

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC KINH TẾ THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG CÔNG NGHỆ VÀ THIẾT KẾ



TIỂU LUẬN CUỐI KỲ

ĐỀ TÀI:

**PHÂN TÍCH DỮ LIỆU CHUỖI CỦA HÀNG CÀ
PHÊ TẠI NEW YORK**

GV HƯỚNG DẪN: Th.s PHẠM THỊ THANH TÂM

LỚP HP: HỆ HỖ TRỢ QUẢN TRỊ THÔNG MINH

MÃ LỚP HP: 24D1INF50908501

NHÓM SINH VIÊN THỰC HIỆN

VÕ NGỌC MỸ KIM 31211027646

VĂN NGỌC NHƯ QUỲNH 31211027669

NGUYỄN NHẬT THẢO VY 31211025542

TP. Hồ Chí Minh, ngày 05 tháng 5 năm 2024

MỤC LỤC

1.TỔNG QUAN.....	13
1.1. Lý do chọn đề tài	13
1.2. Mục đích, mục tiêu của đề tài.....	13
1.3. Đối tượng và phạm vi nghiên cứu	14
1.4. Phương pháp thực hiện.....	14
1.5. Bố cục của đề tài	14
1.6. Phân công công việc	15
2. MÔ TẢ DOANH NGHIỆP	17
2.1. Giới thiệu doanh nghiệp.....	17
2.2. Thực trạng của doanh nghiệp	17
2.3. Bài toán của doanh nghiệp	17
3. QUÁ TRÌNH ETL	19
3.1. Trích xuất dữ liệu (Extract)	19
3.1.1.Mô tả bộ dữ liệu	19
3.1.2. Đánh giá chất lượng của dữ liệu.....	20
3.1.3. Thống kê mô tả.....	22
3.2. Giai đoạn 2: Chuyển đổi dữ liệu (Transform)	49

3.2.1. Tiền xử lý dữ liệu	49
3.2.2. Transform	53
3.3. Giai đoạn 3: Nạp dữ liệu vào kho dữ liệu (Load)	54
3.3.1 Giới thiệu các bảng Dim	54
3.3.2. Giới thiệu bảng Fact	58
3.3.3. Mô hình dữ liệu (Data Model)	61
4. PHÂN TÍCH DỮ LIỆU	62
4.1. Các hàm DAX	62
4.1.1. Hàm DAX cho Column	62
4.1.2. Hàm DAX cho Measure	64
4.2. Overview	72
4.3. Time	81
4.4. Product	88
4.5. Product Group	94
4.6. Product Category	100
4.7. Product Subcategory	108
4.8. Sales Outlet	113
4.9. Pastry Inventory	117
4.10. Customer	122

5. ĐỀ XUẤT GIẢI PHÁP	127
6. KẾT LUẬN	131
6.1. Vấn đề đạt được.....	131
6.2. Vấn đề chưa đạt được	131
6.3. Hướng phát triển	132
TÀI LIỆU THAM KHẢO.....	133

DANH MỤC HÌNH ẢNH

Hình 1. Chỉ số thống kê mô tả của quantity	22
Hình 2. Chỉ số thống kê mô tả của unit_price	23
Hình 3. Chỉ số thống kê mô tả của birth_year	24
Hình 4.. Chỉ số thống kê mô tả của cost	25
Hình 5. Chỉ số thống kê mô tả của start_of_day	26
Hình 6. Chỉ số thống kê mô tả của quantity_sold.....	27
Hình 7. Chỉ số thống kê mô tả của waste	28
Hình 8. Chỉ số thống kê mô tả của %waste	28
Hình 9. Chỉ số thống kê mô tả của total_goal	29
Hình 10. Thống kê mô tả của instore_yn.....	30
Hình 11. Thống kê mô tả của promo_item_yn	31
Hình 12. Thống kê mô tả của gender.....	32
Hình 13. Thống kê mô tả của product_group.....	33
Hình 14. Thống kê mô tả của product_category	34
Hình 15. . Thống kê mô tả của product_type	35
Hình 16. Thống kê mô tả của taxx_exempt_yn.....	36
Hình 17. Thống kê mô tả của promo_yn	37
Hình 18. Thống kê mô tả của new_product_yn.....	38

Hình 19. Thống kê mô tả của sales_outlet_type.....	39
Hình 20. Thống kê mô tả của store_address.....	40
Hình 21. Thống kê mô tả của store_city.....	41
Hình 22. Thống kê mô tả của store_state_province	42
Hình 23. Thống kê mô tả của Neighborhood	43
Hình 24. Thống kê mô tả của quantity sau tiền xử lý.....	44
Hình 25. Thống kê mô tả của unit_price sau tiền xử lý.....	44
Hình 26. Thống kê mô tả của birth_year sau tiền xử lý	45
Hình 27. Thống kê mô tả của cost sau tiền xử lý	45
Hình 28. Thống kê mô tả của start_of_day sau tiền xử lý.....	46
Hình 29. Thống kê mô tả của quantity_sold sau tiền xử lý	46
Hình 30. Thống kê mô tả của waste sau tiền xử lý.....	47
Hình 31. Thống kê mô tả của %waste sau tiền xử lý	48
Hình 32. Thống kê mô tả của total_goal sau tiền xử lý.....	49
Hình 33. Transaction_date trước khi đổi định dạng	50
Hình 34. Kết quả sau khi chỉnh định dạng	50
Hình 35. Cột instore_yn có dữ liệu bị thiếu	51
Hình 36. Customer_id trong bảng sales_receipt.....	52
Hình 37. Bảng dữ liệu sau khi tạo cột mới.....	53

Hình 38. Model schema.....	61
Hình 39. Quan hệ giữa các bảng.....	62
Hình 40. Dashboard Overview	72
Hình 41. Cards trong dashboard Overview	72
Hình 42. Chart Sum of profit and Sum of total_cost by Week_desc	73
Hình 43. Chart sum of sales_amount by Date and store_city	74
Hình 44. Chart Sum of waste and Sum of start_of_day by store_city and store_address	75
Hình 45. Chart sum of sales_amount, Sum of total_cost and Sum of profit by product_group.....	76
Hình 46. Chart sum of sales_amount, Sum of total_cost and Sum of profit by product_group (zoom slider)	77
Hình 47Chart sum of sales_amount, Sum of total_cost and Sum of profit by product_group (zoom slider)	78
Hình 48. Chart Sum of sales_amount by gender	79
Hình 49. Dashboard Analysis by Time	81
Hình 50. Chart Count of customer_id and Sum of quantity by Week_Desc.....	81
Hình 51. Chart Sum of sales_amount and Sum of cost by Date	82
Hình 52. Chart Sum of quantity by transaction_time.....	84
Hình 53. Slicer: group Merchandise.....	85
Hình 54. Slicer group: Whole Beans/Teas	86

Hình 55. Slicer: store_address 32-20 Broadway	87
Hình 56. Dashboard Analysis by Product	88
Hình 57. Card trong dashboard Product	89
Hình 58. Card hiển thị product Civet Cat.....	89
Hình 59. Card hiển thị product hazelnut syrup.....	89
Hình 60. Chart sum of current_retail_price and Sum of quantity by product	90
Hình 61. Sum of current_retail_price and Sum of quantity by product	91
Hình 62. Chart Count of product_id by promo_yn	92
Hình 63. Sum of sales_amount, Sum of total_cost and Sum of profit by product	93
Hình 64. Dashboard Analysis by Product Group	95
Hình 65. Chart Sum of quantity by product_group	95
Hình 66. Chart Count of product by product_group	96
Hình 67. Chart Average of sales_amount and Average of total_cost by product_group.....	97
Hình 68. Chart Average of profit by Date and product_group.....	98
Hình 69. Dashboard Analysis Analysis by Product Category	100
Hình 70. Chart Sum of profit by Date and Product_category	101
Hình 71. Chart Sum of profit by Date and product_category	101
Hình 72. Chart Sum of quantity by product_category and store_address	102

Hình 73. Chart Sum of total_cost by product_category	103
Hình 74. Chart Sum of Sales_amount by Age Group and product_category ..	104
Hình 75. Chart Sum of Sales_amount by Age Group and product_category sử dụng slicer	105
Hình 76. Chart Sum of sales_amount by Age Group and product_category (slicer)	105
Hình 77. Chart Sum of Sales_amount by Age Group and product_category (slicer)	106
Hình 78. Dashboard Analysis by Product Subcategory	108
Hình 79. Chart of quantity of Top 5 product_subcate by quantity	108
Hình 80. Chart Sum of quantity of 5 lowest subcategory by quantity	109
Hình 81. Chart Sum of sales_amount, Sum of profit and Sum of total_cost by product_subcategory of Top 5 subcategory by quantity	110
Hình 82. . Chart Sum of sales_amount, Sum of profit and Sum of total_cost by product_subcategory of 5 lowest subcategory by quantiry	111
Hình 83. Dashboard Analysis by Sales outlet	113
Hình 84. Chart Sum of quantity by store_address and Week_desc.....	114
Hình 85. Chart Sum of profit by store_city and store_address	115
Hình 86. Chart table: sale_outlet_id, sum of sales_amount, total_goal, Completion Percentage.....	116
Hình 87. Chart Neighborhood, total_cost and porfit	116
Hình 88. Dashboard Analysis by Pastry inventory	117

Hình 89. Card trong dashboard Analysis by Pastry inventory	118
Hình 90. Chart Sum of waste and Sum of start_of_day by Day	118
Hình 91. Chart Sum of target_sales and Sum of waste_value by store_address	119
Hình 92. Chart Average of % waste by product.....	120
Hình 93. Chart Sum of waste_value by product.....	120
Hình 94. Dashboard Analysis by Customer	122
Hình 95. Chart Count of transaction_id by Age Group.....	123
Hình 96. Chart Sum of quantity by Age.....	124
Hình 97. Chart Count of product_id by product_group and Age Group.....	125
Hình 98. Chart Sum of sales_amount and Sum of total_cost by Age Group ...	126
Hình 99. Chart Average of profit by Age Group and gender	126

DANH MỤC BẢNG

Bảng 1. Mô tả bộ dữ liệu	19
Bảng 2. Tiêu chí đánh giá bộ dữ liệu.....	20
Bảng 3. Mô tả các bảng dim.....	54
Bảng 4. Mô tả bảng Fact.....	58

LỜI CẢM ƠN

Lời đầu tiên, nhóm chúng em muốn gửi lời tri ân sâu sắc đến ThS. Phạm Thị Thanh Tâm - người hướng dẫn chúng em trong môn Hệ hỗ trợ quản trị thông minh thuộc Khoa Công nghệ thông tin kinh doanh (BIT), Đại học Kinh tế TP. Hồ Chí Minh (UEH).

Qua quá trình học tập và nghiên cứu môn Hệ hỗ trợ quản trị thông minh, nhờ những bài giảng và tài liệu từ Cô, chúng em đã nắm bắt được kiến thức tổng quan về Business Intelligence và hiểu rõ các phương pháp thống kê cơ bản. Điều này đã tạo nền tảng cho chúng em trong việc phát triển tư duy, phân tích dữ liệu thực tế và xây dựng các báo cáo cho tiểu luận cuối kỳ. Bên cạnh đó, Cô luôn sẵn lòng giải đáp mọi thắc mắc và hỗ trợ nhóm chúng em trong suốt quá trình học tập. Những yếu tố này đã đóng góp quan trọng vào sự hoàn thiện của đồ án cuối kỳ trong môn Hệ hỗ trợ quản trị thông minh của nhóm.

Chúng em xin cam đoan rằng bài báo cáo là kết quả nghiên cứu độc lập của nhóm dưới sự hướng dẫn của Cô Phạm Thị Thanh Tâm. Dù đã nỗ lực áp dụng kiến thức đã học trong học kỳ để ứng dụng vào đồ án, nhưng do kinh nghiệm, thời gian còn hạn chế và nguồn tài liệu tham khảo không đầy đủ, đồ án của chúng em có thể có những thiếu sót trong cách trình bày. Chúng em rất mong nhận được sự phản hồi, góp ý và đánh giá từ Cô để có thể rút kinh nghiệm và hoàn thiện đồ án hơn.

Cuối cùng, chúng em xin kính chúc Cô luôn mạnh khỏe và thành công trong sự nghiệp.

Xin trân trọng cảm ơn Cô!

1.TỔNG QUAN

1.1. Lý do chọn đề tài

Ngành bán lẻ cà phê đang phát triển và trở thành một trong những lĩnh vực kinh doanh hàng đầu. Để thành công trong việc xây dựng và phát triển một chuỗi cửa hàng cà phê, việc hiểu rõ nhu cầu và kỳ vọng của khách hàng là một trong những yếu tố vô cùng quan trọng. Thông qua việc phân tích dữ liệu từ một chuỗi cửa hàng cà phê, các nhà quản lý có thể nắm bắt được cách ứng dụng các phương pháp phân tích dữ liệu vào thực tế, từ đó đưa ra các chiến lược kinh doanh sáng tạo và hiệu quả. Kết quả phân tích này không chỉ có thể giúp tối ưu hóa hoạt động kinh doanh, cải thiện chất lượng sản phẩm và dịch vụ mà còn mang lại trải nghiệm tốt hơn cho khách hàng.

Ngoài ra, việc sử dụng công cụ phân tích dữ liệu như Power BI không chỉ giúp nhà quản lý cà phê hiểu rõ hơn về hành vi mua hàng và sở thích của khách hàng mà còn giúp họ dự đoán xu hướng tiêu dùng trong tương lai. Điều này trở nên vô cùng quan trọng trong bối cảnh thị trường cà phê đang phát triển nhanh chóng và cạnh tranh gay gắt. Việc có thông tin chính xác và kịp thời về xu hướng tiêu dùng sẽ giúp nhà quản lý đưa ra các quyết định chiến lược về sản phẩm, vị trí và marketing một cách linh hoạt và hiệu quả hơn. Do đó, nghiên cứu và phân tích xu hướng tiêu dùng của khách hàng và công việc hoạt động kinh doanh của cửa hàng trong ngành cà phê không chỉ mang lại hiểu biết sâu sắc về thị trường mà còn giúp nhóm dự án có cơ hội phát triển các kỹ năng và kiến thức về BI.

1.2. Mục đích, mục tiêu của đề tài

Mục tiêu của đề tài là sử dụng dữ liệu từ chuỗi cà phê hư cấu để và phân tích sâu hơn các yếu tố quan trọng ảnh hưởng đến sự thành công của một doanh nghiệp. Đầu tiên, nhóm phân tích và nghiên cứu hoạt động kinh doanh bao gồm việc theo dõi doanh thu, hiệu suất của các cửa hàng và các yếu tố khác ảnh hưởng đến hoạt động kinh doanh. Điều này giúp nhóm hiểu rõ hơn về cách thức hoạt động của chuỗi cửa hàng cà phê và xác định những điểm mạnh và điểm yếu trong hoạt động kinh doanh của họ. Từ đó, đề xuất những giải pháp hỗ trợ doanh nghiệp giải quyết bài toán kinh doanh của họ.

Bên cạnh đó, nhóm cũng so sánh doanh thu bán hàng giữa các cửa hàng và xác định cửa hàng nào hoạt động hiệu quả nhất để xác định những yếu tố quan trọng đóng góp vào sự thành công của một cửa hàng và áp dụng những điều này cho các cửa hàng khác.

1.3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của đề tài này là một chuỗi cửa hàng cà phê ở New York với các thông tin về thông tin giao dịch, quản lý sản phẩm, mục tiêu bán hàng, khách hàng và điểm bán hàng.

Phạm vi nghiên cứu của đề tài là phân tích và mô tả dữ liệu quản lý bán hàng của một chuỗi cà phê ở New York. Bài nghiên cứu sẽ tập trung vào việc phân tích dữ liệu và đưa ra các kết luận và khuyến nghị để cải thiện hoạt động kinh doanh của chuỗi cà phê này. Nhóm sẽ sử dụng Excel để thực hiện tiền xử lý dữ liệu. Sau đó, thực hiện các bước chuyển đổi và nạp dữ liệu vào Power BI để phân tích và đánh giá bộ dữ liệu.

1.4. Phương pháp thực hiện

Trong đề tài này, nhóm sử dụng phương pháp thống kê mô tả, phương pháp phân tích thống kê và phương pháp phân tích đề xuất.

Nhóm sử dụng phương pháp thống kê mô tả để mô tả ngắn gọn các dữ liệu sau khi thu thập nhằm hiểu rõ hơn về phân phối của dữ liệu.

Tiếp đến, nhóm sử dụng phương pháp phân tích thống kê để khám phá “Những gì đã diễn ra?” thông qua việc sử dụng dữ liệu trong quá khứ dưới dạng dashboard trong phần phân tích dữ liệu.

Sau khi hoàn thành phân tích thống kê, nhóm sử dụng phương pháp đề xuất để kết hợp những phát hiện từ phân tích trước đó, nhằm xác định những vấn đề cần giải quyết và đưa ra quyết định.

1.5. Bố cục của đề tài

Đề tài nghiên cứu bao gồm những nội dung chính sau:

Phần 1: Tổng quan đề tài:

- Chương 1: Tổng quan: giới thiệu lý do chọn đề tài, mục tiêu nghiên cứu, đối tượng và phạm vi nghiên cứu và phương pháp thực hiện đề tài.
- Chương 2: Bài toán doanh nghiệp: bên cạnh việc giới thiệu doanh nghiệp, mục này cũng trình bày bài toán mà doanh nghiệp cần giải quyết.

Phần 2: Thực hiện phân tích:

- Chương 3: Quá trình ETL bao gồm 3 giai đoạn:
 - Giai đoạn 1: Trích xuất dữ liệu (Extract): Tiến hành trích xuất dữ liệu *Coffee Shop Sample Data* lấy từ Kaggle. Sau đó, tiến hành mô tả các thuộc tính có trong bộ dữ liệu, đánh giá bộ dữ liệu và tiến hành thực hiện thống kê mô tả.
 - Giai đoạn 2: Chuyển đổi dữ liệu (Transform): Bao gồm các bước tiền xử lý dữ liệu, chuẩn hóa dữ liệu và gộp các thuộc tính dữ liệu theo yêu cầu phân tích.
 - Giai đoạn 3: Nạp dữ liệu vào kho dữ liệu (Load): Sau khi chuyển đổi dữ liệu, tiến hành nạp dữ liệu vào kho dữ liệu.
- Chương 4: Phân tích dữ liệu: sử dụng công cụ hỗ trợ Power BI để trực quan hóa dữ liệu thành các dashboard, sau đó tiến hành phân tích.
- Chương 5: Đề xuất giải pháp: dựa vào những phân tích các dashboard, tổng kết lại đưa ra các quyết định và đề xuất những giải pháp cho doanh nghiệp để quản lý hiệu quả hơn.

Phần 3: Kết luận:

- Chương 6: Đưa ra kết luận và xác định hướng phát triển trong tương lai.

1.6. Phân công công việc

MSSV	Tên thành viên	Phân công	Đánh giá
31211027646	Võ Ngọc Mỹ Kim	Chương 2, Thống kê mô tả	100%

		Phân tích dashboard Overview, Product, Pastry Inventory Đề xuất và đánh giá giải pháp, DAX column, Vấn đề đạt được, Thiết kế slide thuyết trình	
31211027669	Văn Ngọc Như Quỳnh	Tổng quan, Mô tả bộ dữ liệu, Đánh giá chất lượng dữ liệu, Thống kê mô tả Giới thiệu dim, fact, Phân tích dashboard Group, Category, Đề xuất và đánh giá giải pháp, Viết giải thích hàm DAX, Vấn đề chưa đạt được, Tổng hợp Word	100%
31211025542	Nguyễn Nhật Thảo Vy	Thống kê mô tả, ETL, Data Model, Phân tích theo Time, Subcategory, Sales Outlet, Customer Đề xuất giải pháp, DAX measure, Hướng phát triển, Tổng hợp Word Tổng hợp và thiết kế Dashboard	100%

2. MÔ TẢ DOANH NGHIỆP

2.1. Giới thiệu doanh nghiệp

Một chuỗi cửa hàng cà phê hư cấu đang vận hành tám cửa hàng tại tiểu bang New York. Chuỗi cửa hàng cà phê này được thành lập bởi Amber và Sandeep. Tuy nhiên, trong bảng dữ liệu thống kê giao dịch chỉ có thông tin từ ba cửa hàng được ghi nhận. Với ba cửa hàng đặt tại những vị trí chiến lược trong thành phố, chuỗi cửa hàng đã trở thành một điểm đến không thể bỏ qua cho người dân và cả du khách.

2.2. Thực trạng của doanh nghiệp

Chuỗi cửa hàng cà phê hư cấu này đang hoạt động với 8 cửa hàng tại tiểu bang New York, tuy nhiên chỉ có dữ liệu thống kê giao dịch vào tháng tư năm 2019 từ 3 cửa hàng.

Trong tháng tư năm 2019, doanh nghiệp đã có một kết quả khả quan. Doanh thu và lợi nhuận đều vượt qua mục tiêu đặt ra, một dấu hiệu tích cực cho sự phát triển của doanh nghiệp. Tuy nhiên, sự suy giảm doanh thu ở cuối tháng là một tín hiệu cảnh báo, đòi hỏi sự chú ý đặc biệt từ phía doanh nghiệp. Việc tiếp tục giữ vững và tăng cường các biện pháp để duy trì đà tăng trưởng là rất cần thiết.

Với sự mở rộng của chuỗi cửa hàng, cơ hội để tối ưu hóa hoạt động và tăng cường hiệu suất là vô tận. Sự tiếp cận với dữ liệu giao dịch từ các cửa hàng khác nhau cũng có thể cung cấp thông tin quý báu để định hình chiến lược tương lai và đảm bảo sự thành công liên tục của doanh nghiệp.

2.3. Bài toán của doanh nghiệp

Mặc dù đã đạt được thành công nhất định, chuỗi cửa hàng cà phê này vẫn đang đối mặt với một số thách thức:

- Thiếu dữ liệu: Việc thiếu dữ liệu giao dịch từ 5 cửa hàng còn lại và những tháng hoạt động trước đó khiến cho việc phân tích hiệu quả hoạt động kinh doanh và đưa ra quyết định chiến lược gặp nhiều khó khăn.

- **Cạnh tranh:** Ngành cà phê tại New York vô cùng cạnh tranh, với nhiều thương hiệu lớn nhỏ khác nhau. Việc thu hút và giữ chân khách hàng là một bài toán quan trọng.
- **Tối ưu hóa hoạt động:** Cần tối ưu hóa hiệu quả hoạt động của 3 cửa hàng hiện có để tăng lợi nhuận và giảm chi phí.
- **Mở rộng thị trường:** Mở rộng thị phần và tăng số lượng cửa hàng là mục tiêu quan trọng để phát triển bền vững.

Còn về mục tiêu cần giải quyết, ở đây nhóm đặt ra hai mục tiêu chung và cụ thể:

- **Mục tiêu chung:** chuỗi cửa hàng cà phê hư cấu này sẽ trở thành một thương hiệu cà phê uy tín và được yêu thích tại New York, với hệ thống cửa hàng rộng khắp, doanh thu cao và lợi nhuận bền vững.
- **Mục tiêu cụ thể:**
 - Tăng doanh thu.
 - Mở rộng thị trường:
 - Mở thêm một cửa hàng trong chuỗi vào năm tới
 - xâm nhập vào các khu vực mới trong các thành phố của tiểu bang New York.
 - Phát triển thị trường sang các khu vực lân cận New York.
 - Nâng cao nhận diện thương hiệu:
 - Tăng số lượng người theo dõi trên mạng xã hội lên 50% trong vòng 1 năm.
 - Tăng số lượng bài báo và bài đánh giá tích cực về thương hiệu lên 20% trong vòng 1 năm.
 - Tăng tỷ lệ khách hàng trung thành:
 - Tăng tỷ lệ khách hàng trung thành lên 15% trong vòng 1 năm.
 - Tăng số lượng khách hàng quay lại cửa hàng lên 10% trong vòng 6 tháng.
 - Giảm chi phí vận hành:
 - Giảm chi phí vận hành 5% trong vòng 6 tháng.
 - Tối ưu hóa quy trình quản lý hàng tồn kho.
 - Kiểm soát chi phí nhân công hiệu quả.

3. QUÁ TRÌNH ETL

3.1. Trích xuất dữ liệu (Extract)

Nhóm thu thập dữ liệu gốc từ Kaggle. Dữ liệu gồm 9 file csv riêng lẻ, mỗi file chứa một bảng thông tin.

3.1.1. Mô tả bộ dữ liệu

Bộ dữ liệu “*Coffee shop sample data*” được lấy từ Kaggle, bao gồm 9 bảng, chứa tổng cộng 78 trường dữ liệu (cột) và 49894 bản ghi (dòng). Thông tin về dữ liệu trong từng bảng trùng phần nhiều với các bảng Dimension và bảng Fact mà nhóm sẽ giới thiệu ở mục 3.1 và 3.2, vậy nên, nhóm chỉ tiến hành giới thiệu sơ lược thông tin về các bảng dữ liệu ở mục này.

Bảng 1. Mô tả bộ dữ liệu

STT	Bảng Dữ Liệu	Mô Tả	Ghi Chú
1	sales receipts	Biên lai bán hàng	chứa các thông tin về sản phẩm, thông tin khách hàng, cửa hàng, cũng như các giao dịch, doanh số và giá cả
2	customer	Khách hàng	chứa các thông tin chi tiết về khách hàng thành viên bao gồm tên, địa chỉ email, giới tính và ngày bắt đầu mua hàng.
3	dates	Ngày	cung cấp thông tin về các ngày giao dịch
4	generations	Thế hệ khách hàng	chứa thông tin về thế hệ khách hàng

5	pastry inventory	Kho sản phẩm	chứa các thông tin về ghi nhận nhập và bán bánh ngọt tại các cửa hàng trong ngày
6	product	Sản phẩm	chứa các thông tin về sản phẩm được bán bao gồm tên sản phẩm, nhóm sản phẩm, các danh mục sản phẩm cũng như giá cả của sản phẩm
7	sales target	Mục tiêu bán hàng	chứa các thông tin về mục tiêu bán hàng của từng loại sản phẩm cũng như tổng mục tiêu bán hàng chung cho tất cả các sản phẩm
8	sales outlet	Cửa hàng	cung cấp thông tin về các cửa hàng bao gồm địa điểm, loại cửa hàng,..
9	staff	Nhân viên	chứa các thông tin về nhân viên cửa hàng

3.1.2. Đánh giá chất lượng của dữ liệu

Đánh giá dữ liệu không chỉ là kiểm tra chất lượng dữ liệu mà còn liên quan đến việc xác định liệu dữ liệu có thể đáp ứng các yêu cầu và mục tiêu của dự án hay không. Đồng thời, nó cũng giúp phát hiện các vấn đề tiềm ẩn và tạo điều kiện cho việc xử lý dữ liệu hiệu quả hơn trong các giai đoạn tiếp theo của quy trình làm việc với dữ liệu. Dưới đây là bảng đánh giá dữ liệu cho bộ dữ liệu theo 8 tiêu chí.

Bảng 2. Tiêu chí đánh giá bộ dữ liệu

Tiêu chí	Đánh giá
----------	----------

Reliability (Độ tin cậy)	Bộ dữ liệu trên là bộ dữ liệu minh họa một chuỗi cửa hàng cà phê hư cấu của IBM nhằm phục vụ mục đích luyện tập thực hành phân tích dữ liệu. Do đó, nó không thể thay thế dữ liệu thực tế của một doanh nghiệp và có độ tin cậy ở mức tương đối thấp.
Accuracy (Tính chính xác)	Dữ liệu trong bộ dữ liệu Coffee Shop có độ chính xác tương đối cao, tương đương với dữ liệu quản lý bán hàng của một doanh nghiệp. Tuy nhiên, do không phải là một bộ dữ liệu thực tế, các giá trị dữ liệu có thể không phản ánh chính xác tình hình kinh doanh thực tế của công ty trong ngành bán lẻ.
Accessibility (Khả năng truy cập)	Bất kỳ ai cũng có thể dễ dàng truy cập và sử dụng bộ dữ liệu để thực hiện cho mục đích phân tích và nghiên cứu. Bộ dữ liệu dễ dàng được tìm thấy và tải xuống trên Kaggle. Có thông tin về dữ liệu đầy đủ.
Security and Privacy (Bảo mật và Quyền riêng tư)	Vì Coffee Shop là một bộ dữ liệu hư cấu, do đó không bao gồm các thông tin nhạy cảm về khách hàng, nhân viên, công ty nên trong phạm vi bài báo cáo, yếu tố này không có tác động đáng kể.
Richness (Độ đa dạng)	Coffee Shop Dataset mà nhóm thu thập được có tính đa dạng cao, bao gồm gần như đầy đủ các thông tin cần thiết về dữ liệu khách hàng, sản phẩm, khuyến mãi,... Do đó giúp cho việc phân tích trở nên dễ dàng hơn, phù hợp với mục tiêu nghiên cứu bài báo cáo của nhóm.
Consistency (Tính nhất quán)	Dữ liệu trong bộ Coffee Shop được tạo ra nhằm mục đích hỗ trợ việc phân tích và nghiên cứu về một chuỗi cửa hàng cà phê (ảo) bán các sản phẩm liên quan đến đồ uống và bánh ngọt, do đó có tính nhất quán với nhau ở các bảng dữ liệu, không có mâu thuẫn thông tin khách hàng với thông tin trong đơn hàng.
Timeliness (Tính kịp thời)	Bộ dữ liệu được sử dụng để minh họa cho hoạt động của chuỗi cửa hàng cà phê trong khoảng thời gian từ ngày 1 đến ngày 30 tháng 4 năm 2019, do đó không phản ánh chính xác tình hình thời gian thực.
Relevancy (Độ liên quan)	Bộ dữ liệu thu được chứa thông tin về sản phẩm, khách hàng, địa điểm,... Mục tiêu của bài phân tích này nhằm xác định được hướng quản lý và phát triển doanh nghiệp dựa vào tình hình mua bán ở thời gian tới, vì vậy các thông tin kể trên đều cần thiết và liên quan đến mục đích phân tích.

Vì bộ dữ liệu được lấy từ kaggle và được kaggle kiểm chứng nên bộ dữ liệu có độ chính xác và đáng tin cậy cao. Với 9 bảng và 49894 dòng thì cũng đa dạng, phù hợp với đồ án và phạm vi dữ liệu đủ để nhóm tìm hiểu. Từ những đánh giá theo 8 tiêu chí ở trên, nhóm nhận thấy bộ dữ liệu này phù hợp với việc nghiên cứu đồ án nhóm.

3.1.3. Thống kê mô tả

Phân tích các đại lượng thống kê mô tả là một bước quan trọng trong việc khai thác thông tin từ các biến của bộ dữ liệu.

Nội dung của chương này xoay quanh độ phân tán của các biến cùng với các đại lượng mô tả xu thế trung tâm như trung bình, trung vị, và yếu vị.

❖ Trước tiên xử lý:

- **Biến numeric**

Trước khi tiền xử lý, dữ liệu gốc có nhiều cột dữ liệu thuộc dạng số, trong đó có các cột ID sẽ được chuyển đổi kiểu dữ liệu ở bước tiền xử lý tiếp theo. Hơn nữa bộ dữ liệu cũng có nhiều cột mà nhóm không dùng đến, vì vậy nhóm chỉ phân tích chỉ số thống kê mô tả của những trường dữ liệu sẽ được dùng.

<i>quantity</i>	
Mean	1.438209003
Standard Error	0.002431121
Median	1
Mode	1
Standard Deviation	0.543038704
Sample Variance	0.294891034
Kurtosis	1.09709037
Skewness	0.822015145
Range	7
Minimum	1
Maximum	8
Sum	71758
Count	49894

Hình 1. Chỉ số thống kê mô tả của quantity

Nhận xét: Biến quantity có Mean (Trung vị) là 1.438, cao hơn Median và Mode, cho thấy dữ liệu có xu hướng lệch về phía dương. Trong khi đó, Median (Trung vị) và Mode (Yếu vị) đều bằng 1, cho thấy rằng giá trị phổ biến nhất và giá trị ở vị trí trung tâm của dữ liệu là 1. Còn Skewness > 0 chứng tỏ phân phối dữ liệu có đuôi kéo dài về phía bên phải của trục số, tức là phân phối lệch phải.

<i>unit_price</i>	
Mean	3.384645448
Standard Error	0.012009441
Median	3
Mode	3
Standard Deviation	2.682544712
Sample Variance	7.196046132
Kurtosis	99.8313699
Skewness	8.530730378
Range	44.2
Minimum	0.8
Maximum	45
Sum	168873.5
Count	49894

Hình 2. Chỉ số thống kê mô tả của unit_price

Nhận xét: Biến unit_price có giá trị Mean là 3.385, cho thấy giá trung bình của giá bán lẻ là khoảng 3.38. Median (Trung vị) và Mode (Yếu vị) đều bằng 3, cho thấy giá trị trung tâm và phổ biến nhất của dữ liệu là 3. Skewness gần 8.53, rất cao, chỉ ra rằng phân phối dữ liệu có đuôi kéo dài rất xa về phía bên phải của trục số.

<i>birth_year</i>	
Mean	1978.385574
Standard Error	0.314937302
Median	1981
Mode	1995
Standard Deviation	14.92550309
Sample Variance	222.7706426
Kurtosis	-1.177747272
Skewness	-0.291395709
Range	51
Minimum	1950
Maximum	2001
Sum	4443454
Count	2246

Hình 3. Chỉ số thống kê mô tả của *birth_year*

Nhận xét: Biến *birth_year* có giá trị Mean là 1978.386, cho thấy năm sinh trung bình của khách hàng là khoảng 1978. Median là 1981 và Mode là 1995, giá trị Skewness là -0.291395709, chỉ ra rằng phân phối dữ liệu có đuôi kéo dài về phía bên trái của trục số.

<i>cost</i>	
Mean	3.888409091
Standard Error	0.602258157
Median	1.195
Mode	0.75
Standard Deviation	5.649682299
Sample Variance	31.91891008
Kurtosis	11.61435469
Skewness	2.892291386
Range	35.96
Minimum	0.04
Maximum	36
Sum	342.18
Count	88

Hình 4.. Chỉ số thống kê mô tả của *cost*

Nhận xét: Biến *cost* có giá trị Mean là 3.888, cho thấy chi phí trung bình là khoảng 3.89. Median là 1.195 và Mode là 0.75, cho thấy chi phí trung tâm và phổ biến nhất trong dữ liệu là thấp hơn so với giá trị trung bình. Giá trị Skewness là 2.892, chỉ ra rằng phân phối dữ liệu có đuôi kéo dài về phía bên phải của trục số.

<i>start_of_day</i>	
Mean	24.05863192
Standard Error	0.688495416
Median	18
Mode	18
Standard Deviation	12.06341423
Sample Variance	145.5259628
Kurtosis	0.227848654
Skewness	1.492114891
Range	30
Minimum	18
Maximum	48
Sum	7386
Count	307

Hình 5. Chỉ số thống kê mô tả của *start_of_day*

Nhận xét: Biến *start_of_day* có giá trị Mean là 24.059, cho thấy số bánh ngọt nhập trung bình trong ngày là khoảng 24.05. Median (Trung vị) và Mode (Yếu vị) đều bằng 18, cho thấy thời gian bắt đầu trung tâm và phổ biến nhất trong dữ liệu là 18. Giá trị Skewness là 1.492, chỉ ra rằng phân phối dữ liệu có đuôi kéo dài về phía bên phải của trục số.

