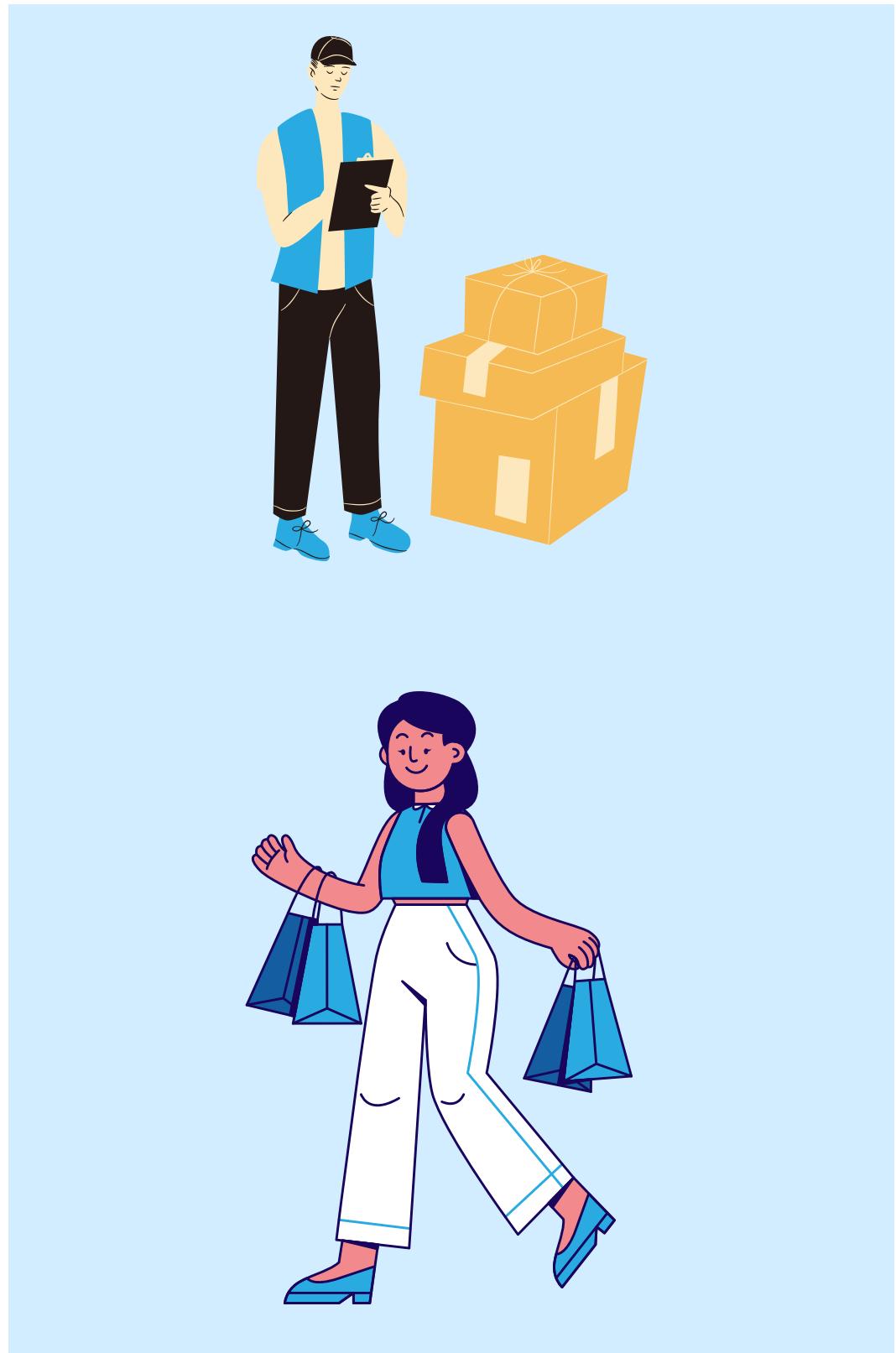
Vấn đề cần giải quyết:

- Số lượng đặt hàng giảm vào cuối tuần
- Thời gian đặt hàng từ sau 9h đến 22h tăng cao trong ngày
- Tốc độ đơn hàng tăng cao vào những giờ cao điểm trong tuần



