

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC KINH TẾ - TÀI CHÍNH THÀNH PHỐ HỒ CHÍ MINH



BÁO CÁO CUỐI KỲ
MÔN THỐNG KÊ MÁY TÍNH
ĐỀ TÀI: PHÂN TÍCH PHÂN KHÚC KHÁCH HÀNG
DỰA TRÊN MÔ HÌNH RFM TRONG THƯƠNG
MẠI ĐIỆN TỬ

Giảng viên hướng dẫn: Ngô Văn Công Bằng

Năm học 2025-2026: Học kì: 1B

Lớp: 251.DAS1101.B01

Nhóm 06

Thành phố Hồ Chí Minh, ngày 13 tháng 1 năm 2026

BÁO CÁO CUỐI KỲ

MÔN THỐNG KÊ MÁY TÍNH

Tên đề tài: Phân tích phân khúc khách hàng dựa trên mô hình rfm trong thương mại điện tử.

Lớp: 251.DAS1101.B01

Ngày nộp báo cáo: 13/01/2026

Nhóm: 06

Thành viên nhóm:

- Hồ Văn Như Ý
- Ngô Minh Triết
- Võ Trí Quyền



UEF

ĐẠI HỌC KINH TẾ TÀI CHÍNH

Bảng 1: Phân công nhiệm vụ

| Thành viên | Nhiệm vụ được phân công | Đánh giá mức độ hoàn thành nhiệm vụ |
|----------------|---|-------------------------------------|
| Hồ Văn Như Ý | <ul style="list-style-type: none">• Thu thập dữ liệu• Làm sạch dữ liệu, mã hóa dữ liệu• Thực hiện kiểm định One - Sample Test và Independent T - Test• Viết kết luận | 100% |
| Ngô Minh Triết | <ul style="list-style-type: none">• Thu thập dữ liệu• Mã hóa các biến• Thực hiện kiểm định Chi - Square, One sample T-test và kiểm định one way ANOVA | 100% |
| Võ Trí Quyền | <ul style="list-style-type: none">• Thu thập dữ liệu• Thực hiện thống kê mô tả• Chiến lược• Viết phần mở đầu, kết luận | 100% |

MỤC LỤC

| | |
|--|----|
| CHƯƠNG I. TÓM TẮT | 7 |
| 1.1 Đề tài..... | 7 |
| 1.2 Mục tiêu | 7 |
| 1.3 Phương pháp nghiên cứu | 7 |
| 1.4 Kết quả nghiên cứu | 8 |
| CHƯƠNG II. ĐẶT VẤN ĐỀ (MỞ ĐẦU) | 9 |
| CHƯƠNG III. PHƯƠNG PHÁP THỰC HIỆN | 10 |
| 3.1 Đối tượng nghiên cứu | 10 |
| 3.2 Mục tiêu nghiên cứu | 10 |
| 3.3 Phương pháp nghiên cứu | 10 |
| 3.4 Quy trình xử lý dữ liệu..... | 11 |
| CHƯƠNG IV. THỐNG KÊ..... | 12 |
| 4.1 Thống kê mô tả | 12 |
| CHƯƠNG V. KIỂM ĐỊNH | 13 |
| 5.1 Phân tích RFM..... | 13 |
| 5.2 Kiểm định mối liên hệ giữa các biến (Chi-Square Test) | 14 |
| 5.3 One Sample T-Test | 19 |
| 5.3.1. Kiểm định giá trị tiêu trung bình của các giao dịch: | 19 |
| 5.3.2. Kiểm định số ngày tính từ lần cuối phát sinh giao dịch của khách hàng | 20 |
| 5.3.3. Kiểm định tần suất phát sinh giao dịch của khách hàng..... | 21 |
| 5.3.4. Kiểm định giá trị trung bình thời gian mua hàng gần nhất (Recency) của khách hàng | 23 |
| 5.3.5. Kiểm định tần suất mua hàng trung bình của khách hàng trên sàn thương mại điện tử | 24 |

| | |
|--|----|
| 5.3.6. Kiểm định mức chi tiêu trung bình (monetary_value) của nhóm khách hàng có rfm_score ≥ 6 và rfm_score < 6 | 25 |
| 5.3.7. Kiểm định thời gian mua gần nhất giữa hai nhóm khách hàng có điểm RFM khác nhau (rfm_score ≥ 6 và rfm_score < 6). | 27 |
| 5.3.8. Kiểm định tần suất mua hàng của hai nhóm khách hàng có điểm RFM khác nhau (rfm_score ≥ 6 và rfm_score < 6). | 28 |
| 5.4 One-way ANOVA | 30 |
| CHƯƠNG VI. KẾT LUẬN | 33 |



UEF

ĐẠI HỌC KINH TẾ TÀI CHÍNH

MỤC LỤC BẢNG

| | |
|--|----|
| Bảng 1: Phân công nhiệm vụ | 3 |
| Bảng 2: customer_segment..... | 12 |
| Bảng 3: Report_monetary_value | 13 |
| Bảng 4: Case Processing Summary | 14 |
| Bảng 5: Chi-Square Tests | 15 |
| Bảng 6: F * M Crosstabulation | 16 |
| Bảng 7: Chi-Square Tests | 17 |
| Bảng 8: Case Processing Summare..... | 18 |
| Bảng 9: Chi-Square Tests | 19 |
| Bảng 10: One - Sample Test monetary_value | 20 |
| Bảng 11: One - Sample Test recency..... | 21 |
| Bảng 12: One - Sample Test frequency..... | 22 |
| Bảng 13: One - Sample Test recency..... | 23 |
| Bảng 14: One - Sample Statistics recency | 23 |
| Bảng 15: One - Sample Test frequency | 24 |
| Bảng 16: One - Sample Statistics frequency | 24 |
| Bảng 17: Independent Samples Test monetary_value..... | 25 |
| Bảng 18: Group Statistics monetary_value | 25 |
| Bảng 19: Independent Samples Effect Sizes monatary_value | 26 |
| Bảng 20: Independent Sample Effect Sizes recency | 27 |
| Bảng 21: Independent Sample Test recency..... | 27 |
| Bảng 22: Group Statistics recency..... | 27 |
| Bảng 23: Group Statistics frequency | 29 |
| Bảng 24: Independent Samples Effect Sizes frequency | 29 |
| Bảng 25: Independent Samples Test frequency..... | 29 |
| Bảng 26: One - way ANOVA (recency - frequency - monetary_value)..... | 31 |
| Bảng 27: Tests of Homogeneity of Variances (recency - frequency - monetary_value)..... | 32 |
| Bảng 28: Robust Tests of Equality of Means (recency - frequency - monetary_value)..... | 32 |

CHƯƠNG I. TÓM TẮT

1.1 Đề tài

Phân tích phân khúc khách hàng dựa trên mô hình RFM trong thương mại điện tử.

