khoanFX20883@funix.edu.vn

DAP305x\_1.0-A\_VN Đồ án cuối khóa - Phân tích dữ liệu

GIỚI THIỆU SẢN PHẨM

Phân Cụm Khách Hàng Trong Chiến Dịch Tiếp Thị Của Ngân Hàng

Mục lục

[Xác định yêu cầu kinh doanh 1](#_Toc168608503)

[Bài toán cần giải quyết: 1](#_Toc168608504)

[Các chỉ số báo cáo phù hợp: 1](#_Toc168608505)

[Mục đích của việc phát triển sản phẩm: 1](#_Toc168608506)

[Mô tả tính năng hoạt động của sản phẩm: 2](#_Toc168608507)

[Phục vụ cho nhu cầu nào của khách hàng: 2](#_Toc168608508)

[Sản phẩm gồm những thành phần nào: 2](#_Toc168608509)

[Tập dữ liệu CSV 2](#_Toc168608510)

[Cơ sở dữ liệu SQL 2](#_Toc168608511)

[Giải pháp SSIS ETL 2](#_Toc168608512)

[Trực quan hóa Power BI 4](#_Toc168608513)

[Sơ đồ hoạt động tổng quan cho sản phẩm: 5](#_Toc168608514)

[Những khó khăn gặp phải khi hoàn thành dự án: 7](#_Toc168608515)

[Nhược điểm của sản phẩm: 7](#_Toc168608516)

## Xác định yêu cầu kinh doanh

Đánh giá tiềm năng của các nhóm khách hàng để điều chỉnh chiến lược tiếp cận đối tượng hiệu quả hơn.

## Bài toán cần giải quyết:

1. So sánh sự khác biệt của tỷ lệ chuyển đổi giữa các nhóm thuộc tính trong bộ dữ liệu.
2. Xác định mức độ ảnh hưởng của các thuộc tính khác có thể dẫn đến việc tăng tỷ lệ chuyển đổi của một nhóm thuộc tính.
3. Tối ưu hiệu quả chiến dịch tiếp thị thông qua chọn lọc các đặc điểm của khách hàng có tỷ lệ chuyển đổi cao để tập trung tiếp cận.

## Các chỉ số báo cáo phù hợp:

* Học vấn
* Nghề nghiệp
* Hôn nhân
* Tín dụng
* Các hoản vay
* Tỷ lệ chuyển đổi

## Mục đích của việc phát triển sản phẩm:

Phân cụm khách hàng trong chiến dịch tiếp thị của ngân hàng với mục tiêu đăng ký gửi tiền có kỳ hạn giúp tối ưu hóa chiến dịch, tăng cơ hội chuyển đổi, cung cấp dịch vụ cá nhân, cải thiện hiệu suất kinh doanh và tăng cường sự thành công của ngân hàng trong thị trường.

## Mô tả tính năng hoạt động của sản phẩm:

1. Chia nhỏ khách hàng thành các nhóm dựa trên học vấn, nghề nghiệp, tình trạng hôn nhân, tín dụng, các khoản vay
   * Mã hóa thông tin khách hàng
   * Phân cụm tự động trên scatter plot của Power BI
2. Phân tích tình trạng hôn nhân và mối tương quan đến tỷ lệ chuyển đổi( nếu có)
   * So sánh sự phân bố của tỷ lệ chuyển đổi từ cao đến thấp có liên quan đến các đặc điểm học vấn, nghề nghiệp, tình trạng hôn nhân, tín dụng, các khoản vay hay không
3. Trực quan hóa các thông tin tổng quan từ dữ liệu và thông tin chi tiết từ các nhóm khách hàng đã được phân cụm
   * Các dashboard với pie chart, bar chart với trend line và scatter plot

## Phục vụ cho nhu cầu nào của khách hàng:

Phân cụm khách hàng giúp ngân hàng tối ưu hóa chiến dịch tiếp thị bằng cách xác định tập trung vào các nhóm khách hàng tiềm năng và có khả năng chuyển đổi cao nhất.

## Sản phẩm gồm những thành phần nào:

### [Tập dữ liệu CSV](bank-additional-full.csv)

* Dữ liệu khách hàng
* Dữ liệu chiến dịch
* Dữ liệu kinh tế thị trường

### Cơ sở dữ liệu SQL

1. Bảng tạm tải dữ liệu chung (Client\_Staging)
2. Bảng dữ liệu quy định mã hóa học vấn (Dim\_Education)
3. Bảng dữ liệu quy định mã hóa việc làm (Dim\_Job)
4. Bảng dữ liệu quy định mã hóa các tình trạng hôn nhân và trạng thái khoản vay (Dim\_Status)
5. Bảng dữ liệu khách hàng đã được mã hóa (Dim\_Client)
6. Bảng dữ liệu chung (Fact\_Campaigne)