Hướng giải quyết

- Tăng cường khuyến mãi vào cuối tuần
- Khung giờ vàng mua sắm (từ 9h-22h)
- Tăng tốc độ xử lý và chất lượng đơn hàng
- Đẩy mạnh quảng bá thương hiệu
- Tích hợp công nghệ và trải nghiệm người dùng

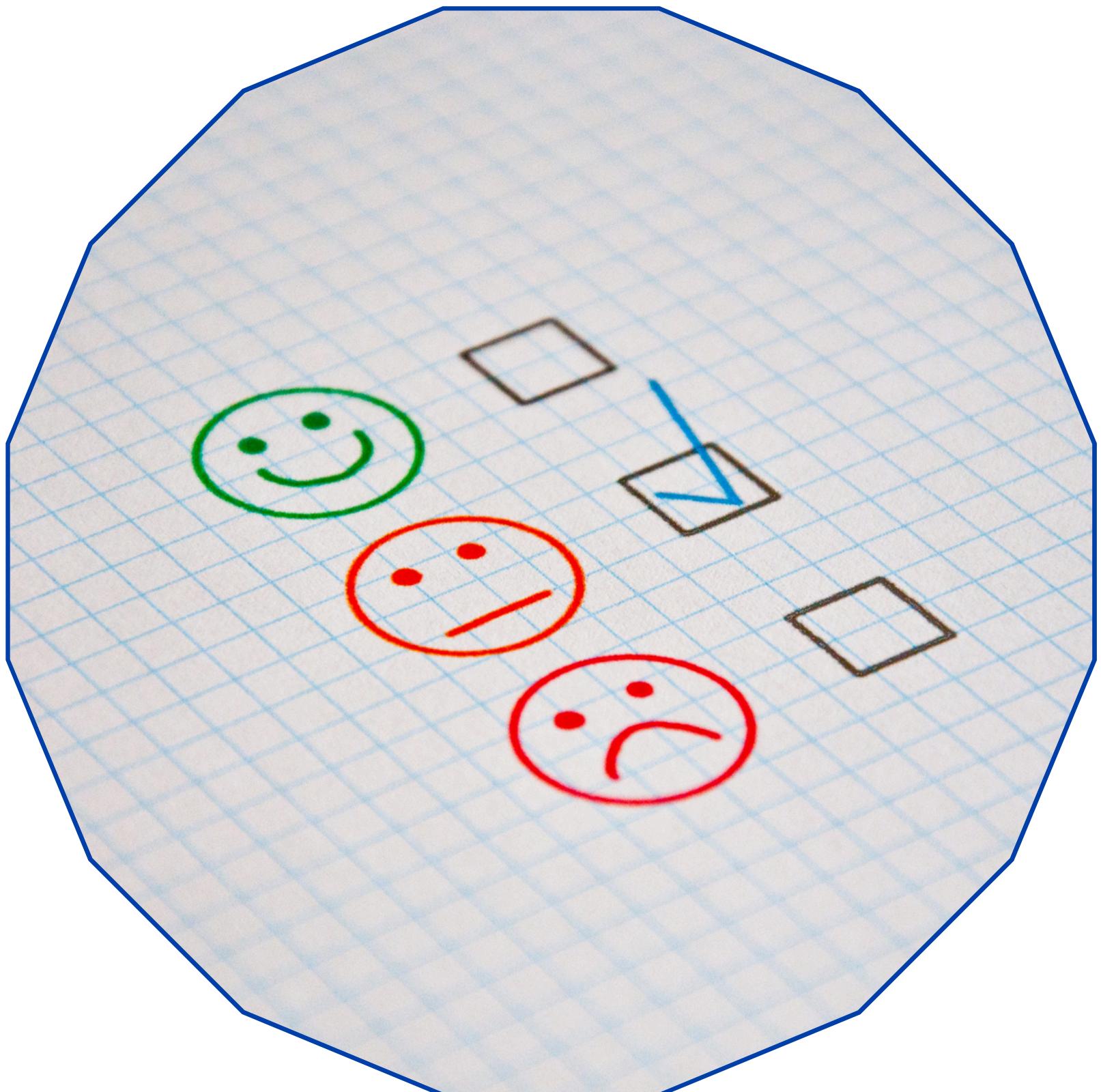


CHƯƠNG 5

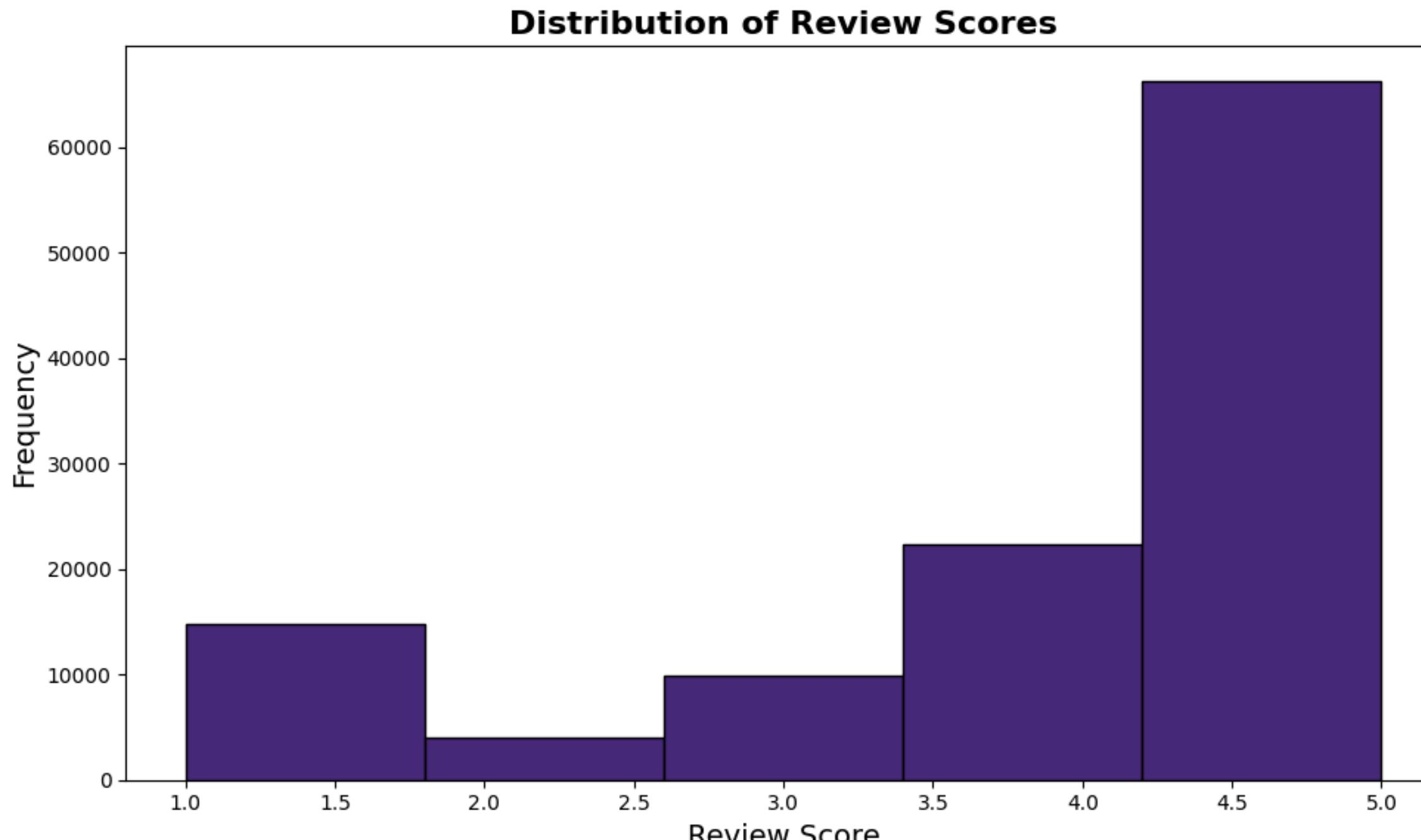
Khách hàng **TRẢI NGHIỆM** và **CẢM NHẬN** như thế nào?