<i>quantity_sold</i>	
Mean	9.296416938
Standard Error	0.31048379
Median	8
Mode	9
Standard Deviation	5.440115477
Sample Variance	29.5948564
Kurtosis	3.144204272
Skewness	1.547494442
Range	32
Minimum	0
Maximum	32
Sum	2854
Count	307

Hình 6. Chỉ số thống kê mô tả của *quantity_sold*

Nhận xét: Biến *quantity_sold* có giá trị Mean là 9.296, cho thấy số lượng bán bánh ngọt đã nhập trung bình là 9.296. Median (Trung vị) bằng 8 và Mode (Yếu vị) bằng 9. Skewness là 1.547, chỉ ra rằng phân phối dữ liệu bị lệch phải, nghĩa là hầu hết các giá trị tập trung ở phía bên trái của phân phối.

waste	
Mean	14.65798046
Standard Error	0.639338052
Median	11
Mode	9
Standard Deviation	11.20210764
Sample Variance	125.4872155
Kurtosis	1.364302624
Skewness	1.551275141
Range	47
Minimum	0
Maximum	47
Sum	4500
Count	307

Hình 7. Chỉ số thống kê mô tả của waste

Nhận xét: Biến waste có giá trị Mean là 14.66, cho biết giá trị trung bình của số hàng bị thừa. Median bằng 11 và Mode bằng 9

AW	AX
% waste	
Mean	0.581074919
Standard Error	0.011905106
Median	0.58
Mode	0.5
Standard Deviation	0.208594309
Sample Variance	0.043511586
Kurtosis	-0.099906512
Skewness	-0.459096439
Range	0.96
Minimum	0
Maximum	0.96
Sum	178.39
Count	307

Hình 8. Chỉ số thống kê mô tả của %waste

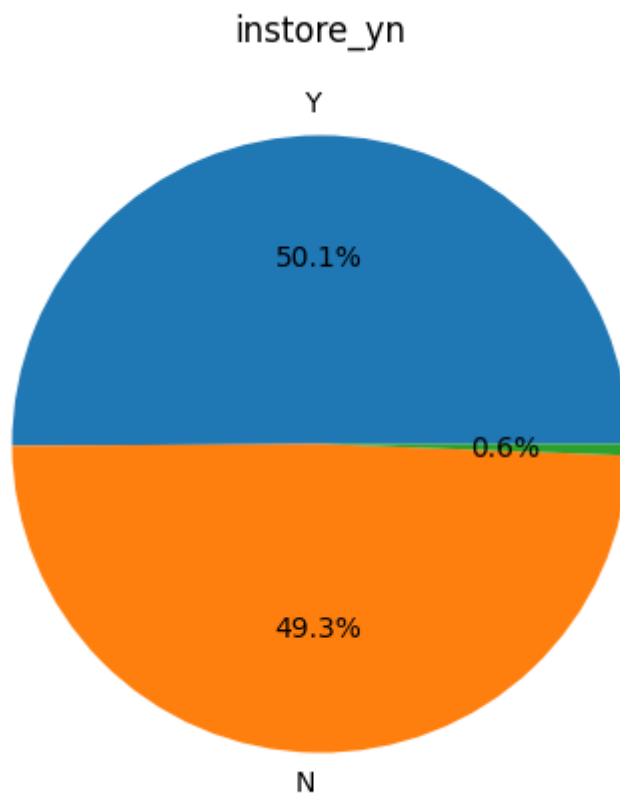
Nhận xét: Biến %waste có giá trị Mean là 0.581, cho biết giá trị trung bình của tỷ lệ số hàng bị thừa. Mean là 0.58 và Mode là 0.5, Skewness là -0.459, chỉ ra rằng phân phối dữ liệu có đuôi kéo dài về phía bên trái của trục số.

	total_goal
Mean	19437.5
Standard Error	970.2609185
Median	18000
Mode	18000
Standard Deviation	2744.3123
Sample Variance	7531250
Kurtosis	1.573664365
Skewness	1.686626812
Range	7000
Minimum	18000
Maximum	25000
Sum	155500
Count	8

Hình 9. Chỉ số thống kê mô tả của total_goal

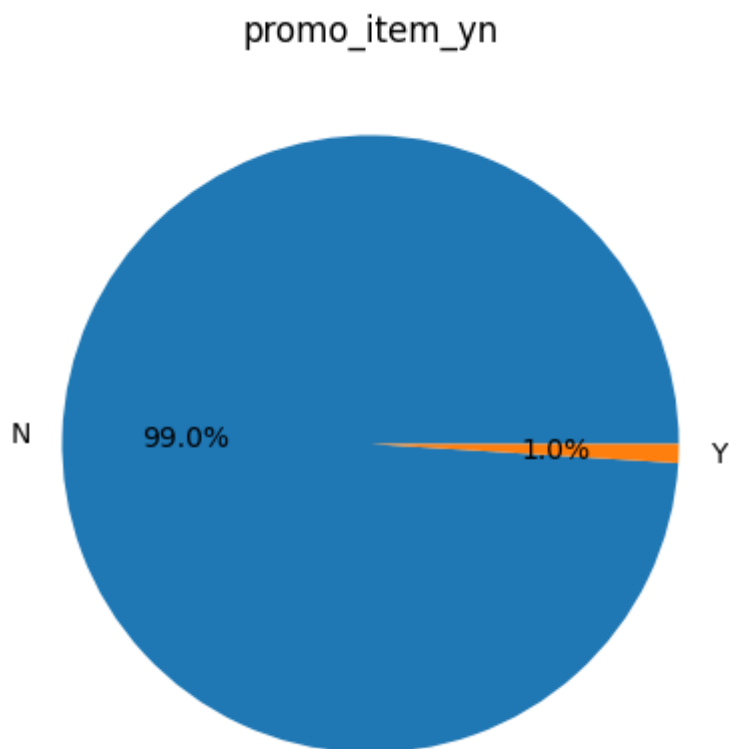
Nhận xét: cột birth_year dữ liệu lệch trái đuôi của phân phối nằm bên trái của trung vị, giá trị trung bình bé hơn trung vị; các cột còn lại đều lệch phải đuôi của phân phối nằm bên phải của trung vị, giá trị trung bình lớn hơn trung vị.

Biến category:



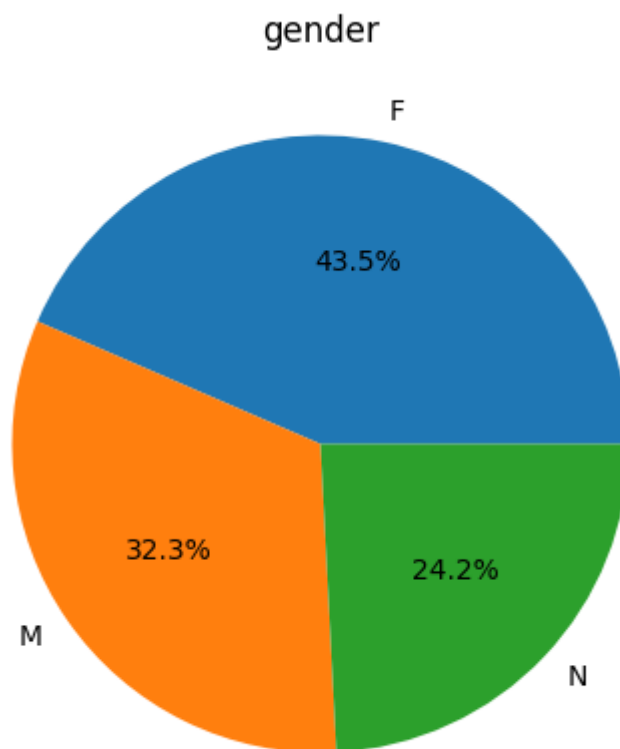
Hình 10. Thống kê mô tả của instore_yn

Nhận xét: Phần Y được tô màu xanh, chiếm đa số với 50,1% tổng số, đại diện cho số lượng phản hồi ‘Có’. Trong khi đó, phần N được tô màu cam, chiếm 49,3% tổng số, đại diện cho số lượng phản hồi ‘Không’. Có một phần nhỏ không xác định được tô màu xanh lá, chiếm 0,6% tổng số.



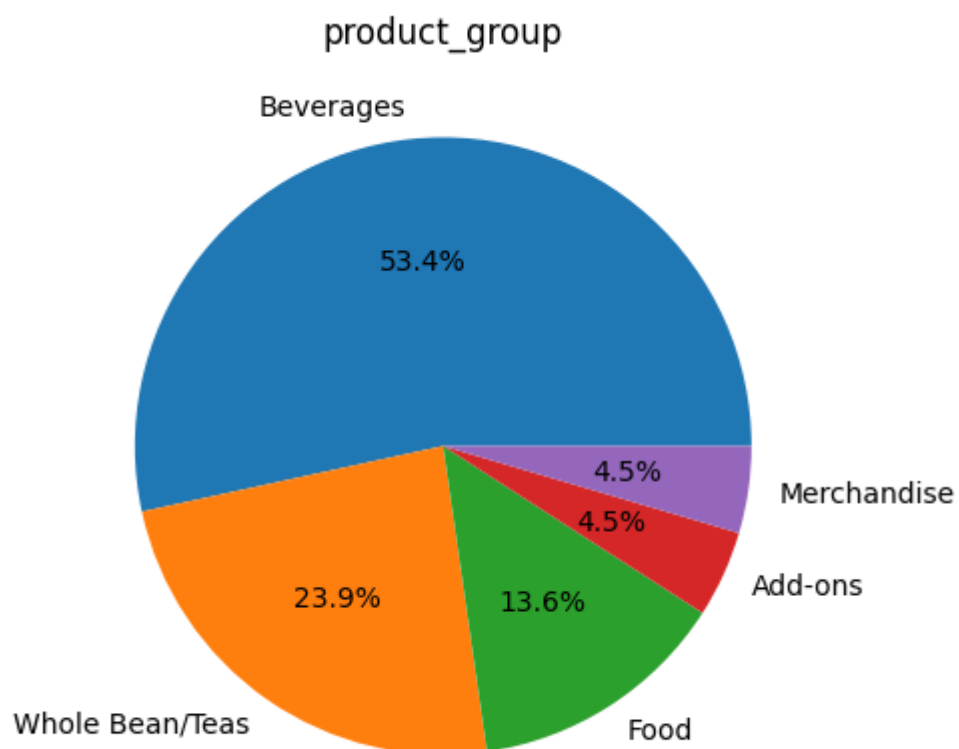
Hình 11. Thống kê mô tả của promo_item_yn

Nhận xét: Phân loại N chiếm phần lớn với 99,0% và được tô màu xanh, trong khi phân loại Y chỉ chiếm 1,0% và được tô màu cam. Biểu đồ này liên quan đến việc phân tích các mặt hàng có khuyến mãi hoặc không. Sự chênh lệch lớn giữa hai phân loại này cho thấy một tỷ lệ rất nhỏ các mặt hàng là khuyến mãi so với số lượng không khuyến mãi.



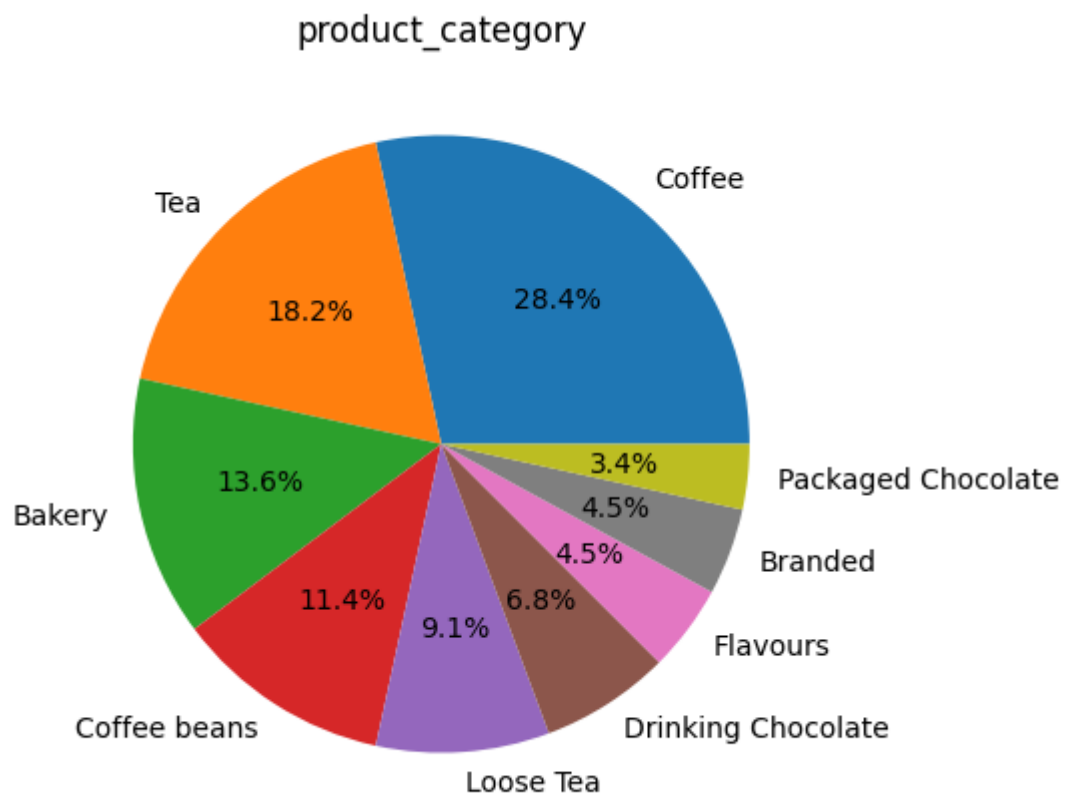
Hình 12. Thống kê mô tả của gender

Nhận xét: Phân loại F chiếm phần lớn nhất với 43,5% được tô màu xanh dương, chỉ giới tính nữ. Phân loại M, màu cam, chiếm 32,3% chỉ giới tính nam. Phân loại N, màu xanh lá cây, chiếm 24,2% chỉ giới tính không xác định.



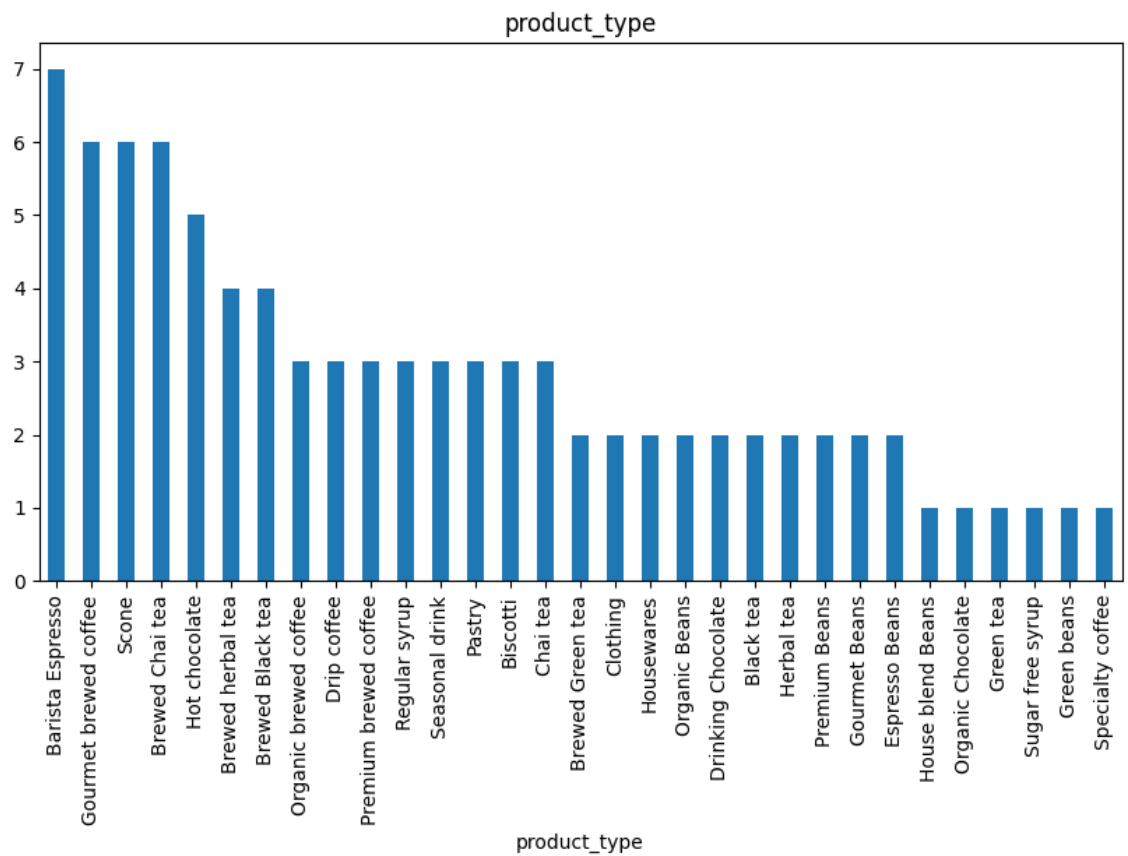
Hình 13. Thống kê mô tả của product_group

Nhận xét: Beverages chiếm phần lớn nhất với 53.4%, tiếp theo là Whole Bean/Teas với 23,9% và hai nhóm nhỏ hơn là Merchandise và Add-ón mỗi nhóm chiếm 4,5.



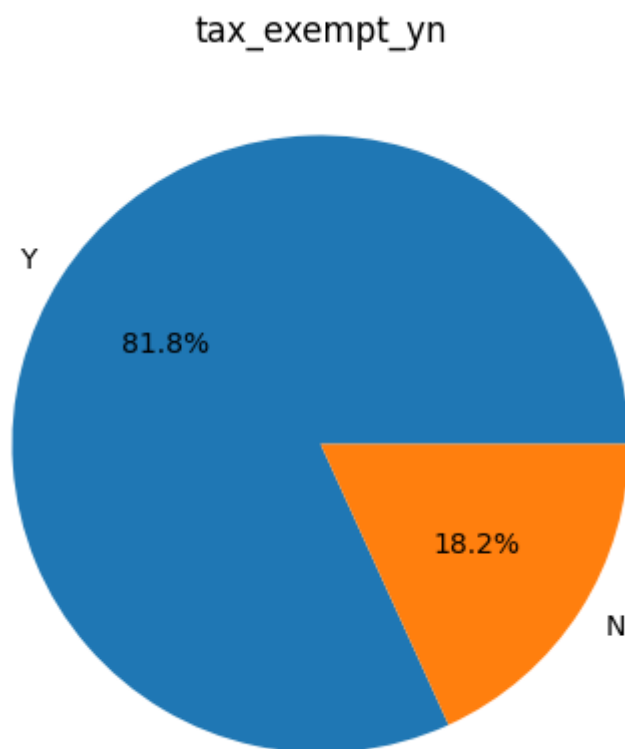
Hình 14. Thống kê mô tả của product_category

Nhận xét: Coffee chiếm phần lớn nhất với 28,4%, theo sau là Tea với 18,2%. Các nhóm khác bao gồm Bakery (13,6%), Packaged Chocolate (3,4%), Branded (4,5%), Flavours (6,8%), Drinking Chocolate (9,1%) và nhóm nhỏ nhất là Coffee Beans với 11,4%.



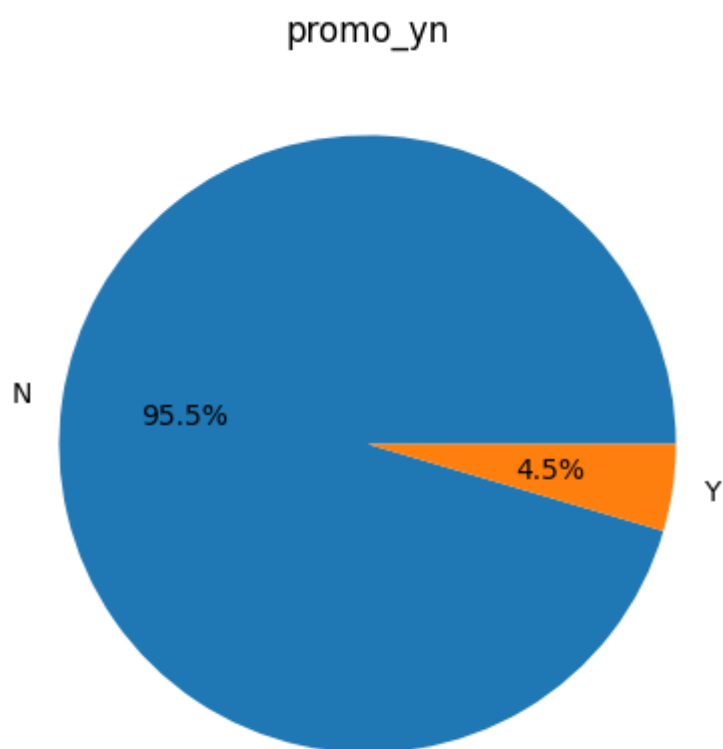
Hình 15. . Thống kê mô tả của product_type

Nhận xét: Barista Espresso có giá trị cao nhất, gần chạm tới 7, tiếp đến là Gourmet brewed coffee vượt quá 6, trong khi Specialty coffee có giá trị thấp nhất gần 1.



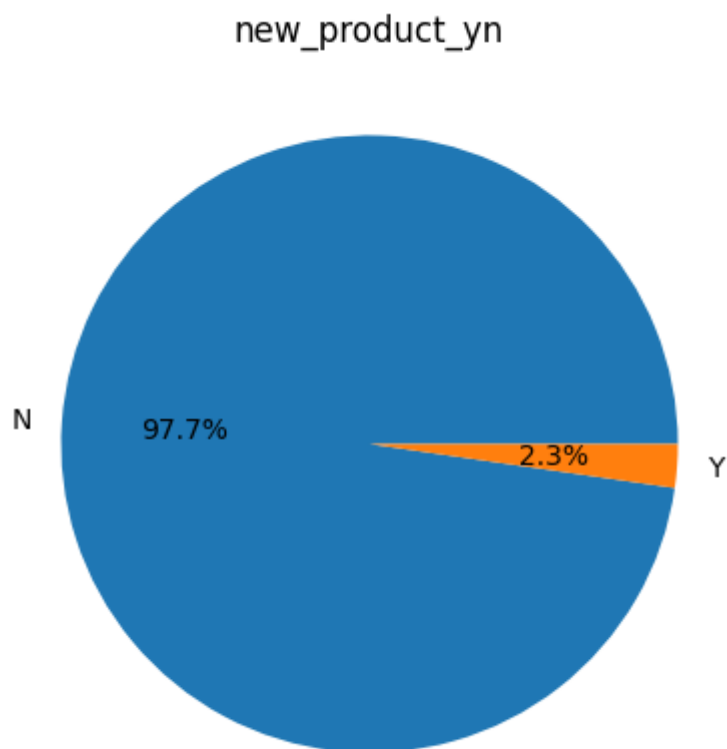
Hình 16. Thống kê mô tả của taxx_exempt_yn

Nhận xét: Có 18,2% sản phẩm không miễn thuế, còn lại 81.8% có miễn thuế.



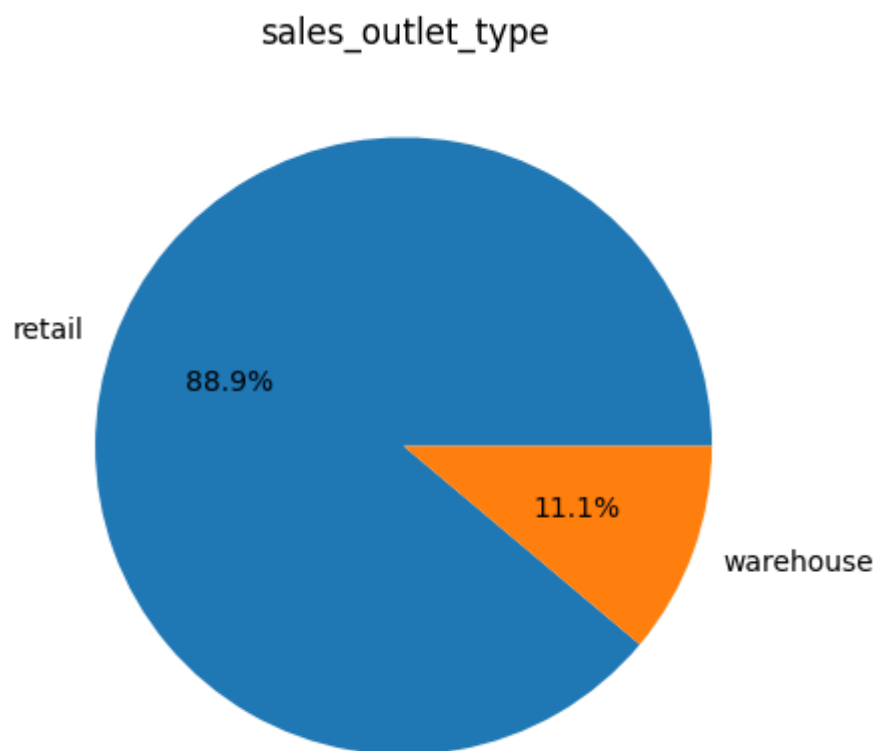
Hình 17. Thống kê mô tả của promo_yn

Nhận xét: Có 4,5% trên tổng số sản phẩm được quảng bá.



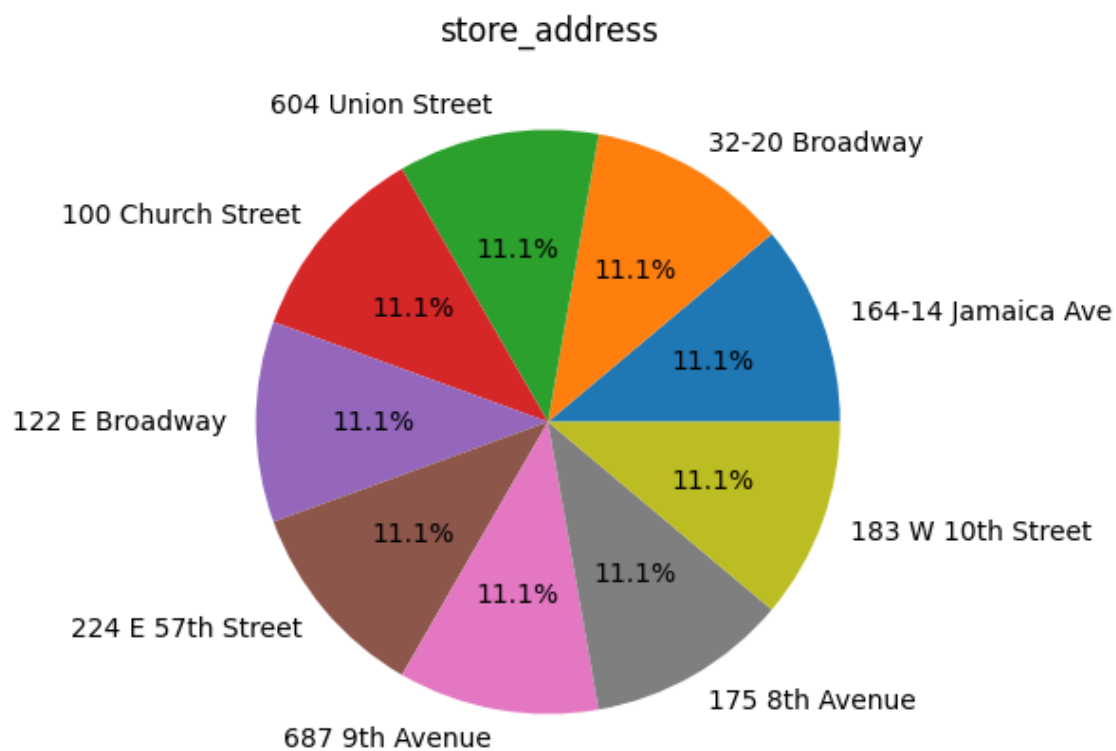
Hình 18. Thống kê mô tả của new_product_yn

Nhận xét: Chỉ có 2,3% trong tổng số là sản phẩm mới, còn lại 97,7% là sản phẩm đã có trước đó rồi.



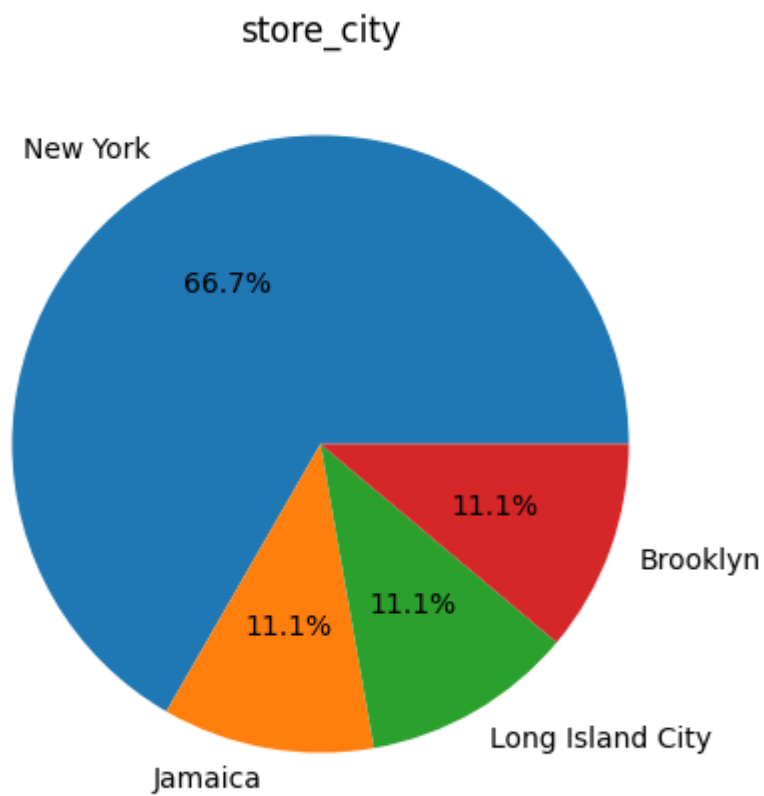
Hình 19. Thống kê mô tả của sales_outlet_type

Nhận xét: 88% trên tổng số các cửa hàng là cửa hàng bán lẻ, còn lại là Warehouse (Kho) chiếm 11.1%.



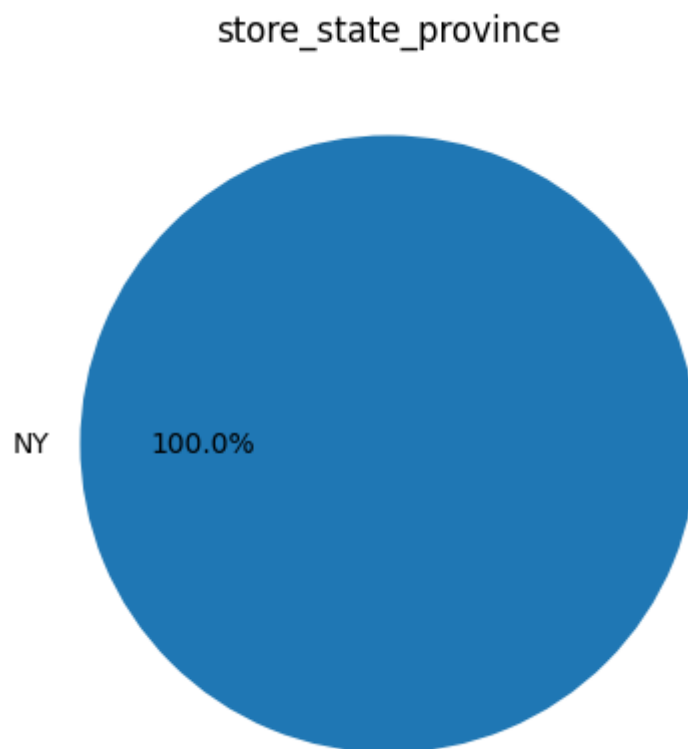
Hình 20. Thống kê mô tả của store_address

Nhận xét: Biểu đồ tròn mô tả sự phân bố đều giữa các địa chỉ cửa hàng, mỗi địa chỉ chiếm 11,1%.



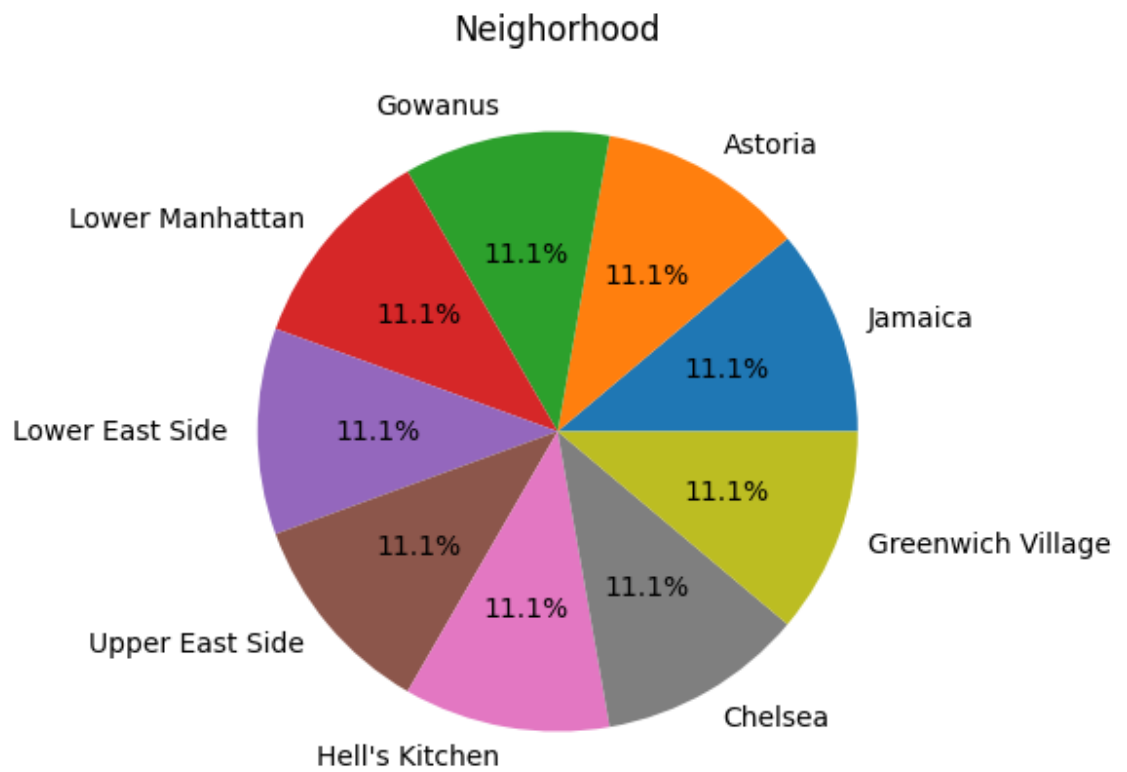
Hình 21. Thống kê mô tả của store_city

Nhận xét: New York chiếm 66.7% tổng số cửa hàng, là thành phố có số lượng cửa hàng nhiều nhất. Brooklyn, Long Island City, và Jamaica thì mỗi thành phố này chiếm 11.1% tổng số cửa hàng.



Hình 22. Thống kê mô tả của store_state_province

Nhận xét: Tất cả cửa hàng đều đặt tại tiểu bang New York 100%.



Hình 23. Thống kê mô tả của Neighborhood

Nhận xét: Biểu đồ tròn mô tả sự phân bố đều giữa các khu vực của cửa hàng, mỗi khu vực chiếm 11,1%.

❖ **Sau tiền xử lý:**

<i>quantity</i>	
Mean	1.438209
Standard Error	0.00243112
Median	1
Mode	1
Standard Deviation	0.5430387
Sample Variance	0.29489103
Kurtosis	1.09709037
Skewness	0.82201515
Range	7
Minimum	1
Maximum	8
Sum	71758
Count	49894

Hình 24. Thống kê mô tả của quantity sau tiền xử lý

Nhận xét: Biến quantity không có biến đổi so với ban đầu

<i>unit_price</i>	
Mean	3.38464545
Standard Error	0.01200944
Median	3
Mode	3
Standard Deviation	2.68254471
Sample Variance	7.19604613
Kurtosis	99.8313699
Skewness	8.53073038
Range	44.2
Minimum	0.8
Maximum	45
Sum	168873.5
Count	49894

Hình 25. Thống kê mô tả của unit_price sau tiền xử lý

Nhận xét: unit_price không có biến đổi.

<i>birth_year</i>	
Mean	1978.38557
Standard Error	0.3149373
Median	1981
Mode	1995
Standard Deviation	14.9255031
Sample Variance	222.770643
Kurtosis	-1.17774727
Skewness	-0.29139571
Range	51
Minimum	1950
Maximum	2001
Sum	4443454
Count	2246

Hình 26. Thống kê mô tả của *birth_year* sau tiền xử lý

Nhận xét: *birth_year* không có biến đổi.

<i>cost</i>	
Mean	3.88840909
Standard Error	0.60225816
Median	1.195
Mode	0.75
Standard Deviation	5.6496823
Sample Variance	31.9189101
Kurtosis	11.6143547
Skewness	2.89229139
Range	35.96
Minimum	0.04
Maximum	36
Sum	342.18
Count	88

Hình 27. Thống kê mô tả của *cost* sau tiền xử lý

Nhận xét: *cost* không biến đổi

<i>start_of_day</i>	
Mean	24.0586319
Standard Error	0.68849542
Median	18
Mode	18
Standard Deviation	12.0634142
Sample Variance	145.525963
Kurtosis	0.22784865
Skewness	1.49211489
Range	30
Minimum	18
Maximum	48
Sum	7386
Count	307

Hình 28. Thống kê mô tả của *start_of_day* sau tiền xử lý

Nhận xét: *start_of_day* không biến đổi

<i>quantity_sold</i>	
Mean	9.29641694
Standard Error	0.31048379
Median	8
Mode	9
Standard Deviation	5.44011548
Sample Variance	29.5948564
Kurtosis	3.14420427
Skewness	1.54749444
Range	32
Minimum	0
Maximum	32
Sum	2854
Count	307

Hình 29. Thống kê mô tả của *quantity_sold* sau tiền xử lý

Nhận xét: quantity_sold không biến đổi

<i>waste</i>	
Mean	14.7752443
Standard Error	0.635902319
Median	11
Mode	9
Standard Deviation	11.14190873
Sample Variance	124.1421303
Kurtosis	1.372911607
Skewness	1.559930382
Range	47
Minimum	0
Maximum	47
Sum	4536
Count	307

Hình 30. Thống kê mô tả của waste sau tiền xử lý

Nhận xét: trung bình tăng lên 14,775. Trung vị và mode giữ nguyên giá trị.

<i>% waste</i>	
Mean	0.587589577
Standard Error	0.011753328
Median	0.6
Mode	0.5
Standard Deviation	0.20593495
Sample Variance	0.042409204
Kurtosis	-0.212732833
Skewness	-0.36949629
Range	1
Minimum	0
Maximum	1
Sum	180.39
Count	307

Hình 31. Thống kê mô tả của %waste sau tiền xử lý

Nhận xét: trung vị tăng lên 0,6; skewness giảm còn -0.36. Dữ liệu lệch trái.