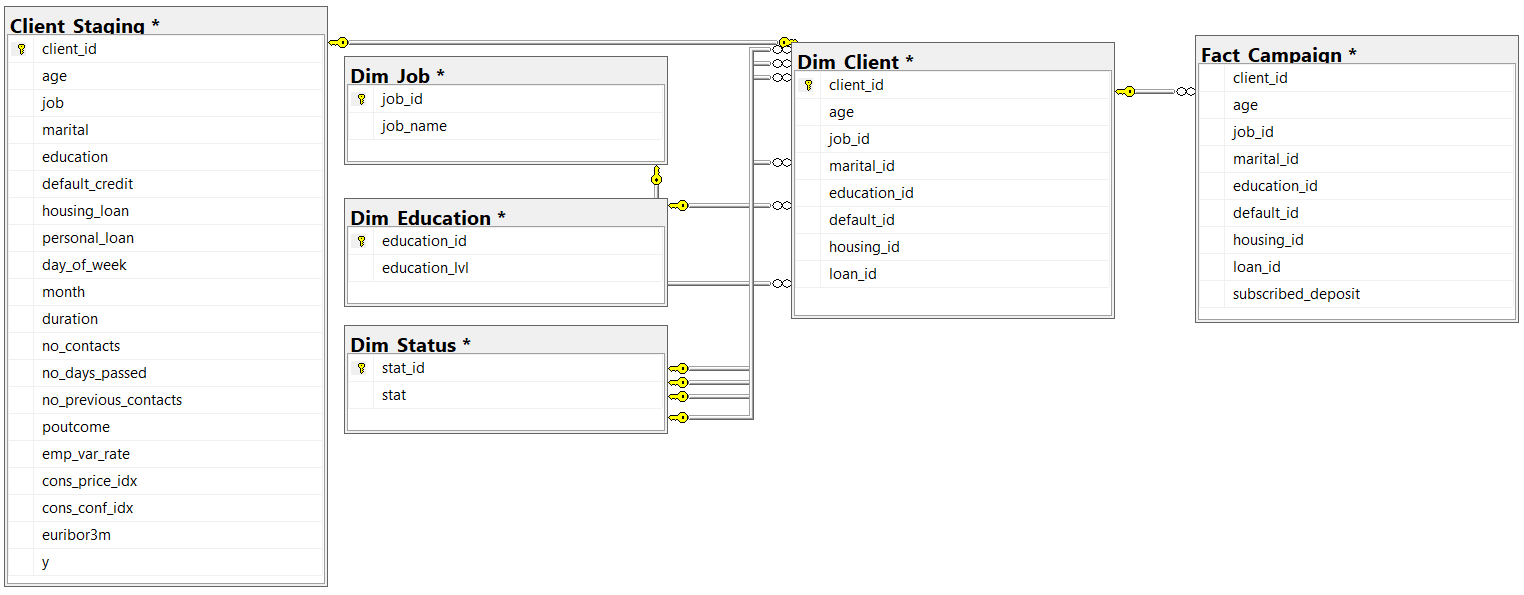


Figure 1. Lược đồ quan hệ của các bảng trong Data warehouse

### [Giải pháp SSIS ETL](bank-campaign-etl/Bank_Marketing_Campaign.sln)

* Tải dữ liệu từ tập tin csv vào bảng tạm (Client\_Staging)

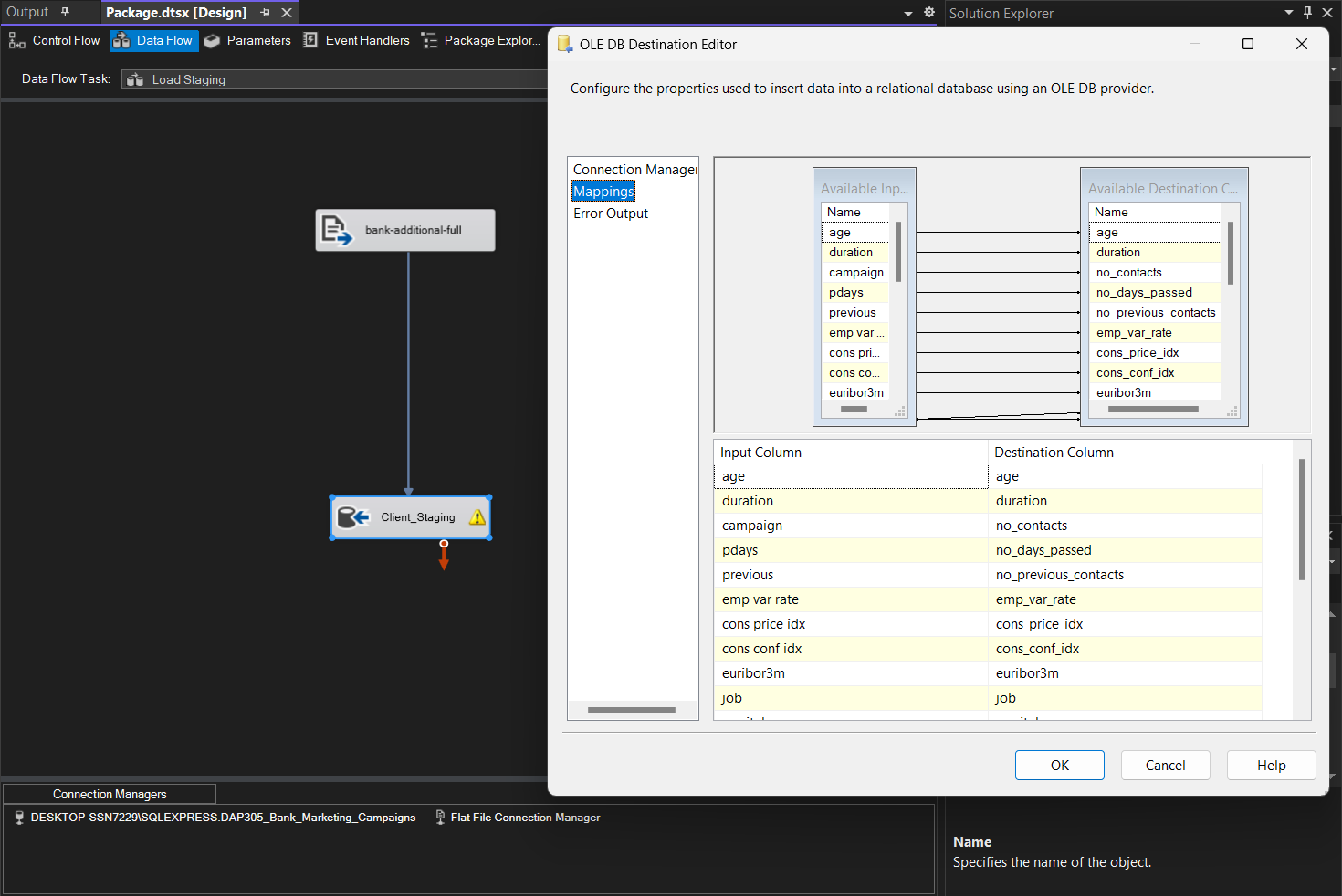


Figure 2. Trích xuất dữ liệu từ file csv

* Mã hóa các thông tin hạng mục và phân tách thành các bảng dữ liệu khách hàng (Dim\_Client)