Nơi phản ánh chân thật nhất về chất lượng và dịch vụ

của từng sản phẩm



Chương 5 : Đánh giá sự hài lòng của khách hàng

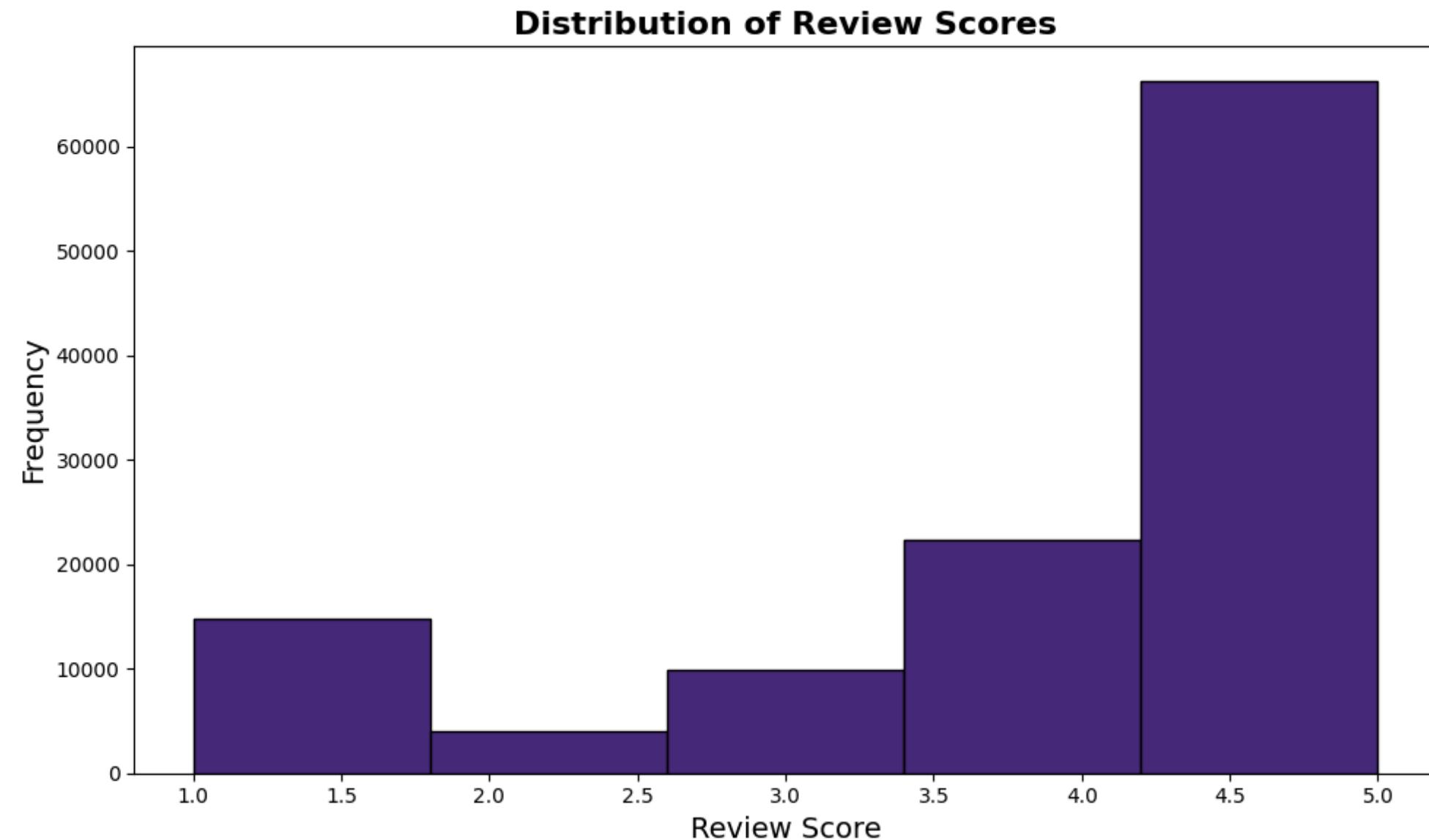


Biểu đồ phân phối đánh giá của khách hàng trên thang điểm 5

Chương 5 : Đánh giá sự hài lòng của khách hàng

Vấn đề cần giải quyết:

- Tổng số đánh giá thấp là con số không hề nhỏ
- Nguyên nhân:
 - 1.Thời gian giao hàng chậm
 - 2.Chất lượng sản phẩm không như mong đợi
 - 3.Dịch vụ hỗ trợ khách hàng yếu
 - 4.Kỳ vọng cao



Chương 5: Đánh giá sự hài lòng của khách hàng

Hướng giải quyết

- Siết chặt quy trình logistics, thời gian giao hàng
- Cải thiện hỗ trợ khách hàng
- Cải thiện trải nghiệm sản phẩm và dịch vụ
- Xây dựng lòng tin với khách hàng