1.2 Mục tiêu

a. Mục tiêu tổng quát:

Phân tích hành vi mua sắm và phân loại khách hàng trên nền tảng thương mại điện tử dựa trên mô hình RFM (Recency – Frequency – Monetary), từ đó làm cơ sở để đề xuất các chiến lược kinh doanh và marketing phù hợp cho doanh nghiệp.

b. Mục tiêu cụ thể:

- **Áp dụng mô hình RFM:** Sử dụng các chỉ số về thời gian mua hàng gần nhất (Recency), tần suất mua hàng (Frequency) và giá trị chi tiêu (Monetary) để tính toán điểm RFM cho từng khách hàng.
- **Phân khúc nhóm khách hàng:** Dựa trên điểm số RFM để phân chia tập khách hàng thành các nhóm riêng biệt như: nhóm khách hàng giá trị cao (VIP), nhóm khách hàng tiềm năng và nhóm khách hàng có nguy cơ rời bỏ hoặc ít hoạt động.
- **Kiểm định sự khác biệt giữa các phân khúc:** Sử dụng các phương pháp thống kê (như Chi-Square, T-Test, One-way ANOVA) để chứng minh sự khác biệt có ý nghĩa thống kê về hành vi và giá trị giữa các nhóm khách hàng khác nhau.
- **Phân tích đặc điểm hành vi:** Làm rõ các đặc điểm cụ thể của từng phân khúc, chẳng hạn như mức chi tiêu trung bình, thời gian quay lại mua hàng và tần suất giao dịch thực tế.
- **Đề xuất chiến lược kinh doanh:** Xây dựng các giải pháp cụ thể nhằm giữ chân khách hàng VIP, khai thác nhóm tiềm năng và tối ưu hóa chi phí đối với nhóm khách hàng ít hoạt động.

1.3 Phương pháp nghiên cứu

- Sử dụng mô hình RFM, gồm có ba chỉ số chính:
 - o **Recency (R):** Đo lường số ngày tính từ lần cuối cùng khách hàng phát sinh giao dịch.
 - o **Frequency (F):** Đo lường tần suất phát sinh giao dịch của khách hàng.
 - o **Monetary (M):** Đo lường giá trị chi tiêu trung bình hoặc tổng số tiền khách hàng đã bỏ ra.

- **Thống kê mô tả:** Dùng để tóm tắt dữ liệu, xác định tỷ lệ phần trăm của từng phân khúc khách hàng.
- **Kiểm định Chi-Square (Chi bình phương):** Sử dụng để kiểm định mối liên hệ giữa các biến định tính, chẳng hạn như mối quan hệ giữa số ngày mua hàng gần nhất và tần suất mua hàng, hoặc giữa tần suất và giá trị chi tiêu.
- **Kiểm định One-Sample T-Test:** Dùng để so sánh giá trị trung bình của tập dữ liệu (về mức chi tiêu, thời gian quay lại, hoặc tần suất) với một giá trị kiểm định cụ thể để xem có sự khác biệt ý nghĩa hay không.
- **Kiểm định Independent Samples T-Test:** Dùng để so sánh sự khác biệt về hành vi (chi tiêu, tần suất, thời gian mua) giữa hai nhóm khách hàng có điểm RFM cao ($\text{rfm_score} \geq 6$) và điểm RFM thấp ($\text{rfm_score} < 6$). Trong đó, **kiểm định Levene** cũng được sử dụng để xem xét tính đồng nhất của phương sai trước khi thực hiện T-Test.
- **Phân tích One-way ANOVA và Welch Test:** Được áp dụng để kiểm định sự khác biệt về giá trị giữa các trường của các biến. Trong trường hợp phương sai không đồng nhất, nhóm đã sử dụng **Welch Test** (Robust Tests of Equality of Means) để thay thế cho ANOVA truyền thống.

1.4 Kết quả nghiên cứu

Xác định được nhóm khách hàng giá trị cao, tiềm năng và nhóm có nguy cơ rời bỏ.

ĐẠI HỌC KINH TẾ TÀI CHÍNH

CHƯƠNG II. ĐẶT VẤN ĐỀ (MỞ ĐẦU)

Trong bối cảnh thương mại điện tử đang phát triển mạnh mẽ, việc thấu hiểu sâu sắc hành vi và giá trị của khách hàng đóng vai trò sống còn trong việc xây dựng các chiến lược kinh doanh hiệu quả. Đối với các doanh nghiệp hiện nay, thách thức không chỉ nằm ở việc thu hút khách hàng mới mà còn là làm thế nào để duy trì và phát triển mối quan hệ bền vững với tệp khách hàng hiện hữu.

Tuy nhiên, thực tế cho thấy không phải tất cả khách hàng đều có hành vi mua sắm giống nhau. Sự khác biệt về nhu cầu và thói quen đòi hỏi doanh nghiệp phải thực hiện phân khúc khách hàng một cách khoa học. Việc phân khúc này giúp doanh nghiệp xác định đúng đối tượng mục tiêu, từ đó tối ưu hóa hoạt động marketing và cá nhân hóa quy trình chăm sóc khách hàng.

Trong số các kỹ thuật phân tích hiện đại, mô hình RFM (Recency – Frequency – Monetary) được xem là một trong những phương pháp phổ biến và hiệu quả nhất để đánh giá định lượng hành vi và giá trị của khách hàng. Xuất phát từ thực tiễn kinh doanh đó, đề tài “Phân tích phân khúc khách hàng dựa trên mô hình RFM trong thương mại điện tử” được thực hiện nhằm ứng dụng các phương pháp thống kê máy tính và công cụ phân tích dữ liệu chuyên sâu.

ĐẠI HỌC KINH TẾ TÀI CHÍNH

CHƯƠNG III. PHƯƠNG PHÁP THỰC HIỆN

3.1 Đối tượng nghiên cứu

Nghiên cứu tập trung vào dữ liệu khách hàng mua sắm trên nền tảng thương mại điện tử. Mục tiêu là phân tích hành vi và phân loại tập khách hàng này thành các nhóm riêng biệt để phục vụ chiến lược kinh doanh.

3.2 Mục tiêu nghiên cứu

Phân tích hành vi mua sắm và phân khúc khách hàng theo mô hình RFM:

- **Recency (R):** Số ngày tính từ lần cuối cùng khách hàng phát sinh giao dịch.
- **Frequency (F):** Tần suất (số lần) phát sinh giao dịch của khách hàng.
- **Monetary (M):** Giá trị chi tiêu (số tiền) của khách hàng trong các giao dịch.

3.3 Phương pháp nghiên cứu

- **Thống kê mô tả:** Xác định tỷ lệ phần trăm và đặc điểm cơ bản của từng phân khúc khách hàng.
- **Kiểm định Chi-Square (Chi bình phương):** Sử dụng để kiểm tra mối liên hệ giữa các biến định tính, cụ thể là mối quan hệ giữa Recency, Frequency và Monetary.
- **Kiểm định T-Test:**
 - *One-Sample T-Test:* So sánh các giá trị trung bình về chi tiêu, tần suất và thời gian mua hàng của mẫu với các giá trị định mức để thấy sự khác biệt.
 - *Independent Samples T-Test:* So sánh sự khác biệt về hành vi giữa hai nhóm khách hàng có điểm RFM cao (≥ 6) và điểm RFM thấp (< 6).
- **Kiểm định Levene:** Dùng để kiểm tra tính đồng nhất của phương sai trước khi thực hiện các so sánh trung bình.
- **Phân tích ANOVA và Welch Test:** Sử dụng One-way ANOVA để kiểm định sự khác biệt giữa các trường của các biến. Trong trường hợp phương sai không đồng nhất, **Welch Test** (Robust Tests of Equality of Means) được sử dụng để thay thế nhằm đảm bảo kết quả chính xác.
- **Đo lường mức độ ảnh hưởng (Effect Size):** Sử dụng các chỉ số như **Cohen's d** và **Glass's delta** để đánh giá quy mô sự khác biệt giữa các nhóm khách hàng trong thực tế kinh doanh.

