**PHÂN TÍCH DỮ LIỆU MARKETING CAMPAIGN CÙNG AUTOVIZ**

**1. TỔNG QUAN**

**1.1. Mục tiêu**

Phân tích hành vi khách hàng và hiệu quả của các chiến dịch marketing để đưa ra insights và recommendations cho chiến lược marketing trong tương lai.

**1.2. Nguồn dữ liệu**

* **Dataset**: Customer Personality Analysis
* **Nguồn**: Kaggle (https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis)
* **Kích thước**: 2,240 khách hàng với 29 thuộc tính
* **Thời gian**: Dữ liệu từ năm 2012-2014

**2. QUY TRÌNH XỬ LÝ DỮ LIỆU**

**2.1. Làm sạch dữ liệu**

* **Xử lý giá trị null**: Phát hiện 24 giá trị null trong cột Income, được điền bằng giá trị median
* **Loại bỏ outliers**: Loại bỏ các khách hàng có tuổi bất thường (< 18 hoặc > 100 tuổi)
* **Chuẩn hóa định dạng**: Chuyển đổi cột Dt\_Customer sang định dạng datetime

**2.2. Feature Engineering**

Tạo các biến mới để phân tích sâu hơn:

| **Feature** | **Công thức** | **Mục đích** |
| --- | --- | --- |
| Age | 2014 - Year\_Birth | Tính tuổi khách hàng |
| Customer\_Days | Ngày hiện tại - Dt\_Customer | Số ngày là khách hàng |
| Total\_Purchases | Tổng các kênh mua hàng | Đo lường tần suất mua |
| Total\_Amount\_Spent | Tổng chi tiêu tất cả sản phẩm | Giá trị khách hàng |
| Total\_Accepted\_Campaigns | Tổng các chiến dịch chấp nhận | Mức độ tương tác |
| Total\_Children | Kidhome + Teenhome | Số con trong gia đình |
| Spending\_Segment | Phân khúc theo chi tiêu | Phân loại khách hàng |

**3. PHÂN TÍCH KHÁM PHÁ DỮ LIỆU (EDA)**

**3.1. Đặc điểm Khách hàng**

**3.1.1. Nhân khẩu học**

* **Độ tuổi trung bình**: ~52 tuổi
* **Thu nhập trung bình**: ~52,000 USD/năm
* **Trình độ học vấn**:
  + Graduation: ~50%
  + PhD: ~21%
  + Master: ~17%
  + 2n Cycle: ~8%
  + Basic: ~4%

**3.1.2. Tình trạng gia đình**

* **Married**: Chiếm tỷ lệ cao nhất (~38%)
* **Together**: ~26%
* **Single**: ~21%
* **Divorced**: ~11%
* **Widow**: ~4%

**3.2. Hành vi Mua hàng**

**3.2.1. Kênh mua hàng ưa thích**

* **Store Purchases** (Mua tại cửa hàng): Kênh phổ biến nhất
* **Web Purchases** (Mua online): Kênh thứ hai
* **Catalog Purchases**: Ít phổ biến hơn
* **Deals Purchases**: Tần suất thấp nhất

**3.2.2. Danh mục sản phẩm**

Chi tiêu theo danh mục (theo tỷ lệ %):

1. **Wines**: ~40-45% tổng chi tiêu
2. **Meat Products**: ~25-30%
3. **Gold Products**: ~10-15%
4. **Fish Products**: ~8-12%
5. **Sweet Products**: ~3-5%
6. **Fruits**: ~2-4%

**Insight**: Rượu và thịt là 2 danh mục sản phẩm chủ lực, chiếm ~70% doanh thu.

**3.3. Phân khúc Khách hàng**

**Theo mức chi tiêu (Spending\_Segment):**

* **Low** (< $100): Khách hàng ít chi tiêu
* **Medium** ($100-$500): Phân khúc lớn nhất
* **High** ($500-$1,000): Khách hàng có giá trị
* **Premium** (> $1,000): Khách hàng VIP, cần chăm sóc đặc biệt