Chương 6

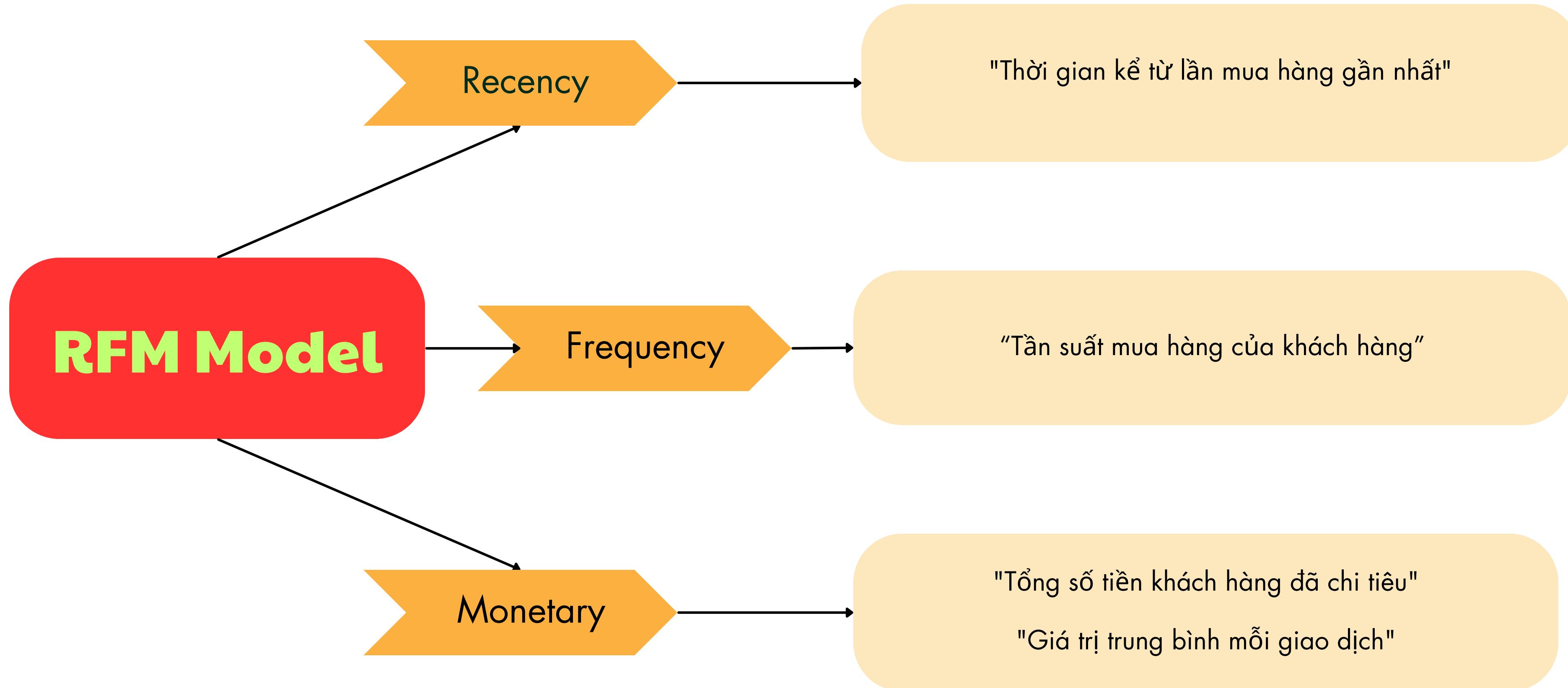
Tối ưu hóa doanh thu hiệu quả thông qua

“PHÂN KHÚC KHÁCH HÀNG”

Làm thế nào để **xác định** được **đối tượng khách hàng** và đưa **giải pháp tiếp cận** đối với từng tệp khách hàng



Mô hình RFM



TIÊU CHÍ PHÂN LOẠI

1

Khách hàng mới: Lần mua hàng gần đây, tiềm năng cho việc tương tác và tiếp thị tiếp theo.

2

Khách hàng tiềm năng: Khách hàng có sự tương tác liên tục với doanh nghiệp.