<i>total_goal</i>	
Mean	19437.5
Standard Error	970.260919
Median	18000
Mode	18000
Standard Deviation	2744.3123
Sample Variance	7531250
Kurtosis	1.57366437
Skewness	1.68662681
Range	7000
Minimum	18000
Maximum	25000
Sum	155500
Count	8

Hình 32. Thống kê mô tả của total_goal sau tiền xử lý

Nhận xét: total_goal không biến đổi đáng kể.

Kết luận: Sau khi tiền xử lý, các chỉ số thống kê có chút thay đổi. Tuy nhiên hình dạng phân phối là không đổi.

Biến category: sau khi tiền xử lý, chỉ số thống kê mô tả của biến category là không đổi.

3.2. Giai đoạn 2: Chuyển đổi dữ liệu (Transform)

Vì dữ liệu nằm ở các file riêng lẻ khó, nhóm tiến hành tổng hợp các file riêng đó vào 1 file chung 9 sheet, mỗi sheet chứa dữ liệu của 1 file. Nhóm tiến hành tiền xử lý trên file vừa tổng hợp được.

3.2.1. Tiền xử lý dữ liệu

Đầu tiên, nhóm chuyển kiểu dữ liệu của tất cả các cột id về dạng ký tự (text).

Kế tiếp, nhóm chỉnh format cột transaction_date (sheet sales, pastry inventory, date). Vì những cột này đang ở dạng mm/dd/yyyy nhưng vì xung đột định dạng nên nhóm đặt lại thành định dạng dd/mm/yyyy.

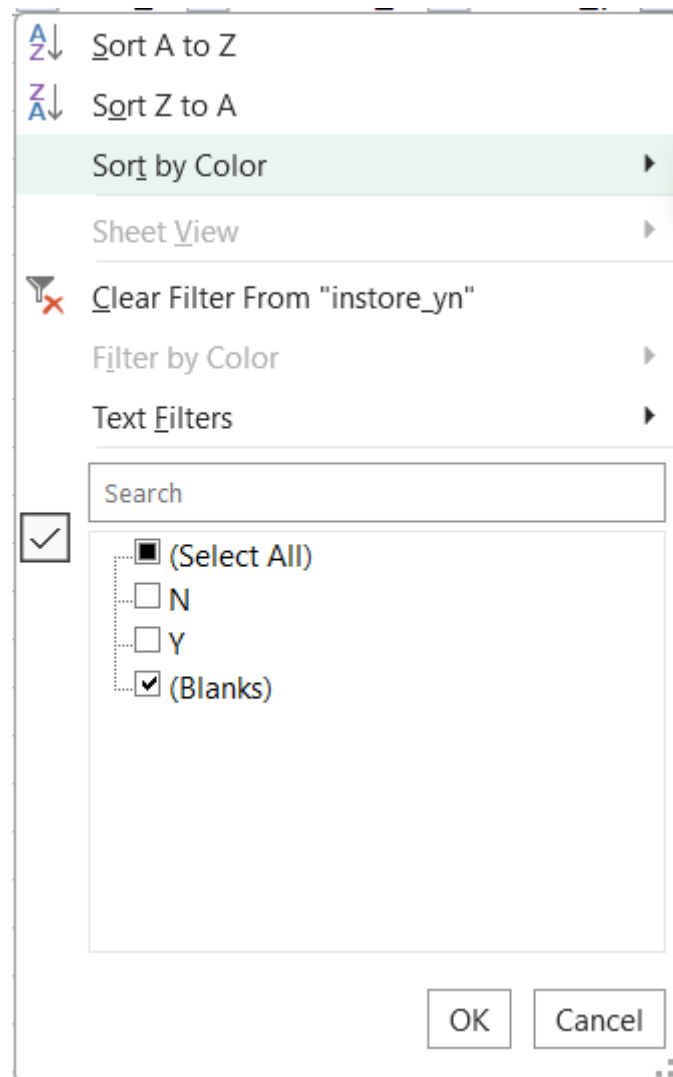
	A	B	C	D	E	F	G	H	I	J
1	transaction_date	Date_ID	Week_ID	Week_Des	Month_ID	Month_Na	Quarter_ID	Quarter_N	Year_ID	
2	04/01/2019	20190401	14	Week 14	4	April	2	Q2	2019	
3	04/02/2019	20190402	14	Week 14	4	April	2	Q2	2019	
4	04/03/2019	20190403	14	Week 14	4	April	2	Q2	2019	
5	04/04/2019	20190404	14	Week 14	4	April	2	Q2	2019	
6	04/05/2019	20190405	14	Week 14	4	April	2	Q2	2019	
7	04/06/2019	20190406	14	Week 14	4	April	2	Q2	2019	
8	04/07/2019	20190407	14	Week 14	4	April	2	Q2	2019	
9	04/08/2019	20190408	15	Week 15	4	April	2	Q2	2019	
10	04/09/2019	20190409	15	Week 15	4	April	2	Q2	2019	
11	04/10/2019	20190410	15	Week 15	4	April	2	Q2	2019	
12	04/11/2019	20190411	15	Week 15	4	April	2	Q2	2019	
13	04/12/2019	20190412	15	Week 15	4	April	2	Q2	2019	
14	4/13/2019	20190413	15	Week 15	4	April	2	Q2	2019	
15	4/14/2019	20190414	15	Week 15	4	April	2	Q2	2019	
16	4/15/2019	20190415	16	Week 16	4	April	2	Q2	2019	
17	4/16/2019	20190416	16	Week 16	4	April	2	Q2	2019	
18	4/17/2019	20190417	16	Week 16	4	April	2	Q2	2019	
19	4/18/2019	20190418	16	Week 16	4	April	2	Q2	2019	

Hình 33. Transaction_date trước khi đổi định dạng

	A	B	C	D	E	F	G	H	I
1	transaction_da	Date_IC	Week	Week_De	Month	Month_Nan	Quarter	Quarter_Nan	Year
2	01/04/2019	20190401	14	Week 14	4	April	2	Q2	2019
3	02/04/2019	20190402	14	Week 14	4	April	2	Q2	2019
4	03/04/2019	20190403	14	Week 14	4	April	2	Q2	2019
5	04/04/2019	20190404	14	Week 14	4	April	2	Q2	2019
6	05/04/2019	20190405	14	Week 14	4	April	2	Q2	2019
7	06/04/2019	20190406	14	Week 14	4	April	2	Q2	2019
8	07/04/2019	20190407	14	Week 14	4	April	2	Q2	2019
9	08/04/2019	20190408	15	Week 15	4	April	2	Q2	2019
10	09/04/2019	20190409	15	Week 15	4	April	2	Q2	2019
11	10/04/2019	20190410	15	Week 15	4	April	2	Q2	2019
12	11/04/2019	20190411	15	Week 15	4	April	2	Q2	2019
13	12/04/2019	20190412	15	Week 15	4	April	2	Q2	2019
14	13/04/2019	20190413	15	Week 15	4	April	2	Q2	2019
15	14/04/2019	20190414	15	Week 15	4	April	2	Q2	2019
16	15/04/2019	20190415	16	Week 16	4	April	2	Q2	2019
17	16/04/2019	20190416	16	Week 16	4	April	2	Q2	2019
18	17/04/2019	20190417	16	Week 16	4	April	2	Q2	2019
19	18/04/2019	20190418	16	Week 16	4	April	2	Q2	2019
20	19/04/2019	20190419	16	Week 16	4	April	2	Q2	2019
21	20/04/2019	20190420	16	Week 16	4	April	2	Q2	2019
22	21/04/2019	20190421	16	Week 16	4	April	2	Q2	2019
23	22/04/2019	20190422	17	Week 17	4	April	2	Q2	2019
24	23/04/2019	20190423	17	Week 17	4	April	2	Q2	2019
25	24/04/2019	20190424	17	Week 17	4	April	2	Q2	2019
26	25/04/2019	20190425	17	Week 17	4	April	2	Q2	2019
27	26/04/2019	20190426	17	Week 17	4	April	2	Q2	2019

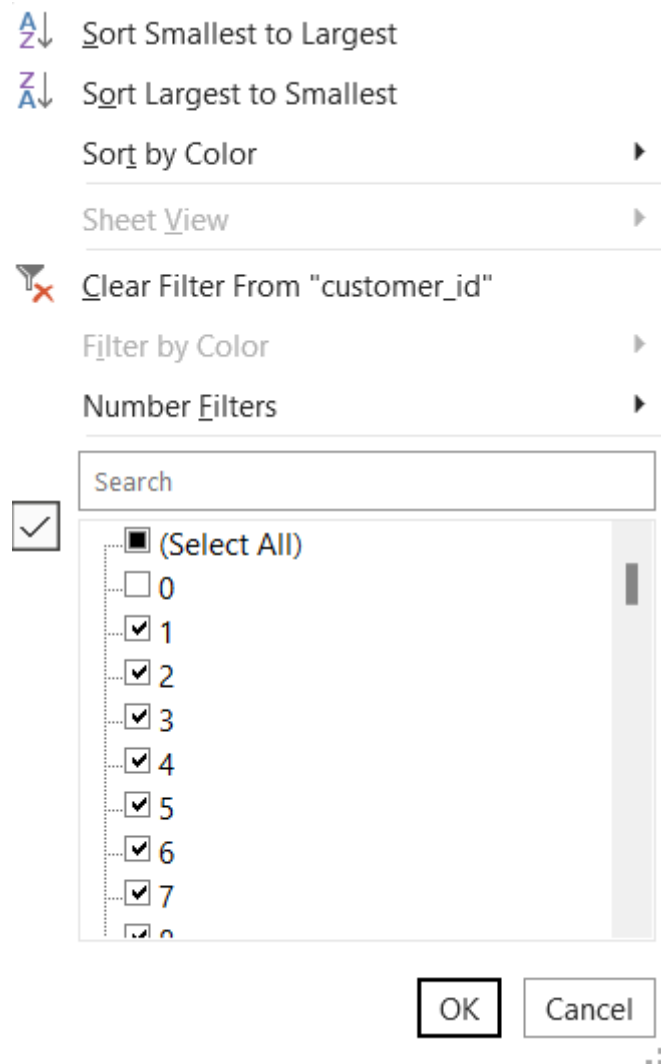
Hình 34. Kết quả sau khi chỉnh định dạng

Bảng sales receipts



Hình 35. Cột *instore_yn* có dữ liệu bị thiếu

Nhóm nhận thấy cột *instore_yn* có dữ liệu bị thiếu ở dạng khoảng trắng không chứa giá trị(blank), tiến hành drop 295 rows này.



Hình 36. Customer_id trong bảng sales_receipt

Tiếp theo, khi tham chiếu đến bảng customer, nhóm nhận thấy không tồn tại khách hàng có id = 0. Tuy nhiên trong bảng sales_receipts ghi nhận rất nhiều giao dịch có customer_id = 0. Tuy nhiên, nhóm không drop để đảm bảo tính toàn vẹn của dữ liệu. Nhóm đưa ra dự đoán rằng đây là những khách hàng chưa tham gia chương trình khách hàng thành viên, vì vậy không có thông tin của họ trong bảng customer. Song đây vẫn là những khách hàng có giá trị đối với doanh nghiệp, vẫn nên giữ lại lịch sử giao dịch này để phân tích ở tương lai. Trong phạm vi của đề án này, nhóm chỉ phân tích dữ liệu của khách hàng đã là khách hàng thành viên, có id ở bảng customer. Những giao dịch có customer_id = 0 sẽ được drop ở bước transform khi tạo bảng dim, fact.

Bảng Product:

Tiếp đến, cột `current_retail_price` đang có ký tự \$ nên nó ở định dạng text, nhóm tiến hành xóa ký tự \$ và đưa về dạng số để đảm bảo khả năng tính toán cho cột dữ liệu.

	D	E	F	G	H	I	J	K	L	M
1	product_type	product	product_description	unit_of_measure	current_wholesale_price	current_retail_price	current_retail_price	tax_exempt_yn	promo_yn	new_product_yn
2	Organic Beans	Brazilian - Organ	It's like Carnival in a cup. 12 oz		14.4	\$18.00	18	Y	N	N
3	House blend Bear	Our Old Time Dir	Out packed blend of bear 12 oz		14.4	\$18.00	18	Y	N	N
4	Espresso Beans	Espresso Roast	Our house blend for a go 1 lb		11.8	\$14.75	14.75	Y	N	N
5	Espresso Beans	Primo Espresso F	Our premium single source 1 lb		16.36	\$20.45	20.45	Y	N	N
6	Gourmet Beans	Columbian Medli A	smooth cup of coffee 1 lb		12	\$15.00	15	Y	N	N
7	Gourmet Beans	Ethiopia	From the home of coffee 1 lb		16.8	\$21.00	21	Y	N	N
8	Premium Beans	Jamaican Coffee Ya	man, it will start your 1 lb		15.8	\$19.75	19.75	Y	N	N
9	Premium Beans	Civet Cat	The most expensive coff. 1 lb		36	\$45.00	45	Y	N	N
10	Organic Beans	Organic Decaf Bl	Our blend of hand picked 1 lb		18	\$22.50	22.5	Y	N	N
11	Green beans	Guatemalan Sust	Green beans you can roa 1 lb		8	\$10.00	10	Y	N	N
12	Herbal tea	Lemon Grass	You will think you are Thi. 9 oz		7.16	\$8.95	8.95	Y	N	N
13	Herbal tea	Peppermint	Cool and refreshing to he. 9 oz		7.16	\$8.95	8.95	Y	N	N
14	Black tea	English Breakfast	The traditional cup to sta. 9 oz		7.16	\$8.95	8.95	Y	N	N
15	Black tea	Earl Grey	A full leaf of Orange Pek. 9 oz		7.16	\$8.95	8.95	Y	N	N
16	Green tea	Serenity Green T	Mountain grown and har 1 oz		7.4	\$9.25	9.25	Y	N	N
17	Chai tea	Traditional Blenc	A traditional blend. 9 oz		7.16	\$8.95	8.95	Y	N	N
18	Chai tea	Morning Sunrise	Fair trade and organic an. 9 oz		7.6	\$9.50	9.5	Y	N	N
19	Chai tea	Spicy Eye Opene	A spicier blend to awaker. 9 oz		8.76	\$10.95	10.95	Y	N	N
20	Drinking Chocolat	Dark chocolate	This drinking chocolate is 1 lb		5.12	\$6.40	6.4	Y	N	N
21	Organic Chocolat	Sustainably Grow	Certified organic contain 1 lb		6.08	\$7.60	7.6	Y	N	N
22	Drinking Chocolat	Chili Mayan	Fragrant with spices, this 1 lb		10.66	\$13.33	13.33	Y	N	N
23	Drip coffee	Our Old Time Dir	An honest cup a coffee. 8 oz		0.4	\$2.00	2	Y	N	N
24	Drip coffee	Our Old Time Dir	An honest cup a coffee. 16 oz		0.5	\$2.50	2.5	Y	N	N
25	Drip coffee	Our Old Time Dir	An honest cup a coffee. 24 oz		0.6	\$3.00	3	Y	N	N
26	Organic brewed c	Brazilian Sm	It's like Carnival in a cup. 8 oz		0.44	\$2.20	2.2	Y	N	N
27	Organic brewed c	Brazilian Rg	It's like Carnival in a cup. 16 oz		0.6	\$3.00	3	Y	N	N
28	Organic brewed c	Brazilian Lg	It's like Carnival in a cup. 24 oz		0.7	\$3.50	3.5	Y	N	N
29	Gourmet brewed	Columbian Medli	A smooth cup of coffee 8 oz		0.4	\$2.00	2	Y	N	N

Hình 37. Bảng dữ liệu sau khi tạo cột mới

Tạo cột mới, điền công thức: `=REPLACE(J2,1,1,"")`. Công thức trên sẽ thay thế ký tự đầu tiên trong chuỗi cũ bằng một chuỗi rỗng, tức là nó sẽ xóa ký tự đầu tiên. Xóa cột cũ, thay thế bằng giá trị cột mới.

Bảng `pastry_inventory`:

Xuất hiện dữ liệu bất thường khi mà `sold = 0` nhưng `waste = 0`. Nhóm điều chỉnh điều chỉnh giá trị cột `waste = 18`.

Bảng `customer`:

Tiếp theo nhóm nhận thấy cột `home_store` của sheet `customer` có dữ liệu tương đương cột `sale_outlet_id` của `sales receipts`, nên tiến hành đổi tên thành `sales_outlet_id` cho đồng bộ.

3.2.2. Transform

Ở phần này nhóm tiến hành tạo bảng `dim`, `fact`. Một số cột, bảng (`staff`, `generation`) không có giá trị phân tích hoặc nhóm không tìm được định nghĩa chính xác của chúng sẽ bị lược bỏ ở bước này:

Bảng Fact:

Tạo thêm cột Date_ID bằng công thức:
VLOOKUP(B2,dates!\$A\$2:\$B\$31,2,FALSE)

Rename cột line_item_amount thành sales_amount

Chỉnh kiểu dữ liệu của transaction_time về dạng time

Loại bỏ customer id không có trong bảng dim_customer bằng hàm: VLOOKUP(F2,dim_customer!\$A\$2:\$A\$2248,1,FALSE). Những dòng customer_id của bảng Fact không có trong dim_customer sẽ ở trả về giá trị N/A. Xóa các dòng này.

Đồng thời xóa các customer_id = 0

Tạo thêm bảng:

Tạo bảng product_group, product_category, product_subcate từ những cột trong bảng product.

3.3. Giai đoạn 3: Nạp dữ liệu vào kho dữ liệu (Load)

Nhóm load data (file sales_reciepts(dim_fact)) vào công cụ Power Bi để tiến hành phân tích.

3.3.1 Giới thiệu các bảng Dim

Bảng 3. Mô tả các bảng dim

Tên thuộc tính	Kiểu dữ liệu (Theo Power BI)	Mô tả	Ghi chú
Customer			
customer_id	text	Mã định danh (ID) cho mỗi khách hàng	

customer_since	date	Thời gian khách hàng bắt đầu tham gia chương trình khách hàng thành viên	
loyalty_card_number	text	Số thẻ thành viên của khách hàng	
birthdate	date	Ngày, tháng, năm sinh của khách hàng	
gender	text	Giới tính của khách hàng	
birth_year	whole number	Năm sinh của khách hàng	
Dates			
transaction_date	date	Ngày, tháng, năm thực hiện giao dịch	
date_id	text	Mã ID của ngày	
week_id	text	Mã ID của tuần	
week_desc	text	Tên mô tả tuần	Ví dụ: week 4
month_id	text	Mã ID của tháng	
month_name	text	Tên tháng	
quarter_id	text	Mã ID của quý	
quarter_name	text	Tên quý	

year_id	text	Mã ID của năm	
date	whole number	Ngày thực hiện giao dịch	
pastry_inventory			
sales_outlet_id	text	Mã định danh (ID) của mỗi cửa hàng	
transaction_date	date	Ngày nhập hàng vào kho	
product_id	text	Mã ID của sản phẩm	
start_of_day	whole number	Số lượng sản phẩm đầu ngày	
quantity_sold	whole number	Số lượng sản phẩm đã được bán ra	
waste	whole number	Số lượng sản phẩm còn tồn lại trong ngày	
%waste	decimal number	Phần trăm tính tổn thất trong ngày	
cost	decimal number	Giá vốn của sản phẩm	
product			
product_id	text	Mã ID của sản phẩm	
product_subcate gory	text	Danh mục con của sản phẩm	

product_subcategory_id	text	Mã ID của danh mục con sản phẩm	
product	text	Tên sản phẩm	
cost	decimal number	Giá vốn của sản phẩm	
current_retail_price	decimal number	Giá bán lẻ sản phẩm	
promo_yn	text	Sản phẩm có được khuyến mãi không?	Có 2 giá trị: yes và no
new_product_yn	text	Có phải là sản phẩm mới không?	Có 2 giá trị: yes và no
sales_outlet			
sales_outlet_id	text	Mã định danh (ID) của mỗi cửa hàng	
store_address	text	Địa chỉ của cửa hàng	
store_city	text	Thành phố đặt cửa hàng	
neighborhood	text	Khu vực đặt cửa hàng	
total_goal	whole number	Tổng mục tiêu bán hàng	
product_group			

product_group	text	Tên nhóm sản phẩm	
product_group_id	text	Mã ID của nhóm sản phẩm	
product_category			
product_group	text	Tên nhóm sản phẩm	
product_group_id	text	Mã ID của nhóm sản phẩm	
product_category	text	Danh mục sản phẩm	
product_category_id	text	Mã ID của danh mục sản phẩm	
product_subcate			
product_subcategory	text	Danh mục con của sản phẩm	
product_subcategory_id	text	Mã ID của danh mục con sản phẩm	
product_category_id	text	Mã ID của danh mục sản phẩm	

3.3.2. Giới thiệu bảng Fact

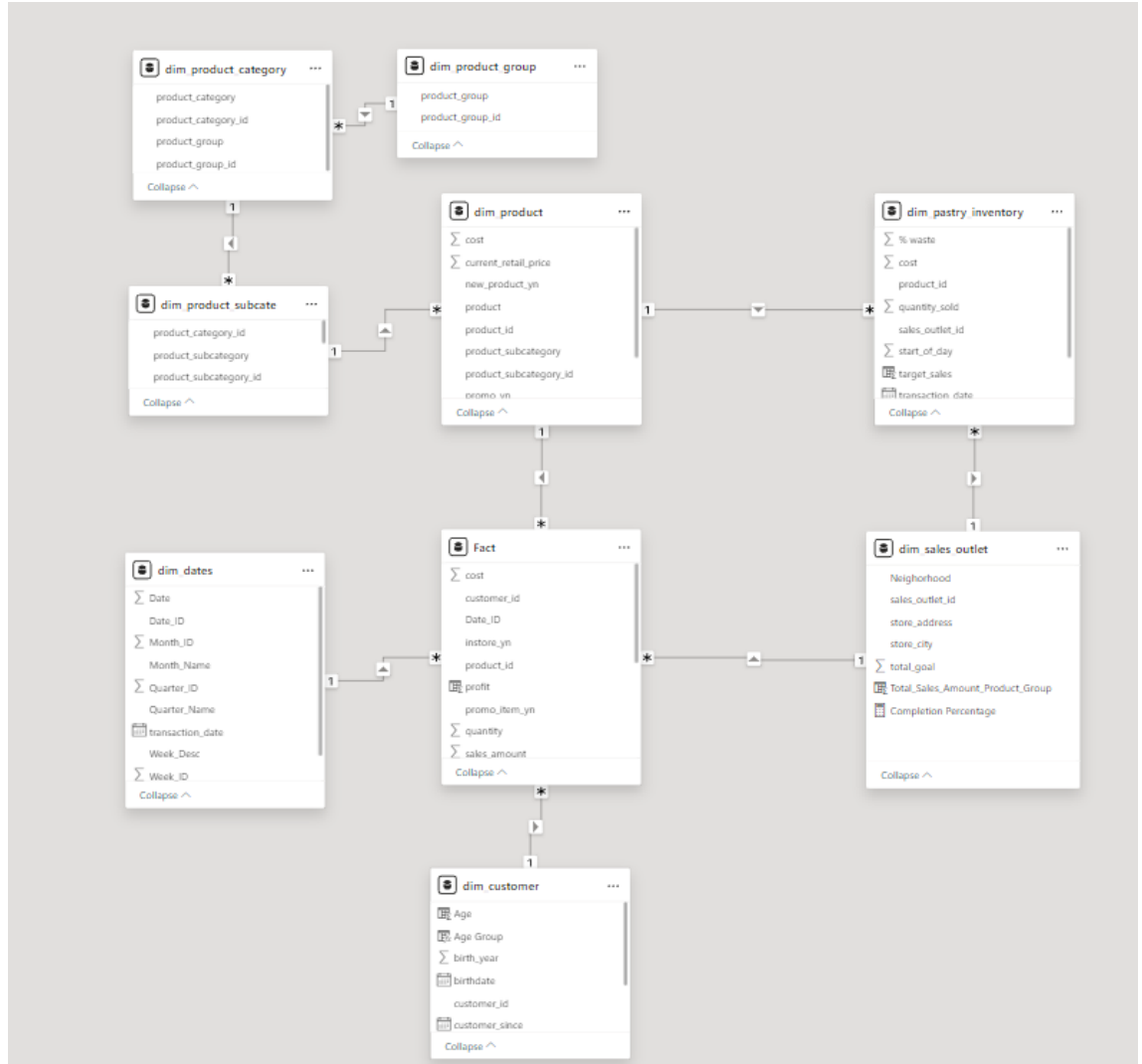
Bảng 4. Mô tả bảng Fact

STT	Tên thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	transaction_id	text	Mã ID của n giao dịch	

2	transaction_date	date	Ngày thực hiện giao dịch	
3	date_id	text	Mã ID của ngày thực hiện giao dịch	
4	transaction_time	time	Thời gian thực hiện giao dịch	
5	sales_outlet_id	text	Mã định danh (ID) của mỗi cửa hàng	
6	customer_id	text	Mã định danh (ID) cho mỗi khách hàng	
7	instore_yn	text	Khách dùng tại quán hay mang về	Có 2 giá trị: yes và no
8	product_id	text	Mã ID của sản phẩm	
9	quantity	whole number	Số lượng sản phẩm bán ra trong giao dịch	

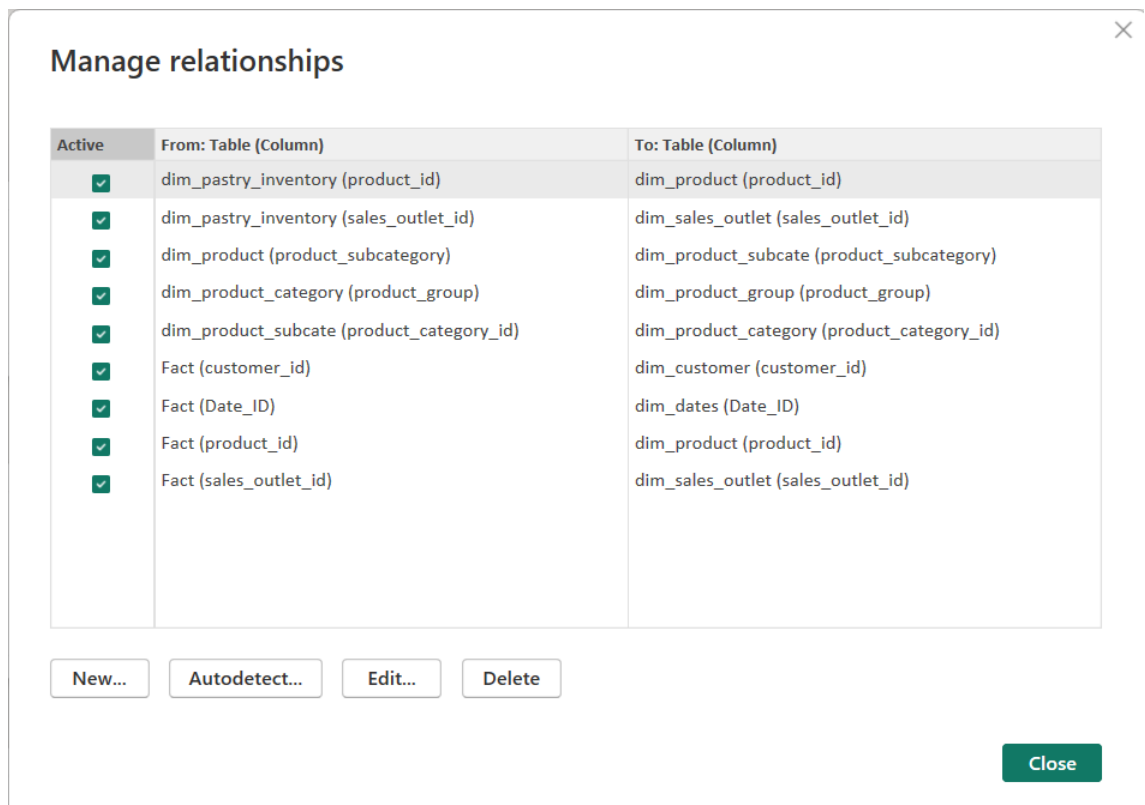
10	sales_amount	decimal number	Doanh thu	
11	unit_price	decimal number	Đơn giá bán lẻ	
12	cost	decimal number	Giá vốn của sản phẩm	
13	promo_yn	text	Sản phẩm có được khuyến mãi không?	Có 2 giá trị: yes và no

3.3.3. Mô hình dữ liệu (Data Model)



Hình 38. Model schema

Mô hình dữ liệu của nhóm được xây dựng theo kiểu mô hình lược đồ bông tuyết (snowflakes schema). Bảng Fact ở giữa là trung tâm, nối với các dimension khác gồm customer, dates, product, sales_outlet. 3 dimension product group, product category, product subcate nối với Fact một cách trung gian thông qua product. Tuy nhiên sự xuất hiện của bảng pastry_inventory có thể xem là bảng fact thứ 2 trong khái niệm multiple lược đồ Constellation.



Hình 39. Quan hệ giữa các bảng

4. PHÂN TÍCH DỮ LIỆU

4.1. Các hàm DAX

4.1.1. Hàm DAX cho Column

target_sales = dim_pastry_inventory[start_of_day] * dim_pastry_inventory[cost]

- dim_pastry_inventory[start_of_day]: Đây là cột start_of_day trong bảng dim_pastry_inventory, thể hiện số lượng bánh ngọt có sẵn ở đầu ngày.
- dim_pastry_inventory[cost]: Đây là cột cost trong bảng dim_pastry_inventory, thể hiện giá của mỗi chiếc bánh ngọt.
- dim_pastry_inventory[start_of_day] * dim_pastry_inventory[cost]: Phép nhân này tính toán doanh số dự kiến cho mỗi ngày dựa trên số lượng bánh ngọt có sẵn ở đầu ngày và giá của mỗi chiếc bánh ngọt.

Kết quả cuối cùng của hàm này là một cột mới “target_sales” trong đó mỗi giá trị là doanh số dự kiến cho mỗi ngày dựa trên số lượng bánh ngọt có sẵn và giá của chúng.

$$\text{waste_value} = \text{dim_pastry_inventory}[\text{target_sales}] - (\text{dim_pastry_inventory}[\text{quantity_sold}] * \text{dim_pastry_inventory}[\text{cost}])$$

- `dim_pastry_inventory[target_sales]`: Đây là cột `target_sales` trong bảng `dim_pastry_inventory`, thể hiện doanh số dự kiến trong ngày.
- `dim_pastry_inventory[quantity_sold]`: Đây là cột `quantity_sold` trong bảng `dim_pastry_inventory`, thể hiện số lượng bánh ngọt đã bán được.
- `dim_pastry_inventory[cost]`: Đây là cột `cost` trong bảng `dim_pastry_inventory`, thể hiện giá của mỗi chiếc bánh ngọt.
- `dim_pastry_inventory[target_sales] - (dim_pastry_inventory[quantity_sold] * dim_pastry_inventory[cost])`: Phép trừ này tính toán khoản lỗ trong ngày, được tính bằng cách lấy doanh số dự kiến trừ đi tổng giá trị của số lượng bánh ngọt đã bán.

Kết quả cuối cùng của hàm này là một cột mới “waste_value” trong đó mỗi giá trị là khoản lỗ trong ngày dựa trên doanh số dự kiến và giá trị thực tế từ số lượng bánh ngọt đã bán.

$$\text{Age} = 2019 - \text{dim_customer}[\text{birth_year}]$$

Hàm này tạo ra một cột mới có tên là “Age”. Giá trị của “Age” được tính bằng cách lấy năm hiện tại (được trả về bởi hàm `TODAY()`) trừ đi năm sinh (cột `birth_year`) trong bảng `dim_customer`. Kết quả là tuổi của khách hàng.

Age Group =

SWITCH (

`TRUE ()`,

```

'dim_customer'[Age] < 15, "<15",
'dim_customer'[Age] >= 15 && 'dim_customer'[Age] < 20, "15-19",
'dim_customer'[Age] >= 20 && 'dim_customer'[Age] < 30, "20-30",
'dim_customer'[Age] >= 30 && 'dim_customer'[Age] < 40, "30-40",
'dim_customer'[Age] >= 40 && 'dim_customer'[Age] < 50, "40-50",
'dim_customer'[Age] >= 50 && 'dim_customer'[Age] < 65, "50-65",
"65+"
)

```

Hàm này tạo ra một cột mới có tên là “Age Group”. Giá trị của “Age Group” được xác định dựa trên giá trị của cột “Age” trong bảng dim_customer. Hàm SWITCH được sử dụng để ánh xạ các khoảng tuổi tới các nhóm tuổi tương ứng. Ví dụ, nếu tuổi của khách hàng nhỏ hơn 15, giá trị của “Age Group” sẽ là “<15”. Nếu tuổi của khách hàng từ 15 đến dưới 20, giá trị của “Age Group” sẽ là “15-19”, và tương tự cho các khoảng tuổi khác. Nếu tuổi của khách hàng lớn hơn hoặc bằng 65, giá trị của “Age Group” sẽ là “65+”.

```
total_cost = 'Fact'[cost] * 'Fact'[quantity]
```

Hàm này tính tổng giá trị của tất cả các sản phẩm dựa trên giá và số lượng của chúng.

```
profit = 'Fact'[sales_amount] - 'Fact'[total_cost]
```

Hàm này tính lợi nhuận từ việc bán hàng sau khi trừ đi chi phí.

4.1.2. Hàm DAX cho Measure

SelectedProduct =

```
VAR SProduct =  
SELECTEDVALUE(dim_product[product])
```

RETURN

```
IF(ISBLANK(SELECTEDVALUE(dim_product[product])), "All  
Products", SProduct)
```

- VAR SProduct = SELECTEDVALUE(dim_product[product]): Đây là phần khai báo biến SProduct sử dụng hàm SELECTEDVALUE để lấy giá trị được chọn từ cột product trong bảng dim_product.
- SELECTEDVALUE(dim_product[product]): Hàm SELECTEDVALUE trả về giá trị duy nhất đã được chọn hoặc lọc trong cột product của bảng dim_product. Nếu có nhiều hơn một giá trị được chọn, hàm này sẽ trả về kết quả là BLANK.
- IF(ISBLANK(SELECTEDVALUE(dim_product[product])), "All Products", SProduct): Hàm IF kiểm tra xem giá trị được chọn từ cột product có phải là BLANK hay không. Nếu là BLANK (tức là không có giá trị nào được chọn hoặc có nhiều hơn một giá trị được chọn), hàm sẽ trả về chuỗi "All Products". Nếu không, hàm sẽ trả về giá trị của biến SProduct (tức là giá trị duy nhất đã được chọn trong cột product).

ProductGroupOfProduct =

```
VAR ProductSubcategoryID =  
MAXX(  
    FILTER(  
        dim_product,  
        dim_product[product] =  
SELECTEDVALUE(dim_product[product])  
    ),  
    dim_product[product_subcategory_id]  
)
```

```

VAR ProductCategoryID =

    CALCULATE(

        VALUES(dim_product_subcate[product_category_id]),

        dim_product_subcate[product_subcategory_id] =
ProductSubcategoryID

    )

VAR ProductGroupID =

    CALCULATE(

        VALUES(dim_product_category[product_group_id]),

        dim_product_category[product_category_id] =
ProductCategoryID

    )

RETURN

    CALCULATE(

        VALUES(dim_product_group[product_group]),

        dim_product_group[product_group_id] = ProductGroupID

    )

```

- VAR ProductSubcategoryID: Đây là phần khai báo biến ProductSubcategoryID. Biến này được tính bằng cách lấy giá trị lớn nhất của product_subcategory_id từ bảng dim_product cho sản phẩm được chọn hiện tại.
- VAR ProductCategoryID: Đây là phần khai báo biến ProductCategoryID. Biến này được tính bằng cách lấy giá trị duy nhất của product_category_id từ bảng dim_product_subcate cho product_subcategory_id tương ứng với ProductSubcategoryID.
- VAR ProductGroupID: Đây là phần khai báo biến ProductGroupID. Biến này được tính bằng cách lấy giá trị duy nhất của product_group_id từ bảng dim_product_category cho product_category_id tương ứng với ProductCategoryID.
- RETURN CALCULATE(...): Phần cuối cùng của hàm này trả về giá trị duy nhất của product_group từ bảng dim_product_group cho product_group_id tương ứng với ProductGroupID.

```

Top5SubcategoryQuantity =

VAR          SSubcategories          =
SELECTEDVALUE(dim_product[product_subcategory])

VAR Top5Subcategories = TOPN(5, SUMMARIZE('Fact',
'dim_product'[product_subcategory], "TotalSubcategoryQuantity",
SUM('Fact'[quantity])), [TotalSubcategoryQuantity], DESC)

VAR Top5TotalQuantity = SUMX(Top5Subcategories,
[TotalSubcategoryQuantity])

RETURN

IF (

ISBLANK ( SSubcategories ),

Top5TotalQuantity,

CALCULATE (

SUM ( 'Fact'[quantity] ),

dim_product[product_subcategory] = SSubcategories

)

)

```

- VAR SSubcategories = SELECTEDVALUE(dim_product[product_subcategory]): Đây là phần khai báo biến SSubcategories. Biến này được tính bằng cách lấy giá trị được chọn từ cột product_subcategory trong bảng dim_product.