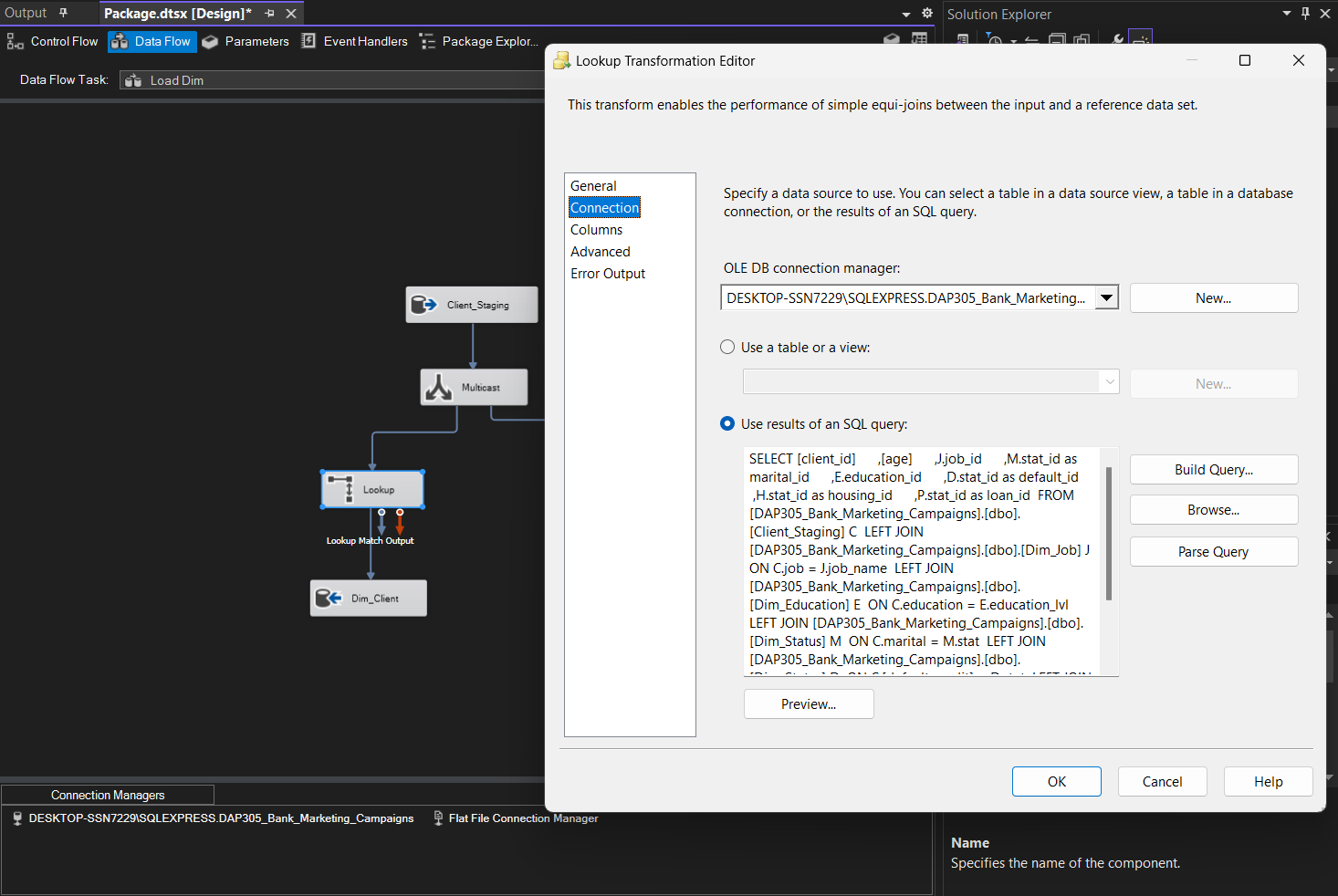


Figure 3. Mã hóa dữ liệu hạng mục vào các Dim Table

* Tải vào những bảng dữ liệu cuối cùng (Fact\_Campaign)