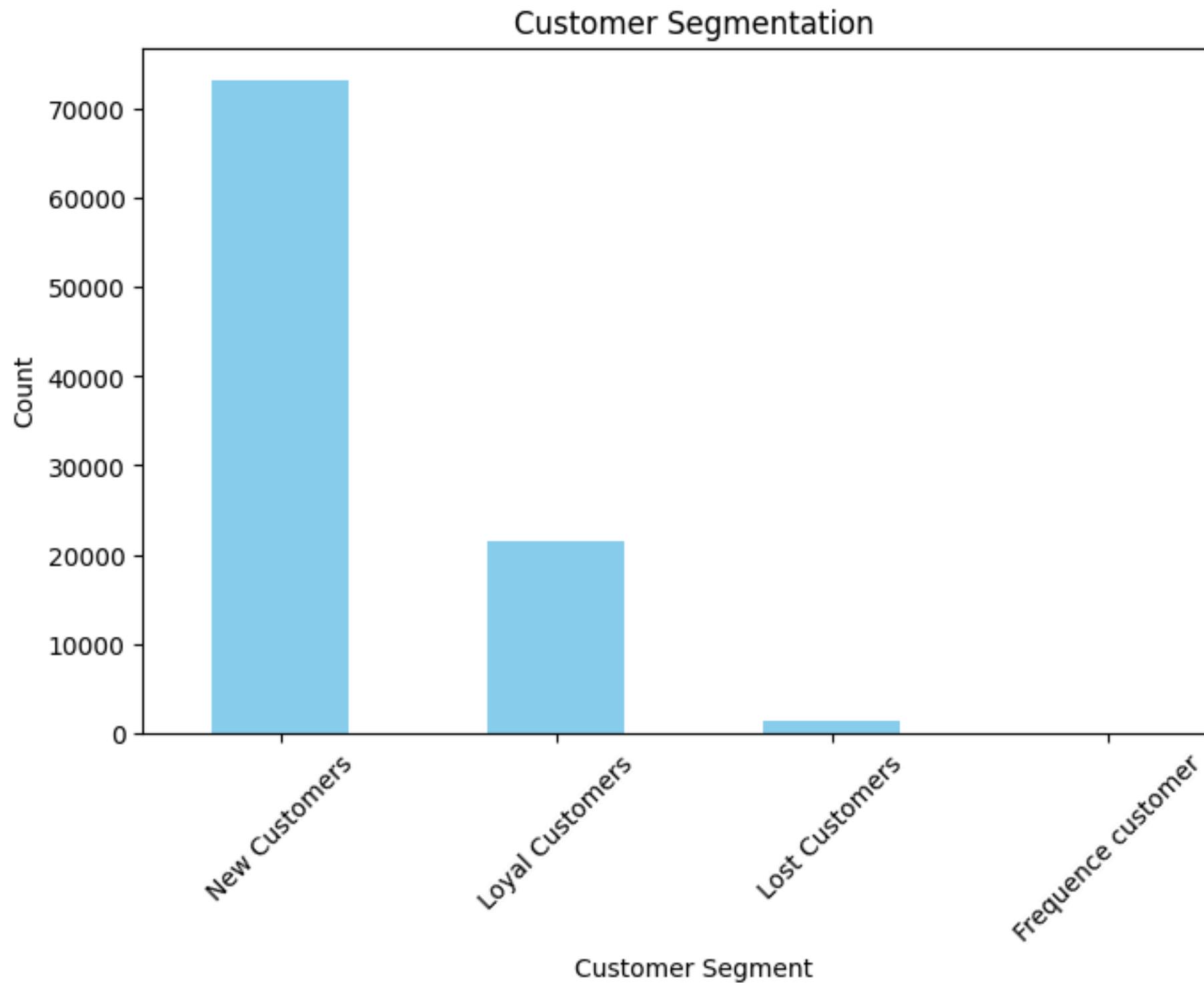
3

Khách hàng có khả năng rời bỏ: Lần mua hàng cách đây lâu, có dấu hiệu suy giảm quan tâm hoặc mất liên hệ.