3.4 Quy trình xử lý dữ liệu

Làm sạch và mã hóa dữ liệu: Dữ liệu thô được xử lý để loại bỏ sai sót và mã hóa các biến số.

Công cụ sử dụng: Nhóm nghiên cứu kết hợp giữa ngôn ngữ lập trình Python (dành cho giai đoạn làm sạch, xử lý dữ liệu) và phần mềm SPSS (dành cho các phân tích thống kê máy tính và kiểm định giả thuyết).



CHƯƠNG IV. THỐNG KÊ

4.1 Thống kê mô tả

| customer_segment | | | | | |
|------------------|-------|-----------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 3 | 985 | 22.8 | 22.8 | 22.8 |
| | 1 | 727 | 16.8 | 16.8 | 39.6 |
| | 2 | 731 | 16.9 | 16.9 | 56.6 |
| | 4 | 703 | 16.3 | 16.3 | 72.8 |
| | 0 | 1174 | 27.2 | 27.2 | 100.0 |
| | Total | 4320 | 100.0 | 100.0 | |

Bảng 2: customer_segment

Segment 0 (Unsteady customers) (27%): đông nhất → nhiều khách ít giá trị / không hoạt động.

Segment 3 (Active customer) (22.6%): nhóm khá lớn → có tiềm năng.

Segment 4 (Top-performing customers) (16.2%): ít hơn nhưng thường là nhóm VIP.

Kết luận: Kết quả thống kê tần suất cho thấy phân khúc khách hàng 0 chiếm tỷ lệ cao nhất với 27%, phản ánh một bộ phận lớn khách hàng có mức độ hoạt động thấp hoặc giá trị không cao. Phân khúc 3 chiếm 22,6%, cho thấy đây là nhóm khách hàng khá phổ biến và có tiềm năng phát triển. Trong khi đó, phân khúc 4 chỉ chiếm 16,2% tổng số khách hàng nhưng thường là nhóm có giá trị cao, đóng góp lớn vào doanh thu. Sự phân bố không đồng đều giữa các phân khúc cho thấy hành vi và giá trị khách hàng có sự khác biệt rõ rệt.

CHƯƠNG V. KIỂM ĐỊNH

5.1 Phân tích RFM

| Report | | | |
|------------------|-------------|------|----------------|
| monetary_value | | | |
| customer_segment | Mean | N | Std. Deviation |
| 3 | 1954.860051 | 985 | 2636.339400 |
| 1 | 383.597455 | 727 | 236.6498295 |
| 2 | 197.947059 | 731 | 91.6316876 |
| 4 | 7149.436230 | 703 | 19383.95980 |
| 0 | 761.209455 | 1174 | 1438.857470 |
| Total | 1914.079947 | 4320 | 8303.245478 |

Bảng 3: Report_monetary_value

Các giá trị trung bình (Mean) của ba biến RFM chính giúp hình dung bức tranh tổng thể về thói quen mua sắm của khách hàng:

- **Mức chi tiêu (Monetary):** Trung bình mỗi giao dịch của khách hàng đạt **1,914.08** đô.
- Tuy nhiên, có sự phân hóa mạnh khi Segment 4 chi tiêu tới **7,149**, trong khi Segment 2 chỉ chi tiêu trung bình **198**.
- **Thời gian mua hàng gần nhất (Recency):** Khoảng thời gian trung bình kể từ lần cuối phát sinh giao dịch là **91.6 ngày** (xấp xỉ 3 tháng)
- Độ lệch chuẩn của biến này rất lớn (**99.615**), cho thấy hành vi khách hàng không đồng nhất: có những người mua hàng rất thường xuyên nhưng cũng có nhiều khách hàng đã lâu không quay lại.
- **Tần suất mua hàng (Frequency):** Tần suất giao dịch trung bình của khách hàng đạt **4.89 lần** (xấp xỉ 5 lần).

Kết quả so sánh giá trị chi tiêu trung bình theo phân khúc cho thấy phân khúc 4 có mức monetary_value cao nhất với trung bình 7.149, chứng tỏ đây là nhóm khách hàng giá trị cao, đóng góp lớn vào doanh thu. Phân khúc 3 đứng thứ hai với mức chi tiêu trung bình 1.954, là nhóm khách hàng tiềm năng cần được khai thác và phát triển. Ngược lại, phân khúc 2 có mức chi tiêu trung bình thấp nhất (198), phản ánh nhóm khách hàng ít giá trị và mức độ đóng góp thấp cho doanh nghiệp.

5.2 Kiểm định mối liên hệ giữa các biến (Chi-Square Test)

Case Processing Summary

| | Valid | | Missing | | Total | |
|-------|-------|---------|---------|---------|-------|---------|
| | N | Percent | N | Percent | N | Percent |
| R * F | 4320 | 100.0% | 0 | 0.0% | 4320 | 100.0% |

R * F Crosstabulation

| | | | F | | | |
|-------|------------|------------|--------|--------|--------|--------|
| | | | 1 | 2 | 3 | Total |
| R | 1 | Count | 1093 | 217 | 105 | 1415 |
| | | % within R | 77.2% | 15.3% | 7.4% | 100.0% |
| | | % within F | 50.8% | 25.1% | 8.1% | 32.8% |
| | | % of Total | 25.3% | 5.0% | 2.4% | 32.8% |
| | 2 | Count | 703 | 328 | 393 | 1424 |
| | | % within R | 49.4% | 23.0% | 27.6% | 100.0% |
| | | % within F | 32.7% | 37.9% | 30.2% | 33.0% |
| | | % of Total | 16.3% | 7.6% | 9.1% | 33.0% |
| | 3 | Count | 355 | 321 | 805 | 1481 |
| | | % within R | 24.0% | 21.7% | 54.4% | 100.0% |
| | | % within F | 16.5% | 37.1% | 61.8% | 34.3% |
| | | % of Total | 8.2% | 7.4% | 18.6% | 34.3% |
| Total | Count | 2151 | 866 | 1303 | 4320 | |
| | % within R | 49.8% | 20.0% | 30.2% | 100.0% | |
| | % within F | 100.0% | 100.0% | 100.0% | 100.0% | |
| | % of Total | 49.8% | 20.0% | 30.2% | 100.0% | |

Bảng 4: Case Processing Summary

| Chi-Square Tests | | | |
|------------------------------|----------------------|----|-----------------------------------|
| | Value | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 969.618 ^a | 4 | <.001 |
| Likelihood Ratio | 1037.786 | 4 | <.001 |
| Linear-by-Linear Association | 955.373 | 1 | <.001 |
| N of Valid Cases | 4320 | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 283.66.

Bảng 5: Chi-Square Tests

Giả thuyết:

- H0: Số ngày tính từ lần cuối mua hàng độc lập với tần suất mua hàng.
- H1: Số ngày tính từ lần cuối mua hàng có liên hệ với tần suất mua hàng.

Ta thấy kết quả kiểm định tỷ lệ ô có tần suất mong đợi thấp, không lớn hơn 20% nên kết quả có ý nghĩa. Với giá trị Sig (<0.001) <0.05 , ta bác bỏ H0 và chấp nhận H1, cho thấy sự liên hệ mật thiết giữa số ngày tính từ lần cuối mua hàng (recency) và tần suất mua hàng (frequency).