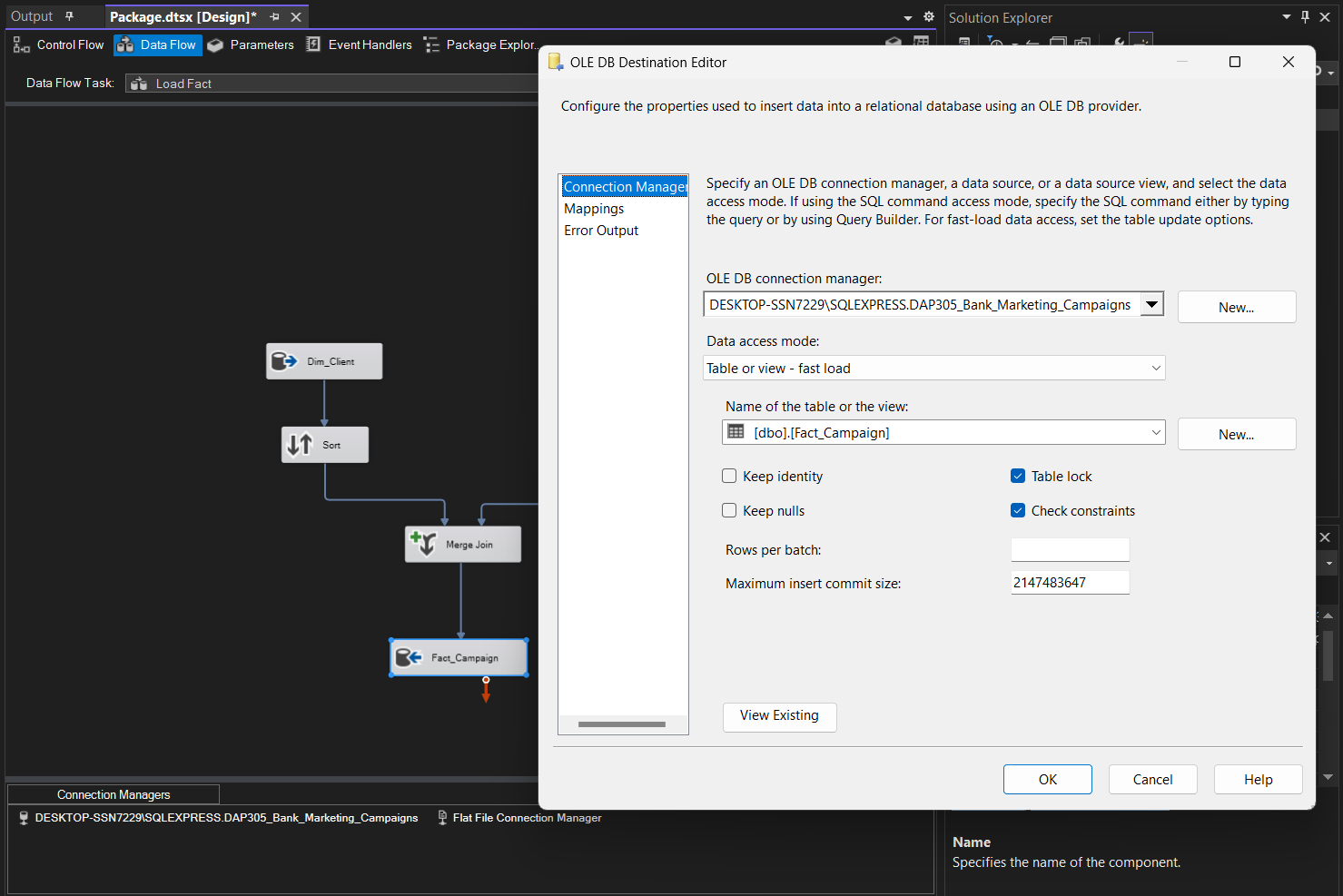


Figure 4. Kết hợp các Dim Table vào Fact Table

### [Trực quan hóa Power BI](bank-campaign-PBI.pbix)

|  |  |
| --- | --- |
| **Tên trang** | **Nội dung trực quan** |
| 1. Client Segments by Age, Marital status & Job | Biểu đồ phân cụm tự động theo Tuổi, Hôn nhân và Nghề nghiệp |
| 1. Subscribed / No Subscribed | So sánh phân bố về hôn nhân, học vấn và nghề nghiệp giữa 2 nhóm đăng ký gửi tiền và không đăng ký gửi tiền |
| 1. Single | Tỷ lệ chuyển đổi giữa các nghề nghiệp theo tình trạng độc thân |
| 1. Not single | Tỷ lệ chuyển đổi giữa các nghề nghiệp theo tình trạng hôn nhân |
| 1. Education Analysis | Thông tin tổng quan của bộ dữ liệu khách hàng về học vấn (đã loại trừ những giá trị ‘unknown’) |
| 1. University degree | Tỷ lệ chuyển đổi giữa các nghề nghiệp theo các trình độ đại học (đã loại trừ những giá trị “illiterate”) |
| 1. Basic 4 years | Tỷ lệ chuyển đổi giữa các nghề nghiệp theo các trình độ khóa đào tạo chuyên nghiệp (đã loại trừ những giá trị “illiterate”) |
| 1. Job Analysis | Thông tin tổng quan của bộ dữ liệu khách hàng về nghề nghiệp (đã loại trừ những giá trị ‘unknown’) |
| 1. Credit & Loan Analysis | Thông tin tổng quan của bộ dữ liệu khách hàng về tín dụng và các khoản vay (đã loại trừ những giá trị ‘unknown’) |
| 1. Housing loan | Tỷ lệ chuyển đổi giữa các nghề nghiệp theo khoản vay mua nhà (đã loại trừ những giá trị ‘unknown’) |
| 1. Personal loan | Tỷ lệ chuyển đổi giữa các nghề nghiệp theo khoản vay cá nhân (đã loại trừ những giá trị ‘unknown’) |