4

Khách hàng không hoạt động: Lần mua chỉ giao động từ 1 đến vài lần và không còn ghé thăm sàn nữa

Chương 6: Phân khúc khách hàng



Thông tin chi tiết

Customr_Segment

New Customer	73213
Loyal Customers	21450
Lost Customers	1430
Inactive customer	2

Vấn đề và giải pháp

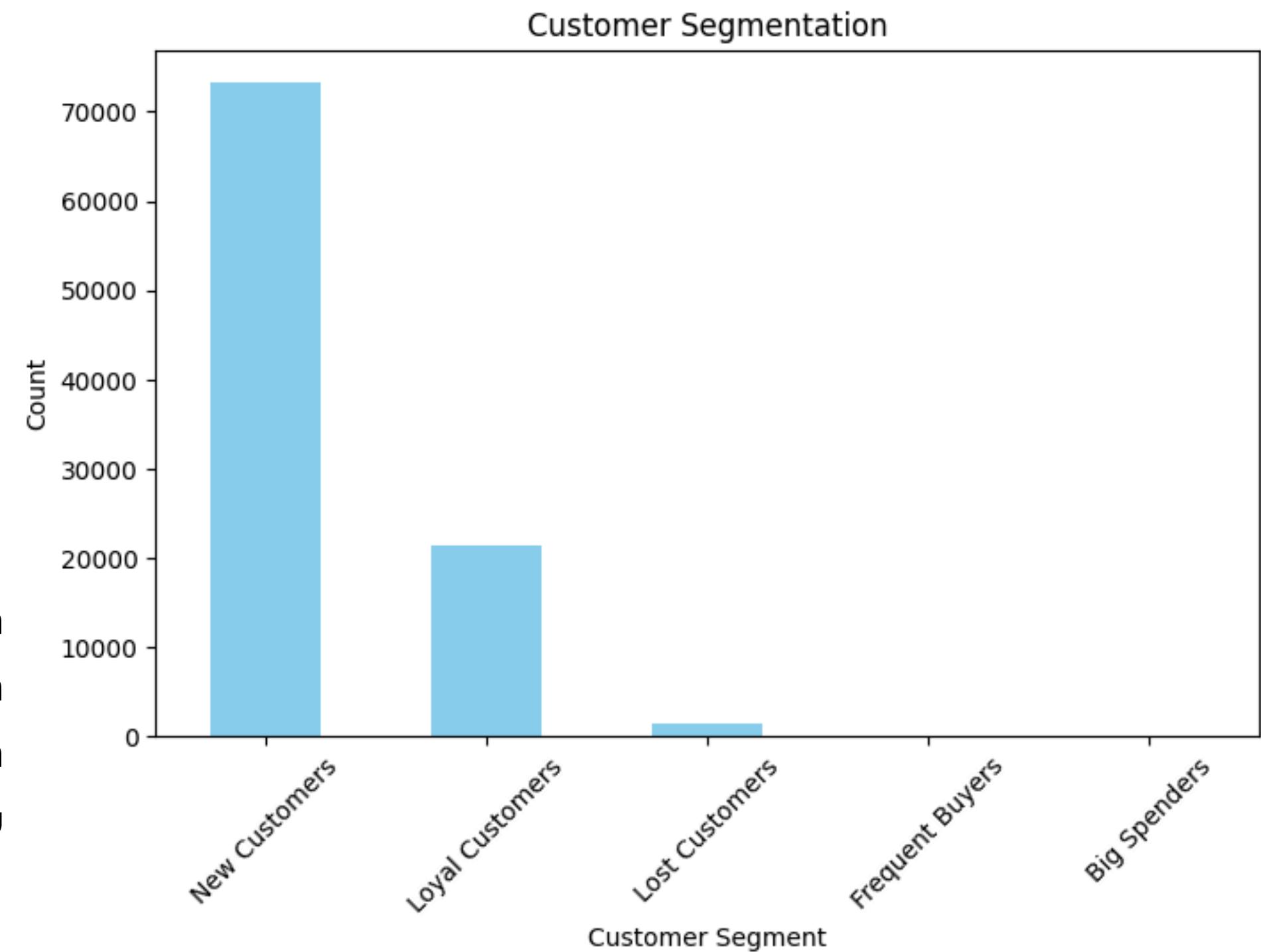
1. Khách hàng bị mất

- Vấn đề: Ngừng mua sắm từ lâu.

2. Khách hàng mới cần duy trì

- Vấn đề: Duy trì sự gắn bó..