**4. PHÂN TÍCH HIỆU QUẢ CHIẾN DỊCH**

**4.1. Tỷ lệ chấp nhận chiến dịch**

| **Chiến dịch** | **Tỷ lệ chấp nhận** | **Đánh giá** |
| --- | --- | --- |
| Campaign 1 | ~6.4% | Trung bình |
| Campaign 2 | ~1.3% | Thấp |
| Campaign 3 | ~7.3% | Cao nhất |
| Campaign 4 | ~7.5% | Cao nhất |
| Campaign 5 | ~7.3% | Cao |
| Last Campaign | ~15% | Rất cao |

**Phát hiện quan trọng**:

* Campaign 2 có hiệu quả kém nhất (1.3%)
* Last Campaign (chiến dịch gần nhất) có tỷ lệ chấp nhận cao nhất (15%)
* Campaigns 3, 4, 5 có hiệu quả tương đương nhau (~7%)

**4.2. Yếu tố ảnh hưởng đến chấp nhận chiến dịch**

Phân tích correlation cho thấy:

* **Thu nhập cao** → Tỷ lệ chấp nhận cao hơn
* **Số con nhiều** → Tỷ lệ chấp nhận thấp hơn
* **Chi tiêu cao** → Xu hướng chấp nhận nhiều hơn
* **Recency thấp** (mua hàng gần đây) → Tỷ lệ chấp nhận cao hơn

**5. PHÂN TÍCH TƯƠNG QUAN**

**5.1. Các yếu tố ảnh hưởng đến Total\_Amount\_Spent**

**Tương quan dương mạnh**:

* **Income** (r ≈ 0.7-0.8): Thu nhập là yếu tố quan trọng nhất
* **Total\_Purchases** (r ≈ 0.6-0.7): Tần suất mua cao → Chi tiêu cao

**Tương quan âm**:

* **Total\_Children** (r ≈ -0.2 đến -0.3): Có con → Chi tiêu giảm
* **NumWebVisitsMonth**: Truy cập web nhiều nhưng không mua → Chi tiêu thấp

**Không tương quan mạnh**:

* **Age**: Tuổi tác không ảnh hưởng nhiều đến chi tiêu
* **Recency**: Thời gian mua hàng gần đây có ảnh hưởng nhỏ

**6. INSIGHTS QUAN TRỌNG**

**6.1. Về Khách hàng**

1. **Khách hàng có giá trị cao**:
   * Thu nhập > 70,000 USD
   * Không có hoặc ít con
   * Học vấn cao (PhD, Master, Graduation)
   * Tình trạng: Married hoặc Together
2. **Khách hàng tiềm năng**:
   * Thu nhập trung bình (40,000-60,000 USD)
   * Mua hàng thường xuyên tại store
   * Ít truy cập web nhưng conversion cao
3. **Khách hàng cần chăm sóc**:
   * Nhiều con, thu nhập trung bình
   * Quan tâm đến deals và promotions
   * Nhạy cảm với giá

**6.2. Về Sản phẩm**

1. **Wines** là danh mục chủ lực, cần:
   * Đầu tư marketing mạnh
   * Phát triển dòng sản phẩm cao cấp
   * Chương trình loyalty cho wine lovers
2. **Meat Products** là danh mục thứ hai:
   * Cross-selling với Wines
   * Bundle promotions
3. **Fruits & Sweets** có doanh thu thấp:
   * Cần đánh giá lại chiến lược
   * Xem xét ngưng hoặc điều chỉnh

**6.3. Về Chiến dịch Marketing**

1. **Campaign 2 thất bại** - Cần phân tích:
   * Thông điệp không phù hợp?
   * Timing không đúng?
   * Targeting sai đối tượng?
2. **Last Campaign thành công** - Học hỏi:
   * Personalization tốt hơn?
   * Offer hấp dẫn hơn?
   * Timing phù hợp hơn?
3. **Multi-channel approach** cần thiết:
   * Store vẫn là kênh chính
   * Web đang phát triển
   * Catalog đang giảm

**7. KHUYẾN NGHỊ CHIẾN LƯỢC**

**7.1. Phân khúc & Targeting**

**Khuyến nghị 1: Focus vào Premium Segment**

* Tập trung 80% ngân sách marketing vào 20% khách hàng chi tiêu cao
* Tạo chương trình VIP với exclusive benefits
* Personalized communication