- VAR Top5Subcategories = ...: Đây là phần khai báo biến Top5Subcategories. Biến này được tính bằng cách lấy 5 danh mục sản phẩm con có tổng quantity lớn nhất từ bảng Fact.
- VAR Top5TotalQuantity = ...: Đây là phần khai báo biến Top5TotalQuantity. Biến này được tính bằng cách lấy tổng quantity của 5 danh mục sản phẩm con được lưu trong biến Top5Subcategories.
- RETURN IF (...): Phần cuối cùng của hàm này kiểm tra xem giá trị của SSubcategories có phải là BLANK hay không. Nếu là BLANK (tức là không có danh mục sản phẩm con nào được chọn), hàm sẽ trả về tổng quantity của 5 danh mục sản phẩm con lớn nhất. Nếu không, hàm sẽ trả về tổng quantity của danh mục sản phẩm con được chọn.

Top5SubcategoryPercentage =

VAR TotalQuantity = SUM('Fact'[quantity])

VAR Top5Subcategories = TOPN(5, SUMMARIZE('Fact',
'dim_product'[product_subcategory], "TotalSubcategoryQuantity",
SUM('Fact'[quantity])), [TotalSubcategoryQuantity], DESC)

VAR Top5TotalQuantity = SUMX(Top5Subcategories,
[TotalSubcategoryQuantity])

RETURN DIVIDE(Top5TotalQuantity, TotalQuantity, 0)

- VAR TotalQuantity = SUM('Fact'[quantity]): Đây là phần khai báo biến TotalQuantity. Biến này được tính bằng cách lấy tổng của cột quantity trong bảng Fact.
- VAR Top5Subcategories = ...: Đây là phần khai báo biến Top5Subcategories. Biến này được tính bằng cách lấy 5 danh mục sản phẩm con có tổng quantity lớn nhất từ bảng Fact.
- VAR Top5TotalQuantity = ...: Đây là phần khai báo biến Top5TotalQuantity. Biến này được tính bằng cách lấy tổng quantity của 5 danh mục sản phẩm con được lưu trong biến Top5Subcategories.

- RETURN DIVIDE(Top5TotalQuantity, TotalQuantity, 0): Phần cuối cùng của hàm này trả về tỷ lệ phần trăm của tổng quantity của 5 danh mục sản phẩm con lớn nhất so với tổng quantity của tất cả các danh mục sản phẩm con.

Kết quả cuối cùng của hàm này là một measure mới “Top5SubcategoryPercentage” trong đó mỗi giá trị là tỷ lệ phần trăm của tổng quantity của 5 danh mục sản phẩm con lớn nhất so với tổng quantity của tất cả các danh mục sản phẩm con.

Bottom5SubcategoryQuantity =

VAR **SSubcategories** =
SELECTEDVALUE(dim_product[product_subcategory])

VAR **Top5Subcategories** = **TOPN**(5, **SUMMARIZE**('Fact',
'dim_product'[product_subcategory], "**TotalSubcategoryQuantity**",
SUM('Fact'[quantity])), [TotalSubcategoryQuantity], **ASC**)

VAR **Top5TotalQuantity** = **SUMX**(Top5Subcategories,
[TotalSubcategoryQuantity])

RETURN

IF (

ISBLANK (**SSubcategories**),

Top5TotalQuantity,

CALCULATE (

SUM ('Fact'[quantity]),

dim_product[product_subcategory] = **SSubcategories**

)

)

- `VAR SSUBCATEGORIES = SELECTEDVALUE(dim_product[product_subcategory])`: Định nghĩa một biến `SSubcategories` bằng cách lấy giá trị được chọn từ cột `product_subcategory` trong bảng `dim_product`.
- `VAR Top5Subcategories = TOPN(5, SUMMARIZE('Fact', 'dim_product'[product_subcategory], "TotalSubcategoryQuantity", SUM('Fact'[quantity])), [TotalSubcategoryQuantity], ASC)`: Định nghĩa một biến `Top5Subcategories` bằng cách lấy 5 danh mục hàng đầu từ bảng `Fact`, được tổng hợp theo `product_subcategory` trong bảng `dim_product` và tính tổng số lượng (`quantity`). Kết quả được sắp xếp theo `TotalSubcategoryQuantity` tăng dần.
- `VAR Top5TotalQuantity = SUMX(Top5Subcategories, [TotalSubcategoryQuantity])`: Định nghĩa một biến `Top5TotalQuantity` bằng cách tính tổng `TotalSubcategoryQuantity` cho tất cả các `Top5Subcategories`.
- `RETURN IF (ISBLANK (SSUBCATEGORIES), Top5TotalQuantity, CALCULATE (SUM ('Fact'[quantity]), dim_product[product_subcategory] = SSUBCATEGORIES))`: Trả về kết quả cuối cùng. Nếu `SSubcategories` trống, nó sẽ trả về `Top5TotalQuantity`. Nếu không, nó sẽ tính tổng `quantity` trong bảng `Fact` cho những dòng mà `product_subcategory` trong bảng `dim_product` bằng `SSubcategories`.

Completion Percentage =

`VAR TotalSales = SUM('Fact'[sales_amount])`

```

VAR TotalGoal =

    SWITCH(

        TRUE(),

        MAX('dim_sales_outlet'[total_goal]) <> BLANK(),
        MAX('dim_sales_outlet'[total_goal]),

        TRUE(),

        CALCULATE(SUM('dim_sales_outlet'[total_goal]),
        ALL('dim_sales_outlet'))

    )

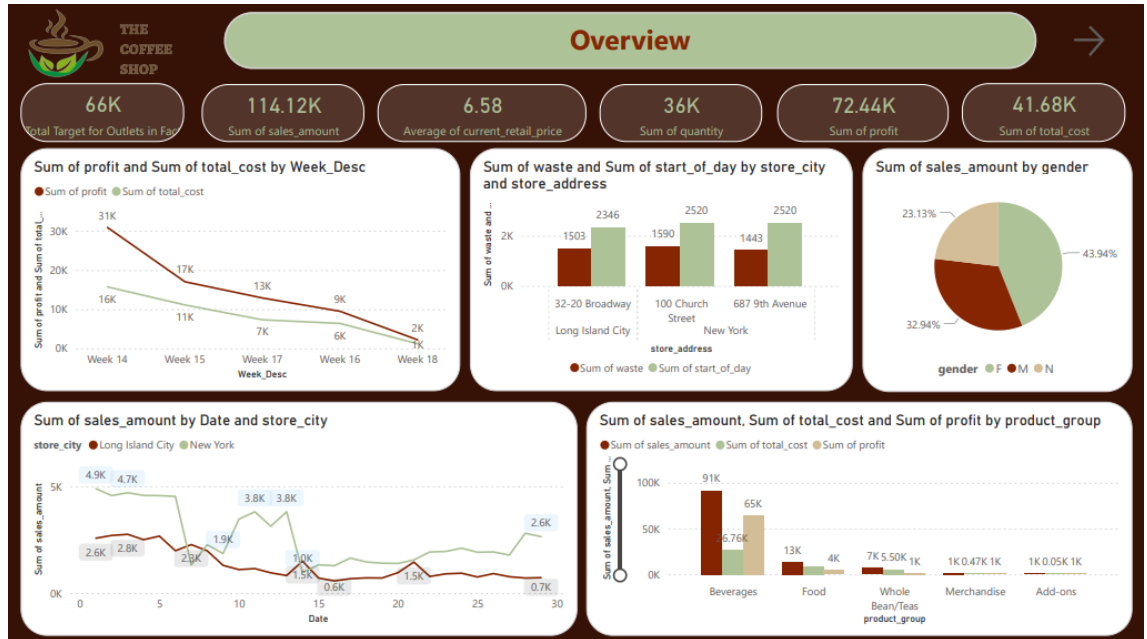
RETURN

    DIVIDE(TotalSales, TotalGoal, 0)

```

- VAR TotalSales = SUM('Fact'[sales_amount]): Đầu tiên, nó tính tổng số tiền bán hàng (TotalSales) từ bảng Fact.
- VAR TotalGoal = SWITCH(...): Tiếp theo, nó xác định mục tiêu tổng cộng (TotalGoal). Nếu mục tiêu tổng cộng (total_goal) trong bảng dim_sales_outlet không phải là trống, nó sẽ lấy giá trị lớn nhất. Nếu không, nó sẽ tính tổng mục tiêu tổng cộng (total_goal) cho tất cả các hàng trong bảng dim_sales_outlet.
- RETURN DIVIDE(TotalSales, TotalGoal, 0): Cuối cùng, nó trả về tỷ lệ giữa tổng số tiền bán hàng (TotalSales) và mục tiêu tổng cộng (TotalGoal). Nếu TotalGoal là 0, nó sẽ trả về 0 để tránh chia cho 0.

4.2. Overview



Hình 40. Dashboard Overview

Trang OverView cung cấp một cái nhìn toàn diện, giúp doanh nghiệp nắm bắt được mọi khía cạnh của khách hàng và sản phẩm một cách chi tiết. Nó không chỉ giúp doanh nghiệp hiểu rõ nhu cầu và hành vi của khách hàng mà còn cung cấp thông tin quan trọng giúp tối ưu hóa chiến lược sản phẩm.

Đầu tiên ta sẽ tiến hành xem xét các chỉ số tài chính tổng quan của doanh nghiệp này.

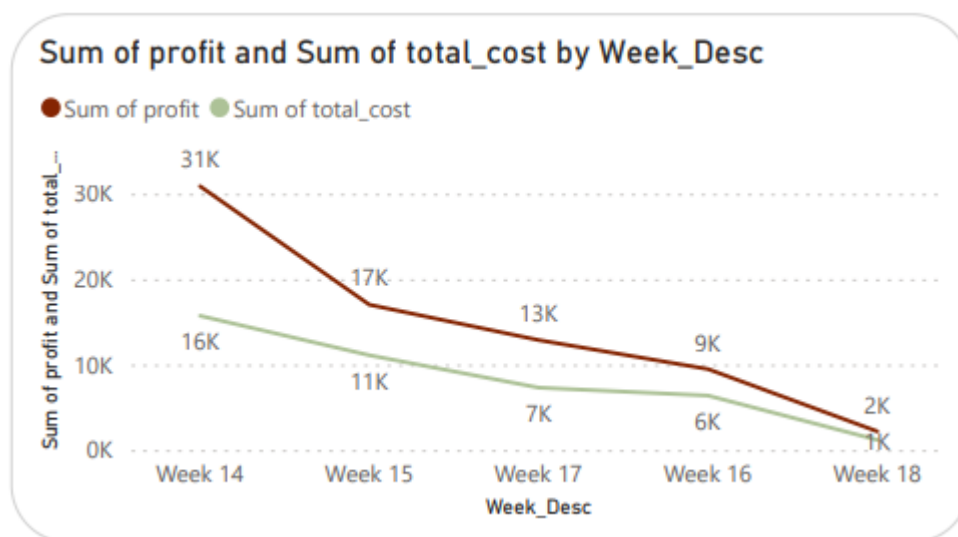


Hình 41. Cards trong dashboard Overview

Từ các chỉ số ở trên ta có thể thấy ba cửa hàng cà phê ở bang New York đang hoạt động khá tốt trong tháng tư. Với doanh thu đạt 114.12 nghìn đô la Mỹ và chi phí bỏ ra chỉ 41.68 nghìn đô la, có nghĩa là doanh nghiệp đã tạo ra lợi nhuận lên đến 72.44 nghìn đô la Mỹ chỉ trong một tháng. Chỉ tiêu cho tháng này là 66

nghìn đô la nhưng mức doanh thu 114.14 nghìn và lợi nhuận 72.44 nghìn đô la Mỹ thì doanh nghiệp hoàn toàn đạt mục tiêu, thậm chí vượt ngưỡng mong đợi. Nguyên nhân doanh nghiệp đạt được lợi nhuận thế này một phần là vì mức giá bán lẻ trung bình của cửa hàng nằm tầm 6.58 đô la Mỹ (trung bình của cả nước, thức ăn, ...). Đây là phân khúc giá hợp lý ở thành phố lớn để thu hút khách hàng, và kết quả là họ đã bán hơn 36 nghìn sản phẩm. Mỗi cốc cà phê không chỉ mang lại về giá trị tài chính mà còn góp phần xây dựng thương hiệu đáng tin cậy. Con số 36 nghìn sản phẩm được bán ra này cho thấy khách hàng không chỉ đến đây mua nước một lần mà có thể là quay lại nhiều lần. Điều này chứng minh sức hút và uy tín của thương hiệu.

Để hiểu rõ hơn những con số tài chính tổng quan trên, ta từng bước tìm hiểu rõ những khía cạnh khác.



Hình 42. Chart Sum of profit and Sum of total_cost by Week_desc

Biểu đồ cho thấy sự sụt giảm đáng kể cho trong cả tổng lợi nhuận và chi phí qua 5 tuần từ tuần 14 đến tuần 18. Tại tuần 14, tức đầu tháng 4, lợi nhuận đạt đỉnh ở 31 nghìn đô la Mỹ và chi phí ở mức 16 nghìn đô. Tuy nhiên một tuần sau đó lợi nhuận đột ngột giảm mạnh chỉ còn 17 nghìn đô la Mỹ, nghĩa là chênh lệch tụt 14 nghìn đô chỉ cách nhau một tuần. Những tuần sau đó lợi nhuận và chi phí của doanh nghiệp giảm đều hơn, với mỗi tuần giảm 4 nghìn đô la. Ở tuần 18 do chỉ có hai ngày 29 và 30 nên lợi nhuận và chi phí thấp hơn hẳn những tuần trước đó.

Tuy chi phí giảm từ từ và không quá mức chênh lệch nhưng sự sụt giảm lợi nhuận từ đầu tháng là 31 nghìn đô và gần cuối tháng lợi nhuận chỉ 9 nghìn đô là một dấu hiệu cảnh báo đỏ dành cho cửa hàng. Doanh nghiệp cần xem xét lại các chiến lược tiếp thị, giới thiệu sản phẩm hoặc ưu đãi đối với khách hàng. Ngoài ra doanh nghiệp còn cần phân tích cụ thể các khía cạnh khác như quản lý tồn kho, năng suất và thái độ của nhân viên hay phản hồi của khách hàng để tìm ra nguyên nhân và mục tiêu cải thiện, đảm bảo sự tăng trưởng bền vững.

Vậy chúng ta sẽ đi sâu vào phân tích các khía cạnh cụ thể hơn.



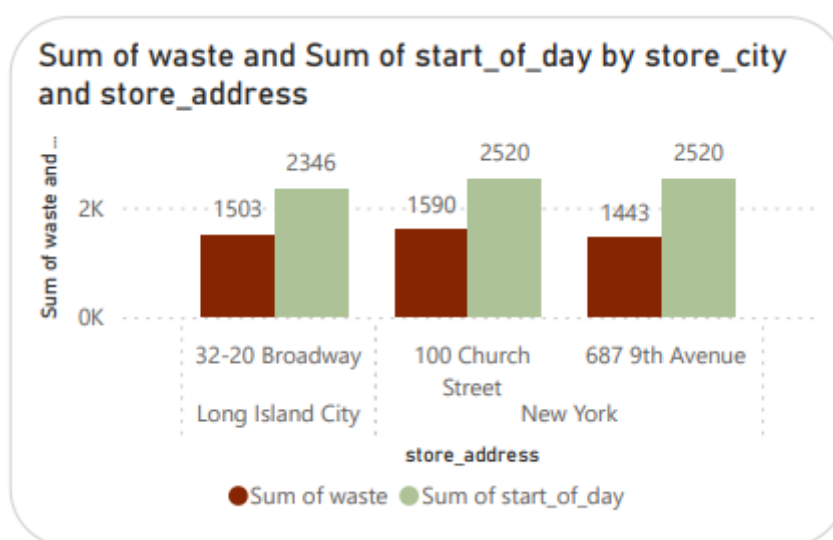
Hình 43. Chart sum of sales_amount by Date and store_city

Đây là biểu đồ biểu diễn tổng doanh thu bán ra theo ngày của các cửa hàng tại hai thành phố là Long Island và New York. Trước tiên nói về cửa hàng ở thành phố Long Island, nó theo xu hướng chung của biểu đồ tổng lợi nhuận và chi phí ở trên là giảm dần đều. Do Long Island chỉ có một cửa hàng nên doanh thu thấp gấp đôi hai cửa hàng ở New York là điều dễ hiểu. Doanh thu của cửa hàng ở đây cũng đạt đỉnh cao vào những ngày đầu, đạt từ 2.3-2.8 nghìn đô la Mỹ. Sau đó doanh thu giảm dần cho tới ngày cuối cùng thu thập dữ liệu là ngày 29 thì doanh thu của cửa hàng này chỉ còn 0.7 nghìn đô la Mỹ.

Còn về doanh thu ở New York ban đầu ở mức 4.7-4.9 nghìn đô la Mỹ, sau đó lại đột ngột giảm mạnh ở ngày 7 của tháng với doanh thu chỉ còn gần 1.3 nghìn

đô la Mỹ. Sau đó lại tăng mạnh vào những ngày kế tiếp lên đến con số 3.8 nghìn đô, rồi lại giảm bất ngờ ở ngày 14, lần này doanh thu thậm chí còn thấp hơn cả ngày thứ 7. Những ngày sau đó tổng doanh thu dần bình ổn và tăng nhẹ lại đến hết tháng với 2.6 nghìn đô la Mỹ. Rõ ràng ta có thể thấy doanh thu ở New York biến động cực mạnh và khó đoán hơn cả.

Cùng với đó, chúng ta chuyển sang phân tích tổng số lượng bánh ngọt được nhập vào hàng ngày và số lượng sản phẩm bị thừa của từng cửa hàng để kiểm soát chi phí và tối ưu hoá lợi nhuận.



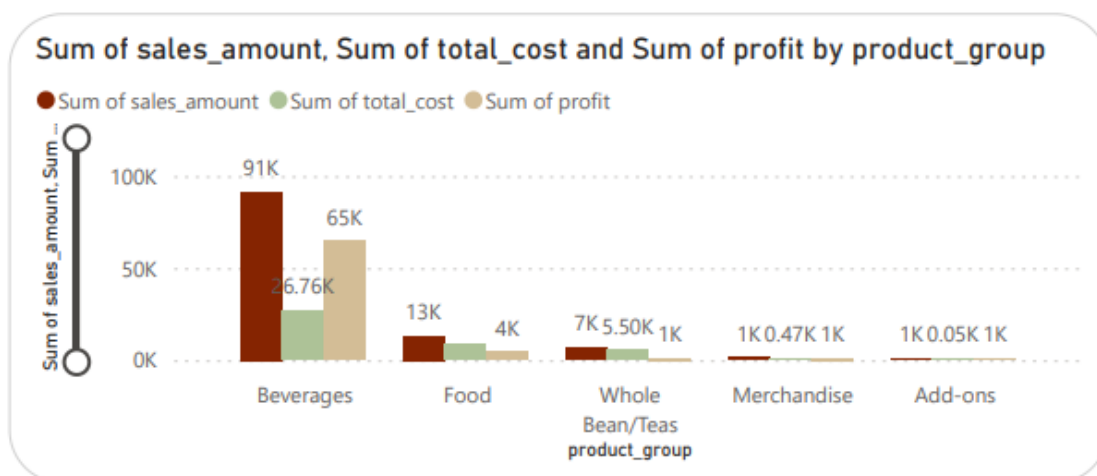
Hình 44. Chart Sum of waste and Sum of start_of_day by store_city and store_address

Chúng ta có thể thấy rằng mỗi cửa hàng đều có một lượng hàng tồn kho không bán được đáng kể so với lượng hàng nhập vào hàng ngày. Cửa hàng ở thành phố Long Island nhập ít bánh ngọt hơn cả nhưng lại có tỷ lệ thừa nhiều nhất với 64% (1503/2346). Hai cửa hàng ở thành phố New York có số lượng nhập hàng bằng nhau nhưng cửa hàng ở địa chỉ 687 9th Avenue có tỷ lệ thừa thấp hơn, và thấp nhất so với cả ba với 57% nhưng sự chênh lệch này là không đáng kể.

Tóm lại, qua phân tích trên, chúng ta có thể thấy rằng doanh thu của cửa hàng ở Long Island có xu hướng giảm dần đều trong khi doanh thu ở New York lại biến động mạnh. Điều này cho thấy sự khác biệt rõ rệt về hiệu suất kinh doanh

của các cửa hàng giữa hai thành phố. Đây là vấn đề đáng quan tâm và cần được nghiên cứu kỹ lưỡng hơn để tìm ra nguyên nhân và giải pháp phù hợp. Bên cạnh vấn đề doanh thu, tỷ lệ thừa bánh ngọt cũng là một vấn đề đáng báo động chung cho cả ba cửa hàng. Tỷ lệ thừa lên đến 60%, cao hơn mức trung bình và gây lãng phí lớn về nguồn lực, ảnh hưởng đến chi phí bỏ ra và lợi nhuận thu vào của chuỗi cửa hàng. Vấn đề này cho thấy yếu kém trong quản lý hàng hoá và chiến lược bán hàng.

Tiếp đến, chúng ta đi sâu vào việc phân tích tổng doanh thu và tổng chi phí theo từng nhóm sản phẩm. Việc phân tích vấn đề này giúp ta có cái nhìn tổng quan về hiệu suất kinh doanh của từng nhóm sản phẩm, từ đó đưa ra các chiến lược kinh doanh phù hợp.

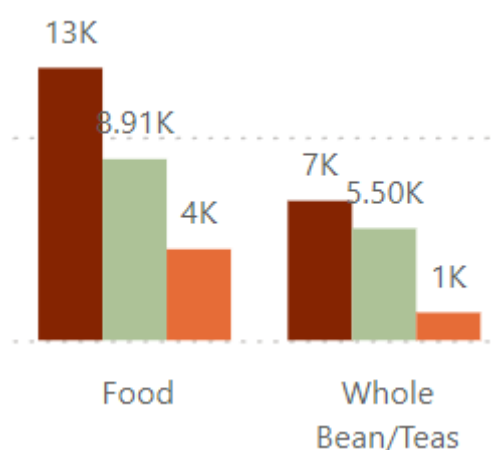


Hình 45. Chart sum of sales_amount, Sum of total_cost and Sum of profit by product_group

Chuỗi cửa hàng bán lẻ cà phê này có 5 nhóm sản phẩm được thể hiện rõ trong biểu đồ là: Beverages (Đồ Uống), Food (Thức Ăn), Whole Bean/Teas (Hạt Cà Phê/Trà), Merchandise (Hàng trưng bày, quà lưu niệm như quần áo, huy hiệu hay những ly uống được thiết kế đẹp...), Add-ons (Những thứ được thêm vào để dùng chung với nước uống hay thức ăn, ví dụ như đường, sữa, trân châu, ...). Nổi bật nhất trong số này là Beverages, đây là nhóm có doanh thu, chi phí và lợi nhuận

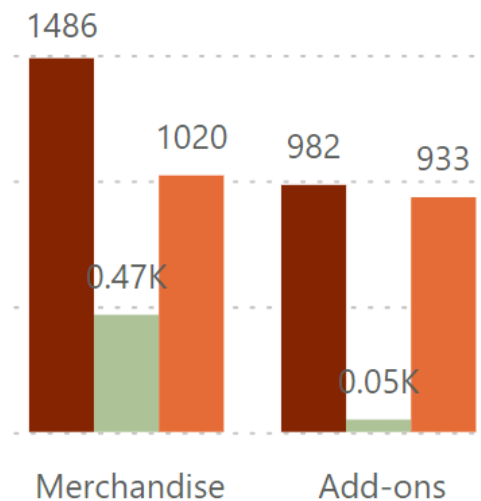
cao nhất, cho thấy đây là mặt hàng chủ lực mang lại lợi nhuận chính cho cửa hàng. Lợi nhuận mang lại của mặt hàng này cực kỳ cao so với chi phí bỏ ra, cụ thể chi phí chỉ có 26.76 nghìn đô và lợi nhuận lên đến 65 nghìn đô la Mỹ (lớn hơn gấp gần 2.5 lần so với chi phí). doanh thu của mặt hàng này cũng cực kỳ áp đảo, gấp 9 lần với Food (nhóm hàng đứng thứ hai sau Beverages).

Với những nhóm sản phẩm còn lại, vì số liệu có chênh lệch lớn nên nhóm dùng Zoom Slider để quan sát chúng rõ hơn.



Hình 46. Chart sum of sales_amount, Sum of total_cost and Sum of profit by product_group (zoom slider)

Ta có thể thấy hai nhóm hàng trên có chi phí bỏ ra còn cao hơn cả lợi nhuận. Về thực phẩm, như đã phân tích ở trên, do vấn đề quản lý hàng tồn kho yếu kém mà chuỗi cửa hàng này đã lãng phí hơn 60% phần bánh ngọt. Do đó chi phí của Food tăng cao dẫn đến lợi nhuận bị kéo xuống theo. Còn về Whole Bean/Teas lợi nhuận chỉ có 1 nghìn đô bởi vì doanh thu của họ chỉ có 7 nghìn đô trong khi đó chi phí lên đến 5.5 nghìn đô la Mỹ. Cho nên về nhóm hàng này, nguyên nhân lợi nhuận ít là do doanh thu bán ra ít và chi phí bỏ ra cao.

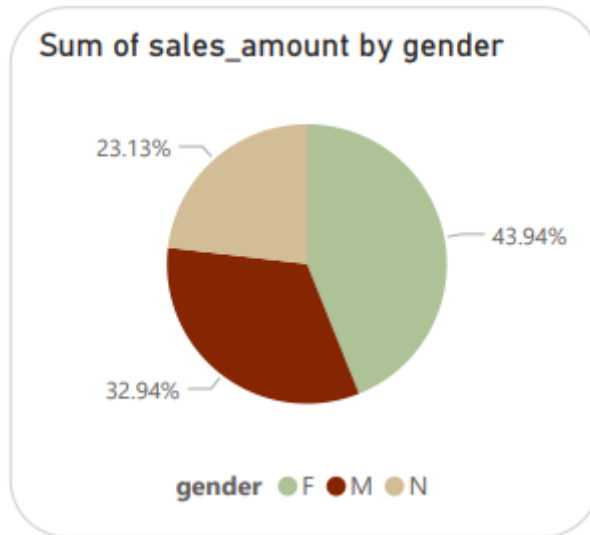


Hình 47 Chart sum of sales_amount, Sum of total_cost and Sum of profit by product_group (zoom slider)

Về hai nhóm hàng có doanh thu thấp nhất ta có vài nhận xét như sau. Merchandise mặc dù có doanh thu ít hơn so với Whole Bean/Teas gấp 7 lần nhưng lợi nhuận mang lại vẫn ngang ngửa (hơn 1 nghìn đô la Mỹ), bởi vì chi phí của mặt hàng này khá thấp, chỉ với 0.47 nghìn đô (tương đương con số lẻ của chi phí mà Whole Bean/Teas bỏ ra). Bên cạnh đó, Add-ons có tổng chi phí thấp nhất (0.05 nghìn đô la Mỹ) nhưng mang về lợi nhuận cũng gần tương đương với Merchandise nhưng doanh thu lại ít hơn gần 500 đô la.

Tóm lại, chuỗi cửa hàng cà phê này có nhóm hàng chủ lực là Beverages mang lại tổng lợi nhuận cao nhất, trong khi Food và Whole Bean/Teas lại có chi phí cao hơn lợi nhuận. Nguyên nhân chính là do vấn đề quản lý hàng tồn kho ở nhóm Food và doanh thu bán ra ít nhưng chi phí bỏ ra cao ở nhóm Whole Bean/Teas/ Trong khi đó, Merchandise và Add-ons dù có doanh thu thấp nhưng vẫn mang lại lợi nhuận tương đương Whole Bean/Teas nhờ vào chi phí thấp.

Cuối cùng ở dashboard tổng quan tình hình kinh doanh này, ta sẽ cùng phân tích đến khía cạnh khách hàng.



Hình 48. Chart Sum of sales_amount by gender

Tổng quan về phân bố khách hàng ta có thể thấy, giới tính nữ chiếm tỷ lệ cao nhất với 43.94%, cho thấy họ là nhóm khách hàng chính. Trong khi đó, nhóm khách hàng giới tính nam chiếm 32.94%, chiếm tỷ lệ nhỏ hơn so với nhóm nữ nhưng vẫn là một nhóm khách hàng tiềm năng. 23.13% còn lại là nhóm khách hàng không xác nhận rõ giới tính, cần nghiên cứu thêm để hiểu rõ và tìm cách tiếp cận nhóm này.

Vậy nên cửa hàng có thể phát triển sản phẩm hoặc dịch vụ đặc biệt nhằm vào mục tiêu là nhóm nữ, đồng thời tạo các chiến dịch quảng cáo để thu hút khách hàng nam. Còn về việc phân tích sâu hơn về nhóm khách hàng không xác định có thể giúp cửa hàng nhận diện được những cơ hội mới và tối ưu hóa chiến lược kinh doanh. Điều này có thể bao gồm việc tạo ra các sản phẩm hoặc dịch vụ không phân biệt giới tính, hoặc thậm chí là các chương trình khuyến mãi đặc biệt.

Kết luận: Những dữ liệu và phân tích trên đã tổng quan và chi tiết về hiệu suất kinh doanh của chuỗi cà phê này. Từ việc xem xét các chỉ số tài chính tổng quan cho đến phân tích sâu về các nhóm sản phẩm, khách hàng để rút ra những điểm mạnh, điểm yếu của doanh nghiệp.

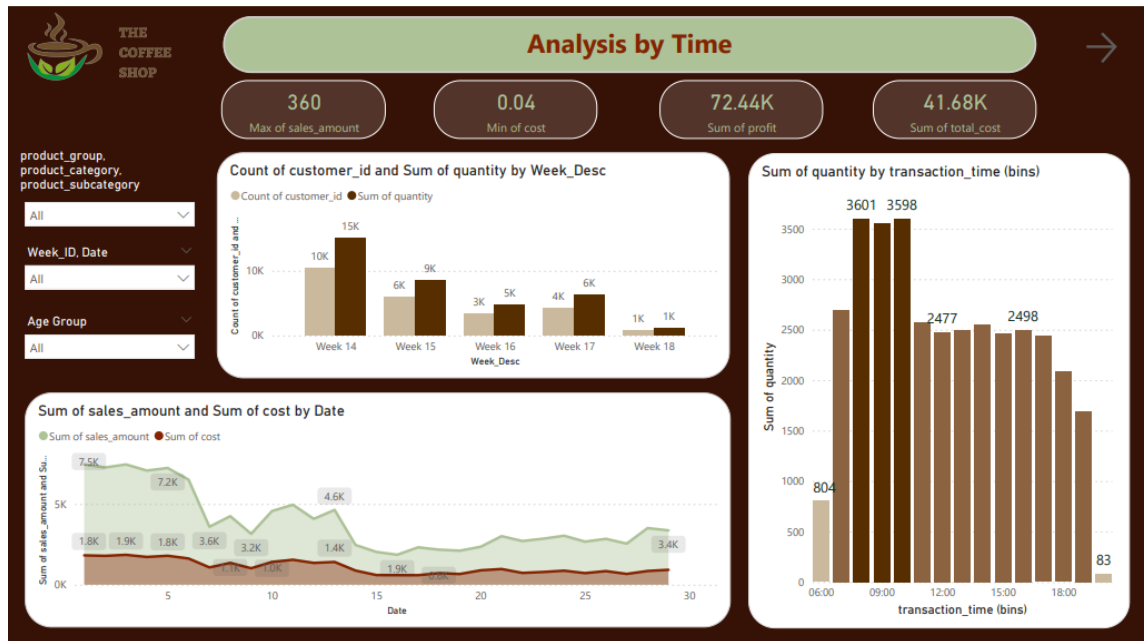
Đầu tiên, có thể nhận thấy rõ trong tháng tư này, doanh nghiệp đã đạt được mục tiêu doanh thu và lợi nhuận, thậm chí vượt qua mong đợi. Tuy nhiên, sự giảm đột ngột trong lợi nhuận từ đầu tháng là một dấu hiệu bất thường.

Việc phân tích về các nhóm sản phẩm cũng đã cho thấy rằng Beverages là mặt hàng chủ lực mang lại lợi nhuận cao nhất, trong khi Food và Whole Bean/Teas có chi phí cao hơn lợi nhuận do vấn đề quản lý hàng tồn kho và doanh thu thấp. Merchandise và Add-ons, mặc dù có doanh thu thấp, nhưng lại mang lại lợi nhuận tương đối cao nhờ vào chi phí thấp.

Cuối cùng, phân tích về phân bố khách hàng đã cho thấy rằng cửa hàng có thể tập trung vào phát triển sản phẩm và dịch vụ để thu hút khách hàng nữ và tăng cường chiến dịch quảng cáo để thu hút khách hàng nam. Đồng thời, việc nghiên cứu sâu hơn về nhóm khách hàng không xác định cũng có thể mở ra những cơ hội mới và tối ưu hóa chiến lược kinh doanh.

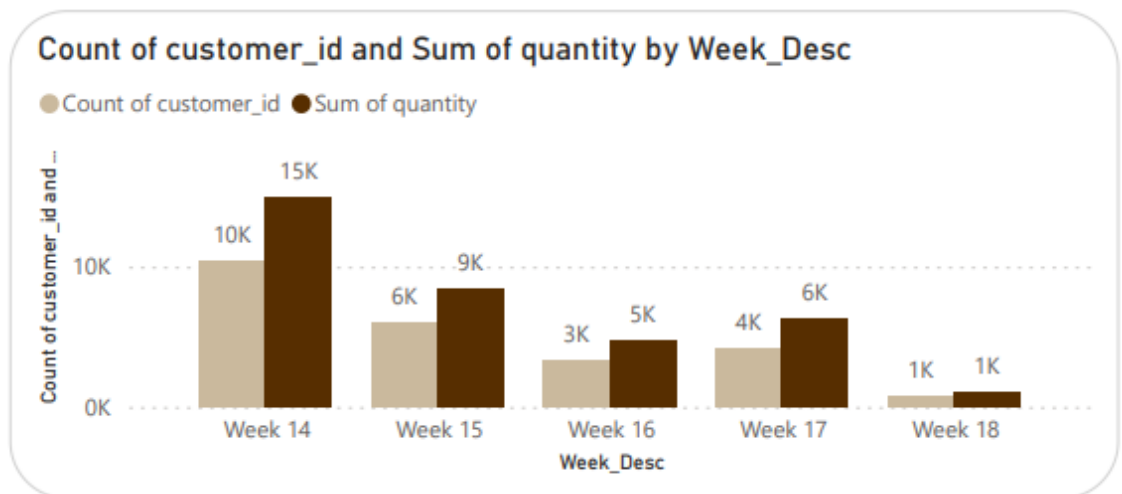
Tóm lại, việc áp dụng các chiến lược như tăng cường tiếp thị, quản lý hàng tồn kho hiệu quả và phát triển sản phẩm đa dạng có thể giúp chuỗi cửa hàng cà phê này duy trì và phát triển sự thành công trong tương lai.

4.3. Time



Hình 49. Dashboard Analysis by Time

Trong tháng 4, tổng doanh số ghi nhận đạt hơn 114 nghìn đô, từ tổng số 36 nghìn đơn vị sản phẩm bán được, thu về tổng lợi nhuận hơn 72 nghìn đô với chi phí bỏ ra là xấp xỉ 42,000\$.

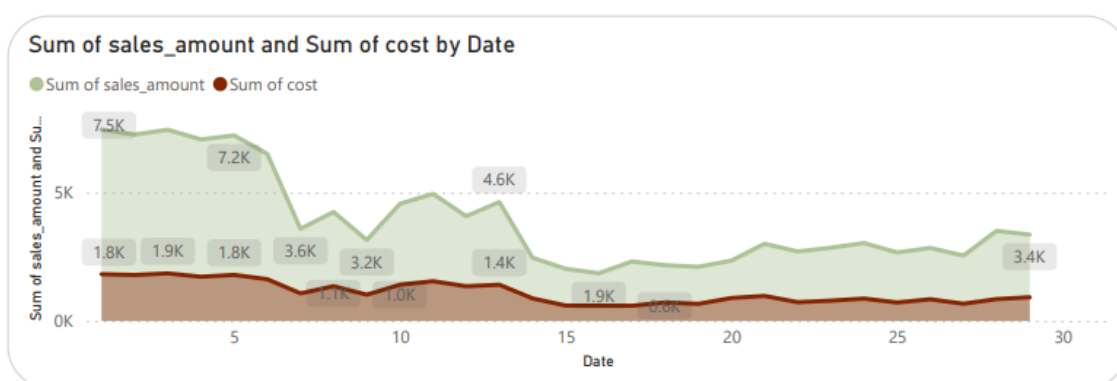


Hình 50. Chart Count of customer_id and Sum of quantity by Week_Desc

Trong biểu đồ đầu tiên, dễ dàng nhận ra chiều hướng giảm dần về số lượng sản phẩm cũng như lượng khách hàng mỗi tuần trong tháng.