ĐẠI HỌC KINH TẾ TÀI CHÍNH

F * M Crosstabulation

| | | | M | | | Total |
|-------|---|------------|--------|--------|--------|--------|
| | | | 1 | 2 | 3 | |
| F | 1 | Count | 1310 | 730 | 111 | 2151 |
| | | % within F | 60.9% | 33.9% | 5.2% | 100.0% |
| | | % within M | 92.3% | 50.3% | 7.7% | 49.8% |
| | | % of Total | 30.3% | 16.9% | 2.6% | 49.8% |
| | 2 | Count | 99 | 508 | 259 | 866 |
| | | % within F | 11.4% | 58.7% | 29.9% | 100.0% |
| | | % within M | 7.0% | 35.0% | 17.9% | 20.0% |
| | | % of Total | 2.3% | 11.8% | 6.0% | 20.0% |
| | 3 | Count | 11 | 212 | 1080 | 1303 |
| | | % within F | 0.8% | 16.3% | 82.9% | 100.0% |
| | | % within M | 0.8% | 14.6% | 74.5% | 30.2% |
| | | % of Total | 0.3% | 4.9% | 25.0% | 30.2% |
| Total | | Count | 1420 | 1450 | 1450 | 4320 |
| | | % within F | 32.9% | 33.6% | 33.6% | 100.0% |
| | | % within M | 100.0% | 100.0% | 100.0% | 100.0% |
| | | % of Total | 32.9% | 33.6% | 33.6% | 100.0% |

Bảng 6: F * M Crosstabulation

| Chi-Square Tests | | | |
|------------------------------|-----------------------|----|-----------------------------------|
| | Value | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 2785.383 ^a | 4 | .000 |
| Likelihood Ratio | 3079.523 | 4 | .000 |
| Linear-by-Linear Association | 2367.010 | 1 | .000 |
| N of Valid Cases | 4320 | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 284.66.

Bảng 7: Chi-Square Tests

Giả thuyết:

- H0: Tần suất mua hàng độc lập với giá trị tiêu.
- H1: Tần suất mua hàng có liên hệ với giá trị tiêu.

Với kết quả kiểm định là 0% số ô có giá trị mong đợi dưới 5 (<20%), kết quả có ý nghĩa và ta tiến hành kiểm định. Giá trị Sig. thu được là $0 < 0.05$, nên ta bác bỏ H0, thừa nhận mối liên hệ giữa tần suất mua hàng và giá trị tiêu.

ĐẠI HỌC KINH TẾ TÀI CHÍNH

Case Processing Summary

| | Cases | | | | | |
|-------|-------|---------|---------|---------|-------|---------|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| R * M | 4320 | 100.0% | 0 | 0.0% | 4320 | 100.0% |

R * M Crosstabulation

| | | | M | | | Total |
|---|-------|------------|--------|--------|--------|--------|
| | | | 1 | 2 | 3 | |
| R | 1 | Count | 780 | 472 | 163 | 1415 |
| | | % within R | 55.1% | 33.4% | 11.5% | 100.0% |
| | | % within M | 54.9% | 32.6% | 11.2% | 32.8% |
| | | % of Total | 18.1% | 10.9% | 3.8% | 32.8% |
| | 2 | Count | 408 | 551 | 465 | 1424 |
| | | % within R | 28.7% | 38.7% | 32.7% | 100.0% |
| | | % within M | 28.7% | 38.0% | 32.1% | 33.0% |
| | | % of Total | 9.4% | 12.8% | 10.8% | 33.0% |
| | 3 | Count | 232 | 427 | 822 | 1481 |
| | | % within R | 15.7% | 28.8% | 55.5% | 100.0% |
| | | % within M | 16.3% | 29.4% | 56.7% | 34.3% |
| | | % of Total | 5.4% | 9.9% | 19.0% | 34.3% |
| | Total | Count | 1420 | 1450 | 1450 | 4320 |
| | | % within R | 32.9% | 33.6% | 33.6% | 100.0% |
| | | % within M | 100.0% | 100.0% | 100.0% | 100.0% |
| | | % of Total | 32.9% | 33.6% | 33.6% | 100.0% |

Bảng 8: Case Processing Summare

Chi-Square Tests

| | Value | df | Asymptotic Significance (2-sided) |
|---------------------------------|----------------------|----|---|
| Pearson Chi-Square | 792.921 ^a | 4 | <.001 |
| Likelihood Ratio | 824.250 | 4 | <.001 |
| Linear-by-Linear Association | 756.781 | 1 | <.001 |
| N of Valid Cases | 4320 | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 465.12.

Bảng 9: Chi-Square Tests

Giả thuyết:

- H0: Số ngày tính từ lần cuối mua hàng độc lập với giá trị tiêu.
- H1: Số ngày tính từ lần cuối mua hàng có liên hệ với giá trị tiêu.
Có 0% số ô có giá trị mong đợi < 5, nên kết quả kiểm định có ý nghĩa.

Với giá trị Sig. thu được là <0.001 (<0.05) nên ta bác bỏ H0, thừa nhận mối liên hệ giữa recency và monetary value.

5.3 One Sample T-Test

5.3.1. Kiểm định giá trị tiêu trung bình của các giao dịch:

Nghiên cứu thực hiện kiểm định One – Sample T – Test để đánh giá liệu mức chi tiêu trung bình của khách hàng trên sàn thương mại điện tử có tương đương với mức kỳ vọng đặt ra hay không.

Giả thuyết :

- H0: Giá trị tiêu trung bình của khách hàng trong từng giao dịch là 1500 đô.

| One-Sample Statistics | | | | | |
|-----------------------|------|-----------|----------------|-----------------|--|
| | N | Mean | Std. Deviation | Std. Error Mean | |
| monetary_value | 4320 | 1914.0799 | 8303.24548 | 126.32986 | |

| One-Sample Test | | | | | |
|-------------------|-------|------|-----------------|-----------------|---|
| Test Value = 1500 | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference |
| | | | | | Lower Upper |
| monetary_value | 3.278 | 4319 | .001 | 414.07995 | 166.4086 661.7513 |

| One-Sample Effect Sizes | | | | | |
|-------------------------|--------------------|---------------------------|----------------|-------------------------|-------|
| | | Standardizer ^a | Point Estimate | 95% Confidence Interval | |
| | | | | Lower | Upper |
| monetary_value | Cohen's d | 8303.24548 | .050 | .020 | .080 |
| | Hedges' correction | 8304.68769 | .050 | .020 | .080 |

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Bảng 10: One - Sample Test monetary_value

Nhận xét:

- Mức chi tiêu trung bình thực tế được tính toàn từ tập dữ liệu là 1914.08 đô.
Nhận thấy Sig. = 0.001 < 0.05, nên ta bác bỏ H0. Chứng minh mức chi tiêu trung bình của khách hàng có sự khác biệt so với ngưỡng 1500 đô với mức ý nghĩa 5%. Thực tế, mức chi tiêu của khách hàng cao hơn đáng kể so với giá trị giả thuyết ban đầu (Mean Difference = 414.07995)

5.3.2. Kiểm định số ngày tính từ lần cuối phát sinh giao dịch của khách hàng
Giả thuyết:

- H0: Trung bình khách hàng sẽ không phát sinh giao dịch 30 ngày.