## Sơ đồ hoạt động tổng quan cho sản phẩm:

1. Tải dữ liệu từ tập dữ liệu vào bảng tạm trong cơ sở dữ liệu

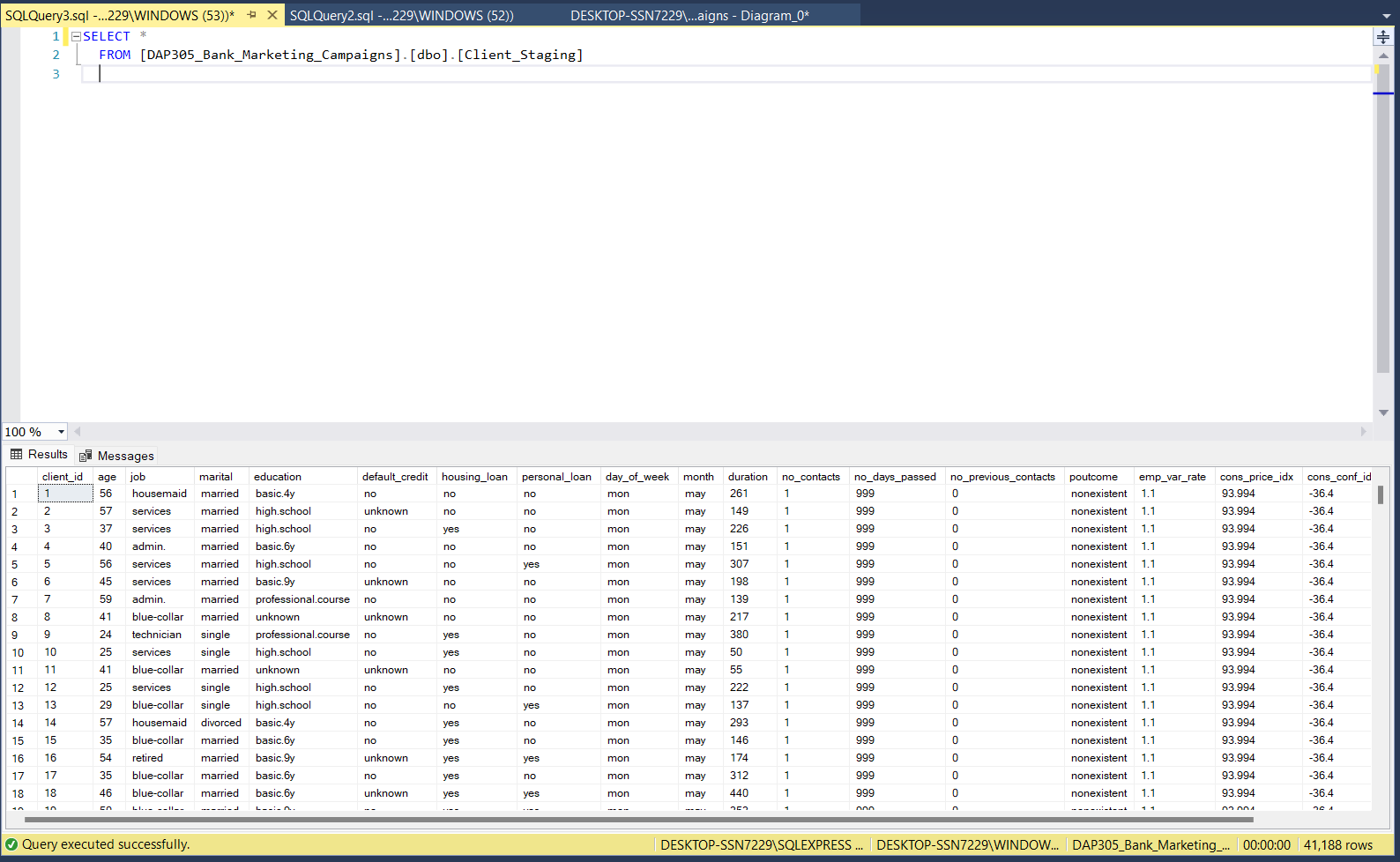


Figure 5. Dữ liệu bảng tạm sau khi tải từ file csv

1. Thực hiện chuyển đổi bằng giải pháp SISS

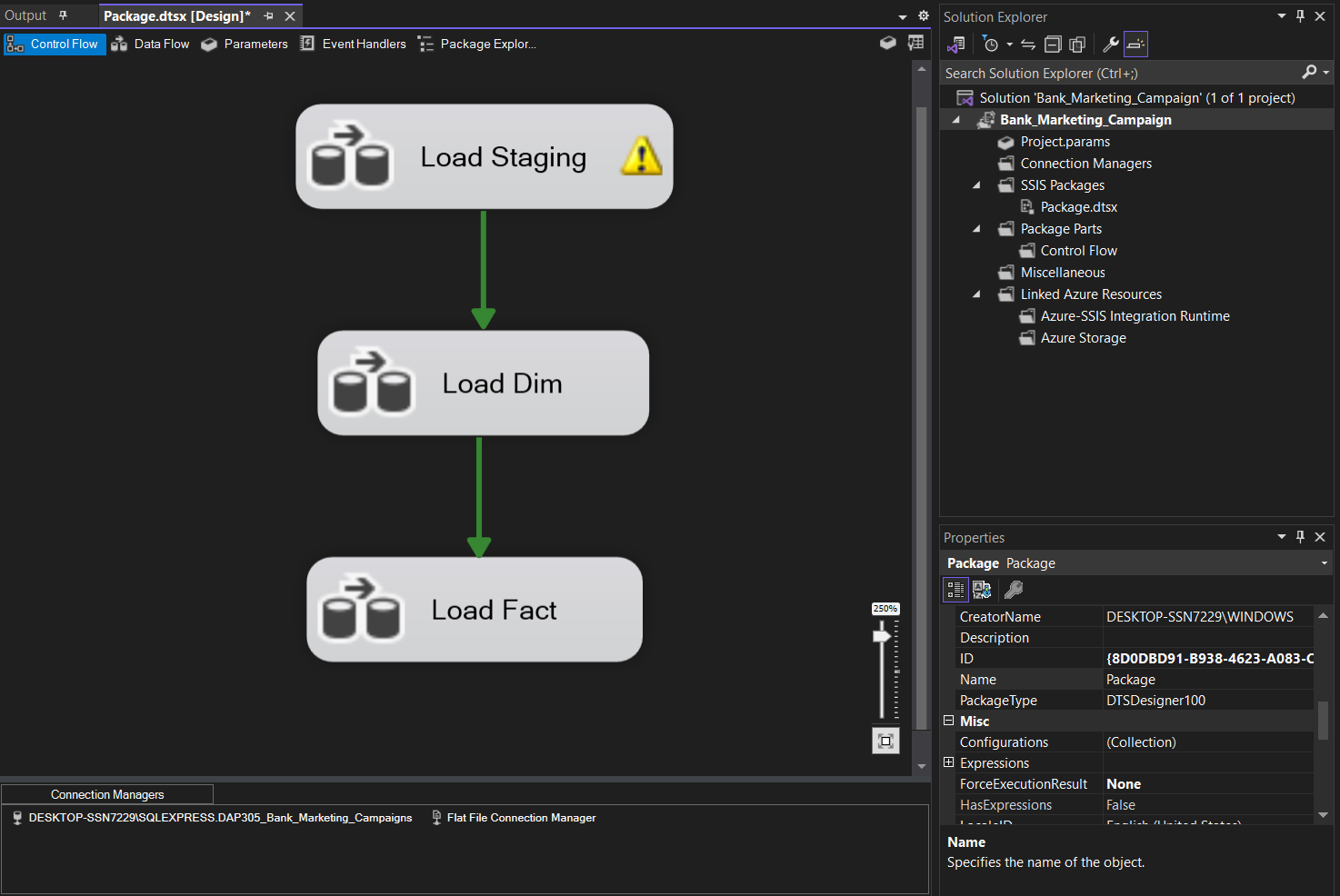