Khoảng thời gian nửa đầu tháng, hiệu suất kinh doanh của các cửa hàng ở mức tốt nhất với hơn 20 nghìn sản phẩm bán được cho hơn 16 nghìn khách hàng. Trong đó, tuần thứ nhất đứng đầu đồng thời về lượng khách hàng và sản phẩm.

Nửa cuối tháng tình hình kinh doanh có phần sụt giảm khi tổng sản phẩm bán được chỉ hơn 11 nghìn với lượng khách hàng thống kê được là khoảng 8000 người. Tuy nhiên, sau khi rơi vào hiệu suất hoạt động thấp nhất ở tuần thứ 3 của tháng (tuần 16 của năm), kinh doanh của các cửa hàng có phần cải thiện, tăng khoảng 1000 đơn vị sản phẩm và lượng khách hàng. 2 ngày cuối tháng rơi vào tuần thứ 18 của năm, do đó số lượng ghi nhận được ít nhất vì chỉ thống kê trong 2 ngày, cần lưu ý đây không phải một dấu hiệu bất thường trong tình hình kinh doanh của thương hiệu.



Hình 51. Chart Sum of sales_amount and Sum of cost by Date

Biểu đồ vùng bên dưới thể hiện tổng số tiền bán hàng và tổng chi phí trong vòng 30 ngày, nó thể hiện rõ biến động kinh doanh của tháng 4.

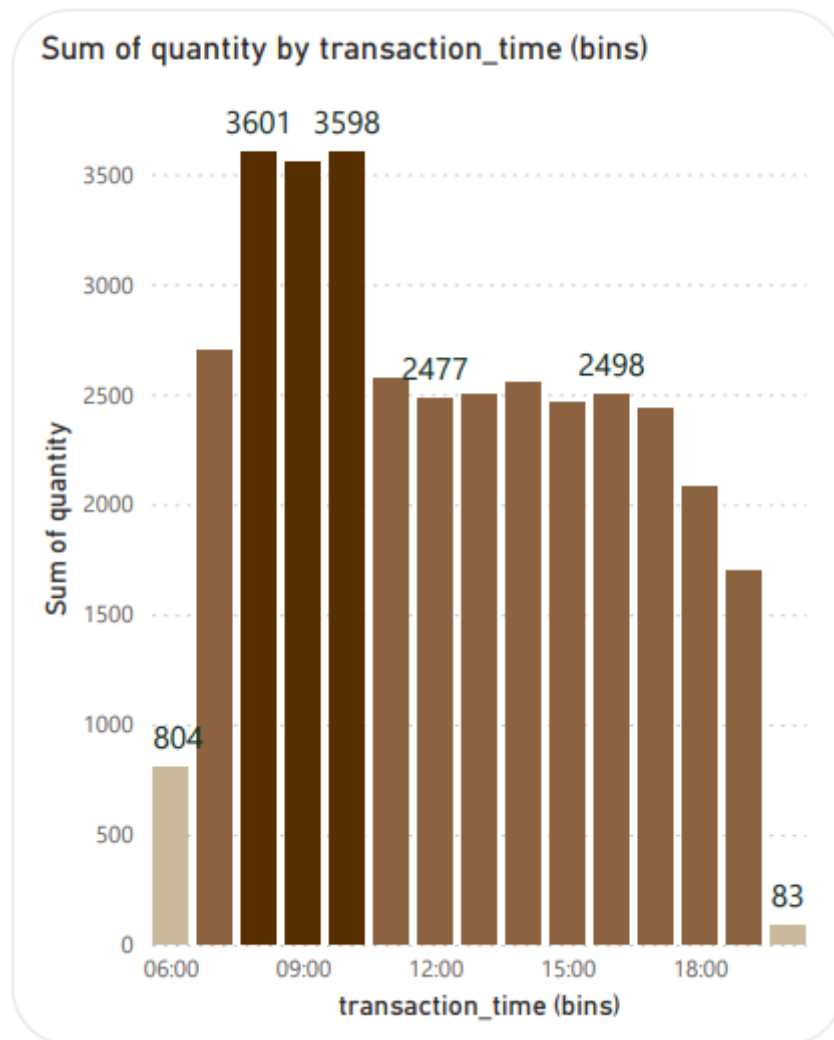
6 ngày đầu tháng dao động trong khoảng 6,5 - 7,5 nghìn đô tổng doanh số, đây là thời điểm doanh số, chi phí và lợi nhuận cao nhất của kỳ kinh doanh này. Hiệu số của doanh thu và chi phí hay chính là lợi nhuận đạt hơn 5000\$. Không có sự chênh lệch rõ rệt trong tổng doanh số, chi phí bán hàng hay lợi nhuận trong

những ngày này. Điều này cho thấy sự ổn định và hiệu quả trong kinh doanh của thương hiệu ở giai đoạn đầu tháng.

Trong 7 ngày tiếp theo, doanh thu không ngừng biến động, thống kê cho thấy sự tăng giảm liên tục. Tuy nhiên, sự giảm sút không mạnh mẽ, tổng doanh thu cao nhất vẫn đạt được gần 5 nghìn đô với chi phí xấp xỉ 1,500, thu về gần 3,5 nghìn đô lợi nhuận.

Bắt đầu từ ngày 14, tình hình kinh doanh đi xuống rõ rệt. Từ giữa tháng cho đến hết tháng 4, tổng doanh số không còn được duy trì ở mức tốt như đầu tháng khi mà doanh số thấp nhất ghi nhận vào ngày 16 là gần 1,900 và cao nhất chỉ đạt hơn 3,500\$ vào ngày 28. Ngày 14, 15 thu được lợi nhuận thấp nhất khi mà chênh lệch giữa doanh thu và chi phí bị thu nhỏ lại.

Nhìn chung, vẫn có sự cách biệt giữa doanh số và chi phí cho thấy hoạt động kinh doanh hiệu quả, nhưng cần theo dõi sát sao các biến động để đảm bảo ổn định.



Hình 52. Chart Sum of quantity by transaction_time

Biểu đồ cột này cho ta hiểu biết về giờ cao điểm trong ngày của doanh nghiệp. Đồng thời cũng có thể làm rõ thời gian mở cửa của các cửa hàng là từ 06:00 - 21:00.

Dễ dàng nhìn ra khoảng thời gian bận rộn nhất trong ngày là khung giờ từ 08:00 - 10:00. Tổng số lượng sản phẩm bán được của mỗi giờ trong cả tháng là hơn 3500 sản phẩm.

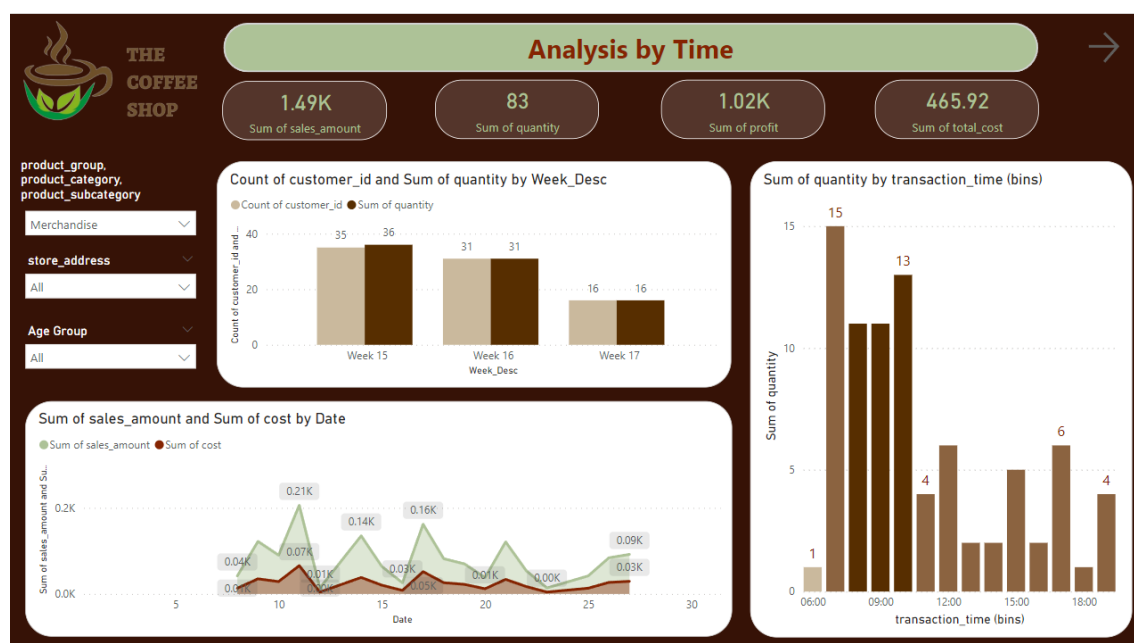
Trong khi đó, 2 thời điểm ghi nhận số lượng bán được ít nhất là 06:00 và 21:00. Điều này tương đối dễ hiểu, vì giờ mở cửa và giờ đóng cửa thường ghi nhận ít giao dịch hơn các khung giờ khác trong ngày. Bên cạnh đó, 6 giờ sáng hoặc 8

giờ tối không phải là thời gian tiêu thụ cafe phổ biến. Vậy ít sản phẩm được bán ra trong 2 khung giờ này trong bối cảnh của thương hiệu cafe cũng là điều tất yếu.

Khi biết được thời gian bận rộn và nhàn rỗi trong ngày, bộ phận quản lý sẽ có được chiến lược phân bổ nguồn lực phù hợp. Đồng thời cũng hỗ trợ nhân viên chuẩn bị tốt hơn cho các thời điểm khác nhau trong ngày, đảm bảo cửa hàng có thể hoạt động trơn tru và phục vụ khách hàng với hiệu suất tốt nhất.

Góc nhìn chi tiết:

Khi cắt lát từng nhóm sản phẩm và cửa hàng cụ thể (slicer: product_group, store_addres), có thể phát hiện ra có sự khác biệt của nhóm sản phẩm Merchandise và Whole Bean/Teas và cửa hàng tại 32-20 Broadway so với tổng thể trong những phân tích theo thời gian nêu trên.

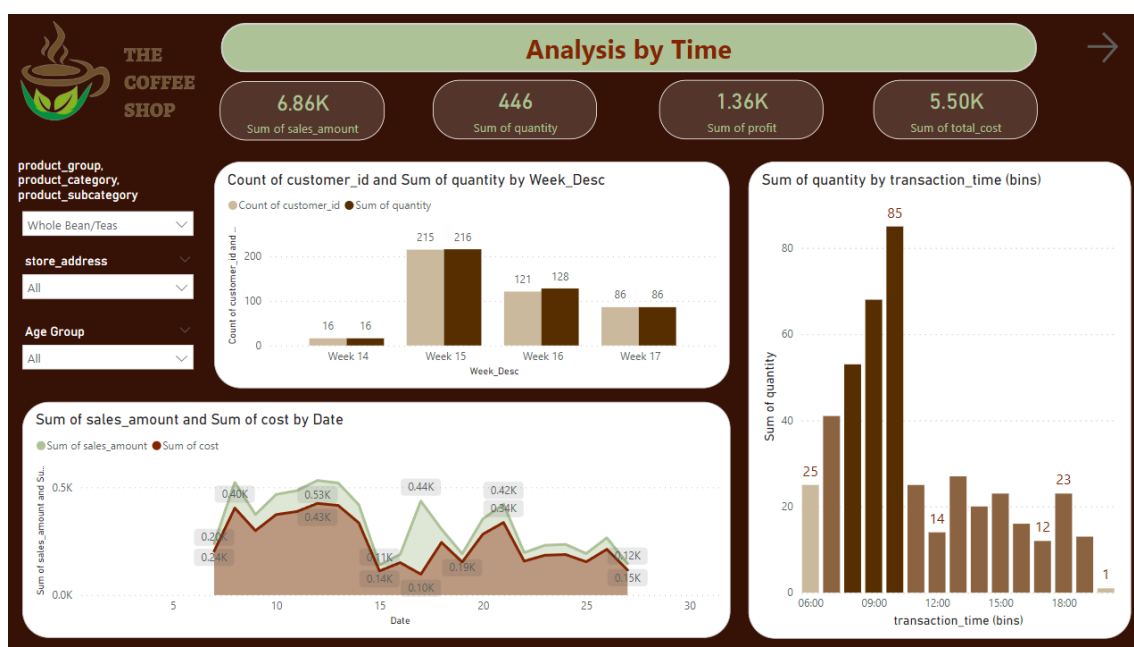


Hình 53. Slicer: group Merchandise

Có thể thấy, nhóm hàng Merchandise được bán nhiều nhất vào tuần thứ 2, thứ 3 của tháng. Đồng thời, 07:00 là thời điểm bán chạy các merchandise nhất. Việc khách hàng tập trung mua sản phẩm sưu tầm của thương hiệu vào một thời

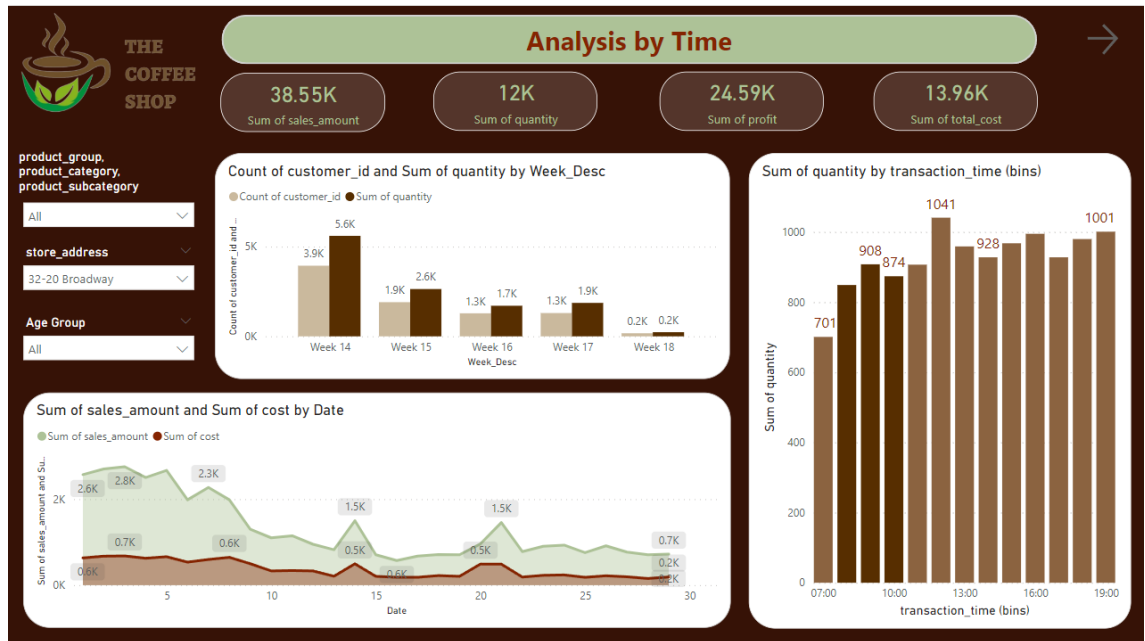
điểm cho thấy thương hiệu có tiềm năng thu hút và tạo cho mình tệp khách hàng trung thành trên thị trường bán lẻ cafe.

Sản xuất merchandise cũng là một chiến dịch kinh doanh thông minh mà nhiều chuỗi cafe nổi tiếng đang áp dụng, nổi bật có thể kể đến Starbucks. Đây là nhóm sản phẩm có biên lợi nhuận cao, đồng thời thúc đẩy việc mở rộng mức độ nhận diện thương hiệu tốt và giúp củng cố lòng trung thành của khách hàng với thương hiệu.



Hình 54. Slicer group: Whole Beans/Teas

Tương tự Merchandise, nhóm hàng Whole Bean/Teas cũng bán chạy nhất vào tuần 15 (tuần thứ 2 của tháng). Tuy nhiên, có thể thấy có những ngày không bán được bất kỳ đơn vị sản phẩm nào, điều này là dễ hiểu vì đây không phải nhóm sản phẩm chính của một thương hiệu bán lẻ cafe. Cùng với đó, Whole Bean/Teas cũng không mang về lợi nhuận cao khi mà doanh thu và chi phí không chênh lệch đáng kể.



Hình 55. Slicer: store_address 32-20 Broadway

Cửa hàng tại 32-20 Broadway có giờ cao điểm khác biệt so với mặt bằng chung. Giờ bán được tổng sản phẩm nhiều nhất của nơi này là 12:00 với 1041 sản phẩm được bán trong tháng. Bên cạnh đó, những khung giờ khác cũng không có quá nhiều cách biệt về số lượng bán ra, cho thấy cửa hàng thường xuyên ở trạng thái bận rộn. Nhân sự của địa chỉ này nên biết được đặc trưng của cửa hàng để xây dựng kế hoạch bán hàng phù hợp.

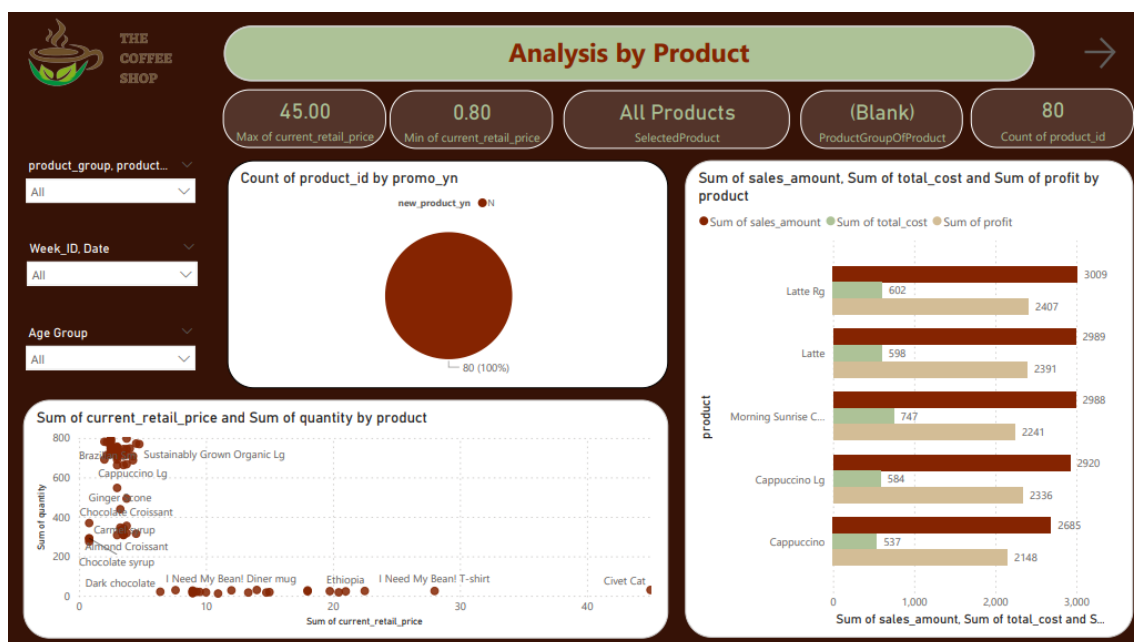
Kết luận: Dữ liệu của dashboard này đang kể câu chuyện về những chuyển động trong “điệu nhảy” của kỳ kinh doanh tháng 4. Ở giai đoạn đầu, tình hình buôn bán “nhảy múa” trong tinh thần tương đối phấn khởi. Tuy nhiên, nửa cuối tháng doanh thu các cửa hàng vận động có phần chậm rãi, vì không duy trì ở hiệu suất tốt như ban đầu. Để duy trì ổn định trong kinh doanh, cần tiến hành nghiên cứu về hành vi tiêu dùng trong tháng của nhóm khách hàng mục tiêu, hay tìm hiểu về xu hướng biến động của thị trường để chủ động ứng phó với những yếu tố có khả năng ảnh hưởng đến doanh số trong tương lai.

Giờ vàng theo mặt bằng chung bắt đầu vào khoảng 8:00 và duy trì trong khoảng 2 tiếng tiếp theo. Tuy nhiên, đặc thù của từng địa điểm là khác nhau, một

số cửa hàng sẽ bận rộn nhất vào đầu giờ chiều và có phần đông đúc trong hầu hết thời gian mở cửa. Nếu những chi nhánh khác trong hệ thống cũng có mật độ đông đúc dần trải như 32-20 Broadway thì sẽ hứa hẹn tiềm năng cải thiện tình hình kinh doanh chung của cả thương hiệu. Tuy nhiên, trong viễn cảnh này, áp lực vận hành và quản lý nguồn lực cũng sẽ tăng theo.

Câu chuyện của dữ liệu về merchandise ở trên đang mở ra một tiềm năng triển vọng về việc tăng cường hình ảnh thương hiệu. Cũng tương tự như các phương thức tiếp thị khác như marketing truyền miệng, marketing qua email, hay marketing qua mạng xã hội, xu hướng quảng bá bằng Branded Merchandise với các vật phẩm gắn liền với thương hiệu cho phép gia tăng độ nhận diện hiệu quả, mở rộng phạm vi tiếp cận khách hàng đến nhiều đối tượng mới, và kết nối với những khách hàng trung thành của mình.

4.4. Product



Hình 56. Dashboard Analysis by Product

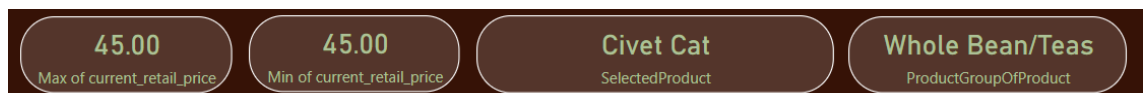
Trang Product cho ta cái nhìn chi tiết về hiệu suất bán hàng của các sản phẩm. Từ đó giúp chúng ta đưa ra các quyết định kinh doanh chính xác và kịp thời.

Như đã phân tích ở Overview, chuỗi cửa hàng cà phê này có trung bình giá ở phân khúc trung bình, chỉ 6.58 đô cho tất cả các mặt hàng. Ở đây ta sẽ xem xét kĩ hơn về mức giá của từng mặt hàng có trong chuỗi cửa hàng này.



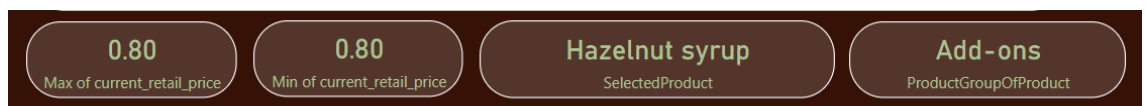
Hình 57. Card trong dashboard Product

Sản phẩm của doanh nghiệp này có mức giá dao động từ 0.8 đô cho đến 45 đô la Mỹ. Cho thấy chuỗi cửa hàng cung cấp nhiều loại sản phẩm (tận 80 loại) với mức giá đa dạng khác nhau từ giá rẻ đến cao cấp, phục vụ nhu cầu của nhiều đối tượng khách hàng.



Hình 58. Card hiển thị product Civet Cat

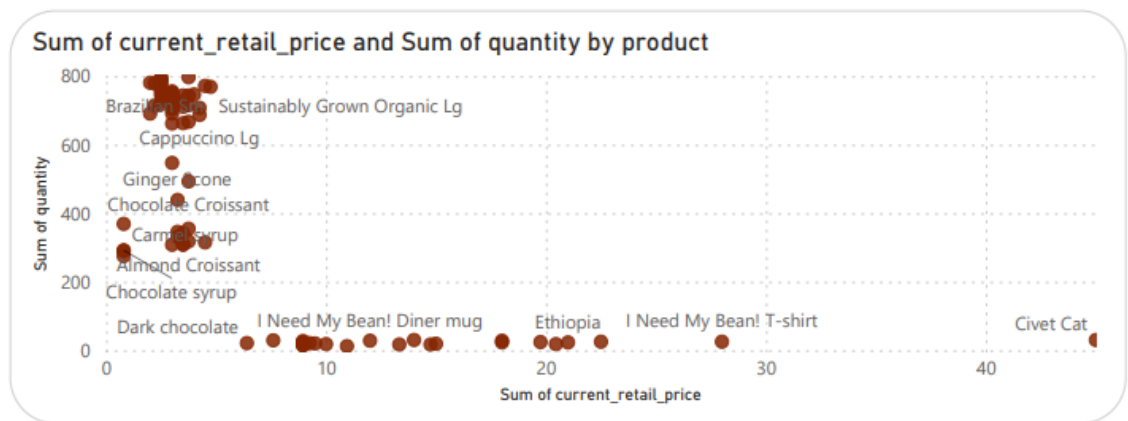
Trong đó sản phẩm mắc nhất với giá 45 đô la Mỹ là Civet Cat, một loại hạt cà phê.



Hình 59. Card hiển thị product hazelnut syrup

Còn sản phẩm rẻ nhất của hàng là Hazelnut syrubb với giá 0.8 đô, một loại siro thêm vào đồ uống để tăng hương vị.

Để xem sự phân bố giá cả và số lượng bán ra chi tiết thế nào, ta quan sát tiếp đến biểu đồ bên dưới.



Hình 60. Chart sum of current_retail_price and Sum of quantity by product

Số lượng mua của những sản phẩm giá dưới 5 đô la cực cao, chia thành hai nhóm. Một nhóm có số lượng tiêu thụ trên 600 với các đại diện là Latte, Sustainably Grown Organic Lg và Cappuccino Lg, cả ba loại này đều thuộc Beverages. Điều này cho thấy sự ưa chuộng của khách hàng với các đồ uống giá rẻ. Nhóm còn lại có số lượng tiêu thụ từ 200 đến 600 sản phẩm với các đại diện là Chocolate Croissant, Cranberry Scone và Almond Croissant, và thuộc nhóm Food. Mặc dù lượng tiêu thụ không cao bằng nhóm các sản phẩm nước uống nhưng nó vẫn cho thấy khách hàng vẫn quan tâm và tiêu dùng cho các loại thức ăn. Vậy phần lớn xu hướng tiêu dùng của khách hàng trong chuỗi cửa hàng này là mua những sản phẩm dưới 5 đô la.

Những sản phẩm bán lẻ có giá hơn 10 đô la Mỹ ta sẽ dùng Zoom slider để quan sát rõ hơn.

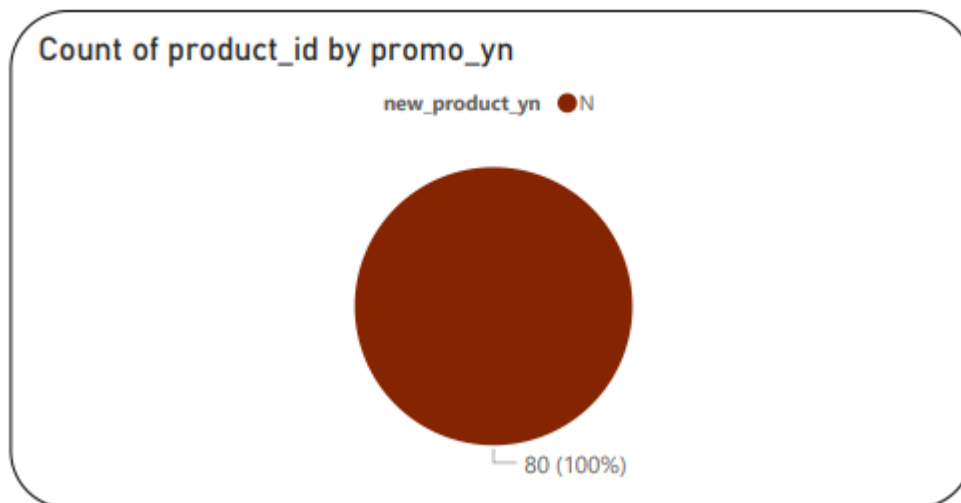


Hình 61. Sum of current_retail_price and Sum of quantity by product

Đa phần những sản phẩm này có số lượng tiêu thụ dưới 40 sản phẩm. Các sản phẩm có thể kể đến gồm “Brazilian - Organic,” “I Need My Bean! T-shirt,” “Civet Cat,” “Ethiopia Organic Decaf Blend,” “Earl Grey,” ... Những sản phẩm này thuộc nhóm Whole Bean/Teas và Merchandise. Những sản phẩm này với giá bán cao hơn và tiêu thụ ít có thể là những mặt hàng cao cấp hoặc chuyên biệt, không phải là lựa chọn hàng ngày của khách hàng. Điều này cũng có thể phản ánh sự nhạy cảm về giá của khách hàng đối với các sản phẩm có giá trên 10 đô la.

Tóm lại, với xu hướng mua hàng của khách hàng trong chuỗi cửa hàng cà phê này thì họ có nhu cầu mua những sản phẩm giá rẻ dưới 5 đô la, và là những đồ ăn thức uống có thể dùng hàng ngày. Thế nên doanh nghiệp nên tập trung vào việc quảng cáo khuyến mãi thị phần này để tối đa hoá doanh số, đồng thời cũng cân nhắc việc đa dạng hoá các sản phẩm để kích thích nhu cầu tiêu dùng. Còn với những sản phẩm giá cao và ít tiêu thụ, doanh nghiệp nên điều chỉnh chiến lược tiếp thị để tăng cường doanh số cho nhóm này hoặc tập trung vào việc phát triển các sản phẩm có giá thấp hơn để đáp ứng nhu cầu của đa số khách hàng.

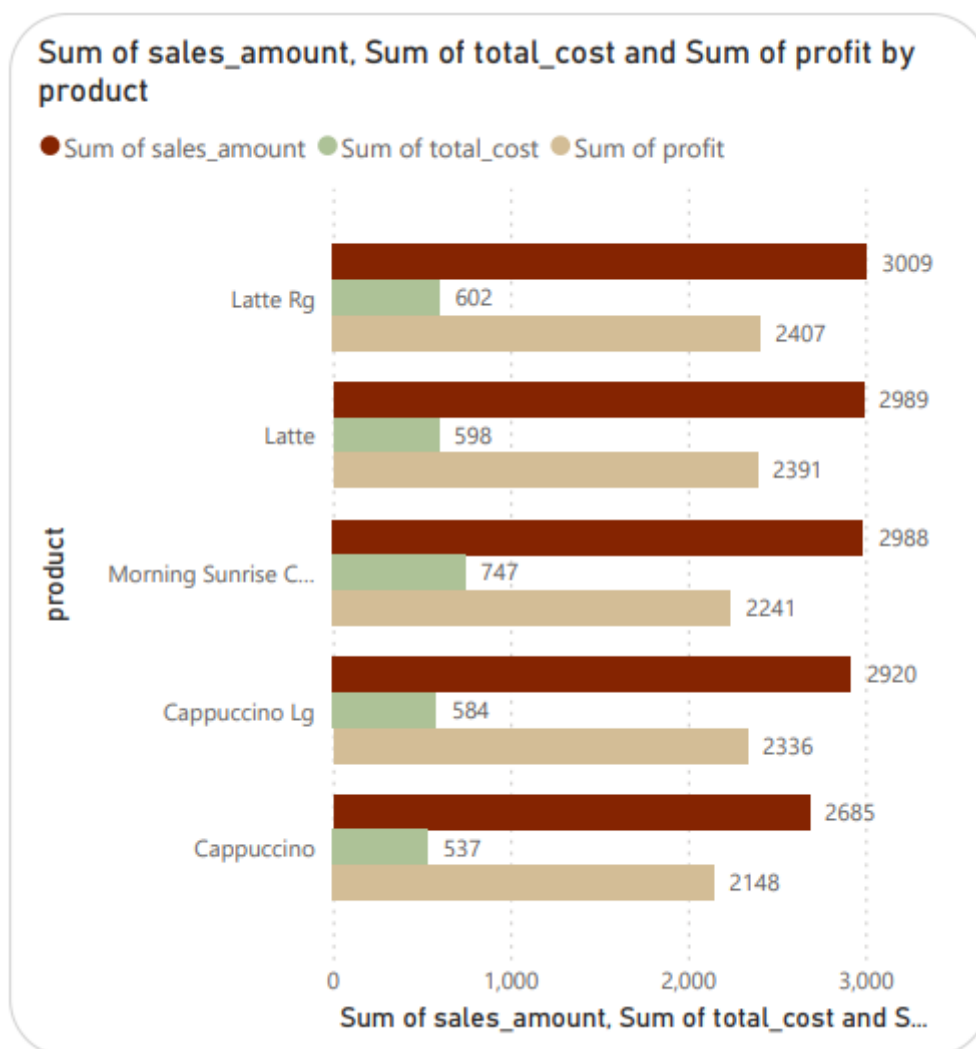
Như đã đề cập bên trên về chiến lược tiếp thị, dưới đây chúng ta sẽ bắt đầu tìm hiểu về thực trạng quảng bá - tiếp thị sản phẩm mới của doanh nghiệp này.



Hình 62. Chart Count of product_id by promo_yn

Từ biểu đồ trên, có thể thấy tất cả các giao dịch được ghi nhận hoàn toàn không xuất hiện những sản phẩm mới (specialty coffee và seasonal drink) của thương hiệu. Điều này phản ánh rằng khả năng giới thiệu dòng sản phẩm mới của các cửa hàng chưa tốt. Cần xem xét lại chiến lược xúc tiến của thương hiệu nói chung và những nỗ lực quảng bá đồ uống mới của những cửa hàng nói riêng để kịp thời cải thiện. Việc thất bại trong khâu tiến hành giới thiệu sản phẩm mới khiến cho nỗ lực làm mới thương hiệu cũng như đa dạng hóa các dòng sản phẩm trở nên vô ích, làm lãng phí tài nguyên và nguồn lực của doanh nghiệp.

Tiếp theo, để nghiên cứu về thị hiếu và đặc điểm của khách hàng, ta sẽ phân tích các chỉ số kinh tế ở những sản phẩm có lợi nhuận cao nhất của cửa hàng.



Hình 63. Sum of sales_amount, Sum of total_cost and Sum of profit by product

Ở đây ta có 5 loại sản phẩm tiêu biểu nhất của cửa hàng bao gồm Latte Rg, Latte, Morning Sunrise Chai Rg, Cappuccino Lg, Cappuccino. Có thể thấy, tổng quan của 5 loại sản phẩm top đầu này đều có doanh thu cao, lợi nhuận cao và chi phí thấp. Chi tiết hơn thì Latte Rg có tổng doanh thu là 3009 đô, tổng chi phí chỉ có 598 đô và lợi nhuận cao nhất trong số các sản phẩm với 2391 đô la Mỹ. Về sản phẩm có lợi nhuận thấp nhất ở đây là Cappuccino, doanh thu 2685 đô, chi phí 537 đô và lợi nhuận ở mức 2148 đô. Các sản phẩm trong biểu đồ này không quá chênh lệch nhau nhiều nhưng cho thấy các sản phẩm chủ lực của chuỗi cửa hàng này trong top đầu đa phần là cà phê và mang lại lợi nhuận đáng kể. Doanh nghiệp nên

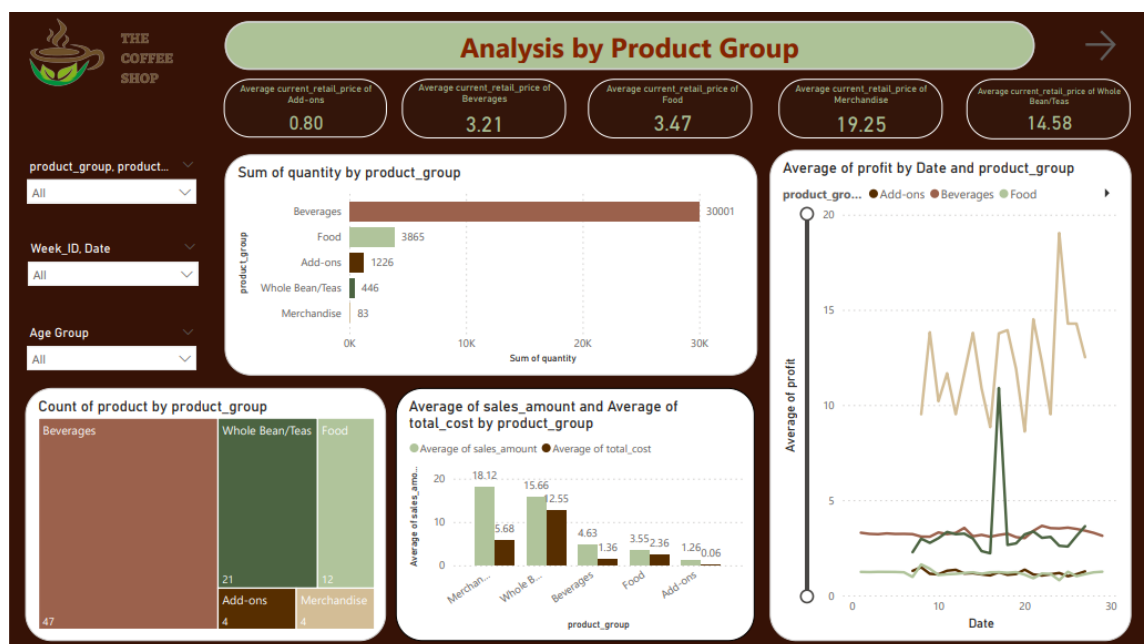
tiếp tục tập trung vào việc quảng bá và cải thiện chất lượng của những sản phẩm này để duy trì sự yêu thích của khách hàng và tăng cường lợi nhuận.