| One-Sample Statistics | | | | | |
|-----------------------|------|-------|----------------|-----------------|--|
| | N | Mean | Std. Deviation | Std. Error Mean | |
| recency | 4320 | 91.60 | 99.615 | 1.516 | |

| One-Sample Test | | | | | |
|-----------------|--------|------|-----------------|-----------------|---|
| Test Value = 30 | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference |
| | | | | | Lower Upper |
| recency | 40.641 | 4319 | <.001 | 61.596 | 58.62 64.57 |

| One-Sample Effect Sizes | | | | | |
|-------------------------|--------------------|---------------------------|----------------|-------------------------|-------|
| | | Standardizer ^a | Point Estimate | 95% Confidence Interval | |
| | | | | Lower | Upper |
| recency | Cohen's d | 99.615 | .618 | .586 | .651 |
| | Hedges' correction | 99.632 | .618 | .586 | .651 |

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Bảng 11: One - Sample Test recency

Nhận xét:

- Số ngày trung bình kể từ lần cuối phát sinh giao dịch thực tế từ dữ liệu là 91.6 ngày.
Giá trị Sig. = $<0.001 < 0.05$ nên ta bác bỏ H0. Chấp nhận khách hàng sẽ không phát sinh giao dịch vào khoảng thời gian khác 30 ngày với mức ý nghĩa 5%.
- Kết quả cho thấy thực tế, khoảng thời gian khách hàng quay lại mua hàng lâu hơn nhiều so với giả thuyết 30 ngày ban đầu. Giá trị trung bình (Mean = 91.60) cũng được ghi nhận là có độ lệch chuẩn rất lớn (Std. Deviation = 99.615), cho thấy hành vi của các khách hàng của doanh nghiệp không đồng nhất, có những khách hàng quay lại mua rất nhanh những cũng có nhiều khách hàng đã “ngủ quên” trong thời gian dài.

5.3.3. Kiểm định tần suất phát sinh giao dịch của khách hàng

Giả thuyết:

- H_0 : Tần suất phát sinh giao dịch trung bình khách hàng bằng 5 lần.

| One-Sample Statistics | | | | |
|-----------------------|------|------|----------------|-----------------|
| | N | Mean | Std. Deviation | Std. Error Mean |
| frequency | 4320 | 4.89 | 8.937 | .136 |

| One-Sample Test | | | | | | |
|-----------------|-------|------|-----------------|-----------------|---|-------|
| Test Value = 5 | | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| frequency | -.822 | 4319 | .411 | -.112 | -.38 | .15 |

| One-Sample Effect Sizes | | | | | |
|-------------------------|--------------------|---------------------------|----------------|-------------------------|-------|
| | | Standardizer ^a | Point Estimate | 95% Confidence Interval | |
| | | | | Lower | Upper |
| frequency | Cohen's d | 8.937 | -.013 | -.042 | .017 |
| | Hedges' correction | 8.938 | -.013 | -.042 | .017 |

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Bảng 12: One - Sample Test frequency

Nhận xét:

- Giá trị trung bình tần suất phát sinh giao dịch kiểm định được là 4.89 lần. Với giá trị Sig. là $<0.001 < 0.05$, ta bác bỏ H_0 . Với mức ý nghĩa 5%, tần suất phát sinh giao dịch trung bình của khách hàng đã giao dịch ít hoặc hơn 5 lần.

ĐẠI HỌC KINH TẾ TÀI CHÍNH

5.3.4. Kiểm định giá trị trung bình thời gian mua hàng gần nhất (Recency) của khách hàng

Giả thuyết:

- H_0 : Giá trị trung bình (Mean) về thời gian mua hàng gần nhất (Recency) của khách hàng **bằng** 90 ngày.
- H_1 : Giá trị trung bình (Mean) về thời gian mua hàng gần nhất (Recency) của khách hàng **khác** 90 ngày.

| One-Sample Statistics | | | | |
|-----------------------|------|-------|----------------|-----------------|
| | N | Mean | Std. Deviation | Std. Error Mean |
| recency | 4320 | 91.60 | 99.615 | 1.516 |

Bảng 14: One - Sample Statistics recency

| One-Sample Test | | | | | |
|-----------------|-------|------|-----------------|-----------------|---|
| Test Value = 90 | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference |
| | | | | | Lower Upper |
| recency | 1.053 | 4319 | .292 | 1.596 | -1.38 4.57 |

Bảng 13: One - Sample Test recency

Nhận xét:

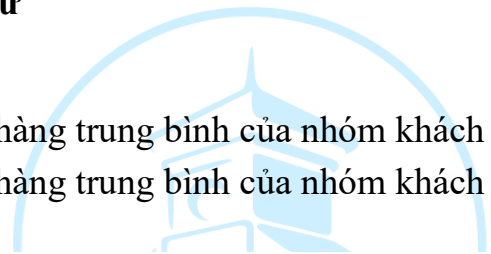
- Giá trị trung bình về thời gian mua gần nhất của các khách hàng trong dữ liệu là 91.60 ngày.
- Nhận thấy Sig. (2-tailed) có giá trị là $0.292 > 0.05$ nên chưa đủ cơ sở bác bỏ giả thuyết H_0 . Điều này nghĩa là giá trị trung bình về thời gian mua gần nhất của các khách hàng khác 90 ngày với mức ý nghĩa là 5%.
- Giá trị trung bình về thời gian mua gần nhất thực tế cao hơn giá trị kiểm định (90 ngày) là 1.60 ngày (Mean Difference = 1.596) nhưng sự chênh lệch này không có ý nghĩa về mặt thống kê do khoảng tin cậy 95% đi từ âm đến dương và chứa giá trị 0 [-1.38,4.57].

- Dữ liệu recency có độ lệch chuẩn (Std. Deviation) = 99.615 cho thấy dữ liệu phân tán rất mạnh
- Vì thế, ta thấy, thời gian quay lại mua hàng của khách hàng khoảng 3 tháng. Hành vi của khách hàng không đồng nhất: có những người vừa mới mua gần đây nhưng cũng có những khách hàng “ngủ quên” làm kéo dẫn độ lệch chuẩn.

5.3.5. Kiểm định tần suất mua hàng trung bình của khách hàng trên sàn thương mại điện tử

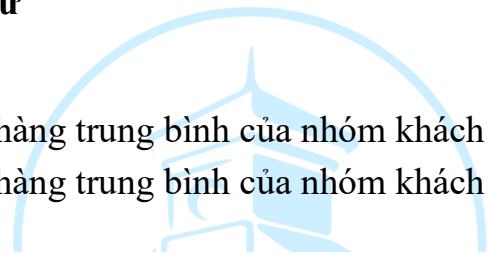
Giả thuyết:

- H_0 : Tần suất mua hàng trung bình của nhóm khách hàng **bằng** 5 lần.
- H_1 : Tần suất mua hàng trung bình của nhóm khách hàng **khác** 5 lần.



| One-Sample Statistics | | | | |
|-----------------------|------|------|----------------|-----------------|
| | N | Mean | Std. Deviation | Std. Error Mean |
| frequency | 4320 | 4.89 | 8.937 | .136 |

Bảng 16: One - Sample Statistics frequency



| One-Sample Test | | | | | | |
|-----------------|-------|------|-----------------|-----------------|---|--------------|
| Test Value = 5 | | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| frequency | -.822 | 4319 | .411 | -.112 | Lower -.38 | Upper .15 |

Bảng 15: One - Sample Test frequency

Nhận xét:

- Tần suất mua hàng trung bình của các khách hàng tại doanh nghiệp là 4.89 lần.
- Nhận thấy Sig. (2-tailed) = 0.411 > 0.05 nên chúng ta chấp nhận giả thuyết H_0 . Điều này cho thấy tần suất mua hàng trung bình của các khách hàng tại doanh nghiệp xấp xỉ 5 lần.