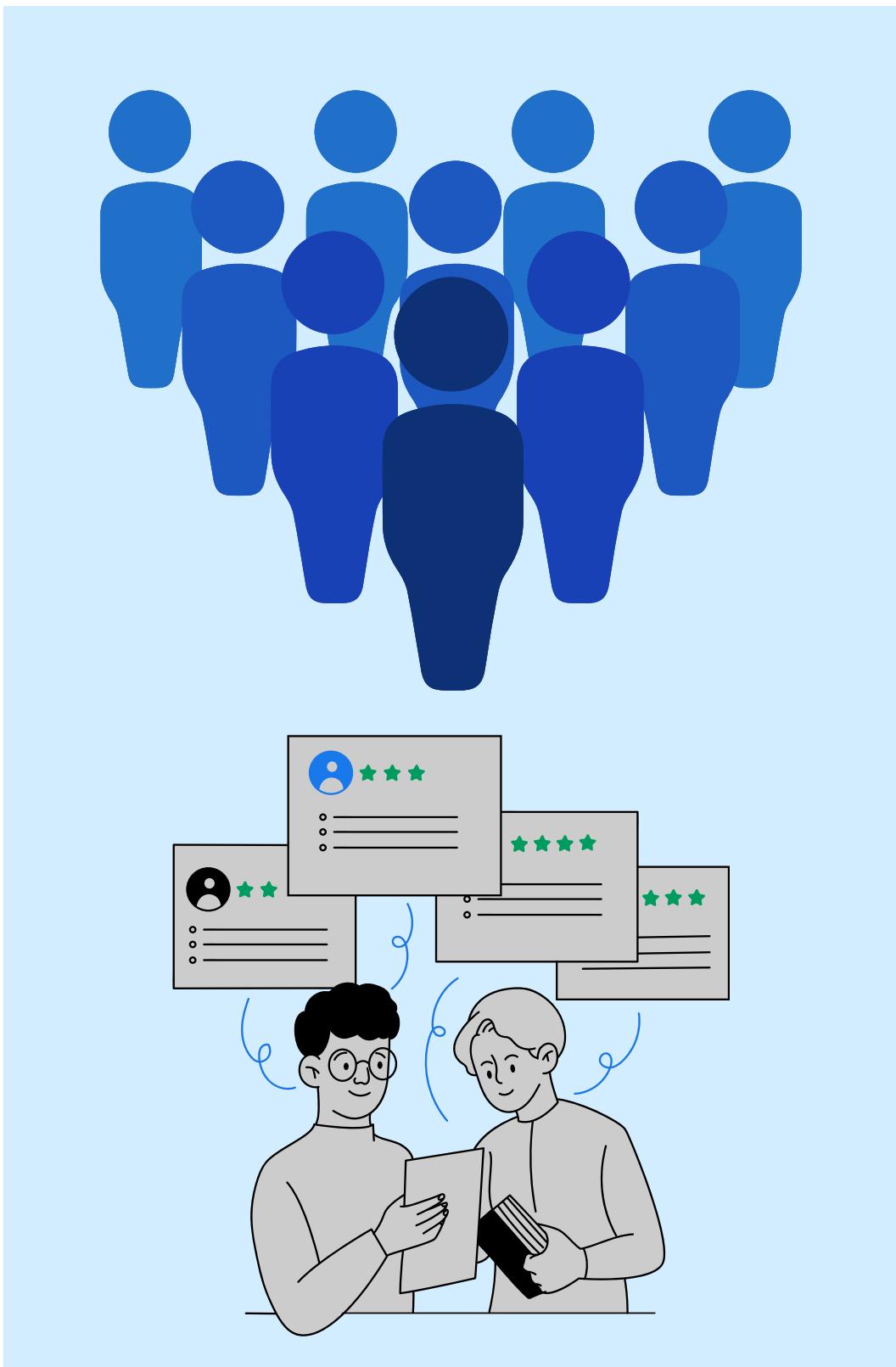
-> Áp dụng khuyến mãi và chương trình khách hàng thân thiết, cùng với dịch vụ chăm sóc khách hàng tốt nhất. Những biện pháp này giúp cải thiện tương tác, thu hút và giữ chân khách hàng hiệu quả.



Chương 6: Phân khúc khách hàng

Hướng giải quyết

- Tăng cường tỷ lệ khách hàng trung thành
- Hạn chế mất khách hàng
- Tối ưu hóa phân khúc khách hàng
- Cá nhân hóa và khai thác dữ liệu khách hàng
- Mở rộng thị trường



HỎI ĐÁP

XIN CẢM ƠN !!!