Kết luận: Có thể nhận thấy doanh nghiệp cung cấp một loạt sản phẩm phong phú với mức giá đa dạng là một chiến lược mang lại hiệu quả, với sự ưa chuộng đặc biệt đối với các mặt hàng giá rẻ dưới 5 đô la Mỹ, đồng thời cũng không quên những sản phẩm cao cấp với giá trị lợi nhuận đáng kể.

Tuy nhiên, việc giới thiệu sản phẩm mới vẫn còn là những điểm mà doanh nghiệp cần tập trung cải thiện. Đồng thời, việc tối ưu hóa chiến lược tiếp thị để tăng cường sự nhận biết và yêu thích của khách hàng đối với các sản phẩm chủ lực cũng là yếu tố không thể bỏ qua.

Cuối cùng, những sản phẩm top đầu không chỉ là chủ lực mà còn là nguồn lợi nhuận đáng kể cho doanh nghiệp. Việc duy trì và nâng cao chất lượng của những sản phẩm này sẽ tiếp tục là một yếu tố quan trọng trong việc đảm bảo sự thành công và tăng trưởng bền vững cho chuỗi cửa hàng trong tương lai.

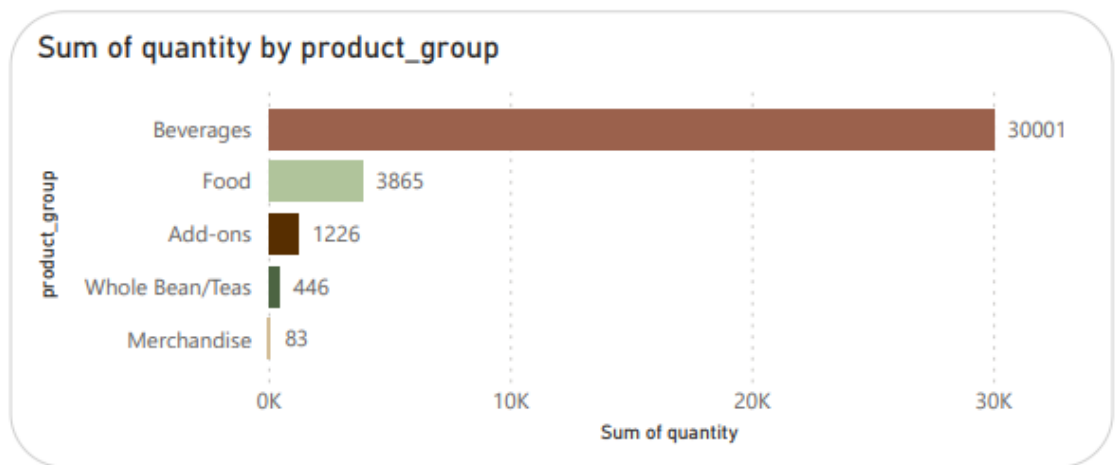
4.5. Product Group



Hình 64. Dashboard Analysis by Product Group

Trang Product Group cung cấp một cái nhìn tổng quan về các nhóm hàng lớn có trong chuỗi cửa hàng. Đây là số liệu mà chúng ta có thể nắm bắt được cái nhìn tổng quan về hiệu suất kinh doanh của từng nhóm sản phẩm, từ đó đưa ra các quyết định quản lý và chiến lược kinh doanh một cách hiệu quả.

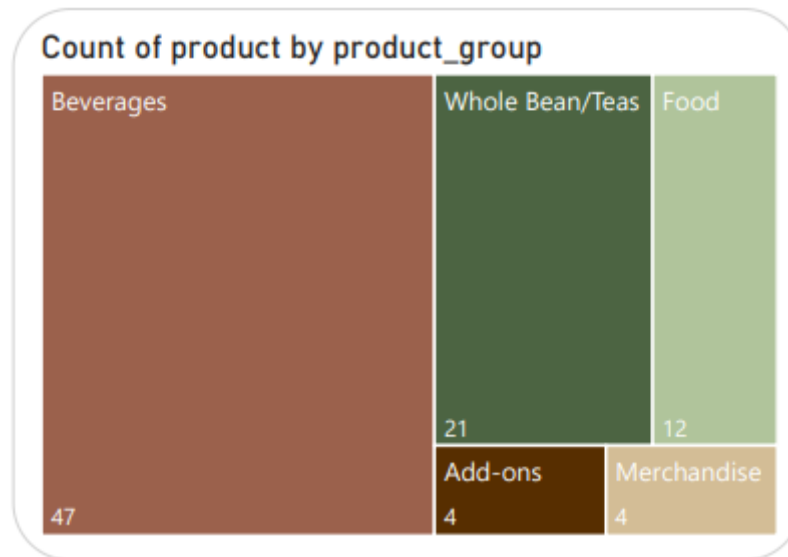
Vì đã phân tích ba chỉ số tổng doanh thu, chi phí và lợi nhuận ở page Overview trước đó, nên mở đầu phân tích về Product Group ta sẽ phân tích luôn chỉ số còn lại là tổng lượng tiêu thụ.



Hình 65. Chart Sum of quantity by product_group

Nhóm Beverages có lượng tiêu thụ hoàn toàn áp đảo những nhóm sản phẩm khác với 30001 nghìn sản phẩm, lần nữa khẳng định rằng đây là nhóm hàng chủ lực của doanh nghiệp thời điểm tháng tư năm 2019 này. Kế đó ta sẽ đi nhanh qua số liệu của các nhóm tiếp theo như Food (3865 sản phẩm), Add-ons (1226 sản phẩm), Whole Bean/Teas (446 sản phẩm) và cuối cùng là Merchandise với 83 sản phẩm.

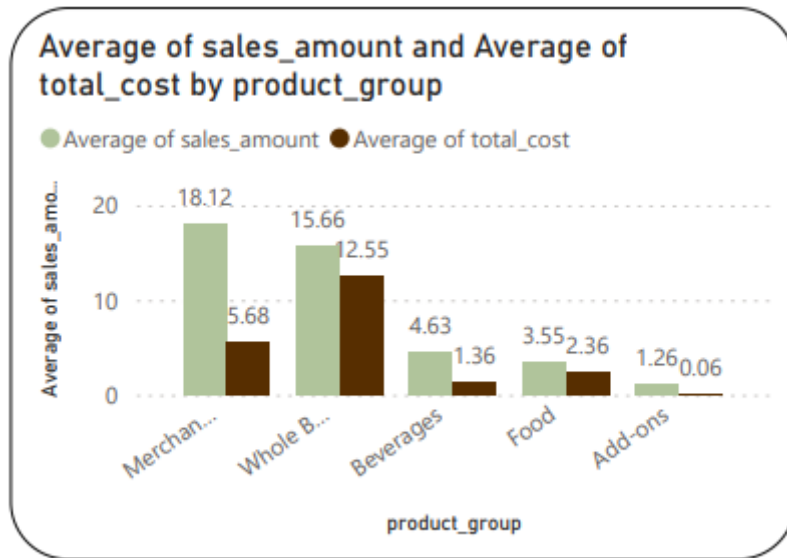
Mức tiêu thụ chênh lệch cực kỳ lớn là như thế, vậy cấu trúc của nhóm sản phẩm này như thế nào, chúng ta cùng phân tích tiếp tổng số sản phẩm có trong mỗi nhóm.



Hình 66. Chart Count of product by product_group

Mặt hàng chủ lực Beverages vẫn đứng đầu về độ đa dạng, có hơn 47 loại sản phẩm trong nhóm này. Whole Bean/Teas đứng thứ hai với 21 loại, thấp hơn gấp 2.5 lần so với Beverages, mặc dù nhóm hàng này chỉ đứng thứ tư về lượng tiêu thụ. Còn nhóm hàng đứng thứ hai về lượng tiêu thụ, Food, chỉ có 12 loại sản phẩm và xếp hạng ba về độ đa dạng sản phẩm trong nhóm. Hai nhóm Add-ons và Merchandise bằng nhau và chỉ có 4 loại sản phẩm.

Trong những phân tích trên về các nhóm sản phẩm, rõ ràng có thể thấy Merchandise và Whole Bean/Teas là những nhóm xếp cuối hoặc là về lợi nhuận, hoặc về doanh thu. Vậy tại sao chuỗi cửa hàng lại lựa chọn vẫn bán mặt hàng này?



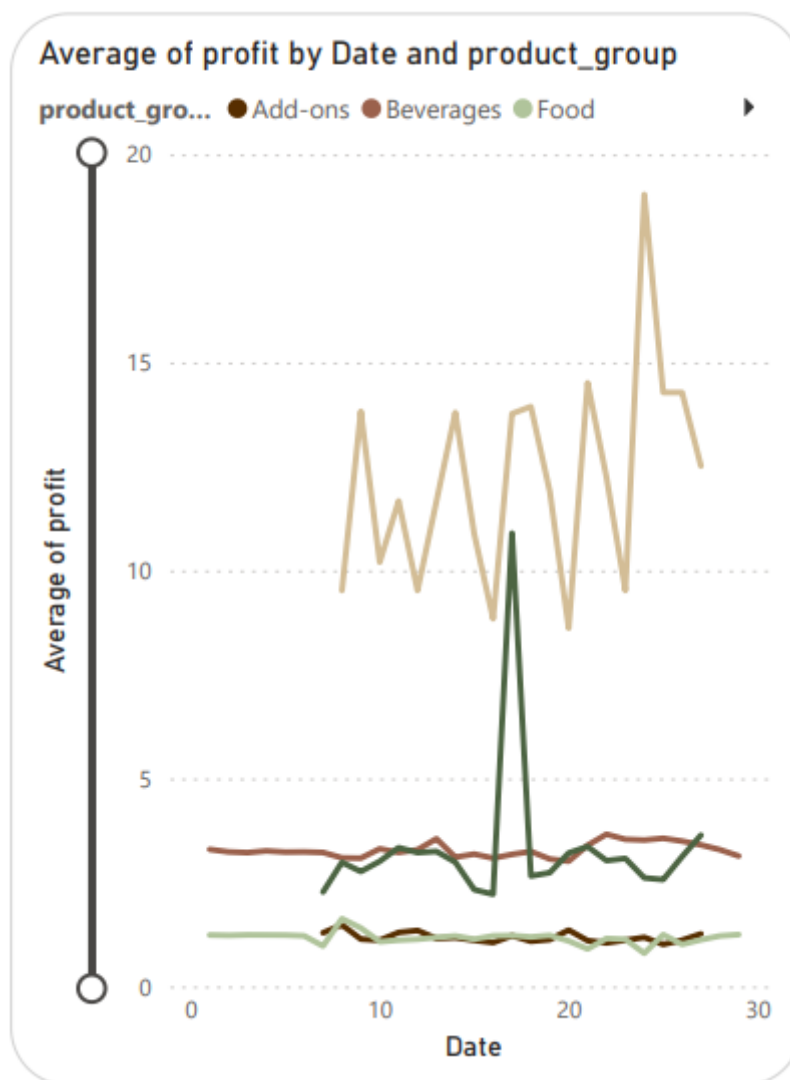
Hình 67. Chart Average of sales_amount and Average of total_cost by product_group

Nếu xét về mức doanh thu và chi phí trung bình, Merchandise - nhóm hàng luôn đứng top cuối, lại dẫn đầu ở khía cạnh này. Mức chi phí bỏ ra trung bình của Merchandise chỉ 5.68 đô nhưng doanh thu lại là 18.12 đô, gấp 3 lần so với chi phí. Đứng thứ hai sau đó là Whole Bean/Teas, chi phí trung bình của nhóm hàng này là cao nhất với 12.55 đô la, bù lại là doanh thu 15.66 đô, chỉ xếp sau Merchandise. Beverages luôn có số liệu áp đảo so với các nhóm hàng khác, nhưng xét về tính trung bình chỉ đứng thứ ba với doanh thu là 4.63 đô và chi phí trung bình chỉ 1.36, dù thế về lợi nhuận thì vẫn hơn Whole Bean/Teas. Còn nhóm Food, vì tình trạng tồn kho gây lãng phí nên dù có tính trung bình hay tổng số thì doanh thu và lợi nhuận của chúng vẫn không cách quá xa là 3.55 và 2.36 đô. Add-ons vẫn đứng cuối với doanh thu và chi phí trung bình rất rẻ là 1.26 và 0.06 đô la Mỹ.

Từ những số liệu trên, chúng ta có thể thấy rằng mỗi nhóm sản phẩm đều có những đặc điểm riêng và đóng góp khác nhau vào doanh thu và lợi nhuận của doanh nghiệp. Nhóm Merchandise dù có số lượng sản phẩm ít nhất nhưng lại mang lại lợi nhuận trung bình cao nhất. Trong khi đó, nhóm Beverages dù có doanh thu và lợi nhuận trung bình không cao nhưng lại chiếm tỷ trọng lớn trong cấu trúc

doanh thu tổng thể nhờ vào lượng tiêu thụ lớn. Nhóm Whole Bean/Teas, Food và Add-ons cần được xem xét kỹ lưỡng hơn để tìm ra cách tối ưu hóa lợi nhuận.

Cuối cùng ta sẽ phân tích trung bình lợi nhuận theo ngày, để xem sự khác biệt giữa tổng lợi nhuận theo ngày.



Hình 68. Chart Average of profit by Date and product_group

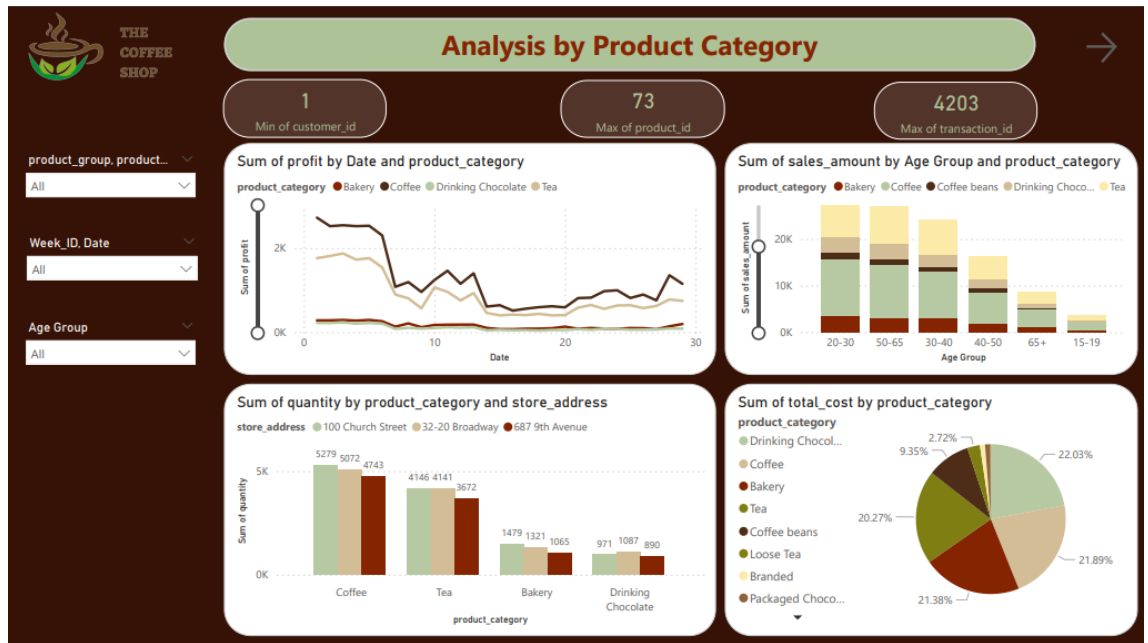
Nhóm Merchandise có lợi nhuận trung bình cao, đường vẽ nằm tách khỏi các nhóm khác, lợi nhuận trung bình dao động mạnh từ 8 đô cho đến 19 đô la Mỹ. Đặc biệt là những ngày đầu tiên của tháng không hề có sự xuất hiện của nhóm sản phẩm này, cho đến tận ngày 8 nó mới xuất hiện. Tương tự như Merchandise, những ngày

đầu tiên Whole Bean/Teas chưa có đường vẽ, mãi đến ngày 7 nó mới bắt đầu xuất hiện, dao động bình thường từ 2-3.5 đô. Tuy nhiên vào ngày 17, lợi nhuận trung bình của nó đột ngột tăng cao lên đến 10.9 đô la, có lẽ vào ngày này có người hoặc nhóm người đã mua số lượng sản phẩm trong nhóm Whole Bean/Teas nhiều hơn ngày thường. Beverage khá ổn định với mức lợi nhuận trung bình rơi vào khoảng 3-3.6 đô. Hai nhóm còn lại là Food và Add-ons có mức lợi nhuận trung bình gần như tương đương nhau và không có biến động nhiều từ 0.8 đến 1.6 đô la Mỹ.

Tóm lại, nhóm Merchandise và Whole Bean/Teas có sự biến động mạnh mẽ trong lợi nhuận trung bình và có mức lợi nhuận trung bình cao hơn, cho thấy sự thay đổi trung bình trong nhu cầu của khách hàng từ ngày này qua ngày khác. Trong khi đó, nhóm Beverages, Food và Add-ons có mức lợi nhuận trung bình ổn định hơn, cho thấy sự ổn định trong nhu cầu tiêu thụ trung bình của các khách hàng.

Kết luận: Thông qua việc phân tích Product Group, chúng ta đã hiểu rõ hơn về cấu trúc và hiệu suất kinh doanh của từng nhóm sản phẩm trong chuỗi cửa hàng. Nhóm Beverages chiếm vị trí hàng chủ lực với lượng tiêu thụ lớn nhất, trong khi đó nhóm Merchandise và Whole Bean/Teas mang lại lợi nhuận trung bình cao hơn nhờ sự biến động trong tiêu dùng trung bình của khách hàng. Các nhóm sản phẩm khác như Food và Add-ons cho thấy sự ổn định trong lợi nhuận trung bình, mặc dù có tỷ trọng trong cấu trúc doanh thu không lớn nhưng vẫn đóng góp vào doanh thu tổng thể của doanh nghiệp.

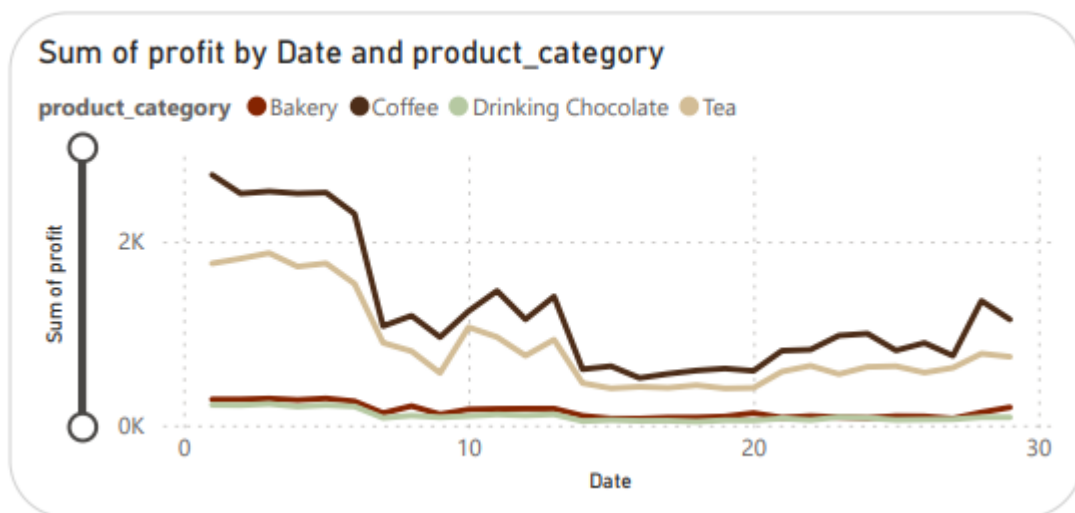
4.6. Product Category



Hình 69. Dashboard Analysis Analysis by Product Category

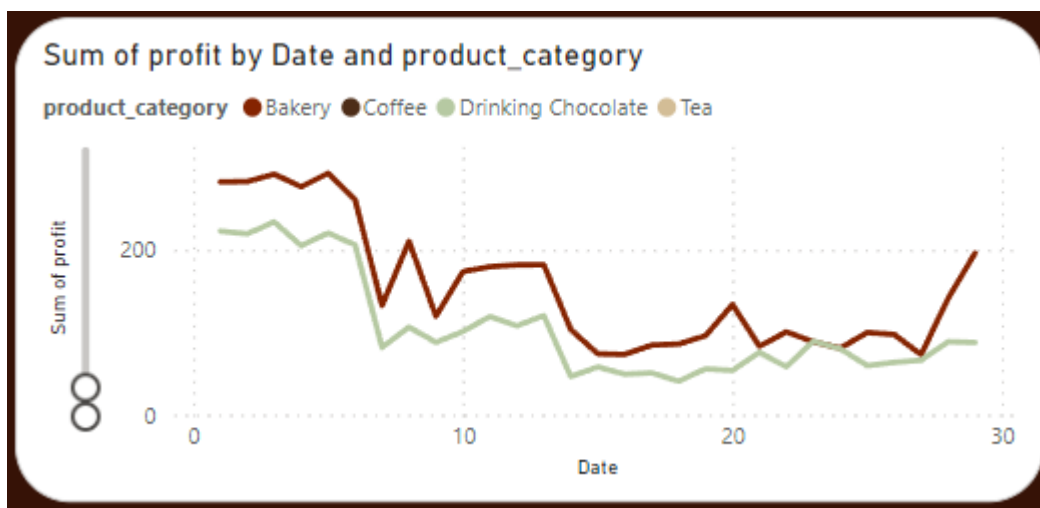
Trang Product Category cung cấp một cái nhìn tổng thể về hành vi tiêu dùng của khách hàng theo doanh thu, số lượng bán ra, lợi nhuận và chi phí của từng danh mục sản phẩm (cụ thể ở đây là 9 danh mục). Qua đó dashboard giúp ta hiểu rõ hơn về xu hướng mua sắm của người tiêu dùng.

Việc đầu tiên ta cần quan tâm là lợi nhuận của từng danh mục hàng ngày của tháng tư này.



Hình 70. Chart Sum of profit by Date and Product_category

Có thể thấy, dẫn đầu lợi nhuận là Coffee trong những ngày đầu tháng với dao động trên 2.5 nghìn đô la Mỹ. Kế đến là danh mục Tea với lợi nhuận dao động gần 2 nghìn đô. Tuy nhiên có sự chênh lệch đáng kể ở những danh mục hàng đầu về lợi nhuận này, hai danh mục Bakery và Drinking Chocolate có đường kẻ nằm sát bên dưới khó quan sát, ta sẽ dùng công cụ Zoom slider để nhìn rõ hơn.



Hình 71. Chart Sum of profit by Date and product_category

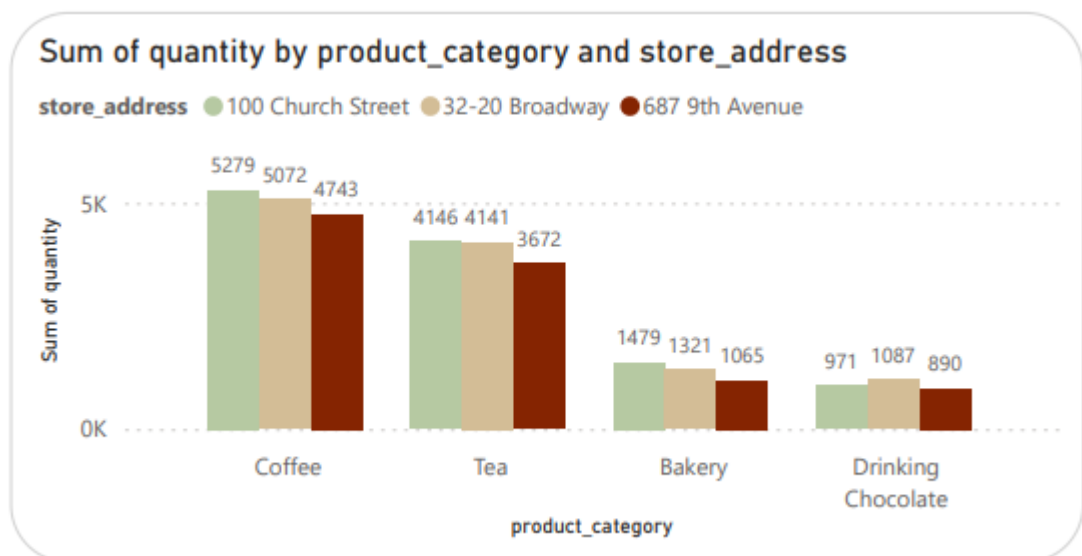
So với lợi nhuận nghìn đô của hai danh mục bên trên, Bakery khởi đầu chỉ xấp xỉ 300 đô và Drinking Chocolate chỉ hơn 200 đô la Mỹ.

Những ngày tiếp theo, tương tự như xu hướng chung, những danh mục này đột ngột giảm lợi nhuận và tiếp tục giảm dần đến cuối tháng. Dù vậy, hai danh mục Coffee và Bakery vào hai ngày cuối tháng lại có xu hướng đi lên rõ rệt.

Danh mục Coffee và Tea là hai danh mục sản phẩm dẫn đầu về lợi nhuận, cho thấy sự ưa chuộng và nhu cầu cao của khách hàng đối với các loại đồ uống này. Trong khi đó, Bakery và Drinking Chocolate có lợi nhuận thấp hơn, có thể gợi lên những cơ hội cải thiện hoặc điều chỉnh chiến lược tiếp thị và sản phẩm để tăng cường doanh số.

Sự biến động trong lợi nhuận của từng danh mục trong tháng cũng cho thấy sự đa dạng và biến động của hành vi tiêu dùng của khách hàng. Việc theo dõi và phân tích chi tiết xu hướng này có thể giúp doanh nghiệp thích nghi và tối ưu hóa chiến lược kinh doanh của mình để đáp ứng nhu cầu thị trường một cách linh hoạt và hiệu quả.

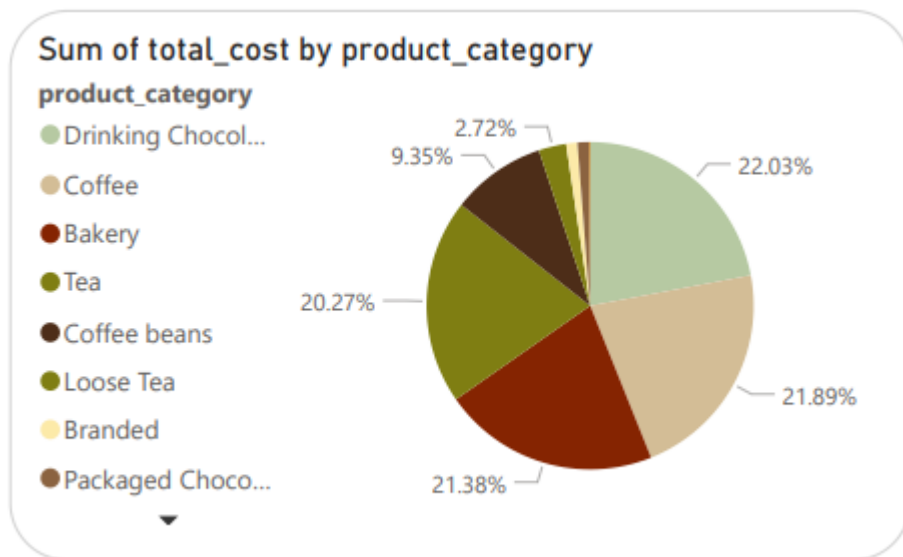
Tiếp đó, chúng ta sẽ đi sâu vào việc phân tích lượng tiêu thụ theo từng danh mục sản phẩm và cửa hàng, để có thể hiểu rõ mức tiêu thụ của khách với các danh mục này.



Hình 72. Chart Sum of quantity by product_category and store_address

Cả ba cửa hàng đều có một vài xu hướng chung. Đầu tiên Coffee là danh mục có lượng tiêu thụ cao nhất, và cao nhất tại 100 Church Street (5279), tiếp đó là 32-20 Broadway (5072), 687 9th Avenue (4743). Danh mục đứng thứ hai là Tea, mặc dù có lượng tiêu thụ thấp hơn Coffee nhưng không quá mức khập khiễng. Dẫn đầu vẫn là cửa hàng ở 100 Church Street, địa chỉ 32-20 Broadway cũng không kém cạnh, cả hai cửa hàng đều có lượng tiêu thụ trà hơn 4140 sản phẩm, xếp sau vẫn là 687 9th Avenue. Hai danh mục Bakery và Drinking Coffee có lượng tiêu thụ thấp hơn so với hai danh mục trên đáng kể, Drinking Coffee có lượng tiêu thụ thấp nhất trong số các danh mục. Tuy nhiên ở danh mục Drinking Coffee, lần đầu cửa hàng tại địa chỉ 32-20 Broadway đã chiến thắng về mức tiêu thụ so với 100 Church Street (1087 và 971 sản phẩm).

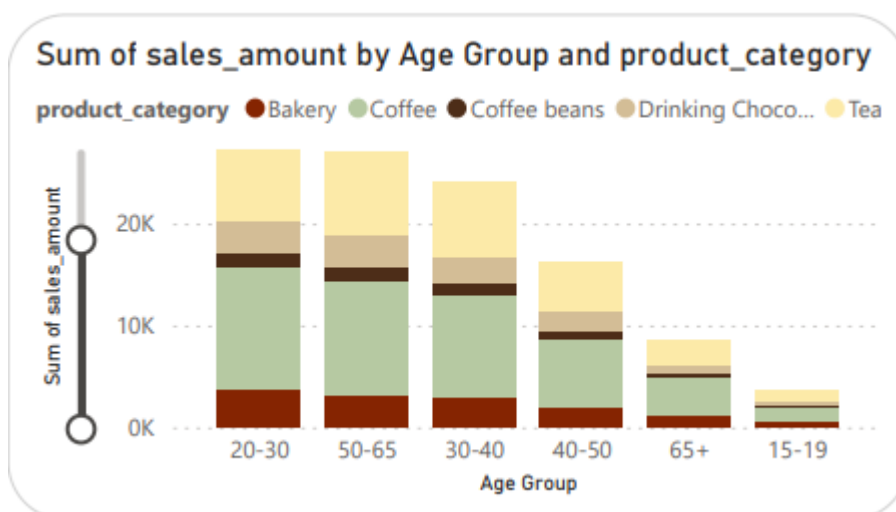
Dựa vào phân tích lượng tiêu thụ theo từng danh mục sản phẩm và cửa hàng, chúng ta nhận thấy một số xu hướng chung. Coffee là danh mục có lượng tiêu thụ cao nhất tại tất cả ba cửa hàng, trong khi Tea đứng ở vị trí thứ hai. Bakery và Drinking Chocolate có lượng tiêu thụ thấp hơn, đặc biệt là Drinking Chocolate có mức tiêu thụ thấp nhất.



Hình 73. Chart Sum of total_cost by product_category

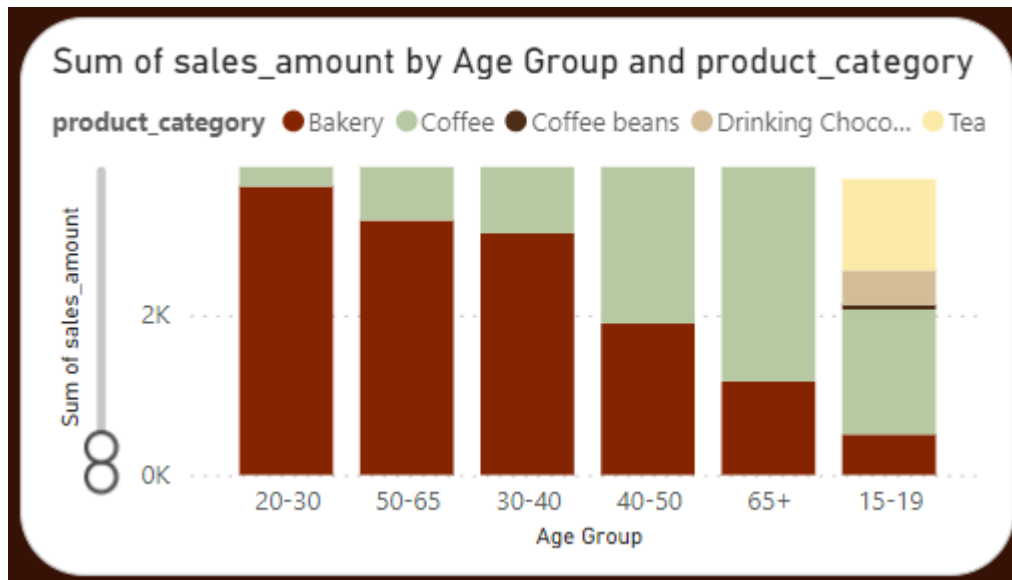
Còn về tổng chi phí theo danh mục thì chiếm tỷ trọng nhiều nhất lại là Drinking Chocolate, trong hai phân tích về lợi nhuận và tiêu thụ trước đó, Drinking Coffee có số liệu thua thiệt hẳn Coffee và Tea, nhưng tổng chi phí của nó lại đứng đầu với 22.03%. Kế đến là Coffee 21.89%, 20.27% là Tea, tuy nhiên Bakery xếp thứ ba trên cả Tea với 21.28% tỷ lệ chi phí, cũng như Drinking Coffee, lợi nhuận và mức tiêu thụ của nó ít hơn hẳn nhưng chi phí chênh lệch với Coffee và Tea rất nhỏ. Bốn danh mục còn lại là Loose Tea, Branded, Packaged Chocolate và Flavours tổng cộng chỉ chiếm hơn 5%, đặc biệt là Flavours chỉ chiếm 0.12% tổng chi phí.

Cuối cùng ta tìm hiểu xu hướng khách hàng theo độ tuổi của từng danh mục. Dưới đây là biểu đồ thể hiện lượng tiêu thụ theo danh mục nhiều nhất của từng nhóm tuổi.



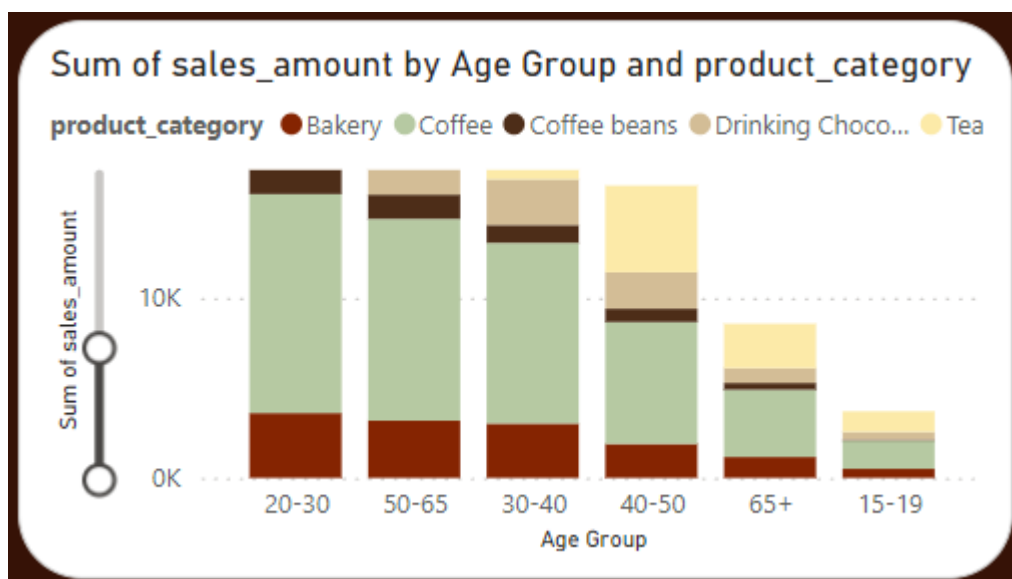
Hình 74. Chart Sum of Sales_amount by Age Group and product_category

Ta sẽ phân tích lượng tiêu thụ từng danh mục theo nhóm tuổi, dùng Zoom slider để nhìn rõ hơn, bắt đầu từ tầng đáy của cột là Bakery.



Hình 75. Chart Sum of Sales_amount by Age Group and product_category sử dụng slicer

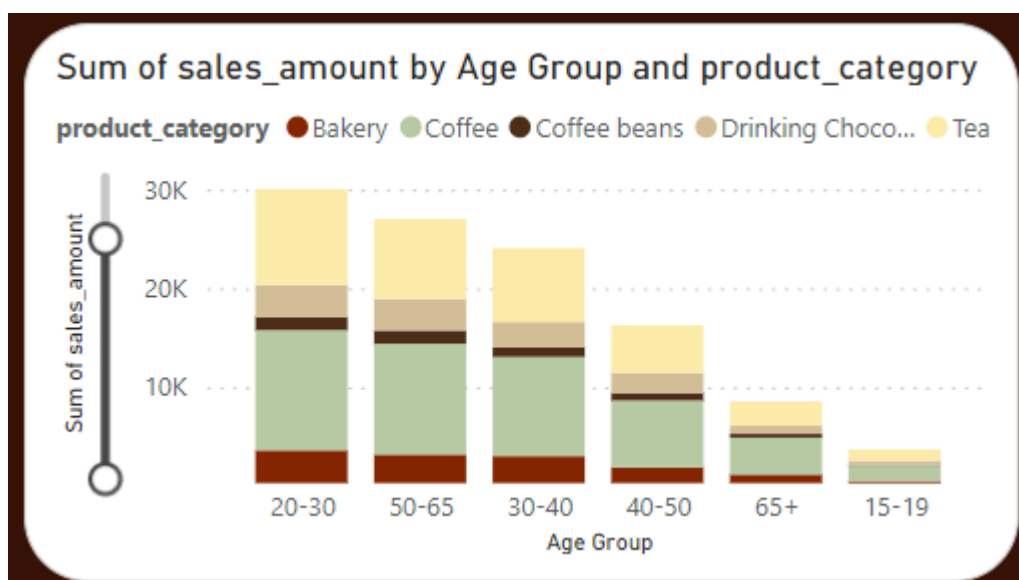
Bakery được nhóm tuổi từ 20-30 ưa chuộng nhất với doanh thu hơn 3.6 nghìn đô la Mỹ, nhóm tuổi 50-65 và 30-40 có doanh thu gần bằng nhau dao động trong khoảng 3.1 nghìn đô la. Xếp hạng kế tiếp lần lượt là nhóm 40-50, 65+ và thấp nhất là nhóm 15-19 với mức doanh thu chỉ với hơn 500 đô la Mỹ.