**Khuyến nghị 2: Re-activate Low Spenders**

* Chiến dịch win-back với special offers
* Survey để hiểu rõ barriers
* Entry-level products với giá hấp dẫn

**7.2. Product Strategy**

**Khuyến nghị 3: Tăng cường Wine Category**

* Wine tasting events
* Wine subscription program
* Educational content về wines

**Khuyến nghị 4: Cross-selling thông minh**

* Bundle: Wine + Meat Products
* Recipe suggestions
* Meal planning content

**Khuyến nghị 5: Đánh giá lại Fruits & Sweets**

* A/B testing với promotions
* Nếu không cải thiện → Consider phasing out
* Reallocate resources sang categories có ROI cao hơn

**7.3. Campaign Optimization**

**Khuyến nghị 6: Áp dụng learnings từ Last Campaign**

* Replicate success factors
* Personalization dựa trên purchase history
* Timing: Sau 30-45 ngày kể từ last purchase

**Khuyến nghị 7: Cải thiện Campaign Acceptance Rate**

* Current rate: ~7-15%
* Target: 20-25%
* Tactics:
  + Better segmentation
  + Personalized offers
  + A/B testing messages
  + Optimal timing

**Khuyến nghị 8: Multi-channel Integration**

* Online-to-Offline strategy
* Click & Collect
* Unified customer experience
* Retargeting web visitors

**7.4. Customer Retention**

**Khuyến nghị 9: Loyalty Program**

* Points-based rewards
* Tiered membership (Silver, Gold, Platinum)
* Exclusive access to new products
* Birthday & anniversary offers

**Khuyến nghị 10: Reduce Churn**

* Identify at-risk customers (high Recency, low purchases)
* Proactive outreach
* Win-back campaigns với incentives

**8. KẾT LUẬN**

**8.1. Tóm tắt**

Phân tích dữ liệu Marketing Campaign đã cung cấp insights sâu sắc về:

* Đặc điểm và hành vi khách hàng
* Hiệu quả của các chiến dịch marketing
* Cơ hội tối ưu hóa doanh thu

**8.2. Key Takeaways**

1. **Thu nhập là yếu tố quyết định** chi tiêu khách hàng
2. **Wines & Meat Products** là 2 danh mục chủ lực
3. **Store channel** vẫn dominant nhưng web đang phát triển
4. **Last Campaign** thành công nhất → cần replicate
5. **Premium segment** (>$1,000) cần chăm sóc đặc biệt

**8.3. Next Steps**

1. **Ngắn hạn** (1-3 tháng):
   * Implement VIP program cho Premium customers
   * Launch improved campaigns dựa trên Last Campaign learnings
   * A/B testing cho Fruits & Sweets category
2. **Trung hạn** (3-6 tháng):
   * Phát triển wine subscription program
   * Tối ưu hóa multi-channel experience
   * Build predictive model cho campaign acceptance
3. **Dài hạn** (6-12 tháng):
   * Customer Lifetime Value (CLV) modeling
   * Advanced personalization engine
   * Automated marketing campaigns

**9. PHỤ LỤC**

**9.1. Công cụ sử dụng**

* **Python**: Ngôn ngữ lập trình chính
* **Pandas & NumPy**: Xử lý và phân tích dữ liệu
* **AutoViz**: Tự động tạo visualizations
* **Matplotlib & Seaborn**: Tạo biểu đồ tùy chỉnh

**9.2. Metrics quan trọng**

* **Total Amount Spent**: Tổng chi tiêu khách hàng
* **Campaign Acceptance Rate**: Tỷ lệ chấp nhận chiến dịch
* **Average Order Value (AOV)**: Giá trị đơn hàng trung bình
* **Purchase Frequency**: Tần suất mua hàng
* **Customer Lifetime Value (CLV)**: Giá trị khách hàng trọn đời

**9.3. Limitations**

* Dữ liệu từ 2012-2014 có thể outdated
* Không có thông tin về customer acquisition cost
* Thiếu data về customer satisfaction
* Không có thông tin về competitors