Figure 6. Control Flow của quá trình ETL

1. Lưu dữ liệu đã được chuyển đổi vào các bảng quan hệ trong cơ sở dữ liệu thông qua ETL

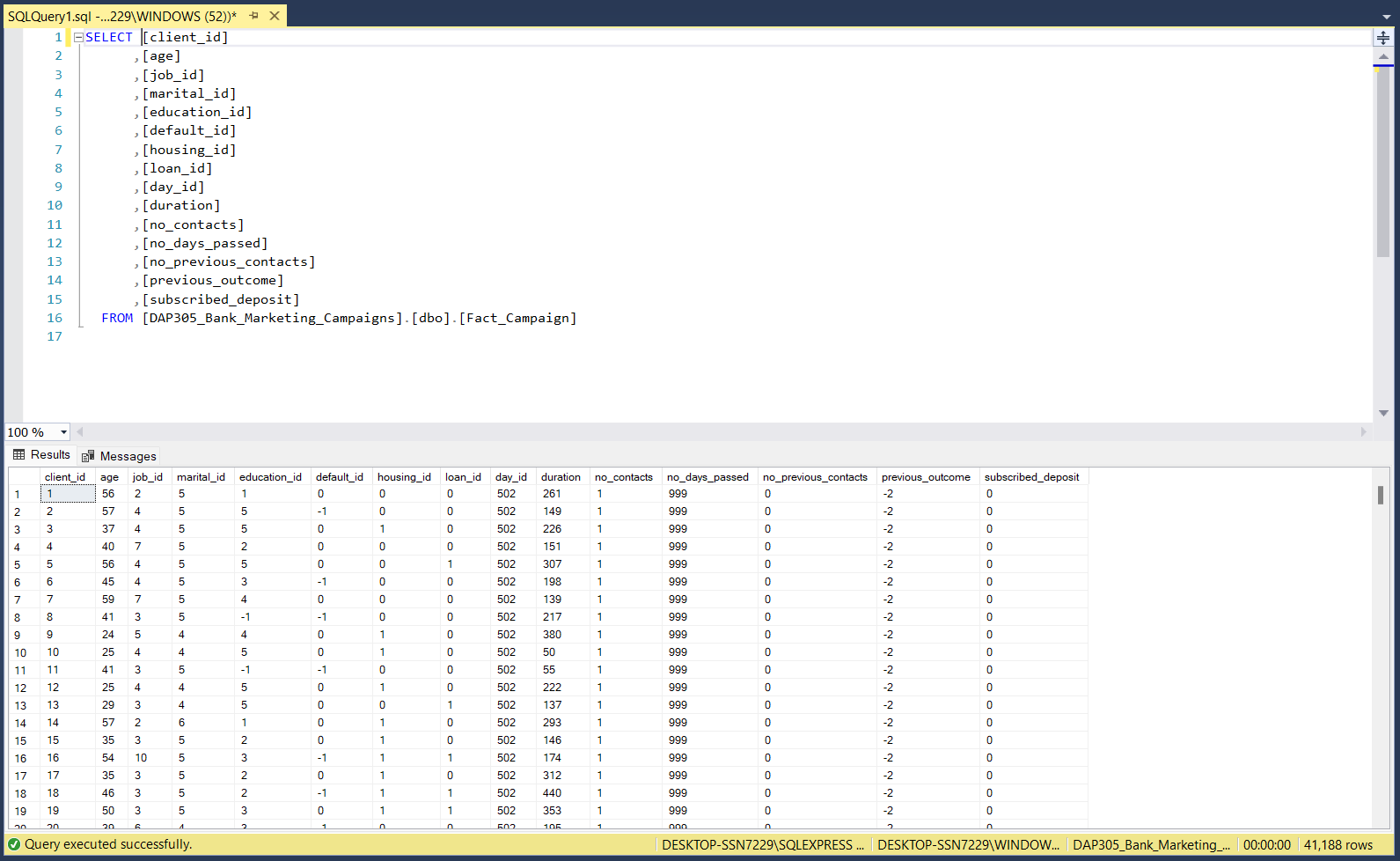


Figure 7. Dữ liệu Fact Table sau khi hoàn thành ETL

1. Tải dữ liệu từ các bảng quan hệ vào data model của Power BI