Hình 76. Chart Sum of sales_amount by Age Group and product_category (slicer)

Về danh mục Coffee có thể thấy rõ hơn sự chênh lệch so với Bakery. Dẫn đầu vẫn là nhóm 20-30 với doanh thu rơi vào khoảng 12.1 nghìn đô la Mỹ. Nhóm tuổi 50-65 cũng chỉ khá mạnh tay với khoảng 11.2 nghìn đô. Hơn 10 nghìn đô la Mỹ là doanh thu của nhóm 30-40 tuổi. Ba nhóm tuổi 40-50, 65+ và 15-19 cộng lại về cuối.

Danh mục Coffee beans có lẽ vì không phải là sản phẩm có thể mua hàng ngày và tiêu dùng dễ dàng như các danh mục khác nên doanh thu thấp nhất trong nhóm danh mục ở đây. Lần này doanh thu nhiều nhất ở danh mục Coffee beans là nhóm 50-65 với 1347 đô la, doanh thu nhóm tuổi 20-30 chỉ thua 10 đô la so với nhóm này. Những nhóm tuổi còn lại là 30-40, 40-50 và 65+ có doanh thu lần lượt là 991, 741, 379 đô la Mỹ. Ở nhóm tuổi 15-19, phải dùng Zoom slider phóng to hết cỡ mới có thể thấy rõ sự biểu diễn của Coffee beans, doanh thu của nhóm thanh thiếu niên này chỉ có 58.5 đô la Mỹ.



Hình 77. Chart Sum of Sales_amount by Age Group and product_category (slicer)

Hai danh mục cuối là Drinking Chocolate và Tea, xu hướng không quá khác biệt với những danh mục còn lại. Dẫn đầu luôn là nhóm 20-30, tiếp đó lần lượt là 50-65, 30-40, 40-50, 65+ và đứng cuối luôn là 15-19.

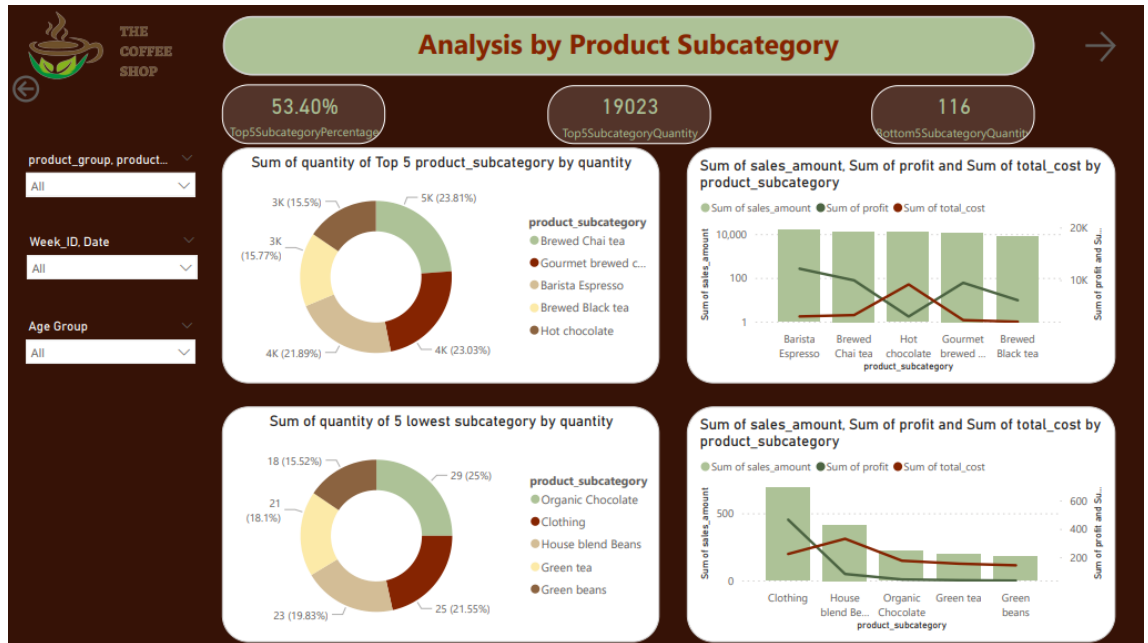
Tóm lại, nhóm 20-30 tuổi là nhóm có sức tiêu thụ mạnh nhất ở tất cả danh mục. Riêng Coffee beans thì nhóm 50-65 đứng đầu, có lẽ vì họ có sự hiểu biết và sở thích chuyên sâu hơn về cà phê. Điều này có thể hiểu là nhóm tuổi trẻ, vừa bắt đầu đi làm có sức tiêu thụ mạnh và bị thu hút bởi những sản phẩm của cửa hàng hơn những nhóm tuổi lớn hơn. Tuy nhiên, có một số đối tượng khách hàng, như nhóm tuổi 15-19, hiển nhiên có mức tiêu thụ thấp hơn đáng kể so với các nhóm khác. Điều này có thể vì vấn đề kinh tế do số tuổi còn quá nhỏ và không có công việc ổn định của nhóm tuổi này. Điều này có thể yêu cầu doanh nghiệp xem xét các chiến lược tiếp thị giá rẻ, nhiều ưu đãi đặc biệt để thu hút và giữ chân khách hàng trong nhóm này.

Kết luận: Thông qua trang Category trên, ta có cái nhìn tổng quan và sâu sắc về hành vi tiêu dùng của khách hàng đối với các danh mục sản phẩm cụ thể. Bằng cách tập trung vào lợi nhuận, lượng tiêu thụ và chi phí của từng danh mục giúp định rõ xu hướng và ưu tiên của khách hàng.

Một điểm chính là sự khác biệt rõ rệt giữa các danh mục sản phẩm, với Coffee và Tea dẫn đầu về lợi nhuận và lượng tiêu thụ, trong khi Bakery và Drinking Chocolate có lợi nhuận thấp hơn và chiếm tỷ lệ chi phí cao. Điều này gợi lên những cơ hội cải thiện chiến lược tiếp thị và sản phẩm để tăng doanh số.

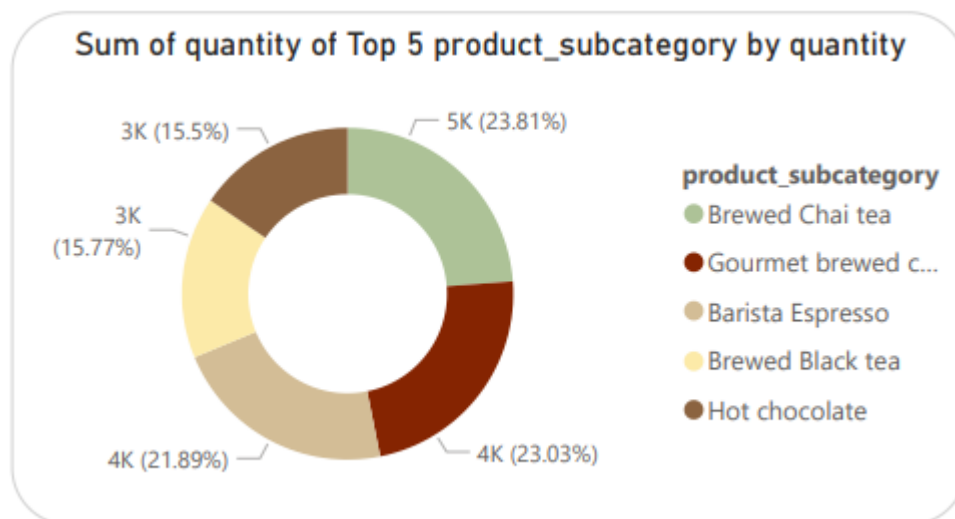
Thêm vào đó, phân tích về lượng tiêu thụ theo từng nhóm tuổi cũng cho thấy nhóm tuổi 20-30 chiếm ưu thế về sức mua mạnh mẽ, trong khi nhóm tuổi 15-19 lại thể hiện mức tiêu thụ thấp hơn đáng kể. Điều này mở ra cơ hội cho các chiến lược tiếp thị đặc biệt nhằm thu hút và giữ chân khách hàng trong nhóm này.

4.7. Product Subcategory



Hình 78. Dashboard Analysis by Product Subcategory

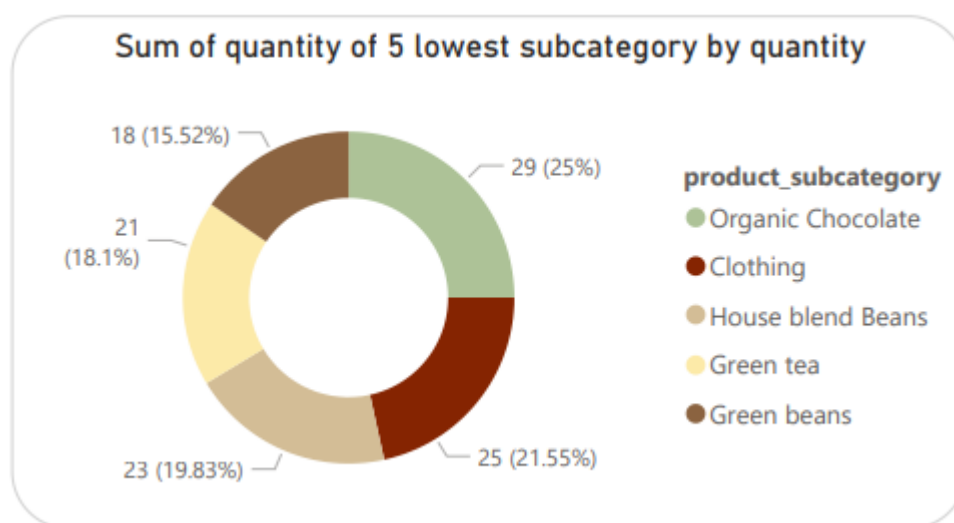
Dashboard này cung cấp cái nhìn tổng quan về hiệu suất bán hàng theo từng subcategory, giúp nhận diện được các sản phẩm chủ lực và những sản phẩm cần cải thiện doanh số. Cụ thể, nhóm chọn ra top 5 subcategory bán chạy nhất và 5 subcategory bán chậm nhất về số lượng để phân tích, so sánh.



Hình 79. Chart of quantity of Top 5 product_subcate by quantity

5 phân loại con dẫn đầu về số lượng trong bộ dữ liệu lần lượt theo thứ tự giảm dần về thứ hạng bao gồm: Brewed Chai tea, Gourmet brewed coffee, Barista Espresso, Brewed Black tea, Hot chocolate. Tỷ lệ của 5 loại này phân bổ tương đối đồng đều với 2 loại đầu khoảng 23% mỗi loại, tương đương 4-5 nghìn sản phẩm được tiêu thụ. Hot chocolate có tỷ trọng thấp nhất cũng chiếm tới hơn 15%. Trong tháng, tổng số lượng bán ra của 5 loại này đạt hơn 19,000 đơn vị, chiếm hơn 53% tổng doanh số được ghi nhận.

Đây là những loại đồ uống được yêu thích nhất của hệ thống. Tuy là chuỗi bán lẻ cafe, doanh nghiệp vẫn bán chạy những loại đồ uống khác như trà, chocolate. Chứng tỏ hiệu quả trong nỗ lực đa dạng hóa sản phẩm của doanh nghiệp mang về hiệu quả và thành công nhất định. Cần tiếp tục những chính sách, chiến lược kinh doanh đồng thời phân bổ nguồn lực thích hợp, đủ để duy trì sự ổn định của những phân loại hàng này.

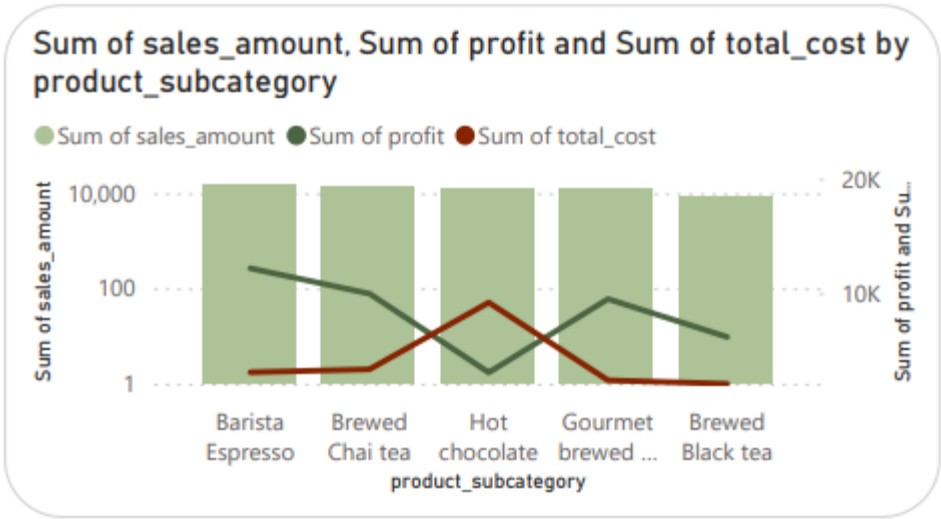


Hình 80. Chart Sum of quantity of 5 lowest subcategory by quantity

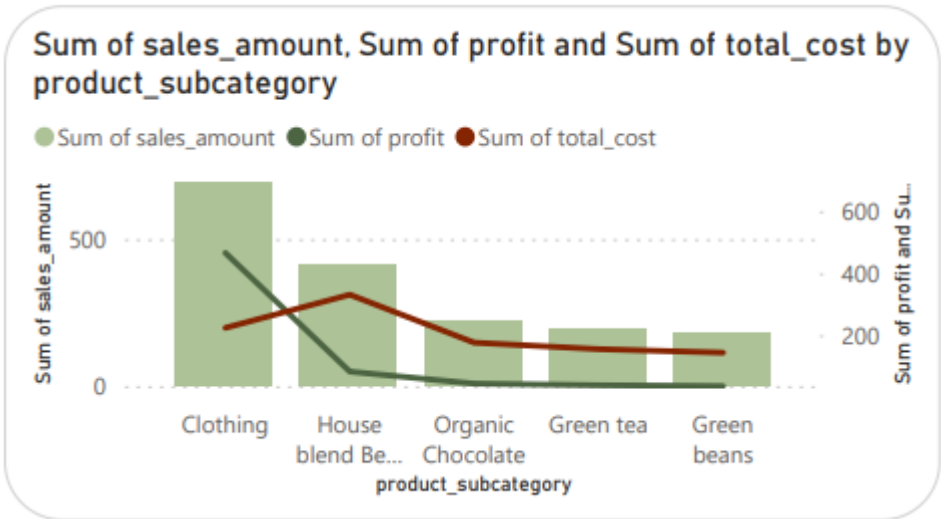
Trong khi đó, nhóm 5 subcategory xếp cuối về số lượng đã bán gồm: Organic Chocolate, Clothing, House blend Beans, Green tea, Green Beans. Trong đó, Green Beans là loại sản phẩm bán được ít nhất chỉ với 18 đơn vị trong tháng. Phần còn lại lần lượt bán được 29, 25, 23 và 21 sản phẩm. Cả 5 loại đều thuộc về

nhóm sản phẩm phụ như các loại hạt hoặc merchandise. Vì vậy, số lượng bán ra của chúng thấp nhất là một điều dễ hiểu.

Để xác định kế hoạch kinh doanh cho những loại sản phẩm trên cần xem xét thêm giá trị mà sản phẩm mang lại cho doanh nghiệp.



Hình 81. Chart Sum of sales_amount, Sum of profit and Sum of total_cost by product_subcategory of Top 5 subcategory by quantity



Hình 82. . Chart Sum of sales_amount, Sum of profit and Sum of total_cost by product_subcategory of 5 lowest subcategory by quantity

Nhóm 5 loại sản phẩm bán chạy có tổng doanh thu khá đồng đều, trong khoảng 9 đến 15 nghìn đô. Tuy nhiên, riêng Hot chocolate ghi nhận tổng lợi nhuận thấp hơn chi phí tương ứng. Đây là một điều đáng lưu tâm khi mà một phân loại sản phẩm bán chạy hàng đầu của hệ thống lại không mang về lợi nhuận tương xứng. Nó đặt ra câu hỏi liệu việc đầu tư tài nguyên và nguồn lực để duy trì số lượng cho phân loại này có còn là quyết định đúng đắn? Trong khi đó, Barista Espresso, Brew Chai tea hay Gourmet brewed coffee và Brewed Black tea đều chứng tỏ chúng xứng đáng là phân loại sản phẩm chủ lực khi mà lợi nhuận thu được luôn cao hơn chi phí, nhất là Barista Espresso, loại có tỷ suất lợi nhuận cao nhất. Để tối ưu hóa lợi nhuận, việc tập trung giảm tổng chi phí cho các sản phẩm phổ biến như “Brewed Chai tea” hoặc tăng doanh số cho các sản phẩm có tỷ suất lợi nhuận cao như “Barista Espresso” có thể mang lại lợi ích.

Ở nhóm đối trọng, Clothing là phân loại bán được tốt hơn cả trong nhóm các phân loại bán được ít nhất. Đây cũng là loại duy nhất có lợi nhuận cao hơn chi phí. Phần còn lại đều có tổng chi phí cao hơn tổng lợi nhuận, cho thấy hiệu quả kinh doanh kém cũng như giá trị mà chúng mang lại là không đáng kể. Việc cân nhắc cắt giảm những sản phẩm như thế này giúp doanh nghiệp tập trung phát triển, duy trì những sản phẩm tiềm năng và ổn định hơn, tối ưu hiệu suất kinh doanh tốt hơn.

Kết luận: Đây là câu chuyện bán hàng cung cấp một cái nhìn tổng quan về hiệu suất bán hàng theo từng phân loại con, mà còn giúp nhận diện được các sản phẩm chủ lực và những sản phẩm cần cải thiện doanh số. Dữ liệu bán hàng của top 5 subcategory bán chạy nhất và 5 subcategory bán chậm nhất đã vẽ ra một bức tranh rõ ràng về sự hiệu quả của các sản phẩm trong danh mục.

Các sản phẩm dẫn đầu về số lượng bán ra trong bộ dữ liệu kể trên, có tỷ lệ tương đối đồng đều trong khi chúng chiếm hơn nửa tổng doanh số. Đây là những

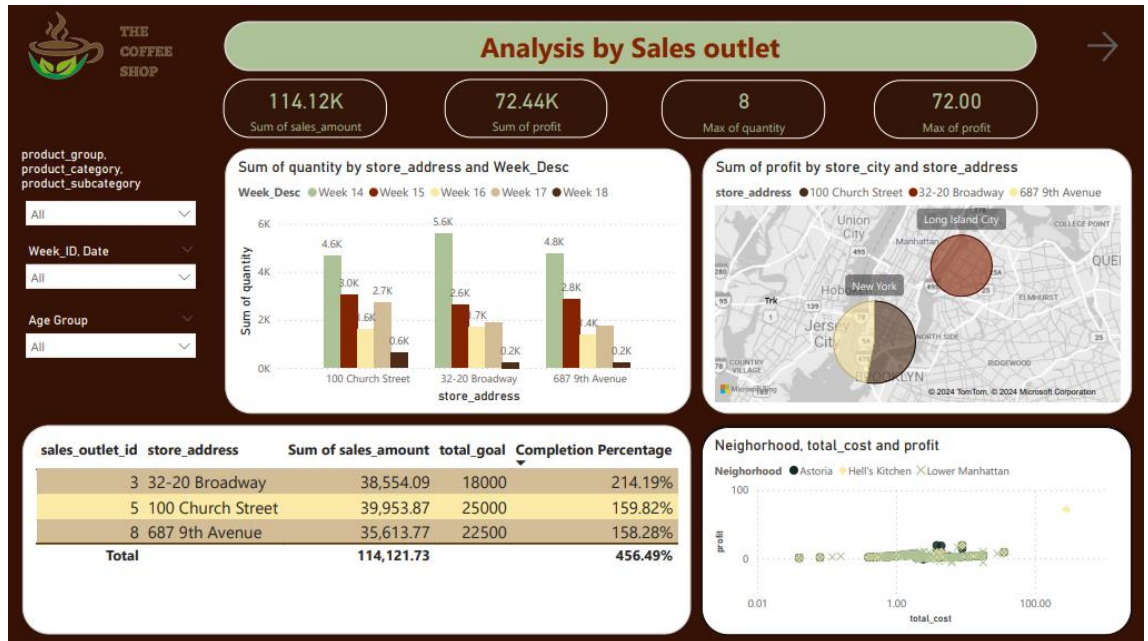
sản phẩm được yêu thích nhất của hệ thống, và việc duy trì sự ổn định của chúng là rất quan trọng.

Mặt khác, nhóm các sản phẩm xếp cuối về số lượng bán thể hiện rằng các sản phẩm phụ như merchandise hoặc Whole Bean/Teas không thể kỳ vọng vào doanh số bán hàng lớn. Việc xem xét giá trị mà sản phẩm mang lại cho doanh nghiệp là cần thiết để xác định kế hoạch kinh doanh cho những loại sản phẩm này.

Trong khi các sản phẩm bán chạy khác thường có tỷ lệ lợi nhuận cao hơn chi phí, sản phẩm Hot chocolate lại ghi nhận tỷ lệ lợi nhuận thấp hơn chi phí tương ứng. Điều này đặt ra câu hỏi về con đường tiếp tục đầu tư tài nguyên và nguồn lực cho sản phẩm này. Việc tập trung vào giảm tổng chi phí cho các sản phẩm phổ biến hoặc tăng doanh số cho các sản phẩm có tỷ lệ lợi nhuận cao có thể mang lại lợi ích.

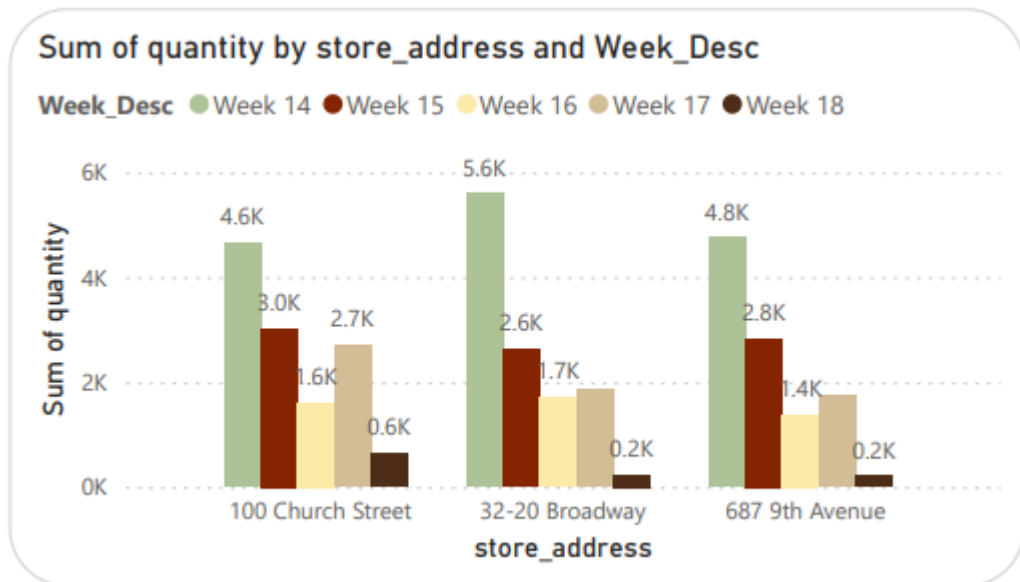
Trong nhóm sản phẩm bán chậm, Clothing là loại sản phẩm bán được tốt nhất với lợi nhuận cao hơn chi phí. Việc cắt giảm các sản phẩm không mang lại giá trị kinh doanh đáng kể giúp doanh nghiệp tập trung vào phát triển và duy trì các sản phẩm tiềm năng và ổn định hơn, từ đó tối ưu hóa hiệu suất kinh doanh.

4.8. Sales Outlet



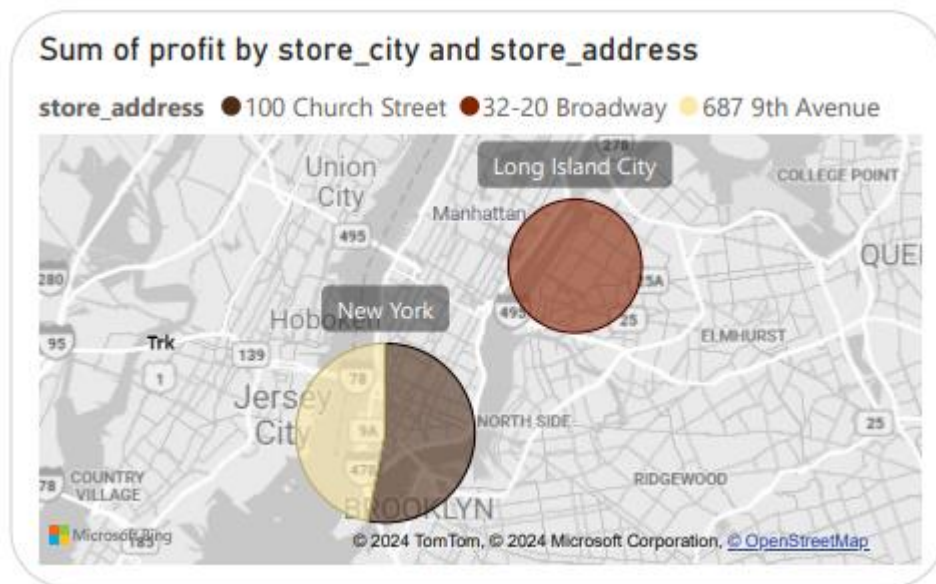
Hình 83. Dashboard Analysis by Sales outlet

Bảng dashboard này cung cấp cái nhìn tổng quan về hiệu suất bán hàng theo từng cửa hàng, giúp nhận diện được những đặc trưng trong vận hành kinh doanh của các chi nhánh trong hệ thống. Bên cạnh đó các cửa hàng còn ghi nhận thông số của các món bánh ngọt được bán trong ngày. Trong tháng, giao dịch có số lượng sản phẩm nhiều nhất là 8 sản phẩm, lợi nhuận cao nhất của một giao dịch là 72\$.



Hình 84. Chart Sum of quantity by store_address and Week_desc

Biểu đồ cột này biểu diễn trực quan số lượng bán hàng của 3 chi nhánh qua các tuần. Sự tương đồng về hình dạng và chiều cao của các cụm cột cho thấy cả xu hướng bán hàng chung của cả 3 cửa hàng trong tháng này. Như những phân tích kinh doanh theo thời gian ở trên, các cửa hàng đều bán được nhiều nhất vào tuần đầu tiên của tháng, trong đó 100 Church Street dẫn đầu về số lượng bán được trong 7 ngày này với doanh thu xấp xỉ 40 nghìn đô, thu về lợi nhuận tháng là 25,42\$. Những tuần còn lại của tháng, doanh số từ các địa chỉ giảm dần, đặc biệt sự sụt giảm đáng kể nhất diễn ra ở tuần thứ 3 (Week 16).



Hình 85. Chart Sum of profit by store_city and store_address

Bản đồ này thể hiện tỷ lệ các vùng nơi có các cửa hàng theo quy mô lợi nhuận của các thành phố. 3 cửa hàng được thống kê trong bộ dữ liệu phân bố ở 2 thành phố là New York và Long Island City cùng thuộc bang New York.

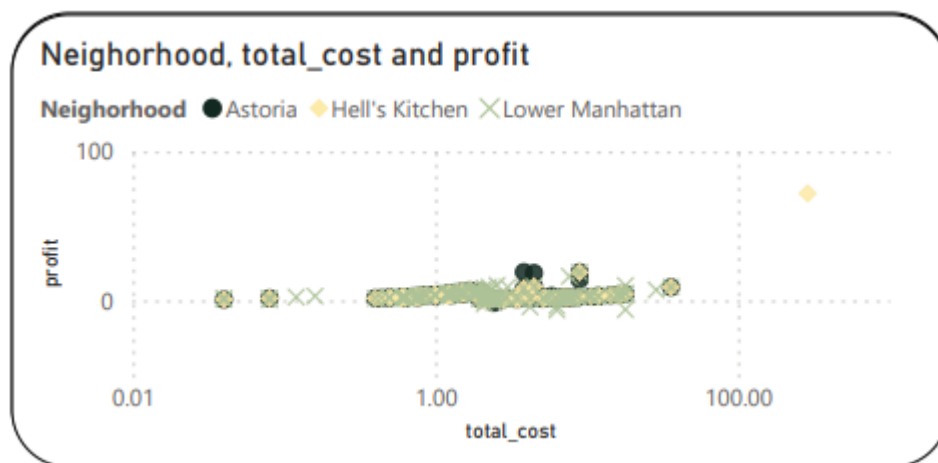
New York là nơi đặt cửa hàng chiếm ưu thế về quy mô lợi nhuận, trong đó vùng Lower Manhattan đạt tổng lợi nhuận là hơn 25 nghìn đô, nhỉnh hơn vùng hơn Hell's Kitchen với 22,000\$.

Trong khi đó, tại Long Island City chỉ có một cửa hàng được thống kê, tuy nhiên, lợi nhuận của riêng của hàng này vẫn rất tốt khi đạt xấp xỉ 24,5 nghìn đô, không kém cạnh vị trí dẫn đầu của cửa hàng 100 Church Street thuộc Lower Manhattan.

sales_outlet_id	store_address	Sum of sales_amount	total_goal	Completion Percentage
3	32-20 Broadway	38,554.09	18000	214.19%
5	100 Church Street	39,953.87	25000	159.82%
8	687 9th Avenue	35,613.77	22500	158.28%
Total		114,121.73		456.49%

Hình 86. Chart table: sale_outlet_id, sum of sales_amount, total_goal, Completion Percentage

Về chỉ tiêu, cả 3 cửa hàng đều vượt kỳ vọng tổng doanh thu được đặt ra. Đặc biệt, chi nhánh tại 32-20 Broadway một lần nữa gây ấn tượng khi doanh thu trong tháng đã đạt hơn 200% so với chỉ tiêu ban đầu. 2 cửa hàng còn lại cũng không kém cạnh khi đạt lần lượt 159,82 và 158,29 phần trăm. Địa chỉ tại 100 Church Street cũng đã có một tháng hoạt động rất năng suất, khi mà chỉ tiêu họ được đặt ra là 25000\$, cao nhất trong 3 cửa hàng, nhưng mức độ vượt chỉ tiêu của họ vẫn có thể xếp hạng nhì. Đây là những con số đầy phấn khởi cho thấy khả năng kinh doanh của các cửa hàng đang biểu hiện rất tốt.



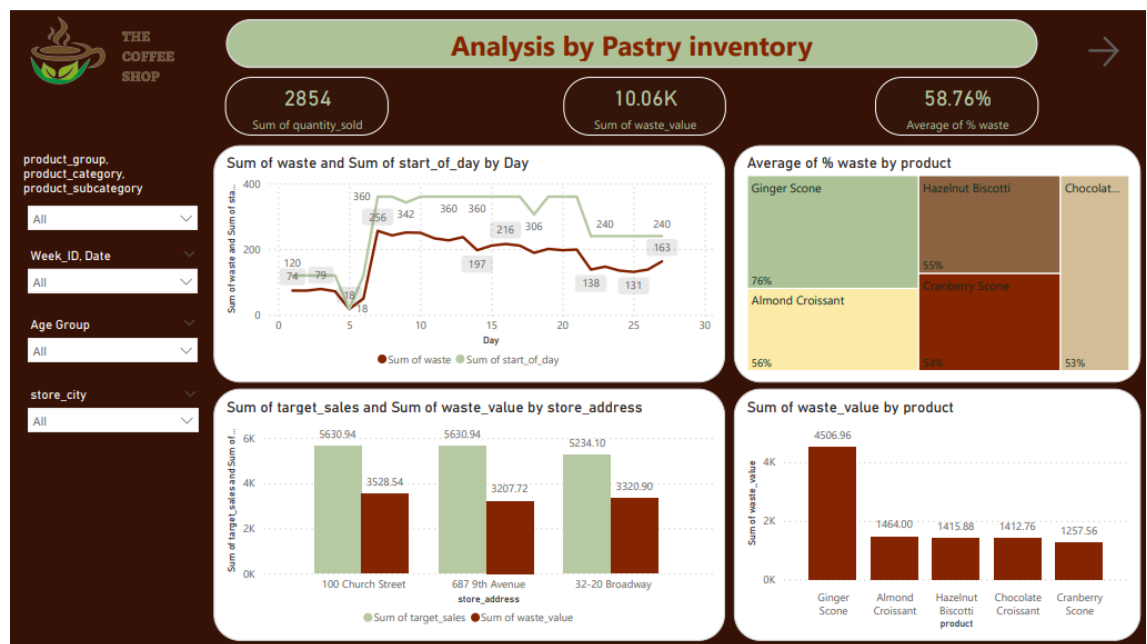
Hình 87. Chart Neighborhood, total_cost and profit

Cuối cùng, là biểu đồ biểu diễn quan hệ tương quan thuận của lợi nhuận và chi phí giữa các khu vực đặt cửa hàng. Tuy không rõ ràng, các điểm dữ liệu có xu hướng tạo thành một đường thẳng đi lên. Điều này nghĩa là chi phí của sản phẩm

cao thì lợi nhuận tương ứng cũng có xu hướng cao. Và các khu vực đều tuân theo xu hướng này. Đặc biệt cửa hàng tại Hell's Kitchen xuất hiện điểm dữ liệu có chi phí và lợi nhuận cao bất thường, có thể là do một giao dịch với số lượng lớn món hàng đã diễn ra trong tháng.

Kết luận: Nhìn chung, các cửa hàng đã có một kỳ kinh doanh tương đối tốt và đồng đều. Các cửa hàng đều sở hữu những điểm sáng trong hiệu suất kinh doanh. Đặc biệt, họ đều thể hiện tốt trong khả năng quản lý kinh doanh và đáp ứng chỉ tiêu doanh số.

4.9. Pastry Inventory



Hình 88. Dashboard Analysis by Pastry inventory

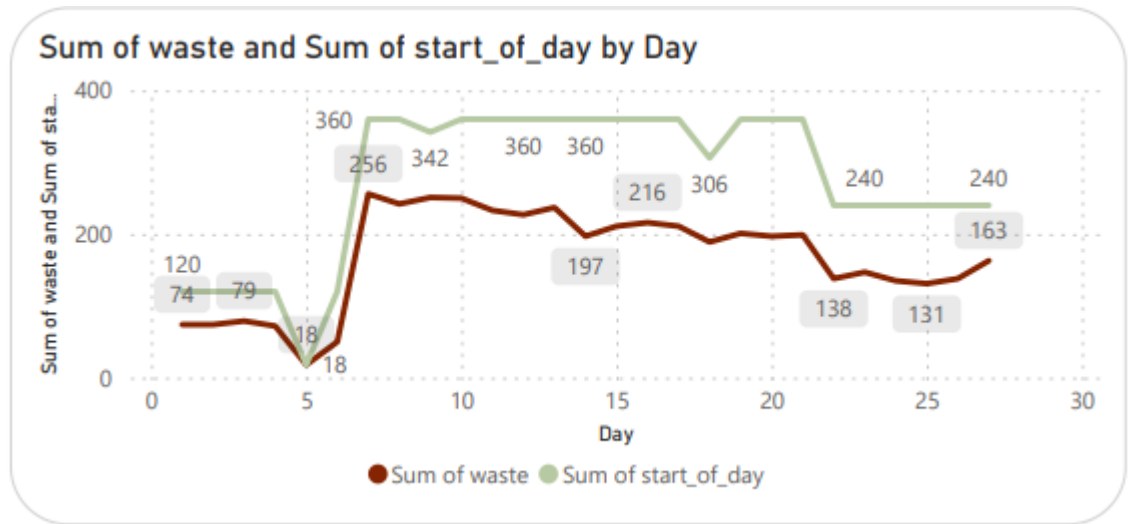
Trang Pastry Inventory cho ta thấy tổng quan tình hình hàng nhập vào và lưu trữ trong kho hàng ngày. Sau khi phân tích khía cạnh này ta có thể sẽ giải quyết được vấn đề lãng phí của nhóm Food.