- Với Mean Difference là $-.112$ và khoảng tin cậy 95% $[-0.38, 0.15]$ có chứa giá trị 0, cho thấy sự khác biệt giữa hai số 4.89 và 5 là không có ý nghĩa thống kê
- Vì thế, ta nhận thấy, tần suất mua hàng trung bình của khách hàng đạt xấp xỉ 5 lần. Dù con số thực tế là 4.89 nhưng về mặt thống kê, con số này được xem là tương đương với cột mốc 5 lần.
- Điều này phản ánh hành vi của khách hàng không đồng nhất, doanh nghiệp đang có một tệp khách hàng trung thành đến giao dịch nhưng cũng có các tệp khách hàng đã lâu không quay lại. Nên những tệp khách hàng đó cần phải có chiến lược cụ thể, hợp lý để tiếp cận và dần chuyển các tệp khách hàng đó thành khách hàng trung thành của doanh nghiệp

5.3.6. Kiểm định mức chi tiêu trung bình (monetary_value) của nhóm khách hàng có rfm_score ≥ 6 và rfm_score < 6 .

Giả thuyết:

- H0-a: Phương sai về mức chi tiêu của nhóm khách hàng có rfm_score ≥ 6 và rfm_score < 6 là như nhau.
- H0-b: Mức chi tiêu trung bình (Monetary Value) của nhóm khách hàng có rfm_score ≥ 6 và rfm_score < 6 là như nhau.

Group Statistics

| | rfm_score | N | Mean | Std. Deviation | Std. Error Mean |
|----------------|-----------|------|-------------|----------------|-----------------|
| monetary_value | ≥ 6 | 2151 | 3454.053268 | 11559.53021 | 249.2412835 |
| | < 6 | 2169 | 386.886487 | 384.0792602 | 8.2469067 |

Bảng 18: Group Statistics monetary_value

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|----------------|-----------------------------|---|-------|------------------------------|----------|-----------------|-----------------|-----------------------|---|-------------|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| monetary_value | Equal variances assumed | 158.370 | <.001 | 12.351 | 4318 | <.001 | 3067.166781 | 248.3428074 | 2580.287348 | 3554.046215 |
| | Equal variances not assumed | | | 12.299 | 2154.708 | <.001 | 3067.166781 | 249.3776832 | 2578.120794 | 3556.212768 |

Bảng 17: Independent Samples Test monetary_value

Independent Samples Effect Sizes

| | | Standardizer ^a | Point Estimate | 95% Confidence Interval | |
|----------------|--------------------|---------------------------|----------------|-------------------------|-------|
| | | | | Lower | Upper |
| monetary_value | Cohen's d | 8161.306610 | .376 | .316 | .436 |
| | Hedges' correction | 8162.724502 | .376 | .316 | .436 |
| | Glass's delta | 384.0792602 | 7.986 | 7.741 | 8.231 |

a. The denominator used in estimating the effect sizes.

Cohen's d uses the pooled standard deviation.

Hedges' correction uses the pooled standard deviation, plus a correction factor.

Glass's delta uses the sample standard deviation of the control group.

Bảng 19: Independent Samples Effect Sizes monetary_value

Nhận xét:

- Kết quả kiểm định Levene: Sig. <.001 (nhỏ hơn 0.05), nên ta bác bỏ H0-a. Điều này có nghĩa là phương sai giữa hai nhóm khách hàng này có sự khác biệt đáng kể (phương sai không đồng nhất). Vì vậy, ta sử dụng kết quả kiểm định ở phần **Equal variances not assumed** (Giả định phương sai không bằng nhau).
- Kiểm định T-Test (phần Equal variances not assumed):** Nhận thấy giá trị **Sig. (2-tailed) < .001** (nhỏ hơn 0.05). Do đó, ta bác bỏ giả thuyết H0-b. Điều này chứng tỏ **có sự khác biệt có ý nghĩa thống kê** về mức chi tiêu trung bình giữa nhóm khách hàng có điểm RFM cao (rfm_score >= 6) và nhóm có điểm RFM thấp (rfm_score < 6).
- Có sự chênh lệch rất lớn về mức chi tiêu trung bình giữa hai nhóm (Mean Difference: **3067.167**). Cụ thể, nhóm có điểm RFM cao chi tiêu trung bình khoảng **3454.05**, trong khi nhóm điểm thấp chỉ chi tiêu trung bình **386.89**.
- Chỉ số **Cohen's d = 0.376** cho thấy mức độ ảnh hưởng (effect size) ở mức trung bình. Nhưng khi xét tới giá trị **Glass's delta** lên tới **7.986** cho thấy khi

so sánh dựa trên độ lệch chuẩn của nhóm đối chiếu, thì sự khác biệt về mức chi tiêu giữa hai nhóm trên là cực kỳ lớn.

5.3.7. Kiểm định thời gian mua gần nhất giữa hai nhóm khách hàng có điểm RFM khác nhau (rfm_score \geq 6 và rfm_score $<$ 6).

Giả thuyết:

- H0-a: Phương sai về thời gian mua hàng gần nhất của nhóm khách hàng có rfm_score \geq 6 và rfm_score $<$ 6 là như nhau.
- H0-b: Thời gian mua hàng gần nhất trung bình (Recency) của nhóm khách hàng có rfm_score \geq 6 và rfm_score $<$ 6 là như nhau.

Group Statistics

| | rfm_score | N | Mean | Std. Deviation | Std. Error Mean |
|---------|-----------|------|--------|----------------|-----------------|
| recency | ≥ 6 | 2151 | 34.61 | 43.335 | .934 |
| | < 6 | 2169 | 148.11 | 107.184 | 2.301 |

Bảng 22: Group Statistics recency

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---------|-----------------------------|---|------|------------------------------|----------|-----------------|-----------------|-----------------------|---|----------|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| recency | Equal variances assumed | 2289.057 | .000 | -45.559 | 4318 | .000 | -113.503 | 2.491 | -118.387 | -108.619 |
| | Equal variances not assumed | | | -45.696 | 2863.164 | .000 | -113.503 | 2.484 | -118.373 | -108.633 |

Bảng 21: Independent Sample Test recency

Independent Samples Effect Sizes

| | | Standardizer ^a | Point Estimate | 95% Confidence Interval | |
|---------|--------------------|---------------------------|----------------|-------------------------|--------|
| | | | | Lower | Upper |
| recency | Cohen's d | 81.873 | -1.386 | -1.453 | -1.320 |
| | Hedges' correction | 81.887 | -1.386 | -1.452 | -1.320 |
| | Glass's delta | 107.184 | -1.059 | -1.126 | -.991 |

a. The denominator used in estimating the effect sizes.

Cohen's d uses the pooled standard deviation.

Hedges' correction uses the pooled standard deviation, plus a correction factor.

Glass's delta uses the sample standard deviation of the control group.