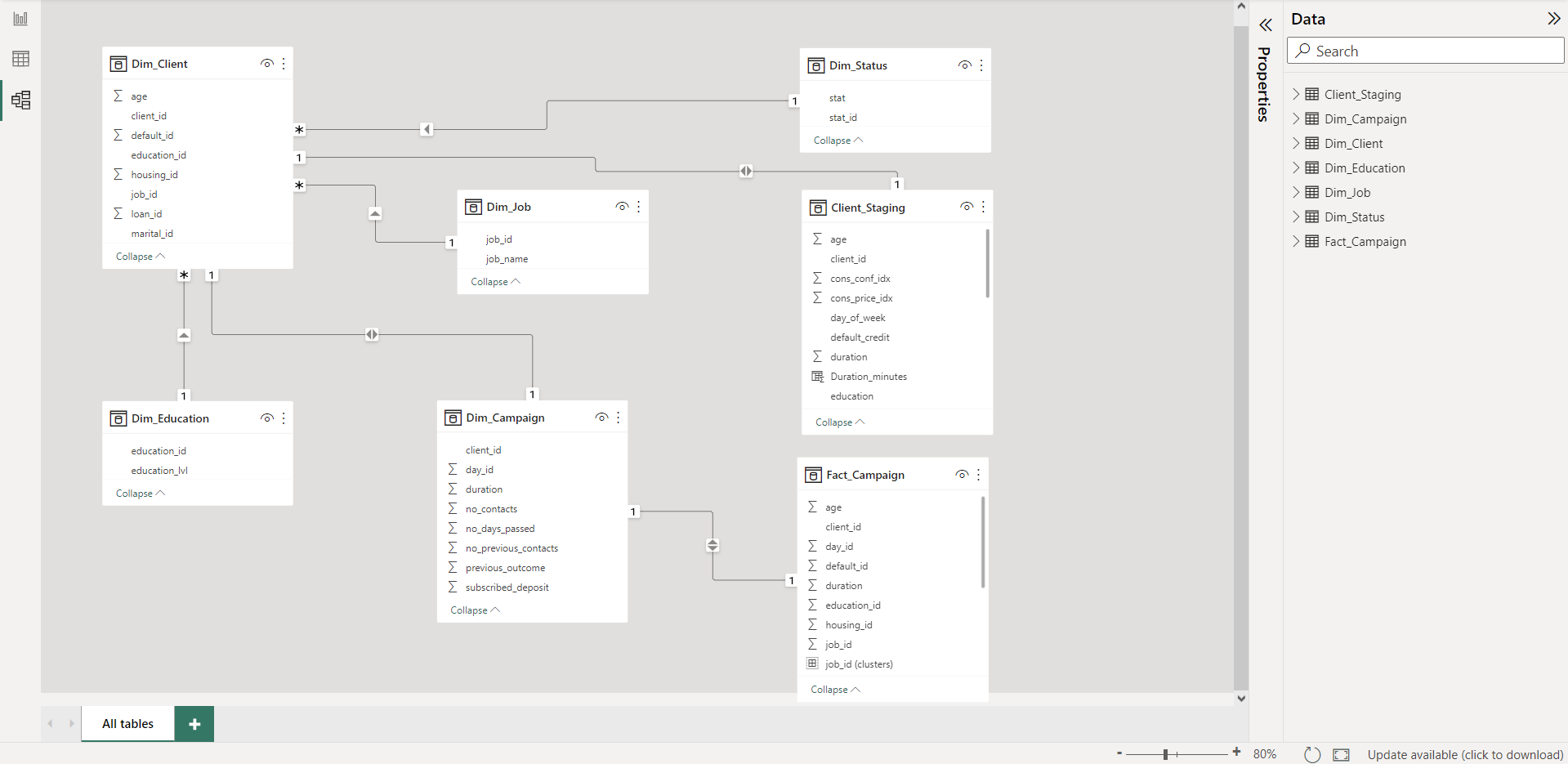


Figure 8. Data Model tải từ SQL Server

1. Loại bỏ các giá trị trùng lặp và thay thế giá trị không xác định ‘unknown’

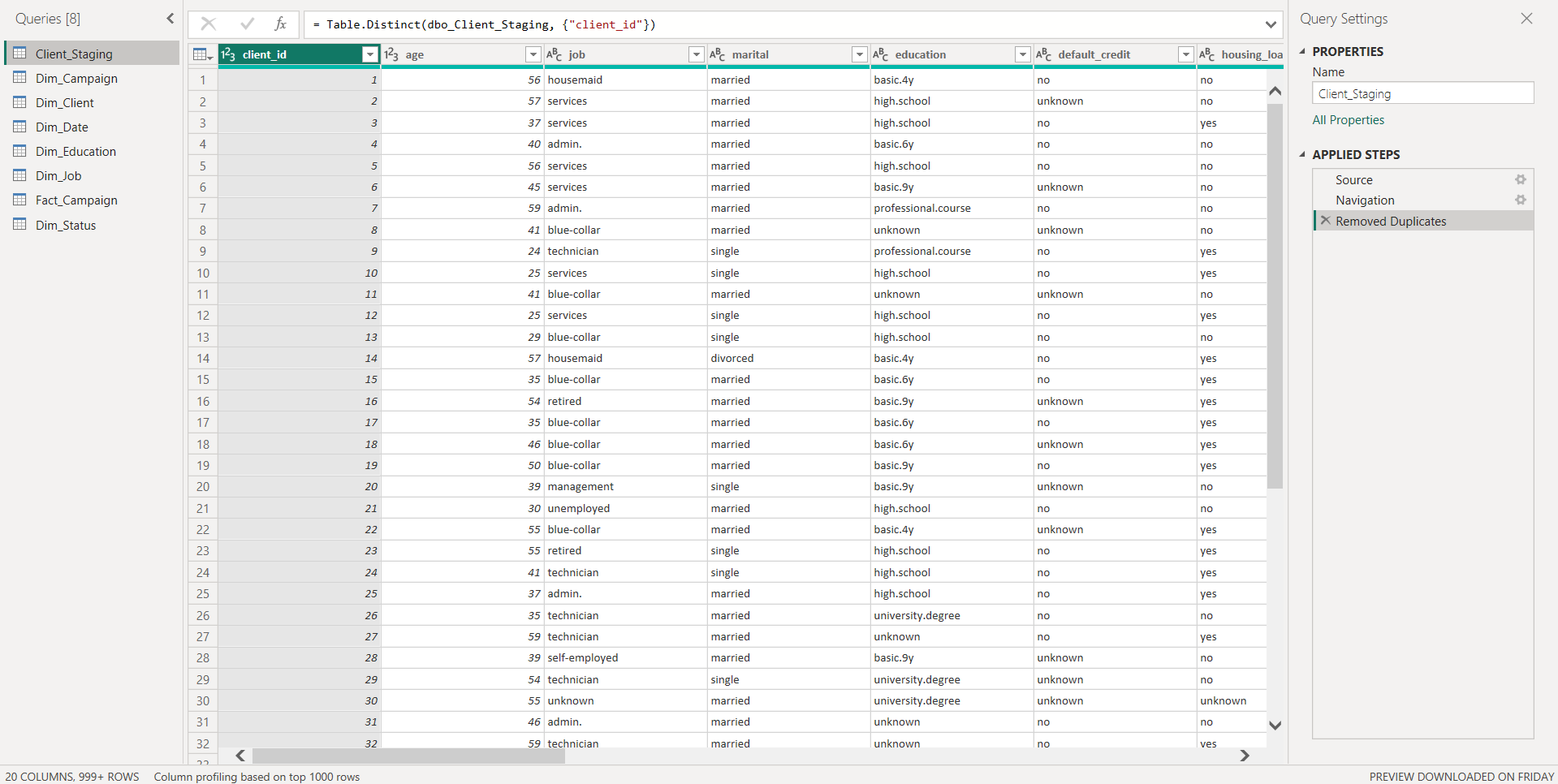
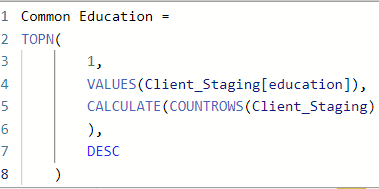