Hình 89. Card trong dashboard Analysis by Pastry inventory

Tổng quan về số liệu, số lượng hàng nhập được bán ra là 2854 sản phẩm, nhưng trung bình thừa lên đến con số 58.75% đồng nghĩa với lãng phí 10.06 nghìn đô của tổng chi phí mà doanh nghiệp bỏ ra trong tháng tư này.

Chi tiết hơn chúng ta sẽ phân tích số lượng thừa bánh ngọt của từng ngày.



Hình 90. Chart Sum of waste and Sum of start_of_day by Day

Bốn ngày đầu tiên, số lượng nhập ổn định là 120 cái, số lượng thừa từ 72 đến 79 cái. Nhưng đến ngày thứ năm, doanh nghiệp không bán được số bánh ngọt nào nên vẫn thừa nguyên là 18 cái. Ngày này số bánh ngọt nhập vào rất ít chỉ 18 cái, có lẽ doanh nghiệp đã định mở cửa bán thời gian ngắn trong ngày. Từ ngày 7 đến ngày 21, số lượng bánh ngọt nhập vào vẫn ổn định từ 306 đến 360 cái tùy ngày, phần lớn là 360 cái. Số lượng bánh ngọt thừa có giảm dần, từ 256 cái chỉ còn 199 cái, nhưng đây vẫn là có số lượng với lượng nhập là 360 cái trong ngày. Sau 21, doanh nghiệp quyết định nhập cố định 240 cái mỗi ngày cho đến cuối tháng. Số lượng thừa cũng giảm theo, dao động trong 131-138 cái, nhưng ngày 27 (ngày cuối cùng được ghi nhận), số lượng bánh ngọt dư tăng lên 163 cái.

Tóm lại, có những ngày lượng nhập và bán ổn định, nhưng cũng có những ngày mà không có bánh ngọt nào được bán ra. Dù vậy thì lượng bánh thừa của doanh nghiệp vẫn rất nhiều mà không có sự điều chỉnh nào phù hợp cả.

Tiếp đến ta sẽ phân tích sự lãng phí của từng cửa hàng trong chuỗi doanh nghiệp ra sao.

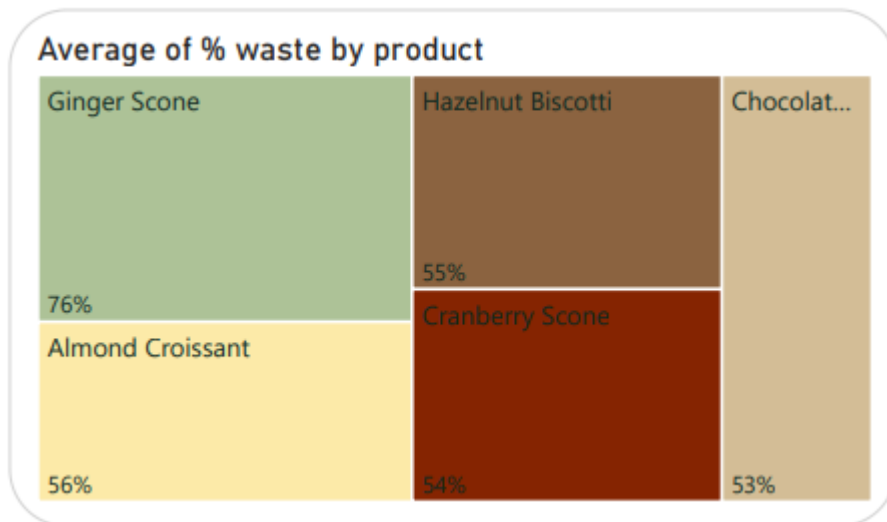


Hình 91. Chart Sum of target_sales and Sum of waste_value by store_address

Target_sales ở đây là tổng chi phí khi đã tiêu thụ hết những bánh ngọt đã nhập vào, còn waste_value là tổng chi phí của những sản phẩm thừa. Ta có thể thấy hai cửa hàng 100 Church Street và 687 9th Avenue có target sales bằng nhau nhưng 100 Church Street lại có waste value cao hơn. Cửa hàng tại 32-20 Broadway có target sales thấp hơn hai cửa hàng kia một chút nhưng waste value lại cao hơn 687 9th Avenue.

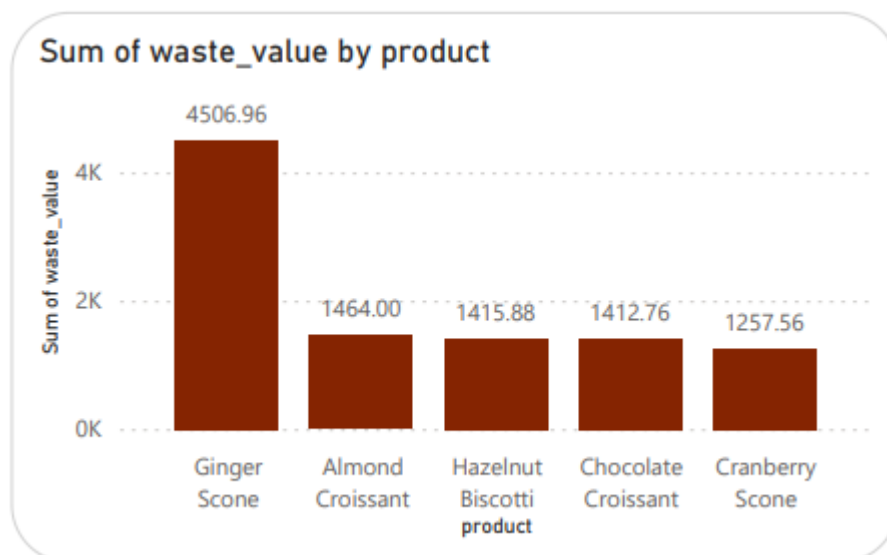
Tóm lại, nhìn chung doanh nghiệp này không hề có một kế hoạch nhập hàng dành riêng cho tình hình kinh doanh của mỗi cửa hàng, vì vậy target sales của 3 cửa hàng gần như bằng nhau, và sự lãng phí cũng dâng lên cao như nhau.

Sau đây ta sẽ phân tích chi tiết hơn doanh nghiệp đã nhập những mặt hàng nào, và chúng đã lãng phí bao nhiêu.



Hình 92. Chart Average of % waste by product

Tỷ lệ lãng phí của 5 sản phẩm được nhập này cực kỳ cao, từ 53% đến 76%, chúng lần lượt là Chocolate Croissant, Cranberry Scone, Hazelnut Biscotti, Almond Croissant, Ginger Scone.



Hình 93. Chart Sum of waste_value by product

Waste value chúng đã gây ra cao nhất là 4506.95 đô của Ginger Scone, các sản phẩm còn lại dao động từ 1257.56 đô đến 1464 đô.

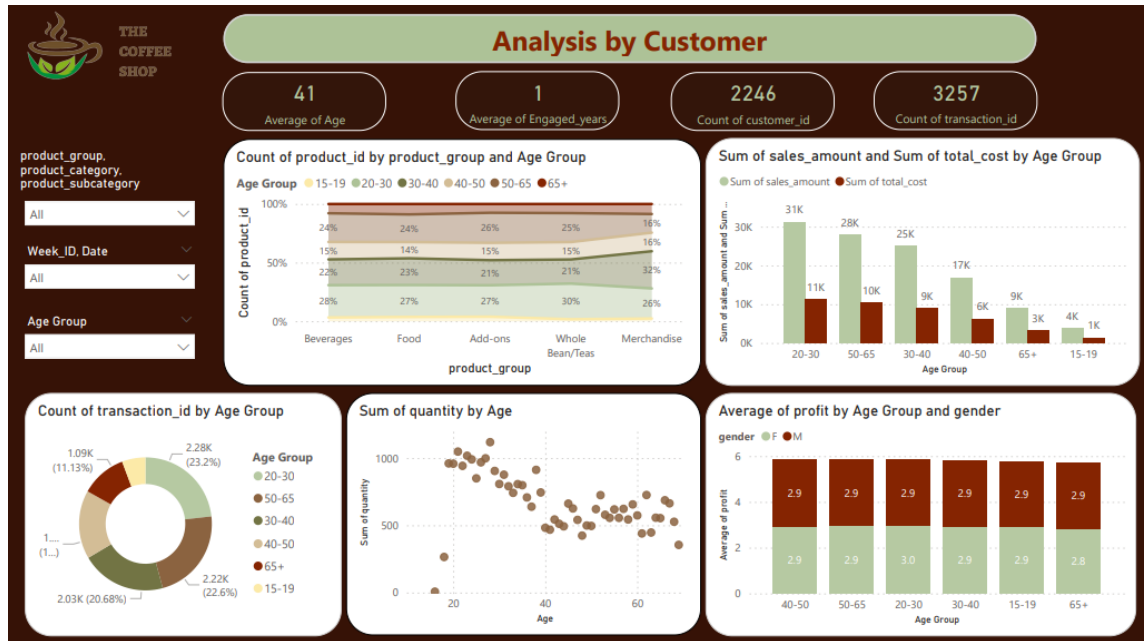
Những sản phẩm này có cả tỷ lệ lãng phí và chi phí lãng phí quá cao, chỉ với 5/12 loại sản phẩm trong nhóm Food mà đã gây ảnh hưởng lớn đến lợi nhuận của tổng thể nhóm Food.

Kết luận: Phân tích tình hình hàng nhập và lưu trữ hàng ngày, chúng ta nhận thấy rằng doanh nghiệp đang phải đối mặt với một vấn đề lãng phí lớn trong nhóm Thức Ăn (Food). Mặc dù có sự biến động (ít) về lượng nhập và bán hàng hàng ngày, nhưng vẫn tồn tại một lượng lớn bánh thừa.

Nguyên nhân của vấn đề này có thể là do sự thiếu điều chỉnh trong kế hoạch nhập hàng và kinh doanh cho từng cửa hàng cụ thể. Target sales không được điều chỉnh phù hợp với nhu cầu tiêu thụ thực tế của từng cửa hàng, dẫn đến sự lãng phí lớn. Điều này càng được thể hiện qua việc waste value của các cửa hàng có sự chênh lệch, mặc dù target sales gần như bằng nhau. Hơn nữa doanh nghiệp cần tìm hiểu kỹ xu hướng thức ăn yêu thích của khách hàng mà chọn loại hợp khẩu vị của người tiêu dùng.

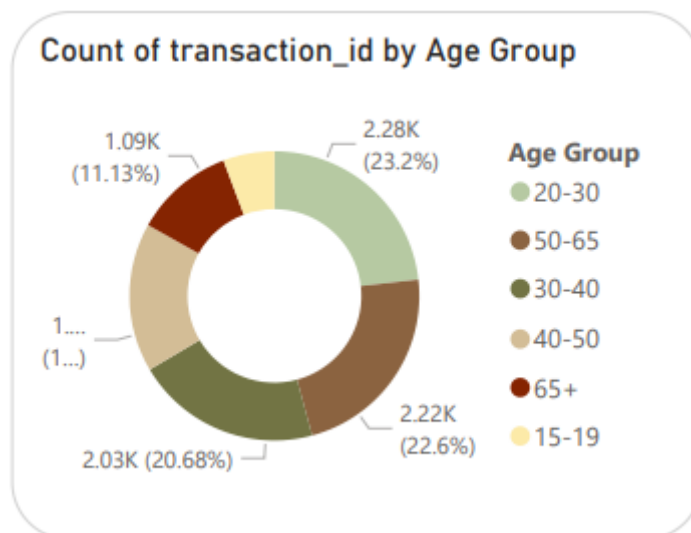
Để giải quyết vấn đề này, doanh nghiệp cần thiết lập một kế hoạch nhập hàng linh hoạt và phù hợp với nhu cầu kinh doanh của từng cửa hàng. Cần phân tích và dự đoán mức độ tiêu thụ hàng hóa tại mỗi cửa hàng, từ đó điều chỉnh lượng nhập hàng sao cho phù hợp. Ngoài ra, cần thiết lập các chính sách quản lý tồn kho hiệu quả, có thể kết hợp với chiến lược giảm giá hoặc chương trình khuyến mãi để thúc đẩy tiêu thụ các sản phẩm còn dư thừa.

4.10. Customer



Hình 94. Dashboard Analysis by Customer

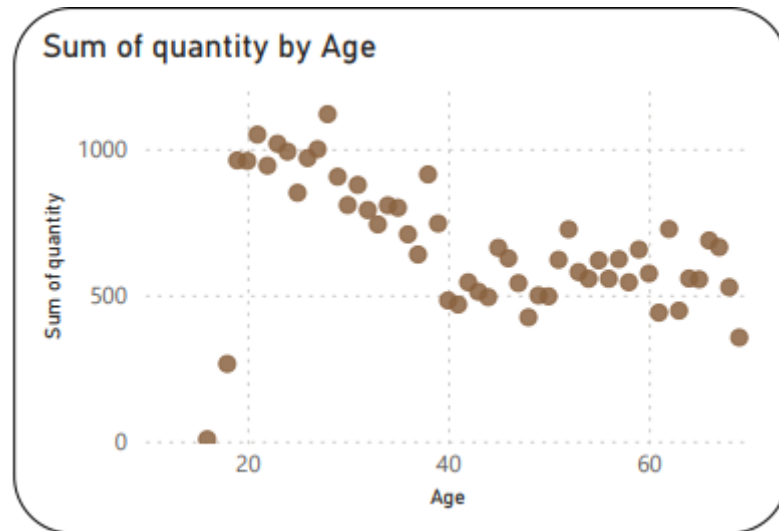
Dashboard này kể câu chuyện về tệp khách hàng mục tiêu của thương hiệu. Những biểu đồ này đi vào phân tích hiệu suất bán hàng đối với từng nhóm khách hàng khác nhau phân loại theo giới tính hoặc độ tuổi. Tổng quan, tuổi trung bình của khách hàng là 41 tuổi, với thời gian trung bình mà họ gắn bó với thương hiệu theo chương trình khách hàng thành viên là khoảng 1 năm. Trong tháng 4 này, 3 cửa hàng ghi nhận 2246 lượt khách hàng thành viên đến mua, thực hiện tổng cộng 3257 giao dịch.



Hình 95. Chart Count of transaction_id by Age Group

Biểu đồ tròn thể hiện tỷ lệ nhóm độ tuổi của khách hàng trong tổng số các giao dịch được thực hiện. Nó trả lời cho câu hỏi trong số những khách hàng đã đến với cửa hàng, đâu là độ tuổi phổ biến nhất? Câu trả lời chính là những khách hàng từ 20-30 tuổi. Nhóm tuổi phổ biến thứ 2 là từ 50-65 tuổi. 2 nhóm tuổi này thực hiện hơn 50% giao dịch trong tháng của các cửa hàng. Nhóm tuổi ít xuất hiện giao dịch nhất chính là nhóm thanh thiếu niên trẻ tuổi, từ 15-19 tuổi. Những người ở tuổi về hưu cũng ít đến với The Coffee Shop. Điều này cho thấy thương hiệu thu hút đa phần là người trưởng thành trong độ tuổi đi học hoặc đi làm.

Tuy nhiên, theo thống kê, 75% người uống cafe ở Mỹ là người từ 55 tuổi trở lên. Trong khi đó, người từ 18-34 tuổi là thế hệ ít uống cafe nhất. Nguyên nhân của sự khác biệt này có thể bắt nguồn từ thói quen tự pha cafe tại nhà của những người uống cafe. Hiệp hội Cà phê Quốc gia, hay NCA, đã thực hiện một cuộc khảo sát cho thấy 79% người Mỹ pha cà phê tại nhà. Thay vì đi ra quán cà phê hay mua cà phê ở cửa hàng tiện lợi hay nhà hàng thức ăn nhanh, hầu hết người Mỹ đều tự pha cà phê cho mình. Chỉ có 36% người tham gia khảo sát cho biết họ uống cà phê ở xa nhà.

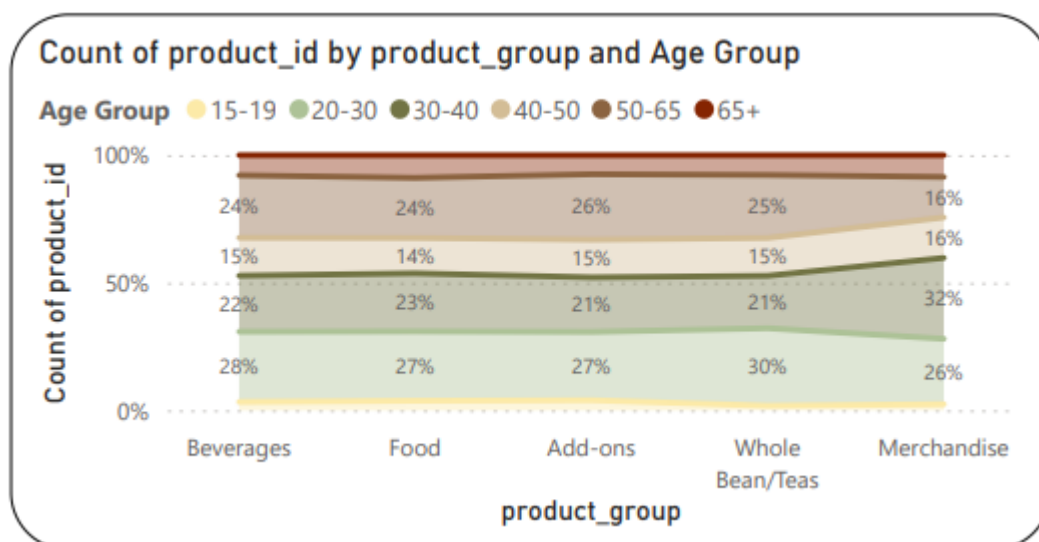


Hình 96. Chart Sum of quantity by Age

Câu hỏi tiếp theo được đặt ra là liệu có tồn tại mối quan hệ nào giữa tuổi của khách hàng và số lượng sản phẩm mà doanh nghiệp bán được? Nhìn vào biểu đồ phân tán, có thể thấy 1 mối quan hệ tương quan nghịch giữa tuổi và số lượng sản phẩm ở độ tuổi từ 20-40. Nghĩa là khi tuổi khách hàng càng tiến dần về 40, số lượng sản phẩm doanh nghiệp bán được cho họ càng giảm.

Ở nhóm tuổi lớn hơn 40, mối tương quan không rõ ràng, các điểm giá trị có xu hướng đi ngang. Những khách hàng lớn hơn 40 tuổi có tổng số lượng sản phẩm đã mua nằm trong khoảng 500-700. Riêng những khách trẻ hơn 20 tuổi, các cửa hàng bán được ít sản phẩm nhất cho nhóm đối tượng này.

Vậy nhìn chung, tuổi của khách hàng và sản phẩm bán được không hoàn toàn là quan hệ đối lập. Tỷ lệ nghịch chỉ đúng với nhóm khách từ 20-40 tuổi của cửa hàng.



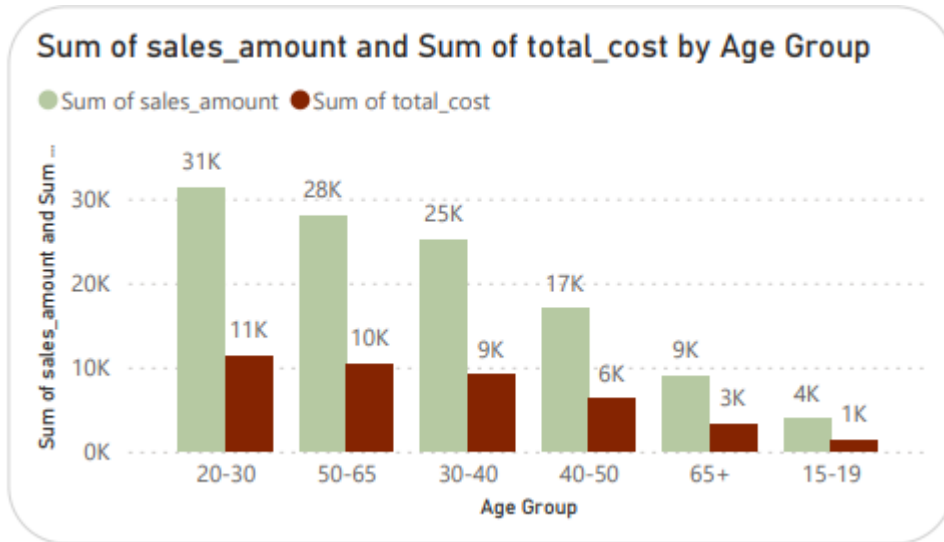
Hình 97. Chart Count of product_id by product_group and Age Group

Để được sự quan tâm của những nhóm tuổi khách hàng khác nhau dành cho từng nhóm sản phẩm mà doanh nghiệp kinh doanh, chúng ta xem qua biểu đồ tỷ lệ trên.

Ở 3 nhóm hàng Beverages, Food và Add-ons, tỷ lệ mua hàng của các nhóm tuổi tương đối tương đồng. Khách hàng trong độ 20-30 tuổi mua nhiều nhất, chiếm phần trăm cao nhất dao động từ 27-28%, kế tiếp là khách trong khoảng 50-65 tuổi chiếm 24-26% và thấp nhất là 15-19 tuổi khi chỉ chiếm hơn 3%. Đối với nhóm hàng Whole Bean/Teas, xu hướng này vẫn giữ nguyên nhưng tỷ lệ có chút khác biệt khi khách trong độ tuổi 15-19 chỉ chiếm hơn 1% và độ tuổi dẫn đầu về sản phẩm tiêu thụ là 20-30 tuổi lên đến 30%.

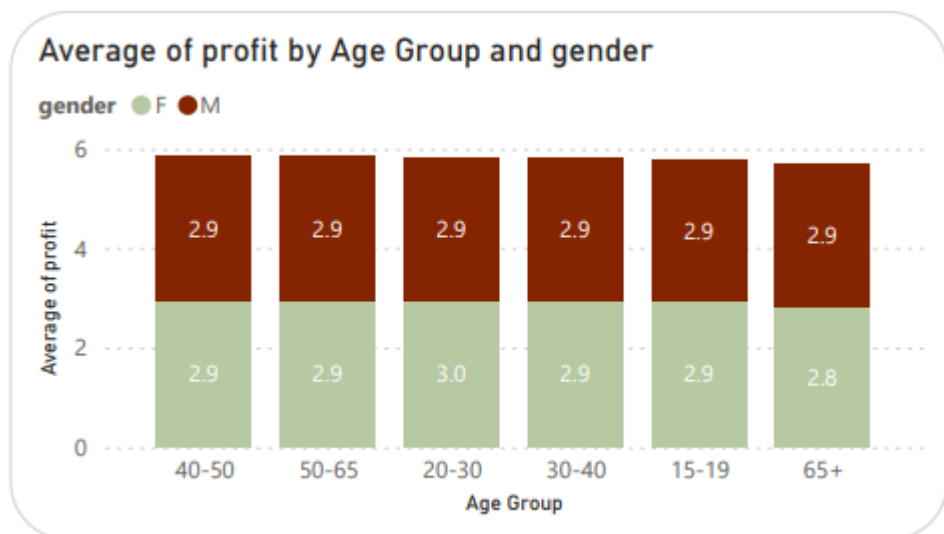
Tuy nhiên, nhóm hàng Merchandise thì không tuân theo tỷ lệ trên. Những khách hàng mua merchandise đa phần ở độ tuổi 30-40 tuổi, họ chiếm 32% tổng lượng khách hàng được ghi nhận. Vị trí thứ 2 thuộc về những khách hàng ở tuổi 20-30.

Từ những phân tích trên, có thể thấy, nhóm khách hàng tiềm năng nhất là 20-40 tuổi, khi mà họ vừa thích đồ uống tại cửa hàng, đồng thời vẫn mua nguyên liệu về tự pha chế, vừa sẵn sàng mua sản phẩm sưu tầm của thương hiệu.



Hình 98. Chart Sum of sales_amount and Sum of total_cost by Age Group

Kế tiếp là phân tích về doanh thu kiếm được từ các độ tuổi và tổng chi phí của những sản phẩm đã bán cho họ. Có thể thấy, thứ tự của những nhóm khách hàng theo doanh thu tương tự với phần phân tích tỷ lệ theo số giao dịch ở trên. Doanh thu cao nhất (31 nghìn đô) mà các cửa hàng kiếm được là từ nhóm khách hàng 20-30 tuổi và thấp nhất là nhóm 15-19 tuổi. Do đó, tổng chi phí của những sản phẩm đã bán cho nhóm khách 20-30 cũng tương ứng, trị giá 11000\$.



Hình 99. Chart Average of profit by Age Group and gender

Về lợi nhuận, nhóm sẽ phân tích lợi nhuận trung bình theo độ tuổi và giới tính (sau khi lọc ra những khách hàng không cung cấp thông tin giới tính). Có thể thấy, lợi nhuận trung bình từ khách hàng nam và nữ ở mọi độ tuổi là gần như bằng nhau, ở mức trên dưới 3 đô cho 1 giao dịch, trong khi số lượng khách hàng nữ nhiều hơn. Vậy có thể sản phẩm bán được cho khách hàng nam có chi phí thấp hơn từ đó thu được lợi nhuận tốt đủ để bù đắp cho khoảng cách về số lượng.

Trong khi doanh số, doanh thu bán hàng cho từng nhóm đối tượng là khác nhau, lợi nhuận trung bình kiếm được từ họ là như nhau. Điều này khẳng định những nhóm tuổi này đều có giá trị đối với doanh nghiệp. Tuy nhiên việc mở rộng tệp khách hàng mục tiêu đòi hỏi nhiều nỗ lực, cần có những nghiên cứu sâu hơn về chiến lược khách hàng để đưa ra phương hướng phù hợp.

Kết luận: Tệp khách hàng mục tiêu của The Coffee Shop có độ tuổi từ 20-65, trong đó nhóm từ 20-30 là mục tiêu tiềm năng hơn cả. Tuy nhiên, không có sự khác biệt nào về lợi nhuận thu được từ những nhóm khách hàng ở các độ tuổi và giới tính khác nhau. Vì vậy, không nên tập trung chú trọng vào một nhóm đối tượng nào mà bỏ qua tiềm năng của những nhóm khách hàng còn lại.

5. ĐỀ XUẤT GIẢI PHÁP

Giải pháp: Áp dụng các chương trình khuyến mãi: mục tiêu là thu hút khách hàng mua sắm, thúc đẩy doanh thu bán hàng vào cuối tháng khi nhu cầu mua sắm thường giảm, hoặc những ngày đặc biệt trong năm. Các hình thức khuyến mãi:

- Giảm giá trực tiếp: Giảm giá % cho một số sản phẩm phù hợp nhất định.
- Mua nhiều giảm giá: Giảm giá khi khách hàng mua nhiều sản phẩm cùng lúc.
- Khuyến mãi combo: Tặng kèm sản phẩm khác khi mua sản phẩm chính.
- Tích điểm đổi quà: Khách hàng tích điểm khi mua sắm và đổi quà vào cuối tháng.

- Giảm giá theo đối tượng khách hàng: Giảm giá cho khách hàng thân thiết, sinh viên, học sinh, ...

Đánh giá:

- Thời gian: Cần thời gian để lên kế hoạch và triển khai các chương trình khuyến mãi, bao gồm việc thiết kế, thực hiện và quản lý chương trình. Một số chương trình khuyến mãi như giảm giá trực tiếp, mua nhiều giảm giá có thể triển khai nhanh chóng, chỉ trong vài ngày. Tuy nhiên các chương trình như tích điểm đổi quà, xây dựng hệ thống VIP có thể cần nhiều thời gian hơn để thu hút khách hàng tham gia và tạo dựng hiệu quả.
- Chi phí: Một số chương trình khuyến mãi như giảm giá có thể thực hiện với chi phí thấp, chỉ cần thay đổi giá sản phẩm. Nhưng các chương trình khuyến mãi phức tạp như tặng kèm sản phẩm có thể tốn kém hơn do chi phí quà tặng.

Giải pháp: Nâng cao hiệu quả quảng bá sản phẩm mới: mục tiêu là giới thiệu sản phẩm mới đến khách hàng tiềm năng, thu hút họ mua sắm. Cần nêu bật điểm độc đáo của sản phẩm mới, sử dụng hình ảnh, video chất lượng cao để thu hút sự chú ý. Bên cạnh đó, kêu gọi hành động rõ ràng để thúc đẩy khách hàng mua sắm cũng là một nội dung quảng bá đáng chú ý. Có thể quảng bá với các hình thức sau:

- Marketing qua mạng xã hội: Facebook, Instagram, TikTok,...
- PR & Media: Bài viết trên báo chí, website, blog,...
- Influencer marketing: Hợp tác với KOLs để quảng bá sản phẩm.

Đánh giá:

- Thời gian: Các kênh quảng bá online như mạng xã hội, influencer marketing có thể triển khai nhanh chóng và tiếp cận lượng lớn khách hàng tiềm năng trong thời gian ngắn. Nhưng những kênh quảng bá offline như PR, quảng cáo trên báo chí có thể cần nhiều thời gian để lên kế hoạch, thực hiện và đánh giá hiệu quả.
- Chi phí: Một số hình thức quảng bá online như đăng bài trên mạng xã hội có thể thực hiện với chi phí thấp hoặc miễn phí nếu tận dụng nguồn lực có sẵn như trang mạng xã hội của thương hiệu. Tuy nhiên, các hình thức quảng bá offline có thể tốn kém hơn do chi phí quảng cáo cao

Giải pháp: Thăm dò ý kiến khách hàng: mục tiêu là hiểu rõ nhu cầu, mong muốn của khách hàng để cải thiện sản phẩm/dịch vụ và nâng cao chất lượng phục vụ. Các hình thức thăm dò:

- Phỏng vấn trực tiếp: Phỏng vấn khách hàng tại cửa hàng, điểm bán hàng.
- Khảo sát trực tuyến: Tạo khảo sát trên website, ứng dụng di động, mạng xã hội.
- Phân tích đánh giá khách hàng: Phân tích đánh giá trên website, mạng xã hội, diễn đàn,...

Đánh giá:

- Thời gian: Các hình thức thăm dò trực tuyến như khảo sát, đánh giá trên mạng xã hội có thể thu thập ý kiến khách hàng nhanh chóng nhưng không chính xác cao. Còn phỏng vấn trực tiếp có thể tốn nhiều thời gian hơn do cần sắp xếp lịch hẹn, thực hiện phỏng vấn và phân tích dữ liệu.
- Chi phí: Các hình thức thăm dò trực tuyến có thể thực hiện với chi phí thấp hoặc miễn phí bằng các công cụ online. Còn phỏng vấn trực tiếp có thể tốn kém hơn do chi phí trả cho nhân viên phỏng vấn.

Giải pháp: Cải thiện chất lượng sản phẩm/dịch vụ: yếu tố quan trọng nhất có lẽ là hiểu rõ nhu cầu và mong muốn của khách hàng. Chất lượng sản phẩm phụ thuộc lớn vào nguyên liệu chất lượng cao từ các nhà cung ứng. Việc thực hiện kiểm tra chất lượng định kỳ có thể giúp phát hiện và giải quyết các vấn đề chất lượng một cách nhanh chóng.

Đánh giá:

- Thời gian: Việc cải thiện chất lượng sản phẩm/dịch vụ có thể diễn ra nhanh chóng hoặc lâu dài tùy thuộc vào bản chất thay đổi.
- Chi phí: Chi phí cải thiện chất lượng sản phẩm/dịch vụ có thể thay đổi tùy thuộc vào mức độ cải thiện, nguồn lực cần thiết và chi phí đầu tư.

Giải pháp: Quản lý kho hiệu quả:

- Xây dựng chiến lược nhập hàng linh hoạt:
 - Phân tích nhu cầu tiêu thụ: Đầu tiên, cần phân tích và dự đoán mức độ tiêu thụ hàng hóa tại mỗi cửa hàng.
 - Điều chỉnh lượng nhập hàng: Dựa trên kết quả phân tích, điều chỉnh lượng nhập hàng sao cho phù hợp với nhu cầu thực tế.
- Thiết lập chính sách quản lý tồn kho hiệu quả:
 - Phân loại hàng hoá, dự báo nhu cầu, kiểm kê định kỳ...
 - Kết hợp chiến lược giảm giá: Có thể kết hợp với chiến lược giảm giá hoặc chương trình khuyến mãi để thúc đẩy tiêu thụ các sản phẩm còn dư thừa.

Đánh giá:

- Thời gian: Một số cải tiến quản lý kho như áp dụng phần mềm quản lý có thể thực hiện nhanh chóng. Mặc dù vậy, việc tối ưu hóa quy trình kho hàng có thể cần nhiều thời gian hơn để phân tích, triển khai và đánh giá hiệu quả.

- Chi phí: Chi phí quản lý kho hiệu quả có thể thay đổi tùy thuộc vào quy mô kho hàng, mức độ tự động hóa và phần mềm sử dụng.

Tóm lại, mỗi giải pháp đều có một mức độ khả thi về thời gian và chi phí nhất định. Để chọn ra phương án triển khai phù hợp cần nhiều nghiên cứu và đánh giá thấu đáo từ nhà quản trị doanh nghiệp.

6. KẾT LUẬN

6.1. Vấn đề đạt được

Trong quá trình hoàn thiện bài đồ án, nhóm đã hiểu rõ được các quy trình phân tích đối với từng loại thông tin của dữ liệu. Đồ án lần này giúp nhóm nâng cao khả năng tự tìm hiểu, tự nghiên cứu và sử dụng thành thạo các phần mềm xử lý dữ liệu, phân tích dữ liệu dựa trên các dashboard, nắm rõ được insights cũng như các vấn đề mà chuỗi cửa hàng cà phê đang gặp phải.

Xuyên suốt thời gian thực hiện bài đồ án, nhóm đã thực hành áp dụng được các kiến thức cũng như công cụ để phục vụ cho đồ án. Nhóm đã hiểu hơn về bộ dữ liệu thông qua việc phân tích mô tả thống kê dữ liệu bằng công cụ Microsoft Excel. Đồng thời, đồ án lần này của nhóm cũng sử dụng công cụ Power BI và các hàm DAX chuyên dụng để tạo nên những dashboard trực quan nhất. Từ những dashboard này, nhóm cũng tiến hành phân tích sâu vào các khía cạnh vấn đề nhằm rút ra những thông tin quan trọng. Từ những phân tích đó, nhóm cũng đã đề xuất những chiến lược phù hợp cho vấn đề kinh doanh của chuỗi cửa hàng cà phê.

6.2. Vấn đề chưa đạt được

Trong quá trình thực hiện đồ án, nhóm đã gặp phải một số khó khăn thách thức do còn hạn chế về kiến thức, nguồn lực cũng như thời gian thực hiện. Những khó khăn này đã làm cho bài báo cáo chưa đạt đến độ hoàn thiện mong muốn. Bộ dữ liệu mà nhóm sử dụng là bộ dữ liệu hư cấu về một chuỗi cửa hàng cà phê, chưa

thể phản ánh chính xác thực tế, không thể thay thế hoàn toàn cho một doanh nghiệp hoặc công ty thực sự. Bên cạnh đó, những đề xuất mà nhóm đưa ra có thể vẫn còn nhiều yếu tố chủ quan và thiên về lý thuyết. Tuy nhiên, dù gặp phải nhiều hạn chế và thách thức, nhóm vẫn không ngừng cố gắng và học hỏi để hoàn thiện bản thân, nhằm hướng đến những kết quả tốt hơn trong các dự án sắp tới.

6.3. Hướng phát triển

Dựa trên những phân tích đã được thực hiện, nhóm đã nắm bắt được các kiến thức cần thiết cũng như biết cách ứng dụng BI vào để phân tích và đưa ra các giải pháp cho một chuỗi cửa hàng cà phê. Chính vì thế, nhóm tin rằng mình hoàn toàn có khả năng sử dụng một bộ dữ liệu thực tế của doanh nghiệp để phân tích tình hình kinh doanh và đưa ra các đề xuất chiến lược kinh doanh phù hợp giúp doanh nghiệp ra quyết định nhằm cải thiện hiệu suất kinh doanh của doanh nghiệp trong tương lai. Hơn nữa, đồ án phân tích của nhóm còn có tiềm năng để phát triển và hoàn thiện hơn trong tương lai, bằng cách phân tích chuyên sâu và tiến hành thực hiện các chiến lược đã đề xuất ở chương 5 một cách khách quan và thực tế hơn để áp dụng vào thực tiễn.

TÀI LIỆU THAM KHẢO

1. Phạm Thị Thanh Tâm. Tài liệu giảng dạy môn Hệ hỗ trợ quản trị thông minh.
2. Vu, U. (2023, July 31). Branded Merchandise Là Gì? Xu Hướng Bộ Sưu Tập Vật Phẩm Của Starbucks, Cheese Coffee, Highlands Coffee, Mixue,... | FnB Việt Nam. Fnb Vietnam. Retrieved May 3, 2024, from https://fnbvietnam.vn/branded-merchandise-fnb/#2_Loi_ich_cua_Branded_Merchandise_trong_viec_tiep_thi_t_huong_hieu
3. Wise, J. (2014, March 13). Coffee Statistics 2024: Demographics, Industry Size in the US & Worldwide. EarthWeb. Retrieved May 5, 2024, from <https://earthweb.com/coffee-statistics/>