Bảng 20: Independent Sample Effect Sizes recency

Nhận xét:

- **Kiểm định Levene:** Giá trị **Sig. = .000 < 0.05**, bác bỏ H0-a. Vậy phương sai giữa hai nhóm điểm RFM có sự khác biệt rõ rệt, ta sẽ sử dụng kết quả kiểm định ở phần **Equal variances not assumed**.
- **Kiểm định T-Test (phần Equal variances not assumed):** Nhận thấy giá trị **Sig. (2-tailed) = .000 < 0.05** nên bác bỏ giả thuyết H0-b với mức ý nghĩa 5%. Điều này chứng tỏ **có sự khác biệt có ý nghĩa thống kê** về thời gian mua hàng gần nhất trung bình giữa nhóm khách hàng điểm cao và nhóm khách hàng điểm thấp.
- Có sự chênh lệch rất lớn về thời gian mua hàng gần nhất giữa hai nhóm (Mean Difference: **-113.503**). Cụ thể, nhóm có điểm RFM cao (>6) có thời gian quay lại mua hàng trung bình rất ngắn, chỉ khoảng **34.61 ngày**, trong khi nhóm điểm thấp (< 6) lên tới **148.11 ngày**.
- Chỉ số **Cohen's d = -1.386** cho thấy mức độ ảnh hưởng (effect size) là **rất lớn**. Điều này khẳng định rằng điểm RFM là một chỉ số phân loại cực kỳ hiệu quả để tách biệt nhóm khách hàng đang hoạt động tích cực và nhóm khách hàng đã lâu không quay lại.

5.3.8. Kiểm định tần suất mua hàng của hai nhóm khách hàng có điểm RFM khác nhau (rfm_score ≥ 6 và rfm_score < 6).

Giả thuyết:

- H0-a: Phương sai về tần suất mua hàng của nhóm khách hàng có rfm_score ≥ 6 và rfm_score < 6 là như nhau.
- H0-b: Tần suất mua hàng trung bình (Frequency) của nhóm khách hàng có rfm_score ≥ 6 và rfm_score < 6.

| Group Statistics | | | | | |
|------------------|-----------|------|------|----------------|-----------------|
| | rfm_score | N | Mean | Std. Deviation | Std. Error Mean |
| frequency | >= 6 | 2151 | 8.28 | 11.704 | .252 |
| | < 6 | 2169 | 1.53 | .764 | .016 |

Bảng 23: Group Statistics frequency

| Independent Samples Effect Sizes | | | | | |
|----------------------------------|--------------------|---------------------------|----------------|-------------------------|-------|
| | | Standardizer ^a | Point Estimate | 95% Confidence Interval | |
| | | | | Lower | Upper |
| frequency | Cohen's d | 8.276 | .815 | .753 | .877 |
| | Hedges' correction | 8.278 | .815 | .753 | .877 |
| | Glass's delta | .764 | 8.828 | 8.558 | 9.097 |

- a. The denominator used in estimating the effect sizes.
 Cohen's d uses the pooled standard deviation.
 Hedges' correction uses the pooled standard deviation, plus a correction factor.
 Glass's delta uses the sample standard deviation of the control group.

Bảng 24: Independent Samples Effect Sizes frequency

| Independent Samples Test | | | | | | | | | | |
|---|-----------------------------|---------|-------|------------------------------|----------|-----------------|-----------------|-----------------------|---|-------|
| Levene's Test for Equality of Variances | | | | t-test for Equality of Means | | | | | | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| frequency | Equal variances assumed | 492.052 | <.001 | 26.790 | 4318 | <.001 | 6.747 | .252 | 6.253 | 7.241 |
| | Equal variances not assumed | | | 26.680 | 2168.186 | <.001 | 6.747 | .253 | 6.251 | 7.243 |

Bảng 25: Independent Samples Test frequency

Nhận xét:

- **Kiểm định Levene:** Giá trị **Sig. < .001** (nhỏ hơn 0.05), do đó bác bỏ H_0 -a. Vậy phương sai giữa hai nhóm giới tính (điểm RFM) có sự khác biệt, ta sẽ sử dụng kết quả kiểm định ở phần **Equal variances not assumed**.
- **Kiểm định T-Test (phần Equal variances not assumed):** Nhận thấy giá trị **Sig. (2-tailed) < .001** (nhỏ hơn 0.05) nên bác bỏ giả thuyết H_0 -b với mức ý nghĩa 5%. Điều này chứng tỏ **có sự khác biệt có ý nghĩa thống kê** về tần suất mua hàng trung bình giữa nhóm điểm cao và nhóm điểm thấp.
- Có sự chênh lệch rõ rệt về tần suất mua hàng giữa hai nhóm (Mean Difference: **6.747**). Cụ thể, nhóm có điểm RFM cao (≥ 6) mua hàng trung bình **8.28 lần**, trong khi nhóm điểm thấp (< 6) chỉ mua trung bình **1.53 lần**.
- Chỉ số **Cohen's d = 0.815** cho thấy mức độ ảnh hưởng (effect size) là **lớn**. Đặc biệt, giá trị **Glass's delta = 8.828** khẳng định sự khác biệt về tần suất mua hàng giữa hai nhóm này là cực kỳ đáng kể trong thực tế kinh doanh. Doanh nghiệp cần có một chính sách chăm sóc phù hợp với tệp khách hàng của mình.

5.4 One-way ANOVA

Phương pháp One – way ANOVA được áp dụng nhằm kiểm tra có sự khác biệt có ý nghĩa thống kê về giá trị trung bình của các chỉ số recency (r), frequency (f) và

monetary(m) giữa các phân khúc khách hàng.

| Descriptives | | | | | | | | | |
|----------------|-------|------|-------------|----------------|-------------|----------------------------------|-------------|-----------|-------------|
| | | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
| recency | 0 | 1174 | 69.40 | 68.626 | 2.003 | 65.47 | 73.33 | 1 | 373 |
| | 1 | 727 | 128.49 | 93.881 | 3.482 | 121.66 | 135.33 | 26 | 374 |
| | 2 | 731 | 239.15 | 80.062 | 2.961 | 233.33 | 244.96 | 93 | 374 |
| | 3 | 985 | 39.83 | 37.373 | 1.191 | 37.50 | 42.17 | 1 | 307 |
| | 4 | 703 | 9.61 | 6.907 | .260 | 9.10 | 10.12 | 1 | 25 |
| | Total | 4320 | 91.60 | 99.615 | 1.516 | 88.62 | 94.57 | 1 | 374 |
| frequency | 0 | 1174 | 2.35 | 1.218 | .036 | 2.28 | 2.42 | 1 | 10 |
| | 1 | 727 | 1.47 | .631 | .023 | 1.42 | 1.51 | 1 | 4 |
| | 2 | 731 | 1.19 | .393 | .015 | 1.16 | 1.22 | 1 | 2 |
| | 3 | 985 | 5.81 | 3.793 | .121 | 5.58 | 6.05 | 1 | 65 |
| | 4 | 703 | 15.22 | 17.971 | .678 | 13.89 | 16.55 | 5 | 239 |
| | Total | 4320 | 4.89 | 8.937 | .136 | 4.62 | 5.15 | 1 | 239 |
| monetary_value | 0 | 1174 | 761.209455 | 1438.857470 | 41.9936600 | 678.818380 | 843.600530 | 30.0000 | 39619.5000 |
| | 1 | 727 | 383.597455 | 236.6498295 | 8.7768564 | 366.366407 | 400.828504 | 5.9000 | 1126.0000 |
| | 2 | 731 | 197.947059 | 91.6316876 | 3.3891204 | 191.293473 | 204.600644 | 2.9000 | 370.8000 |
| | 3 | 985 | 1954.860051 | 2636.339400 | 84.0007580 | 1790.018833 | 2119.701269 | 191.1700 | 60369.9300 |
| | 4 | 703 | 7149.436230 | 19383.95980 | 731.0798903 | 5714.071242 | 8584.801219 | 1139.7300 | 278778.0200 |
| | Total | 4320 | 1914.079947 | 8303.245478 | 126.3298569 | 1666.408569 | 2161.751324 | 2.9000 | 278778.0200 |