Figure . Các bước chuyển đổi dữ liệu trong Power Query

1. Thêm các Measure về tỷ lệ chuyển đổi, giá trị phổ biến của các thuộc tính trong bảng có liên quan.



1. Tương tác với các phần trực quan hóa trên dashboard

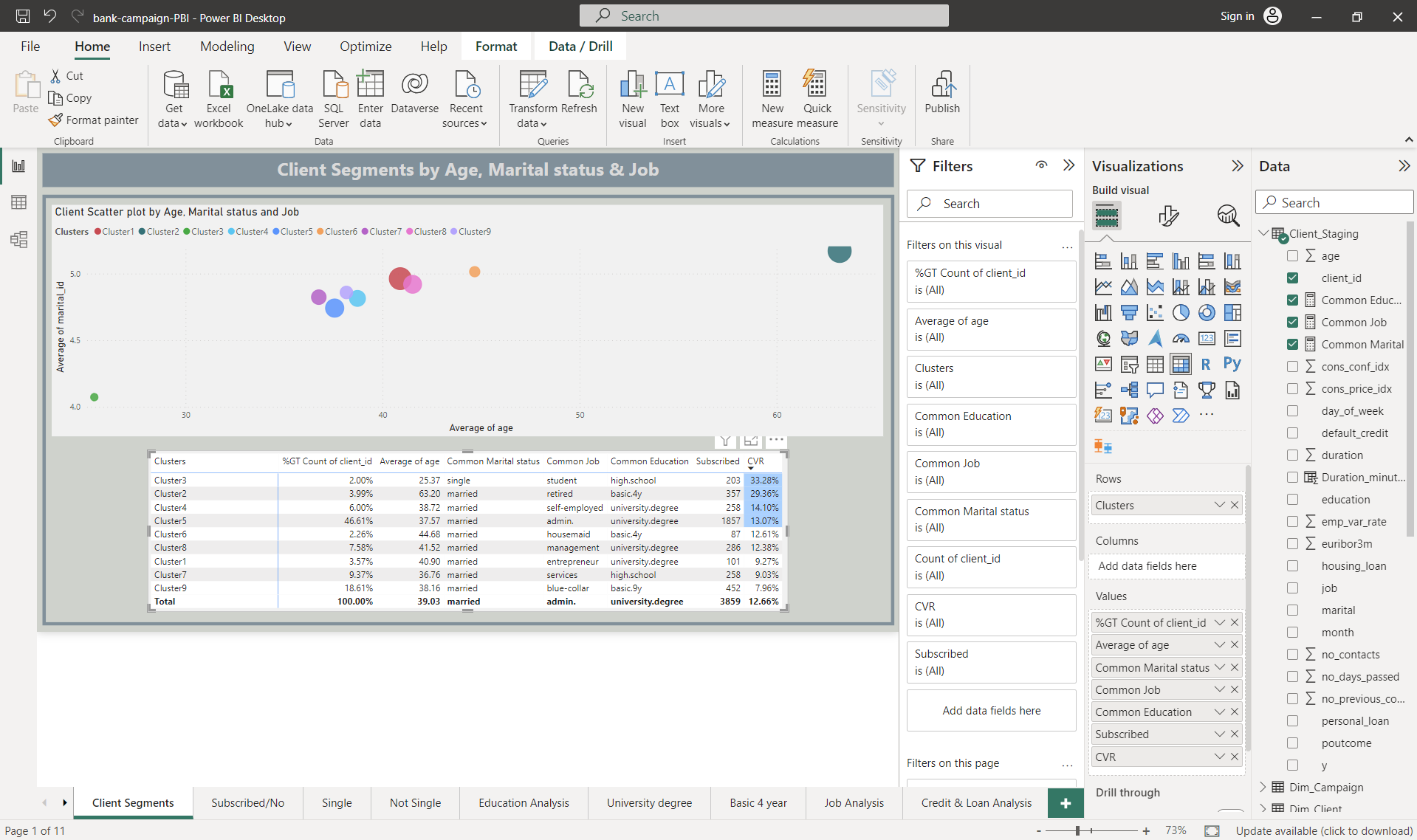


Figure 10. Report View của dữ liệu trong Power BI Desktop

## Những khó khăn gặp phải khi hoàn thành dự án:

* Thiếu dữ liệu về khách hàng đã tham chiến dịch trước đó
* Thiếu dữ liệu về chi phí của chiến dịch và giá trị giao dịch gửi tiền có kỳ hạn

## Nhược điểm của sản phẩm:

* Chưa đủ dữ liệu về chiến dịch trước đó để tính tỷ lệ rời bỏ và tỷ lệ giữ chân (churn rate, retention rate)
* Chưa thể tính toán đánh giá hiệu suất chiến dịch dưa trên các chỉ số đo lường hiệu quả của một chiến dịch như ROI và CPA ( Return on Investment và Cost per Acquisition)