Bảng 26: One - way ANOVA (recency - frequency - monetary_value)

| Tests of Homogeneity of Variances | | | | | |
|-----------------------------------|--------------------------------------|------------------|-----|----------|-------|
| | | Levene Statistic | df1 | df2 | Sig. |
| recency | Based on Mean | 453.855 | 4 | 4315 | .000 |
| | Based on Median | 292.854 | 4 | 4315 | <.001 |
| | Based on Median and with adjusted df | 292.854 | 4 | 2781.033 | <.001 |
| | Based on trimmed mean | 412.855 | 4 | 4315 | <.001 |
| frequency | Based on Mean | 244.544 | 4 | 4315 | <.001 |
| | Based on Median | 161.732 | 4 | 4315 | <.001 |
| | Based on Median and with adjusted df | 161.732 | 4 | 787.037 | <.001 |
| | Based on trimmed mean | 183.107 | 4 | 4315 | <.001 |
| monetary_value | Based on Mean | 105.823 | 4 | 4315 | <.001 |
| | Based on Median | 53.590 | 4 | 4315 | <.001 |
| | Based on Median and with adjusted df | 53.590 | 4 | 748.083 | <.001 |
| | Based on trimmed mean | 59.180 | 4 | 4315 | <.001 |

Bảng 27: Tests of Homogeneity of Variances (recency - frequency - monetary_value)

| Robust Tests of Equality of Means | | | | | |
|-----------------------------------|-------|------------------------|-----|----------|-------|
| | | Statistic ^a | df1 | df2 | Sig. |
| recency | Welch | 2093.003 | 4 | 1787.745 | .000 |
| frequency | Welch | 667.503 | 4 | 1968.215 | .000 |
| monetary_value | Welch | 266.305 | 4 | 1838.475 | <.001 |

a. Asymptotically F distributed.

Bảng 28: Robust Tests of Equality of Means (recency - frequency - monetary_value)

Giả thuyết:

- H0-a: Phương sai các trường bên trong các biến (r ,f, m) là như nhau. 3 giá trị sig của cả 3 biến r,f,m lần lượt là 0, <0.001,<0.001, đều < 0.05, nên ta bác bỏ H0, thừa nhận phương sai của các trường bên trong các biến không đồng nhất, điều này có nghĩa là ta phải kiểm định bằng 1 cách khác chứ

không sử dụng bảng ANOVA. Lúc này nhóm đã tiến hành chạy lại dữ liệu và tiến hành thực hiện Welch Test (Robust Tests of Equality of Means).

- H0-b: Không có sự khác biệt về giá trị trong các trường của các biến. Với giá trị sig trong bảng Robust Tests of Equality of Means của 3 giá trị r,f,m lần lượt là 0,0,<0.001, đều < 0.05, ta bác bỏ H0-b, rằng có sự khác biệt về giá trị trong các trường của các biến. Điều này là đúng khi ta xem xét về bản chất dữ liệu là dữ liệu về các phân khúc khách hàng.
- Kết luận: Kiểm định cho thấy sự khác biệt rõ rệt về hành vi và giá trị (thời gian mua hàng gần nhất, tần suất giao dịch và mức chi tiêu) giữa các phân khúc khách hàng được phân loại.

CHƯƠNG VI. KẾT LUẬN

6.1 Kết quả về phân khúc khách hàng

Việc áp dụng mô hình RFM đã giúp doanh nghiệp nhận diện rõ rệt sự phân hóa trong tập khách hàng của mình:

- **Nhóm khách hàng VIP (Segment 4):** Chiếm 16.2% nhưng đóng góp lớn nhất vào doanh thu với mức chi tiêu trung bình lên tới 7,149 đô. Đây là nhóm có tần suất mua hàng cao và thời gian quay lại nhanh nhất.
- **Nhóm khách hàng tiềm năng (Segment 3):** Chiếm 22.6%, là phân khúc quan trọng cần được đầu tư để chuyển đổi thành nhóm khách hàng trung thành.
- **Nhóm khách hàng ít hoạt động/nguy cơ rời bỏ (Segment 0 và 2):** Chiếm tỷ lệ lớn nhất (27% ở Segment 0), phản ánh bộ phận khách hàng có mức độ tương tác thấp hoặc đã lâu không phát sinh giao dịch.

6.2 Sự tương quan giữa các hành vi mua sắm

Các kiểm định thống kê (Chi-Square) đã khẳng định rằng các chỉ số Recency (R), Frequency (F) và Monetary (M) không tồn tại độc lập mà có mối liên hệ mật thiết với nhau. Một khách hàng mua sắm thường xuyên hơn (Frequency cao) có xu hướng chi tiêu nhiều hơn (Monetary cao) và có khoảng thời gian quay lại mua hàng ngắn hơn (Recency thấp).

6.3 Hiệu quả của chỉ số RFM Score trong phân loại

Kết quả kiểm định Independent Samples T-Test cho thấy sự khác biệt cực kỳ lớn giữa nhóm có rfm_score ≥ 6 và nhóm có rfm_score < 6:

- Nhóm điểm cao chi tiêu gấp gần 9 lần so với nhóm điểm thấp (3,454.05 so với 386.89).
- Tần suất mua hàng của nhóm điểm cao gấp hơn 5 lần nhóm điểm thấp (8.28 lần so với 1.53 lần).
- Thời gian quay lại mua hàng trung bình của nhóm tích cực chỉ là 34.61 ngày, trong khi nhóm còn lại lên tới 148.11 ngày. Điều này chứng minh điểm RFM là một thước đo chính xác để doanh nghiệp ưu tiên nguồn lực chăm sóc khách hàng.

6.4 Đề xuất chiến lược kinh doanh

Dựa trên các bằng chứng thống kê, nghiên cứu đề xuất hướng đi cho doanh nghiệp:

- **Giữ chân nhóm VIP:** Tập trung vào các chương trình đặc quyền, tri ân để duy trì giá trị chi tiêu cao.
- **Khai thác nhóm tiềm năng:** Sử dụng các chiến dịch marketing cá nhân hóa để khuyến khích họ mua sắm thường xuyên hơn.
- **Tối ưu chi phí với nhóm ít hoạt động:** Thay vì đầu tư dàn trải, cần có các biện pháp tái kích hoạt chọn lọc hoặc giảm bớt chi phí marketing cho nhóm khách hàng không còn giá trị.

Kết quả phân tích cho thấy hành vi mua sắm của khách hàng có sự khác biệt rõ rệt giữa các phân khúc. Nhóm khách hàng giá trị cao có tần suất mua lớn, chi tiêu cao và mua hàng gần đây, trong khi nhóm khách hàng không hoạt động có mức độ tương tác thấp. Trên cơ sở đó, nhóm đề xuất các chiến lược tập trung vào việc giữ chân khách hàng giá trị cao, phát triển khách hàng tiềm năng và tối ưu chi phí đối với nhóm khách hàng ít hoạt động.